Identification of mitochondrial RNA processing factors in *Arabidopsis thaliana*

DISSERTATION

zur Erlangung des Doktorgrades Dr. rer. nat.

der Fakultät für Naturwissenschaften

der Universität Ulm

vorgelegt von

CHRISTIAN JONIETZ

aus Bobingen

2010

Die Arbeiten im Rahmen der vorgelegten Dissertation wurden in der Zeit von August 2007 bis Dezember 2010 im Institut Molekulare Botanik der Universität Ulm durchgeführt und von Herrn Prof. Dr. Stefan Binder betreut.

Amtierender Dekan der Fakultät für Naturwissenschaften:

Prof. Dr. Axel Groß

Erstgutachter:

Prof. Dr. Stefan Binder

Zweitgutachter:

Prof. Dr. Anita Marchfelder

Tag der Promotion:

23.03.2011

Table of Contents

1. Introduction 6

1.1 Mitochondrial RNA expression in Arabidopsis thaliana 6

1.2 Aim of the project 9

2. Materials & Methods 11

2.1 Materials 11

2.1.1 Plant material 11

2.1.2 Consumables 11

2.1.3 Nucleic Acids and Markers 12

2.1.3.1 Oligonucleotides 12

2.1.3.2 Markers 17

2.1.3.3 Vectors/Plasmids 17

2.1.4 Enzymes 17

2.1.4.1 DNA Polymerases 17

2.1.4.2 RNA Polymerases 17

2.1.4.3 Further Enzymes 17

2.1.5 Miscellaneous 18

2.1.6 Kits 18

2.1.7 Solutions 18

2.1.7.1 Overproduction of RPF2 and RPF3 18

2.1.7.2 In vitro transcription (IVT) 18

2.1.7.3 SDS PAGE and Immunodetection analysis 19

2.1.7.4 Eletrophoretic mobility shift assay (EMSA) 19

2.1.7.5 Primer Extension Analysis 19

2.2 Methods 20

2.2.1 Circularized RNA RT-PCR (CR-RT-PCR) 20

2.2.2 Overproduction of RPF2 and RPF3 20

2.2.3 In vitro transcription (IVT) 21

2.2.4 Electrophoretic mobility shift assay (EMSA) with recombinant RPF2 protein 22

2.2.5 Primer Extension analysis of Terminator exonuclease-treated RNA 22

3. Results 24

3.1 Characterization of RPF2 24

3.1.1 Screening of T-DNA insertion lines reveals an aberrant cox3 mRNA phenotype in rpf2-1 (SALK_146198) 24

3.1.2 Overproduction of RPF2 26

3.1.3 EMSA with recombinant RPF2 27

3.2 Summary of the results obtained in the RPF2 analyses 27

3.3 Identification and characterization of RPF3 28

3.3.1 Screening of T-DNA insertion lines reveals an almost complete absence of mature ccmC transcript in mutant line SAIL18E04 28

3.3.2 The function of RPF3: transcription initiation or 5‘ processing 29

3.3.2.1 Primer extension analysis with terminator exonuclease-treated RNA 30

3.3.2.2 The search for a ccmC leader molecule 31

3.3.2.2.1 Detection of the ccmC 5‘ leader molecule by CR-RT-PCR 31

3.3.2.2.2 Detection of the ccmC 5‘ leader by RT-PCR 32

3.3.3 Accession specific cis elements lead to the formation of different ccmC mRNA ends 33

3.3.4 Introduction of Col RPF3 does not restore the ccmC phenotype in No-0 38

3.3.5 Overproduction of RPF3 in E.coli 40

4. Discussion 42

4.1 RF-like PPR proteins in Arabidopsis thaliana 42

4.2 The physiological relevance of 5‘ mRNA processing 44

4.3 Cis elements of mitochondrial mRNA end processing in Arabidopsis 45

4.4 RPF3 might fulfill different roles in different Arabidopsis accessions 47

5. Summary 50

5. Zusammenfassung 51

6. References **52**

7. Appendix **55**

7.1 Supplemental Figures ***55***

7.2 List of Publications ***61***

7.2 Publications/Manuscripts ***62***

7.3 Danksagung ***89***

7.4 Curriculum vitae ***90***

7.5 Erklärung über die in Anspruch genommenen Hilfen ***91***

1. Introduction

1.1 Mitochondrial RNA expression in Arabidopsis thaliana

Like allmost all eukaryotic cells plant cells contain mitochondria. The origin of mitochondria most likely goes back to a phagocytotic event in which an aerobic eubacterium got engulfed by an anaerobic ancestral eukaryotic cell. This view has been contested in recent times by the hydrogen hypothesis. According to this theory a hydrogen dependant archea engulfed a facultatively anaerobic eubacterium that was able to produce hydrogen. These events took place about 1.5 billion years ago and explain why mitochondria have their own genome and machinery for expressing their genetic information.

The main function of mitochondria was discovered as early as 1949 when it was shown that the oxidative energy metabolism takes place within this organelle (Kennedy and Lehninger 1949). In the following years several other vital functions performed by mitochondria were uncovered. Substantial progress in understanding mitochondrial function and organization was made possbile when genome sequencing approaches revealed the nuclear and mitochondrial genomes of several species.

The mitochondrial genome of the flowering plant *Arabidopsis thaliana,* now referred to as *Arabidopsis,* is about 360 kb in size and contains 57 genes (Unseld et al, 1997). Those genes either encode for subunits of the respiratory chain complexes I to IV, proteins necessary for cytochrome c maturation, ribosomal proteins, ATP synthesis as well as 3 rRNAs and 22 tRNAs. The vast majority of proteins needed for mitochondrial biogenesis and function is encoded in the nucleus and has to be imported into the organelle. This includes allmost all proteins required for the maintenance and the realization of the mitochondrial genetic information. Mitochondrial genes are found solitary or clustered throughout the mitochondrial genome. The corresponding transcription units are separated by large intergenic regions that account for about 330 kb of the genome (Unseld et al, 1997).

Gene expression in plant mitochondria is a complex process that requires concerted action of nuclear and mitochondrial factors and is far from being fully understood. Transcription is initiated from a multitude of promoters, often several active ones are found upstream of a given gene (Tracy and Stern 1995, Kühn et al. 2005, Kühn et al. 2009). Most of these promoters contain a YRTA, RRTA or ATTA motif (Y = T or C and R = A or G) directly upstream of the initiation site (Kühn et al, 2005, Kühn et al. 2009). RNA synthesis is most likely carried out by two nuclear encoded phage-type RNA polymerases (Hedtke et al.,

2000) and leads to mono- or polycistronic transcripts that undergo numerous maturation steps. These include RNA editing, *cis* or *trans* splicing of group II introns and processing at the 5‘ and 3‘ extremities of the transcript (Gagliardi and Binder 2007).

In recent years only a few factors taking part in mRNA maturation have been identified. Apart from the two 3‘-5‘ exonucleases polynucleotide phosphorylase (PNPase), an RNase R homolog 1 (Perrin et al. 2004a, 2004b), one RNase P homologue called proteinaceous RNase P (PRORP1) (Gobert et al. 2010) and two DEAD-box proteins (Matthes et al. 2007) all other factors characterized so far are Penta tricopeptide repeat (PPR) proteins.

The PPR proteins were found to be greatly overrepresented in plants with about 450 members in *Arabidopsis* compared to only a few in other eukaryotes and none in bacteria. Proteins of the PPR family consist of 2-26 tandem arrays of a loosely conserved 35 amino acid motif (P), each folding into two α-helices, probably resulting in a superhelix. These structural features allow the formation of a hydrophilic grove that is charged at its base. These properties enable PPR proteins to either directly interact with RNA or at least mediate interaction of other factors with RNA (Small and Peters, 2000). Variations in length of this motif with either 31(S) or 36 (L) amino acids in length are also possible. According to the presence of these motifs in the protein the PPR genes have been divided into two main classes: the P class proteins consisting exclusively of canonical PPR motifs and the PLS class proteins containing a succession of the P, the L and the S form. The PLS class is further subdivided by presence of certain plant specific C terminal domains. Subgroups either have no additional domains (PLS), an E domain only (E subgroup), an E and E+ domain (E+ subgroup) or E, E+ and DYW domains (DYW subgroup).

Recent studies have shown that PPR proteins participate in many steps of organellar mRNA maturation in plants. Most of the PPR genes involved in mitochondrial RNA metabolism encode for RESTORERS OF FERTILITY (RF) in different plant species where cytoplasmatic male sterility (CMS) occurs. The CMS phenotype results from the inability to produce viable pollen and is caused by a polypeptide that is encoded by a chimeric mitochondrial gene. Most RF proteins restore the phenotype by either leading to cleavage of the CMS causing transcript or blocking its translation (reviewed in Schmitz-Linneweber and Small 2008). Other PPR proteins that have been characterized in detail include the ORGANELLAR TRANSCRIPT PROCESSING DEFECT43, which is needed for trans-splicing of intron 1 of the *nad1* mRNA in mitochondria of *Arabidopsis* (de Longevialle et al. 2007) and several MITOCHONDRIAL RNA EDITING FACTORs (Zehrmann et al. 2009; Verbitskiy et al. 2010; Takenaka et al. 2010; Bentolila et al. 2010).

At the beginning of the here presented work hardly any factors taking part in mitochondrial mRNA processing had been published (reviewed in Binder et al. 2010). Studies with the aforementioned mitochondrial 3‘-5‘ exonucleases PNPase and RNR1 demonstrated the importance of 3‘ processing of mitochondrial mRNAs as knockouts of these exonucleases lead to embryolethality or severely retarded growth (Walter et al. 2002, Holec et al. 2006). More insight on 5‘ and 3‘ processing was obtained when tRNA-like structures, so called t-elements, were found at the 5‘ and 3‘ ends of some mitochondrial transcripts. These secondary structures most likely act as *cis* elements in the recognition of specific processing sites, probably by the mitochondrial tRNA processing enzymes RNase Z and PRORP1 (Forner et al. 2007; Gobert et al. 2009). Indeed it was found that these enzymes are able to cleave synthetic RNAs containing t-elements *in vitro* at the same positions as found in the corresponding mRNA *in vivo* (Canino et al 2009, Gobert et al., 2009). Taken together these findings strongly suggest that these endonucleases take part in mRNA maturation in mitochondria of *Arabidopsis* most likely guided to their respective target mRNAs by t-elements or tRNA acceptor-stem-like stem loop structures. However, these secondary structures are only scarcely present in mitochondrial mRNAs with seven found at 5‘ ends and only two found near 3‘ ends so far. The processing sites of the *ccmFc*, *rps3*, *atp6-1* and *atp6-2* mRNAs are determined by real tRNAs encoded in direct vicinity of these genes (Forner et al.,2007). For the remaining mitochondrial transcripts no such secondary structures could be detected raising the question of how processing sites are recognized when no obvious t-RNA-like structures seem to be present. In chloroplasts of maize it was shown that a primary sequence motif gets recognized by PPR5. Binding of PPR5 to the unspliced precursor of the *trnG*-UCC RNA leads to its stabilization by protecting them from endonucleolytic degradation (Beick et al. 2008). Similarly, PPR10 from maize binds to a 25 nucleotide motif present at 5‘ and 3‘ ends of a number of chloroplast mRNAs, building a barrier for exonucleolytic attack and thereby defining the 5‘ and 3‘ ends of the mature transcripts (Pfalz et al. 2009). Binding to primary sequence elements might also be feasible in mitochondrial mRNA processing. So, when no secondary structures are present, primary structures might be recognized by PPR proteins which, in turn, direct effector complexes to the respective processing site.

In a comprehensive approach major 5‘ and 3‘ ends of mitochondrial protein coding transcripts in the three *Arabidopsis* accessions Columbia (Col), C24 and Landsberg *erecta* (L*er*) have been mapped using circularized RNA reverse transcription PCR (CR-RT-PCR) (Forner et al., 2007). Virtually no variation was found at the 3‘ ends of mRNAs of the investigated ecotypes. Apart from the ccmC and nad6 transcripts, where 3‘ ends are

located within the coding region (Forner et al. 2007), all 3‘ ends have been found 30-500 nucleotides downstream of the translation stop codon. Also only one common 3‘ end per gene transcript was found in most of the cases. In contrast, a lot more complex situation was observed at the 5‘ extremeties. Here often mRNAs with several different 5‘ ends per gene are present. These either result from transcription inititation at various promoters upstream of a given gene or they are generated by several processing events. Comparison of 5‘ ends from Col, C24 and *Ler* also revealed a number of accession-specific polymorphisms. Those were either due to differences in the mitochondrial DNA configurations present in these ecotypes or they could be attributed to variations in the nuclear genomes (Forner et al. 2005, 2007). In the case of the *cox3* mRNA a mitochondrially encoded 5‘ end polymorphism between Col and C24 was observed. Here sequence differences several hundred nucleotides upstream of the mature Col seem to influence the 5‘ end processing pattern (Forner et al. 2005). In the case of the *nad4* and *nad9* transcripts nuclear factors were shown to be responsible for the observed phenotypical differences (Forner et al. 2008).

1.2 Aim of the project

The aim of the here presented studies was to identify and characterize the nuclear factor accountable for the *nad9* mRNA polymorphism between Col and C24 by linkage based cloning. To this end an F2 mapping population of about 200 plants was phenotyped and genotyped. The RPF2 locus mapped to an interval on chromosome 1 between 22.3 and 24 mbp (Jonietz, 2007). Within this region a cluster of PPR proteins with strong similarity to RF proteins was chosen as the best candidates for taking part in mRNA processing. Suprisingly these genes are present in *Arabidopsis* even though CMS has not been reported here. This indicates a function for these proteins apart from fertility restoration. A set of homozygous T-DNA insertion lines harboring their respective T-DNAs within the open reading frames of those RF like PPR genes was established. Phenotyping of those T-DNA insertion lines by CR-RT-PCR and a complementation experiment unambiguously identified At1g62670 as RNA PROCESSING FACTOR2 (RPF2). For further characterization of RPF2 it should be investigated how the mature *nad9* mRNAs are generated. To establish a model of the mode of function of RPF2 its RNA binding capability was to be checked by an eletrophoretic mobility shift assay (EMSA). Another goal was to get more insight into the evolution of RF-like PPR proteins in Arabidopsis by comparison of RPF2 alleles from different accessions. It was also of interest whether RPF2 or other RF-

like PPR proteins might have other functions in mitochondrial mRNA processing. Therefore the screening of T-DNA insertion lines was to be expanded to cover all mitochondrial protein coding transcripts. This screening gave hints on another function of RPF2 in the processing of *cox3* mRNAs. It also lead to the identification of another RNA processing factor which is needed to generate mature *ccmC* mRNA in Col and to less extent also in C24. This factor, termed RPF3, was also to be characterized further in a similar fashion as RPF2. A complementation experiment was to be carried out to unambiguously demonstrate its role in the generation of mature *ccmC* mRNA. The subcellular localization and function of RPF3 in other *Arabidopsis* accessions was also to be studied. Most of the results of these experiments can be found in a publication and a manuscript in the appendix while additional results have been summarized below.

2. Materials & Methods

2.1 Materials

2.1.1 Plant material

- 14-21 day old seedlings of *Arabidopsis thaliana* accessions Bayreuth (Bay-0), Bologna (Bl-1), Blanes (Bla-10), Calver (Cal), Las Palmas (Can-0), Coimbra (Co), Columbia (Col), Cap Verde Islands (Cvi), Dijon (Di-1), East Melling (Ema-1), Enkheim (En-1), Estland (Est), Kashmir (Kas-1), Landsberg (La-1), Mertubal/Cyrenaika (Mt-0), Niederzenz (Nd), Nossen (No-0), Oystese (Oy-1), Pitztal/Tirol (Pi-2), Shakdara/Pamiro-Alay (Sha), Sorbo/ Kurmatov (Sorbo), Tsu (Tsu-1), UofBC/B.C. (Van-0), Wassileskija (Ws), Yosemite National Park (Yo-0), Landsberg *erecta* (L*er*), C24.
- 14-21 day old seedlings of T-DNA insertion lines SALK_061670, SALK_020638, SALK_146198, SALK_139736, GABI_337F04, SALK_022265C, SALK_020230, FLAG_288A04, SAIL_1282C03, SAIL_18E04, SALK_128068, SALK_087235
- leaves from 21-30 day old complemented (+At1g62930 from Col) plants from the accession No-0
- *Arabidopsis thaliana* hypocotyl cell suspension culltures from the accessions Col and C24
- mitochondria prepared from cell suspension cultures of the accessions Col and C24

2.1.2 Consumables

Consumables were ordered from the following companies:

- Carl Roth
- Merck
- Sigma Aldrich
- Normapur
- Rectapur
- Duchefa Biochemie
- Sarstedt

2.1.3 Nucleic Acids and Markers

2.1.3.1 Oligonucleotides

RPF2 related work:

Primername	Sequence (5‘-3‘)
Atnad9-1	agagggatcatcaactccgc
Atnad9-2	cttgggtaaagtctcccaac
Atnad9-3	cgattttgctagtccttggg
Atnad9-7	ttggtccaaccaagaaaggc
Atnad9-NS.H	atggataaccaattcattttc
Atnad9-NS.R	ttatccgtcgctacgctg
Atnad9-EMSA.H	ggtcgagttaagtaaagactggttacc
Atnad9-EMSA.R	taatacgactcactataggggcgagtcgtctagggcatc
Atatp9-2	ttgcaccttctaacatctcg
Atcox3-2	gagtgaacccgaaataggcc
Atcox3-23	gatctatcgagtaaccaggaaagag
Le-cox3-HindIII.H	atataagctttccataggcttggccctag
Le-cox3-HindIII.R	ctaagaagcttcgaataccctttaccaag
Atcox3-3‘Primersonde	cgactttggtatgatgtcct
Atcox3-Mega3‘	cacataaatgaaggaatcgaagagattatggcag
Atcox3-Mega5‘	aacggagacaccaattaaggaacgcac
Atcox3-5'-Primersonde	gattgaatctcagaggcattc
Atcox3-3'-Primersonde	cgactttggtatgatgtcct
At1g62670.H	ggttgaattttacagagcttccgaatcagagagc
At1g62670.R	gatatggtttgctcgactattctctcgactcaagg
At1g62670-Kompl.H	atgcagtcgagcggcgcgccttgtaattgttactgtgtttgtgcatttcccc
At1g62670-Kompl.R	tgactccgaggcttaattaattccagcacctgatgagttttctcatcc
At1g62670-Kompl.H3	ttcacaagtacccgggcctatgaggatatcgtttgcgattgc
At1g62670-Kompl.R5	tgactccgaggcttcccgggggaagggaagggacgagacttg
At1g62670-Kompl.R3	tgactccgaggcttcccgggcgaagaaagcatatcgaggaagct
At1g62670UEx.H2	atgcagtcgagcggatccggtaatgctgcaactgtttctcc
At1g62670UEx.R	gacttagagtttctgcagttaagaaagcatatcgaggaagcttttgtc
Atccb2-8	aaaggtgcaggaggaaaagg
Atccb2-Mega5'	gccgcgtaacatcatcttcaaccaacc
Atccb2-Mega3'	gttcgtatctctttttccaatttcggtctcga
Atrpl5-1	cacacaatccaggaacttcc
Atrpl5-5	tttgagcaacagatcttgacg
Atcob-3	attctgggacgagttggaag
Atatp6-1-NS.R	ttatatccgccgcctgtcccac
Atatp6-3	accggacgcttgaccactag
Atatp6-2	ggtctggaattaggtgtagc
Atatp8-pu2	tacaccattgggatacttcg
Atatp8-1	gtactccatctccatcattgc
Atatp8-4	agtggctcacgaggaatgg
Atnad7-2	tatgtggttccgcacgttcc

Atnad7-7	tgcacctaaaggagaatttggtgtctttctgg
Atnad7-8	ttttaatttgccttttcctagtcgtcatggg
Atrpl2-2	gtctaagtgctcttgctctc
Atrpl2-3	gaggaagtgcttcgtttcag
AtorfX-3	cacagctgctcttttcacac
Atnad4-3	gactgcatgcatacatccg
Atnad4-5	ttgtatccgaggaacaggag
Atnad4-15	aacccttcccgtgtatacg
Atatp9-14	agagcaaagcccaaaatgg
Atatp9-Mega.5'	agagcaaagcccaaaatggcataacc
Atatp9-Mega.3'	gccccaatgatggccttttgatcttattcg
Atrps7-2	gctttatattctctacggcg
Atrps7-7	atcaatagtttttgctcaccatccaagccc
Atrps7-8	gaagtttcgcgcatttcagatggtgg
Atcox1-1	cccaccaagaatttgatcg
Atcox1-2	tacgaatcagtactgagaagc
Atcox1-4	gaatggatggtacaaagtcc
Atnad2ab-1	acccacattactggctaacg
Atnad2ab-2	aaagatctctggggaaaccg
Atnad2cde-3	gtcgttactactagcaatgac
Atnad1a-2	gtaacaatccgaacgatccc
Atnad1a-Mega.5‘.B	ttccaagtatttcagctggaacagctatgtaca
Atnad1de-Mega.3‘.B	tggtcttttagtcacctttcaatggcttcc
Atatp1-3	caatcccatctccaactgag
Atatp1-8	ttcgcgtaaaagttcctaattcgactttcg
Atatp1-9	gcagctgtcaatggattctgtgatcg
Atatp6-2.Mega5'	taatatatccggcctgaatcactccctcg
Atatp6.Mega3'	tcttgcattaaccggtctggaattaggtg
Atnad3-1	aaacgacttctggcatcacc
Atnad3-5	aaggatcgaaaccacattcgtaggcc
Atrps12-6	aggagtaaaggatttgatgggaattccgg
Atccb6n2-1	ttcgaattggttcctggtcc
Atccb6n2-5	tcctgatcagagaccactgtgttcgc
Atccb6n2-6	agcaggcatcggtccgtagaacg
Atccb3-5	caataatactcatccgagcc
Atccb3-Mega3'nah	tgggagcattagccgatttggtacatc
Atccb3-Mega5' nah	cttgtgcgtggcactccttgct
Atccb6n1-1	caaaatgcaagtgctgcacc
Atccb6n1-Mega.5'	tgtaagtgaatgcaacgaaaagacccaga
Atcccb6n1-Mega.3'	agcgagaagccactcgactaaaaggtactg
Atnad5ab-1	ttccgatgagatccatggag
Atnad5ab-2	atgaaacgcacgtagtggtc
Atnad5de-4	tgcctttgcaatgttacttgg
Atnad4L-1	agctgccaccgttagaacc
Atnad4L-2	aggaggattccccaaatacc
Atorf25-3	tcattcgagagagcttggtg

Atrps4-2	ctctgttccgaacatttcctg
Atrps4-4	taggtcatatccctcacgac
Atnad6-1	agtaaacctgaagtgtcgcg
Atnad6-2	gcacgtacaaccatcaaacc
Atnad6-4	ttagtagccatgattggggc
Atccb6-1	catttcaatagtacgggtg
Atccb6-2	gaatagggcgacttgctctg
Atccb6-3	gagaaagttccgtggagttc
Atcox2-5	aaggttccttcattttctgc
Atcox2-6	ttgtatggcaggaataccac
Atrps3-1	tacacctaccgagacgaaag
Atrps3-2	aaggtgagtctcataggtgg
Atrpl16-3	ccaagtttgttcagtggtcg
CER450727.H	gaagattccagatttggcgttgtctc
CER450727.R	agaaagctcagaaaggaaaacgcacag
CER452163.H	aatcgacgaataccactattttgggctc
CER452163.R	tctgccagaagttgctcgttaaagacac
CER452601.H	ccattgaagcaagaaatttgaggaagatgg
CER452601.R	tggatgcctaataccgggactatttcctg
CER458022.H	tcattcttatagcaaagctttcgatccata
CER458022.R	gcagttgggagtctcatgtattctatgataggc
CER458676.H	ttcactgaaatcctggctatctaatgcg
CER458676.R	ccattctgccaaaagatattattcagggtc
CER460683.H	cgcaaccactatatacaaaatgatgcaagg
CER460683.R	cttgaaaggagacgcgagacgc
CER464632.H	tctgcgaagtggaagagaaaaaagaactc
CER464632.R	gttttgccggtgattgtgacctaagc
CER479928.H	cagttcccaatgagcaatgtacacatgg
CER479928.R	ccttgttggatgtaaaatagtttcgggttcc
CER449403.H	tgattccagcgggtaggcatgacc
CER449403.R	cgaaacattcaaaaagtggttatgggagag
CER449629.H	ccaaaaggcaacatgaatgacatgg
CER449629.R	ttttggggtttttgcagaaacatgg
CER452047.H	caaaagccacactaagggacaacaatcc
CER452047.R	tgcaccattgagacaagctctatacaatctg
CER458621.H	cagcctcattgcctgatgtcaagaatgc
CER458621.R	cgggtgaatgggcagcatgg
CER460534.H	catggtcaatgcaacgttacaaagtgc
CER460534.R	cgaaacgcccaaagcgtaatctcc
CER452265.H	aatgagcgtcgagcttatacaatgatatgc
CER452265.R	tgtctcacgcgatgcgtggg
CER454209.H	tgtggtgatgtagcaaccactggttctactac
CER454209.R	cactctcttctaaatggtttgagtcgttcg
CER460458.H	ctgaaaatccttgcgtgagtgtttatgag
CER460458.R	ttttagtttgtgacgaatagtgaaaggagagaag
CER464890.H	cggtactgctcgaaggaacgagactg

CER464890.R	gcctctgtggctgctactgaattcg
CER464940.H	gcagacgatctgttagctaggcgaagtg
CER464940.R	cgtctcgcgcccagtacacg
CER464947.H	tttggttgctttgaagaacaaatgttatatcg
CER464947.R	cgtgctgtgtgcgcatgc
CER470109.H	ttttgtgattctatgttggcatcagtgg
CER470109.R	caaaccgtagtgcaaattagagcatagtttcc
CER452226.H	gagttcgacggtggaggagattgc
CER452226.R	tgttgagttactttaagctctgtgtacattctccag
CER453202.H	gaggaggaagacaaggtaattggagttgg
CER453202.R	tcaacgttttgcaaatttattgaatgtgc
CER453947.H	cgtgtattgttgcccacttgtgtgg
CER453947.R	catgaactgcattaacatggaacattgc
CER460302.H	gtcggcttacaaacgcgtttatgtcc
CER460302.R	ctgctaagtcgatcaaagccagatcg
CER460528.H	caatgatttggtgggtcaagagtcaagac
CER460528.R	gcaaaagtcattacggacaataccaaacg
CER454633.H	ggtcttgctgaggttagtgtctttgtgtgtg
CER454633.R	ggatcatagcaagtaaaacacttcttcaaagagc
CER455160.H	tgggttaaaataacaagggctttgcag
CER455160.R	cgtgttattggctctcttaaggtcaatctgg
CER456317.H	cattgccccgcggtaataagctg
CER456317.R	gaagatcttcttttcatatgagctgcttgctc
CER456556.H	cggtttttaatgggcttcacgtgttg
CER456556.R	cacactctgctacgaattttctcgaactcag
CER456657.H	ccttcagaatcaccaactgtggagtcc
CER456657.R	gaaagaaaccatatgcatcaaatctaccttcg
CER456772.H	ccatgtgacatgcacttacacatgtagtg
CER456772.R	ccactccacaattctttcttcttcattgg
CER459335.H	ctcgggcacaaaaccctcagg
CER459335.R	gctttaagagagtgatggcgatgttgg
CER449062.H	gtacatacgaggctccggaccatcc
CER449062.R	tgttcaaaccacataatacgtaaccatgattcg
CER449068.H	gcaaagcaagacacatcttcagacactgc
CER449068.R	tccggtaccactctaactctattgccatcc
CER449079.H	gatctttaacacacgtagatgcaacaaacatcc
CER449079.R	tgtggcttaattattttgcaagtcaacctacgtg
CER452585.H	ccaatatagtgggggaaacaattgggatgac
CER452585.R	cctcatatggtcagtcaacaacatatgctcagtg
CER464632.H	ctgtattctggaaatatctgcgaagtggaagag
CER464632.R	ttcattagttttgccggtgattgtgacc
CER464694.H	tgagaaacatcatcggatgttgatagtacatcg
CER464694.R	gctcttagttgttgtaatgaatatgaaatcttagtcg
CER452694.H	tgactccgtcttccagagtttgactctg
CER452694.R	tgcagaagcaggtaaggagaagaccag
CER465522.H	gatatcccccgccggtgattc

CER465522.R	gcgggattggatggctaaattaactagatctagtc
CER465523.H	tgaaatgttaaggccaaaatacagctgacg
CER465523.R	gagggaagaatgaagaacatgtgcttgg
CER474029.H	cccttgctgccttatatgcaattctgtc
CER474029.R	gcgcaagaaatttcatttctgctagcag
CER479915.H	tgttccaattcaaaagaagttgtgctctctg
CER479915.R	ctttgccgagatatacttggatcatactgatctc
CER449342.H	ggttcttaaatatttgcggctgaaagacaagg
CER449342.R	gaggtggctcaccctgaaactctcc
CER450582.H	ggttagtggaagttaatgtggtatctcacgatacg
CER450582.R	catctccgatgtattttgctaattgaggttatgg
CER450591.H	tggattggtgaggtgagtctgtggac
CER450591.R	cacgtttatccctttctcatcaagcagagg
CER459065.H	ctctatcgatttcatgggggagaaactagc
CER459065.R	cagctattttatcgtggtgcgccatc
CER464751.H	gctaagagcactctgtcttagaaacatgttgacc
CER464751.R	caaggacattcacacaattgtaaaacgttgg
MASC00581.H	ttttttggttatgtagctagcggttcttgatttcaatgc
MASC00581.R	gcctaaccgtttttctcactctctgctc
MASC04181.H	ggggtcagttacattactagc
MASC04181.R	gtacggttcttcttccctta
MASC03744.H	caattctctgataatctaatttaaaagacaaaaaaattacaaataaaattttcga
MASC03744.R	tcgaattctaacatggtttctccagaccac

RPF3 related work:

Primername	**Sequence [5'-3']**
Atccb3-Mega5'.nah	cttgtgcgtggcactccttgct
Atccb3-Mega3'.nah	tgggagcattagccgatttggtacatc
Atccb3-Mega5'.fern	tcttacgccgatcgatatcgatcaactc
Atccb3-3	tgttctggaaacacgtcttc
Atccb3-12	gtacaaggaaagtggctcaaaacgcc
Atccb3Sonde5'	tcgctatgcgccgctccg
Atccb3Sonde3'	aaagagagcgagtgaactagg
At26S-Adapter.R2	ttgccctttaggaatcttcagtcaaggtacg
Atccb3QPCR.1	atcccacacccaaaaggtc
Atccb3QPCR.2	aaccggtatagaaatgggtgc
Atccb3QPCR.3	aaacacatgaactcttagtatccgaag
Atccb3QPCR.4	catttcatgttcagtcatcacaac
At1g13320.1	taacgtggccaaaatgatgc
At1g13320.2	gttctccacaaccgcttggt
At4g27960.1	tcacaatttccaaggtgctgc
At4g27960.2	tcatctgggtttggatccgt
At18S-Adapter.R1	ctggcacggagttagcc
At18S-Adapter.R3	cattggcttggtcagcc
At18S-Adapter.R4	cctagctccccgagaacaacgttcg
smartcDNA3‘	gaattcatgtcgacggtctca

At1g62930Kompl.H	atgcagtcgagcggcgcgcccctgaacaaaagagacattaactgaac
At1g62930Kompl.R	gacttcgagtttaattaattcgaaccgacaaattgtttgtc
At1g62930UEx.H2	atgcagtcgagcggatccatgagtttggcgaagagatttgc
At1g62930UEx.R	gacttagagtttgagctctcaagacagcatttccagatagctttttcc
At1g62930UEx.H	atgcagtcgagcggatccatgacaagttgtgttcatttagggattg
At1g62930GFP.R2	gacttagagtttcccgggtttattgaattcaacaatggaaggtaaagg
Atatp9-Mega.5‘	agagcaaagcccaaaatggcataacc
Atatp9-Mega.3‘	gccccaatgatggccttttgatcttattcg
KAN1	gaacaagatggattgcac
KAN2	gcgatagaaggcgatg

2.1.3.2 Markers

- 1kB DNA ladder (peqLab Biotechnologie)
- pUC19 *MspI* digested
- pGEM®3 Vector (Novex Electrophoresis) labelled with ^{32}P
- PageRuler™ Prestained Protein Ladder (Fermentas)

2.1.3.3 Vectors/Plasmids

- pET32a (Novagen)
- pCR®2.1-TOPO® (Invitrogen)

2.1.4 Enzymes

2.1.4.1 DNA Polymerases

- GoTaq® DNA Polymerase (Promega)
- Phusion® Flash High-Fidelity PCR Master Mix (Biozym)

2.1.4.2 RNA Polymerases

- M-MLV Reverse Transcriptase RNase H Minus Point Mutant (Promega)
- T7 RNA Polymerase (Fermentas)

2.1.4.3 Further Enzymes

- T4 RNA Ligase (Fermentas)
- T4 Polynucleotide Kinase (Fermentas)

- Terminator™ 5‘-Phosphate-Dependant Exonuclease (Epicentre)
- Recombinant Enterokinase (Novagen)

2.1.5 Miscellaneous

- S- protein Agarose (Novagen)
- EKapture Agarose (Novagen)
- S-protein HRP Conjugate (Novagen)
- Amicon Ultra 0,5 ml, Amicon Ultra 4 ml 30kDa (Millipore)
- ECL Detection System (Novagen)

2.1.6 Kits

- GFX ™ PCR DNA and Gel Band Purification Kit (GE Healthcare)
- Rediprime™ II Random Prime Labelling System (Amersham Biosciences) for labelling of DNA probes
- Spectrum™ Plant Total RNA Kit (Sigma Aldrich)
- GenElute™ Plant Genomic DNA Miniprep Kit (Sigma Aldrich)
- High Pure Product Purification Kit (Roche)

2.1.7 Solutions

2.1.7.1 Overproduction of RPF2 and RPF3

- 10 x Bind/Wash Buffer (1.5 M NaCl, 200mM Tris-HCl, 1 % Triton ® X-100, pH 7.5)

2.1.7.2 In vitro transcription (IVT)

- **5 x Transcription Buffer:** 200 mM Tris pH 7.9, 50 mM NaCl, 30 mM $MgCl_2$, 10 mM Spermidin, 0.25 % (v/v) Tween 20
- **10 x TBE Buffer (1 l):** 180 g Tris, 55 g Boracic Acid, 40 ml 0.5 M Na_2EDTA (pH 8.0), dd H_2O ad 1 l
- **4% Polyacrylamide-Gel (40 ml):** 16.8 g Urea, 4 ml 10 x TBE buffer, 4 ml PAA 40 % (w/v), dd H_2O ad 40 ml, beofre pouring the gel add 30 μl TEMED, 150 μl 10 % APS
- **RNA Extraction Buffer:** 0.5 M NH_4OAc, 0.1 M EDTA, 0.1 % SDS

2.1.7.3 SDS PAGE and Immunodetection analysis

- **2 x Lower Tris**: 1.5 M Tris, 0.2 % SDS, pH 8.85
- **4 x Upper Tris**: 0.5 M Tris, 0.4 % SDS, pH 6.8
- **Separating Gel (15 % PAA)**: 5 ml dd H_2O, 20 ml 2 x Lower Tris, 15 ml PAA 40 % (w/v), before pouring the gel add 30 µl TEMED, 175 µl 10 % APS
- **Loading Gel (4.5 % PAA)**: 7 ml dd H_2O, 2.5 ml 4 x Upper Tris, 1 ml PAA 40 % (w/v), before pouring the gel add 10 µl TEMED, 45 µl 10 % APS
- **SDS Running Buffer**: 0.25 M Tris, 1 % SDS, 1.9 M Glycine
- **SDS Sample Buffer**: 1 x Upper Tris, 20 % Glycerole, 4 % SDS, 100 mM DTT, small amount of Bromphenolblue
- **Transfer Buffer:** 20 mM Tris, 150 mM Glycine, 0.2 % SDS, 20 % (v/v) MeOH
- **TTBS Buffer:** 100 ml 2 M Tris HCl pH 7.4, 62 ml 5M NaCl, 2 ml Tween 20, dd H_2O ad 2 l
- **Membrane:** Roti® PVDF (Roth)

2.1.7.4 Eletrophoretic mobility shift assay (EMSA)

- **Binding Buffer B1**: 20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.2 mM EDTA, 4 mM DTT, 0.4 mg/mL BSA, 5 % (v/v) glycerol
- **Binding Buffer B2:** 50 mM Na phosphate, pH 8.0, 12.5 mg/mL heparin, 20 mM NaCl, 0.2 mM EDTA, 4 mM DTT, 0.4 mg/mL BSA
- **10 x THE Buffer:** 340 mM Tris-HCl pH 7.5, 660 mM HEPES, 10mM EDTA

2.1.7.5 Primer Extension Analysis

- **10 x TBE Buffer:** see IVT
- **5% Polyacrylamide-Gel (40 ml):** 16.8 g Urea, 4 ml 10 x TBE buffer, 5 ml PAA 40 % (w/v), dd H_2O ad 40 ml, beofre pouring the gel add 30 µl TEMED, 150 µl 10 % APS
- **IVT Loading Buffer:** 98 % (v/v) Formamide, 2 % (v/v) 0.5 M EDTA pH 8.0, small amounts of Bromphenolblue and Xylene cyanol

2.2 Methods

Basic methods of molecular biology were performed following standard protocols (Sambrook et al., Weigel and Glazebrook, 2002). In silico sequence analysis were carried out using various BLAST tools at NCBI servers (McGinnis and Madden, 2004) and the server at http://predictprotein.org for structural predictions of RPF2 and RPF3. DNA sequencing was commercially obtained (4base lab, Reutlingen and GATC Biotech, Konstanz).

2.2.1 Circularized RNA RT-PCR (CR-RT-PCR)

CR-RT-PCR was used for analysis of 5‘ and 3‘ mRNA extremities. RNA was prepared from 14-20 day old seedlings or leaves using the Plant Spectrum Total RNA Kit (Sigma). 5-15 µg of total RNA was ligated in a bulk reaction using 10 U of T4 RNA Ligase (Fermentas) in a buffer supplied by the manufacturer. Ligated RNA was desalted and purified by centrifugation at 13.800 x *g* for 10 min using Amicon Ultra 0.5 ml 30K columns (Millipore). 2-5 µg of purified ligated RNA was used for cDNA synthesis with 200 u of M-MLV Reverse Transcriptase (RNase H Minus Point Mutant) (Promega) in a buffer supplied by the manufacturer and random hexamer oligonuleotides. The reaction was carried out for 1 h in a total volume of 25 µl. Synthesis was stopped by the addition of 5 µl 1 M NaOH and incubation at room temperature for 10 min. Afterwards the pH was neutralized by the addition of 5 µl 1 M HCl. After purification with illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) cDNAs were eluted in 100 µl dd H_2O. For amplification 2-10 µl of cDNA was used in a 25 µl PCR reaction using GoTaq DNA Polymerase in a supplied buffer (Promega) according to the manufacturers instructions.

2.2.2 Overproduction of RPF2 and RPF3

The coding region of the *RPF2* gene excluding the predicted mitochondrial target sequence of 24 aa was amplified with primers At1g62670UEx.H2 and At1g62670UEx.R by standard PCR using total DNA from Col WT as a template. Likewise, the coding region of the *RPF3* gene was amplified from Col-0 total DNA with primers At1g62930UEx.H2 and At1g62930UEx.R leaving out the predicted mitochondrial target sequences of 23 aa. Both PCR products were cloned into the overexpression vector pET32a (Novagen) using the *SacI* and *BamHI* restriction sites. For overexpression of the recombinant proteins these constructs were transformed into *E.coli* strain BL21AI (Invitrogen). Each of the reading

frames is N-terminally fused with a thioredoxin tag, an S-tag and a His-tag resulting in proteins with expected sizes of about 86 kDa in both cases. The thioredoxin tag increases the solubility of the recombinant proteins. The S- tag and His-tag can be used for affinity purification as well as for immunodetection of recombinant proteins. After transformation cells were grown at 37°C to an OD_{600} of about 0,4 and then transferred to either 4°C, 20°C, 30°C or 37°C to find the best conditions for overproduction. After 30 min protein expression was induced by addition of 0.2 % (w/v) L-arabinose and 0.1 mM IPTG. In the case of RPF2 overexpression was carried out for both 2 hours and 20 hours, while RPF3 overexpression was performed for 2 hours. Afterwards the cells were harvested by centrifugation at 16.000 x *g* for 5 min and resuspended in 1/10 of the culture volume of 1 x S-Tag B/W buffer. Bacteria were disrupted by ultrasonifying each sample 5-10 times with a Branson SONIFIER 250 with a microtip set at 90 % output. Soluble and insoluble fractions were then separated by 3 rounds of centrifugation at 25.000 x *g* for 20 min. All fractions were finally analyzed by SDS-PAGE and immunodetection with an S-protein horseradish peroxidase (HRP) conjugate raised against the S- tag.

2.2.3 *In vitro* transcription (IVT)

To test whether the RPF2 protein is able to bind *nad9* RNA a synthetic RNA fragment was transcribed *in vitro*. This RNA contained the region from -296 to -124 in relation to the *nad9* translation start codon. To generate the template for RNA synthesis a PCR was performed with primers Atnad9-EMSA.H and Atnad9-EMSA.R on total DNA from Col WT. The latter oligonucleotide contained a T7 promoter sequence. About 1 µg of PCR product in 25 µl of dd H_2O was mixed with 10 µl of 5 x transcription buffer (see materials) for T7 RNA polymerase, 2 µl of a 3 NTP mix (10 mM each), 5 µl of 100 µM UTP, 5 µl [α-^{32}P]UTP (50 µCi), 40 u Ribolock (Fermentas), 40 u T7 RNA polymerase (Fermentas) in a total volume of 50 µl. After incubation for 90 min at 37°C the RNA was purified using the High Pure Product Purification Kit (Roche) following the instructions of the manufacturer. After concentrating the RNA to a final volume of 20 µl in a vacuum concentrator the *in vitro* transcripts were loaded onto a 4% PAA gel containing 7 M Urea. The gel was run at 10 W for 50 min before it was exposed to an X-ray film (GE Healthcare). The piece of gel containing the RNA was excised with a scalpel and incubated overnight at 4°C in RNA extraction buffer (see materials) to let the RNA diffuse out of the gel. Afterwards the *in vitro* transcripts were Ethanol precipitated, resolved in 20 µl dd H_2O and incorporation of

[α-^{32}P]UTP was measured in a scintillation counter. 10000 cpm of the labelled transcript were used per binding reaction in subsequent gel shift assays.

2.2.4 Electrophoretic mobility shift assay (EMSA) with recombinant RPF2 protein

Recombinant RPF2 containing the S-tag was enriched by S-tag affinity chromatography using S-protein Agarose (Novagen) according to the manufacturers instructions. At the end of this step the protein was eluted from the S-protein Agarose by cleaving off the thioredoxin, S- and His-tag with recombinant Enterokinase (Novagen). To remove the recombinant Enterokinase the protein lysate was mixed with 500 µl of EKapture™ Agarose (Novagen) and incubated for 10 min at room temperature on an orbital shaker. After centrifugation at 500 x *g* for 5 min the supernatant was transferred to a new tube. The agarose pellet was washed again with 1-2 ml of 1x Bind/Wash buffer (see materials) again followed by centrifigation at 500 x *g.* The supernatants containing the recombinant RPF2 protein were pooled then concentrated and further purified by centrifugation at 4000 x *g* in a swing bucket rotor using a 30kDa Amicon Ultra-4 column (Millipore). Centrifugation was carried out until the sample volume was reduced to 300 µl. Likewise, a control lysate was obtained from a culture containing an empty pET32a vector. EMSAs were performed with two different buffer systems (see materials). 150 pM of RNA (around 10000 cpm) was denatured at 94°C for 2 min, cooled on ice for 2 min and then incubated with 150-1000 ng of RPF2 or control lysate in EMSA binding buffer 1 or 2 for 30 min at 25°C. The binding reactions were carried out in a total volume of 10 µl. After addition of IVP loading buffer (see materials) samples were loaded on a 5% PAA gel containing 1 x THE buffer. The gel was run at 10-12 W for 30 min at 4°C. After electrophoresis the gel was transferred onto Whatman® paper and the RNA was visualized by exposition of an X-ray film (GE healthcare).

2.2.5 Primer Extension analysis of Terminator exonuclease-treated RNA

This experiment was performed to discriminate between primary (e.g. originating from transcription initiation) and secondary (derived from processing) RNAs.
8 µg of mtRNA from Col and C24 was incubated for 60 min at 30°C in a thermocycler with a heated lid with 1 u of Terminator™ 5'-Phosphate-Dependant Exonuclease (Epicentre) in a total reaction volume of 20 µl. Control RNAs were incubated under indentical conditions without exonuclease. Afterwards the mtRNAs were purified and concentrated in an Amicon

Ultra 0.5 ml 30K column by centrifugation at 13.800 x *g* (Millipore) until a volume of 10 µl was reached.

1 µl of a 20 µM dilution (20 pmol) of the oligonucleotide used for Primer Extension (Atccb3Mega-5‘.nah) was 5‘ labelled with 30µCi (3 µl) of [γ-^{32}P]ATP and 1 µl (10 u) of Polynucleotide Kinase (Fermentas) in a buffer supplied by the manufacturer and 13 µl dd H_2O. The labelled primer (~450.000 cpm) was used to initiate cDNA synthesis with 200 u of M-MLV Reverse Transcriptase, RNase H Minus, Point Mutant (Promega) according to the manufacturers instructions. The reaction was stopped by addition of 5 µl 1 M NaOH and incubation at room temperature for 10 min which hydrolizes the RNA. Afterwards the reaction mixture was neutralized by adding 5 µl 1 M HCl. cDNA products were purified with the use of the illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) according to the manufacturers instructions and eluted in 100 µl of dd H_2O. After concentration to a final volume of 8 µl in a vacuum concentrator and addition of IVT loading buffer the samples were loaded on a 5 % (v/v) PAA gel and electrophoresed alongside a radioactively labelled pGEM® marker. The gel was transferred onto a Whatman® paper and cDNA products were visualized by exposition to an X-ray film (GE healthcare).

3. Results

3.1 Characterization of RPF2

3.1.1 Screening of T-DNA insertion lines reveals an aberrant *cox3* mRNA phenotype in *rpf2-1* (SALK_146198)

After RPF2 was identified and its role in *nad9* mRNA generation had been established (Jonietz et al. 2010) it seemed likely that the other RF-like PPR proteins might also fulfill functions in mitochondrial mRNA processing. Therefore a CR-RT-PCR screening for abnormal mitochondrial RNA phenotypes of all protein coding genes was carried out. This analysis included twelve *A. thaliana* T-DNA insertion lines containing T-DNAs within RF-like PPR genes, including *rpf2-1,* defective in the generation of the *nad9* -202 RNA ends. (Table 1).

Lane	T-DNA insertion line	affected gene
1	SALK_061670	At1g62590
2	SALK_020638	At1g63080
3	SALK_146198	RPF2
4	SALK_139736	At1g62680
5	GABI_337F04	At3g22470
6	SALK_022265C	At1g64100
7	SALK_020230	At1g63130
8	FLAG_288A04	At5g41170
9	SAIL_1282C03	At1g63230
10	SAIL_18E04	At1g62930
11	SALK_128068	At1g63070
12	SALK_087235	At1g64580

Table 1: Overview of the twelve T-DNA insertion lines used in this study.

For most of the mature mitochondrial mRNAs no changes in the CR-RT-PCR product patterns have been observed in the mutant lines. Investigation of the transcript of *cox3,* however, revealed abnormal phenotypes in two lines: FLAG288A04 and SALK146198 (*rpf2-1*). In Col wildtype (WT) a single major *cox3* 5‘ end was mapped -380 nucleotides upstream of the *cox3* ATG represented by a CR-RT-PCR product of 230 bp. In FLAG line 288A04, which has a Wassilewskija (Ws) nuclear background, an abnormal product pattern was apparent (Figure x, lane 8). This pattern is similar to the one known from C24 with product sizes of 230 bp, 300bp and 390 bp corresponding to 5‘ ends at -380, -448 and -540 (Forner et al. 2005), raising the possibility that the pattern observed in FLAG288A04 is due to the Ws nuclear background and not the T-DNA insertion in

At5g41170. Indeed, identical *cox3* PCR products in Ws wildtype showed that the oberserved products are accession-specific and independant from the T-DNA insertion in the RF-like gene At5g41170.

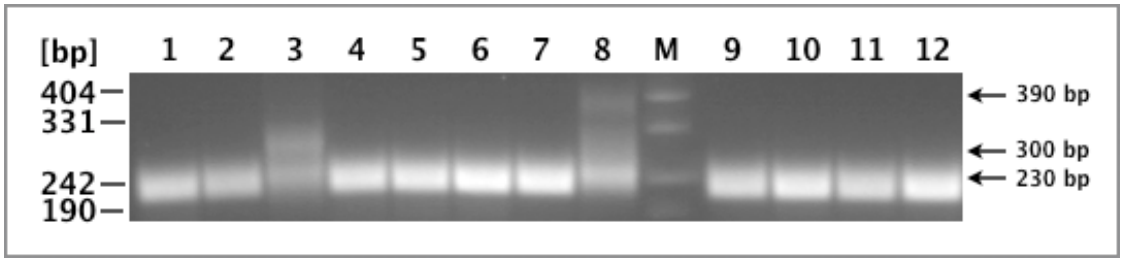


Figure 1: CR-RT-PCR analysis of the *cox3* mRNA extremeties in twelve T-DNA insertion lines with primers Atcox3-Mega5‘.nah and Atcox3-Mega3‘.nah. Numbering of the investigated lines corresponds to Table1. PCR products are separated alongside a marker (M). A single PCR product of 230 bp indicates a Col wildtype configuration. Additional PCR products of 300 bp in line 3 as well as 300 bp and 390 bp in line 8 contain additional *cox3 5‘* mRNA ends.

Surprisingly an additional product of about 300 bp was also visible in the *rpf2-1* mutant line (Figure 1, lane 3). Sequencing of this cDNA product revealed a *cox3* 5‘ end around position -448 suggesting a function of RPF2 in *cox3 5‘* mRNA end processing. The correlation of this phenotype with the knockout of RPF2 was confirmed by complementation of the mutant with the Col allele of At1g62670, which restored the mutant *cox3* mRNA phenotype to wildtype (Jonietz et al. 2010, Figure 4).

As C24 wildtype shows a similar *cox3* PCR product pattern as the one observed in the *rpf2-1* mutant the *cox3* mRNA phenotype of complemented C24 plants was also examined by CR-RT-PCR, primer extension and northern blot analyses. In the *rpf2-1* mutant, a complete restoration of the phenotype could be achieved while in the complemented C24 plants the additional larger mRNAs were still visible to some extent (Jonietz et al. 2010, Figure 4). This indicates that RPF2-mediated *cox3* mRNA processing is less efficient in C24. The reason for that can most likely be found in the differing *cox3* upstream regions in these accessions which might contain different *cis* elements (Forner et al. 2005). As stated in the introduction not much is known about the *cis* elements of mitochondrial mRNA end processing. No t-elements are present at the *nad9* and *cox3* mRNA ends. Consequently, RPF2 should recognize the target RNAs by other means, i.e. other secondary structures or even sequence elements with somewhat conserved primary structures. As a first step towards the identification and characterization of a potential RPF2 binding site, the RNA binding capacity of RPF2 should be investigated.

3.1.2 Overproduction of RPF2

PPR proteins have been predicted to bind RNA due to their superhelical structure (reviewed in Nakamura et al. 2004). RPF2 is a P class PPR protein of about 71kDa comprising 15 canonical PPR repeats. The secondary structure of RPF2 is predicted to be about 67 % α- helical (http://predictprotein.org) which is in agreement with the current model of PPR protein structure. To study the interaction of RPF2 with its potential target RNA *in vitro,* RPF2 was overexpressed in E.*coli*. As PPR proteins tend to accumulate in inclusion bodies with only very few recombinant protein remaining soluble overproduction was carried out under different temperatures to optimize expression conditions. Figure 2 shows an immunodetection analysis of recombinant RPF2 protein N-terminally fused with a Thioredoxin tag, a His-tag and an S-tag resulting in a protein with a molecular mass of about 87 kDa.

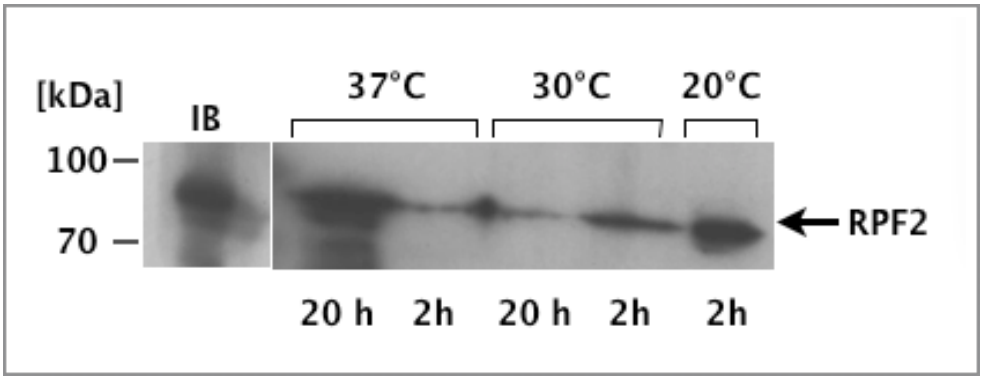


Figure 2: Immunodetection analysis of RPF2 in an inclusion body (IB) fraction and soluble protein fractions from overexpression of RPF2 in E.*coli* strain BL21AI at different temperatures. Overproduction was carried out for either 2h or 20h at different temperatures. Binding of S-protein cross-linked to horseradish peroxidase to the S-Tag of the recombinant protein leads to signals at around 90kDa, which is consistent with the calculated mass of recombinant RPF2. The yield of soluble protein was best when overproduction was performed at 37°C for 20h or 20°C for 2h.

Highest yield of soluble protein was obtained when the overexpression was carried out at either 20°C for two hours or at 37°C for 20 hours. For convenience the former condition was chosen for large scale overproduction of RPF2.

3.1.3 EMSA with recombinant RPF2

After removal of the tags by enterokinase and purification of the recombinant protein, electrophoretic mobility shift assays (EMSA) were performed to check the ability of recombinant RPF2 to bind its target RNA *in vitro* (Figure 3).

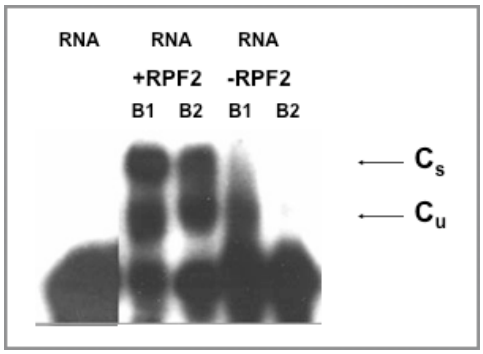


Figure 3: Electrophoretic mobility shift assay (EMSA) performed with recombinant RPF2 and an *in vitro* transcribed *nad9* RNA substrate (RNA). Recombinant RPF2 (+RPF2) and an *E.coli* control lysate (-RPF2) were added to synthetic *nad9* RNA spanning the region from -296 to -124 relative to the *nad9* ATG. This RNA comprises the 5‘ processing site at around -202. Two different buffer systems (B1, B2) were used. The formation of a specific binding complex (C_s) can only be observed when the RPF2 lysate is used while an unspecific binding complex (C_U) is also formed with the control lysate.

Therefore a synthetic *nad9* was transcribed from a PCR product *in vitro*. This RNA covers a region surrounding the *nad9* 5‘ processing site at -202. A lysate purified from *E.coli* containing the empty expression vector was used as a negative control. A specific protein-RNA complex could only be seen when the RPF2 lysate was added to the synthetic *nad9* RNA. When the *E.coli* lysate was used only an unspecific binding complex was visible. These results demonstrate that recombinant RPF2 is indeed able to bind to a *nad9* RNA *in vitro*.

3.2 Summary of the results obtained in the RPF2 analyses

RNA PROCESSING FACTOR 2 is a nuclear encoded mitochondrial protein with high similarity to RF proteins. It belongs to the P subclass of PPR proteins and is required for efficient generation of distinct *nad9* and *cox3* transcript ends in the accessions Col and C24. This has been demonstrated by complementation of the *rpf2-1* mutant and C24 with the intact allele of RPF2 from Col WT. It most likely fulfills these functions by binding to the target RNA and recruiting an endonuclease to the cleavage sites. The efficiency of this process is influenced by natural genetic variation of the potential *cis* elements. Investigation of other *Arabidopsis* accessions has shown that several inactive versions of

the *RPF2* gene exist that most likely originate from recombinations with other RF-like PPR genes (Supplemental Figure S1). Furthermore, a knockout of RPF2 does not have any severe physiological consequences. In the mutant, NAD9 and COX3 proteins accumulate to comparabale levels as in WT plants (Jonietz et al. 2010, Figure 5). Also the abundance of the most prominent mitochondrial proteins does not seem to be altered as indicated by results of 2D gel eletrophoresis experiments (Supplemental Figure S2).

3.3 Identification and characterization of RPF3

3.3.1 Screening of T-DNA insertion lines reveals an almost complete absence of mature *ccmC* transcript in mutant line SAIL18E04

Another abnormal CR-RT-PCR product pattern was found when the *ccmC* transcript was investigated in the mutant lines listed in Table 1. In Col WT a single major 5‘ and 3‘ end is represented by a CR-RT-PCR product of 650 bp. In two mutant lines (Figure 5, lanes 8 and 10) the product patterns differed from Col WT.

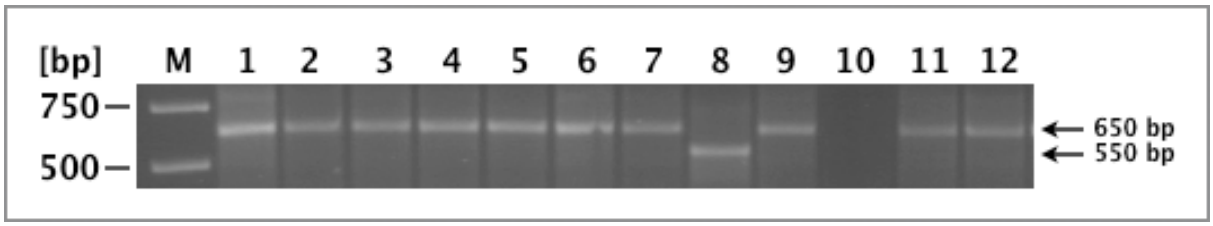


Figure 5: CR-RT-PCR analysis of the *ccmC* mRNA extremeties in twelve T-DNA insertion lines with primers Atccb3-Mega5‘.nah and Atccb3-Mega3‘.nah. Numbering of the lanes corresponds to lines listed in Table 1. PCR products are separated alongside a marker (M). A single PCR product of 650 bp indicates Col wildtype configuration of the *ccmC m*RNA. The PCR product of 550 bp in lane 8 represents a different *ccmC 5‘* mRNA end in Ws. In lane 10 no products were visible.

In the FLAG line 288A04 (lane 8) a CR-RT-PCR product of about 550 bp was amplified. Like in the case of *cox3* this deviation stems from natural genetic variation between Col and Ws and is not caused by the T-DNA insertion present in the gene At5g41170.

In lane 10 no CR-RT-PCR products were visible suggesting a participation of At1g62930 in the generation or stabilization of the *ccmC* mRNA. This mutant was reexamined by CR-RT-PCRs with different primers (Figure 6). First an alternative 3‘ UTR annealing primer was used. This PCR indicated that the mature *ccmC* transcript is almost completely lost in the mutant while longer PCR products accumulate to higher levels as in Col WT (Figure 6A). To check whether the larger products were amplified from *ccmC* precursor RNAs, a 5‘ primer annealing upstream of the main 5‘ end was used. Again a higher accumulation of CR-RT-PCR products was seen in the mutant (Figure 6B), further substantiating that higher amounts of *ccmC* precursor RNAs are present here. Indeed, sequencing of the 300

bp and 350 bp products confirmed both of them to be *ccmC*-specific and revealed a 5‘ end of a *ccmC* precursor RNA around -677 in the shorter product.

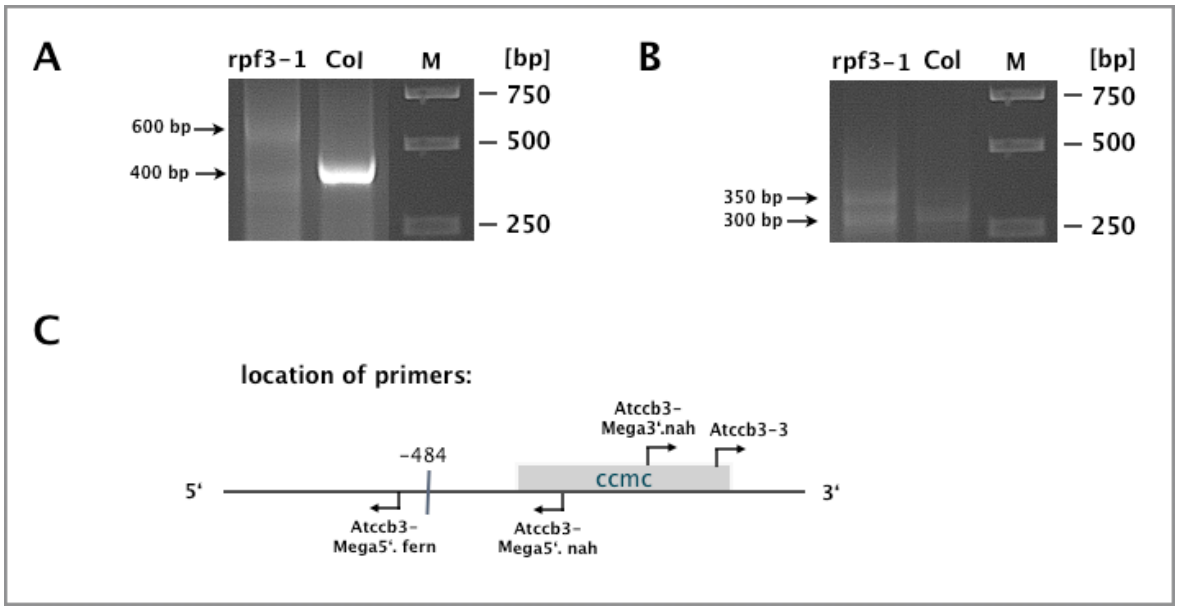


Figure 6: CR-RT-PCR analysis of *ccmC* transcript termini in the rpf3-1 mutant and Col WT. Amplification reactions were performed with different sets of primers. PCR products were separated alongside a marker (M) on agarose gels. A: The primers Atccb3-3 and Atccb3-Mega5‘.nah allow the amplification of cDNAs containing 5‘ ends of mature and precursor *ccmC* RNA. In Col, a strong product of around 400 bp corresponds to the mature *ccmC* transcript. In addition, a very weak product of around 600 bp is visible. The 400 bp product is almost completely absent in the *rpf3-1* mutant while more of the 600 bp product was amplified. B: Location of the primers Atccb3-3 and Atccb3-Mega5‘.fern excludes the mature *ccmC* transcript from amplification and only precursor molecules should be amplified. In the *rpf3-1* mutant PCR products of 300 bp and 350 bp accumulate to higher amounts than in Col WT. C: Schema showing the location of the primers.

Since the generation of the mature *ccmC* mRNA is almost completely abolished in the mutant RPF3 could either be a transcription initation factor or an RNA processing factor. A function in RNA stability can be excluded since it would not explain the concomitant increase in precursor molecules upon knockout of RPF3.

3.3.2 The function of RPF3: transcription initiation or 5‘ processing

Previous analyses revealed that several RNAs with 5‘ ends upstream of the *ccmC* CDS exist in *Arabidopsis thaliana* mitochondria. These 5‘ ends have been found to be surrounded by promoter-like sequences (Kühn et al. 2009). This is also the case for the mature *ccmC* transcript with 5‘ ends around -482. Here a CGTA motif is present a few nucleotides upstream of the mRNA ends. This opens up the possibility that these transcript ends originate directly from transcription initiation and that no further processing is needed. In this case, RPF3 would function as a transcription initiation factor rather than an RNA

processing factor. To tackle the question whether RPF3 is a transcription or a processing factor several independant approaches were carried out. Two of these, the northern blot analysis and the quantitative Real-Time RT-PCR are discussed in detail in the RPF3 manuscript, which can be found in the appendix. In both experiments putative *ccmC* precursor molecules were found to be enriched in the *rpf3-1* mutant in comparison to Col WT. However, those results are still insufficient to unambiguously exclude a function of RPF3 in transcription initiation. Another experiment, the primer extension analysis with terminator exonuclease-treated RNA will be described here.

3.3.2.1 Primer extension analysis with terminator exonuclease-treated RNA

To distinguish between possible functions of RPF3 mtRNA from Col WT was treated with terminator exonuclease. This enzyme degrades RNAs with monophosphate groups at their 5‘ ends, i.e. which derive from processing events. Molecules carrying triphophosphate groups at the 5‘ ends derive directly from transcription initiation and are resistent to digestion by the exonuclease. Since primary RNAs carry a triphosphate group and secondary RNAs carry a monophosphate group only the latter RNAs should be degraded. After treatment with the exonuclease the mtRNA was used as template for primer extension analysis (Figure 7).

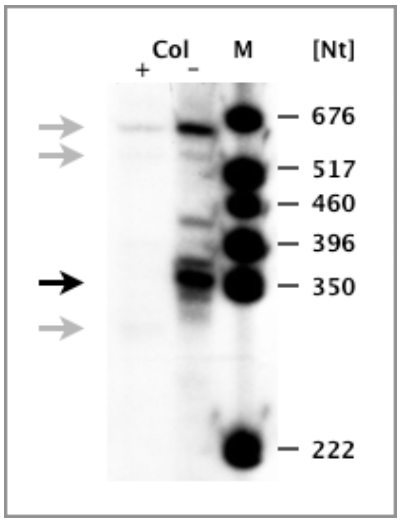


Figure 7: Primer extension analysis with terminator exonuclease-treated mtRNA. Extension products from untreated (-) and exonuclease treated (+) mtRNA separated alongside a marker (M) can be seen. The products corresponding to the -484/-482 ends with a size of around 370 Nt are marked by a black arrow. No product is visible after treatment with terminator exonuclease demonstrating that these ends are generated by posttranscriptional processing. Grey arrows mark products that are still visible after treatment indicating that those originate from primary transcripts.

Despite the presence of a potential promoter motif in direct vicinity of the -482/-484 5‘ ends (Kühn et al. 2009) this experiment demonstrates that these ends derive from processing and not from transcription initiation. The extension products of around 370 Nt are no longer generated after digestion with the exonuclease. This clearly speaks in favor of RPF3 functioning as an RNA processing factor. Apparently several longer and a shorter primary *ccmC* RNA exist (Figure 7, grey arrows) which is in agreement with the analyses of primary 5‘ ends in *Arabidopsis* mitochondria (Kühn et al. 2009).

3.3.2.2 The search for a *ccmC* leader molecule

As it is now established that the mature *ccmC* mRNA end derives from processing, information should be obtained about the mode of processing. In general there are two possibilities: 5‘ exonucleolytic processing or endonucleolytic cleavage. The former possibiity seems rather unlikely since no 5‘-3‘ exonucleolytic activity has been described in plant mitochondria yet. Accordingly, endonucleolytic processing is alot more probable. If a *ccmC* precursor RNA gets cleaved by an endonuclease two molecules will be left behind: the 5‘ leader and the mature mRNA (Figure 8A). Ideally the 3‘ terminal nucleotide of the 5‘ leader should map to a position one nucleotide upstream of the 5‘ end of the mature mRNA. Thus, the presence of such a 5‘ leader molecule would be indicative for the involvement of an endonuclease in the generation of the -482/-484 *ccmC* RNA 5‘ end.

3.3.2.2.1 Detection of the ccmC 5‘ leader molecule by CR-RT-PCR

As a first approach to detect a putative 5‘ leader and determine its 3‘ end, CR-RT-PCR was performed using mtRNA from Col WT and the rpf3-1 mutant. Since such a leader molecule might potentially be really short, self-ligation could be problematic. To cope with that the CR-RT-PCR strategy was slightly modified (Figure 8). Instead of detecting self-ligated molecules we wanted to find 5‘ leader fragments that have been ligated to 18S rRNA. This was chosen as an endogenous anchor molecule since it is very abundant in the organelle. Consequently an 18S rRNA specific primer was used to initiate cDNA synthesis.

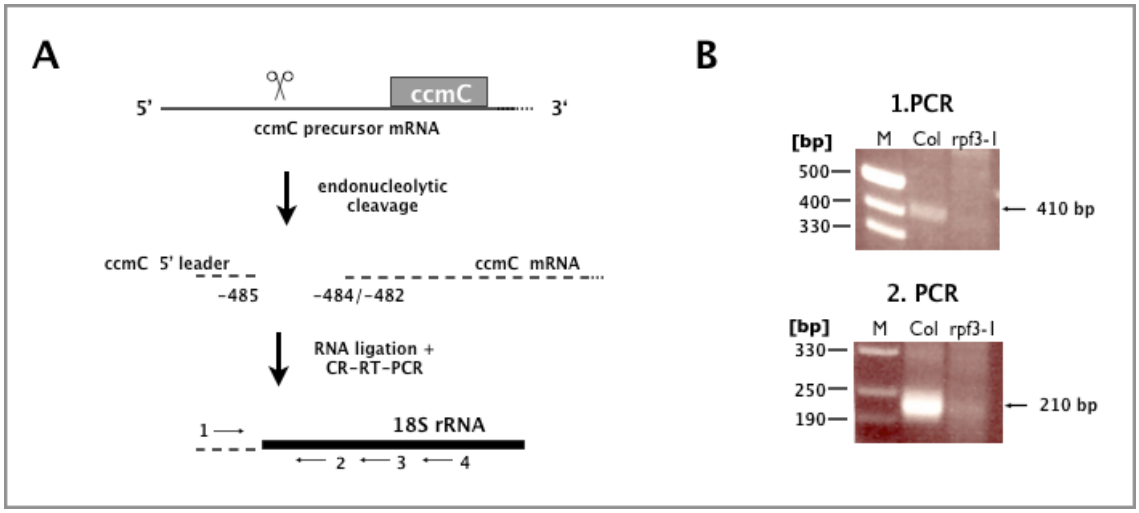


Figure 8: CR-RT-PCR amplification of the *ccmC* 5‘ leader molecules. A: Schema showing the endonucleolytic generation of the -482/-484 *ccmC* 5‘ ends resulting in a 5‘ leader and the mature mRNA and of the experimental strategy. The scissors symbolize an endonuclease. Primer 4 (At18S-Adapter.R1) was used for cDNA synthesis after RNA ligation. Primers 1 (Atccb3-12), 2 (At18S-Adapter.R4) and 3 (At18S-Adapter.R3) were used in subsequent CR-RT-PCR analyses. B: The 410 bp PCR product from the first CR-RT-PCR with primers 1 and 3 in Col has been used as a template in a half nested PCR with primers 1 and 2 leading to a PCR product of 210 bp in Col. In both PCRs cDNA from the *rpf3-1* mutant was used as a negative control since processing is almost completely abolished here. Accordingly, only very weak products have been amplified.

Subsequent PCR amplifications with 18S rRNA-specific primers and an oligonucleotide annealing to the putative *ccmC* leader fragment (Figure 8A) amplified PCR products of 410 bp and 210 bp in the first and second PCR, respectively. Since processing is severely impaired in the *rpf3-1* mutant only faint PCR products could be generated (Figure 8B). The 210 bp Col PCR product was further analyzed by direct sequence analysis or by sequencing of individual cloned products. The corresponding results are shown in Figure 9. The 3‘ ends of the detected leader molecules cluster a few nucleotides upstream of the -484/-482 ends in a region between -497 and -492 relative to the *ccmC* translation start codon. Only one 3‘ end of a single clone was found further upstream, at position -512.

3.3.2.2.2 Detection of the ccmC 5‘ leader by RT-PCR

The 3‘ ends of 5‘ leader molecules determined by CR-RT-PCR map a few nucleotides upstream of the expected position at -485/-483. It is possible that the 3‘ terminal nucleotides of the leader molecules are rapidly removed by one of the mitochondrial 3‘-5‘ exonucleases RNR1 and/or PNPase. To investigate the involvement of PNPase, cDNA from a respective knockdown mutant (PNP^-) (generously gifted by Dominique Gagliardi) was used for detection of the *ccmC* 5‘ leader fragment. So far, all substrates of PNPase were found to be polyadenylated. If the *ccmC* 5‘ leader gets degraded by PNPase it

should also be polyadenylated prior to degradation. Consequently, the cDNA from PNP- mutant was synthesized using an oligo-d(T) adapter primer. This oligonucleotide carries an adapter at its 5‘ end that can be used for subsequent PCR amplification. Two rounds of PCR were performed with one primer annealing to the putative *ccmC* leader fragment and an oligonucleotide annealing to the adapter region of the cDNA synthesis primer resulting in a weak product of the expected size (data not shown). Sequencing of this PCR product revealed a 3‘ end at position -495 relative to the *ccmC* ATG (Figure 9).



FIgure 9: Sequencing results from amplification of *ccmC* 5‘ leader molecules by CR-RT-PCR and RT-PCR. The nucleotide sequence surrounding the -484/-482 5‘ ends is shown. 3‘ ends of leader molecules either from bulk sequencing of the CR-RT-PCR product (black arrow below the sequence), the RT-PCR product (grey arrow, below the sequence) as well as 9 individual CR-RT-PCR clones (indicated by numbers) are shown. The 3‘ end of clone number 10 was detected further upstream at position -512 relative to the *ccmC* ATG (data not shown).

In summary, the experiments described in the last paragraphs support the hypothesis that the -484/-482 *ccmC* mRNA ends derive from endonucleolytic cleavage. No precursor molecules with 3‘ ends directly upstream of the 5‘ mRNA ends have been found. However, the existence of molecules with 3‘ ends very close to the expected site as well as the preferential amplification of these molecules in Col makes 5‘ exonucleolytic processing highly unlikely and rather suggests the involvement of endonucleolytic cleavage in the generation of mature *ccmC* 5‘ mRNA ends. Again the question arises how the cleavage site is determined.

3.3.3 Accession specific *cis* elements lead to the formation of different *ccmC* mRNA ends

At the 3‘ end of the *ccmC* mRNA a t-element is present but no obvious secondary structures could be detected at the 5‘ end (Forner et al. 2007). Analyses of the *ccmC* mRNA end phenotypes in reciprocal F1 hybrids of a crossing between Col and C24

showed that a mitochondrial *cis* element was responsible for the different phenotypes that have been observed (Forner et al. 2008). To get more insight into the nature of this *cis* element the DNA sequences surrounding the Col and C24 5‘ mRNA ends were compared (Figure 10). The nucleotide sequence was found to be almost identical apart from a stretch of nucleotides ranging from positions -578 to -513 upstream of the *ccmC* translation start codon. Here sequence identity only was 63.6% suggesting that the nucleotide exchanges present in this region are responsible for the distinct accession specific transcript termini.

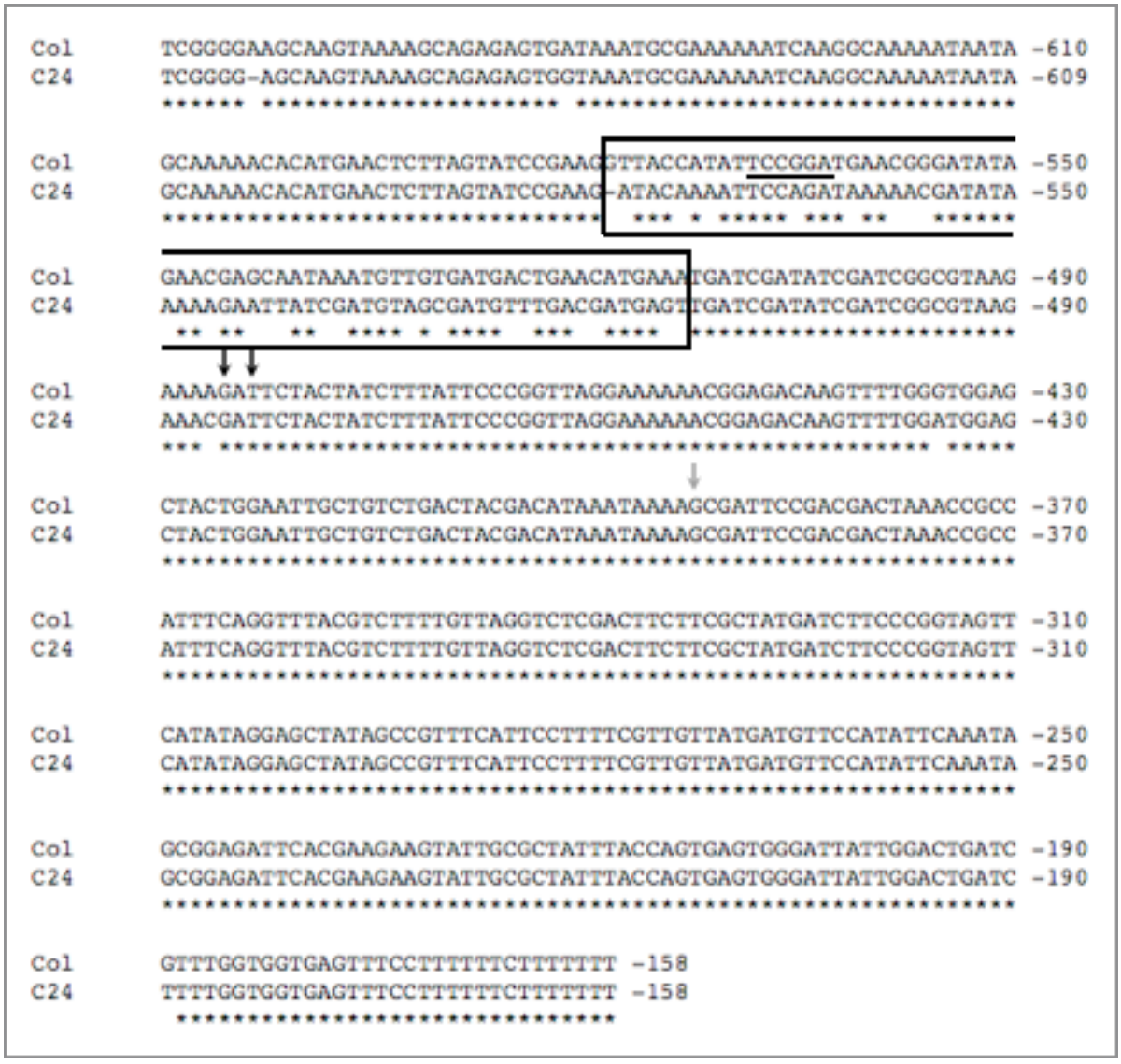


Figure 10: Clustal W allignment of the mitochondrial *CcmC* upstream sequences from 240952 to 241419 (-670 to -158 in relation to ATG) in Col and C24. Stars mark sequence identity between Col and C24. The Kpn21 restriction site in Col is underlined. The -390 mRNA end in C24 (grey arrow) and the -484/-482 mRNA ends in Col (black arrows) are shown.

To support the potential correlation of the mitochondrial differences and the *ccmC* 5‘ end polymorphism the corresponding mtDNA region was analyzed in 26 accessions (see

Materials) making use of a *Kpn21* restriction site polymorphism (Figure 11) and compared to the *ccmC* transcript phenotypes that were determined by CR-RT-PCR (Figure 12).

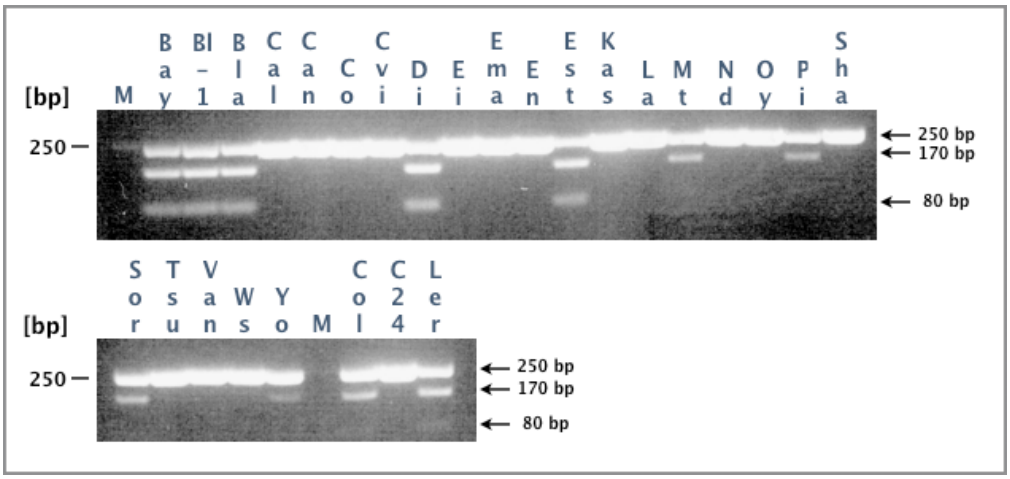


Figure 11: Mitochondrial *ccmC* upstream DNA configurations in different accessions.
Restriction analysis of a PCR product generated by PCR amplification of the region from -488 to -734 in relation to the *ccmC* translation start codon with the primers Atccb3-Mega5‘.fern and Atccb3-12. Digestion with *Kpn21* leads to fragments of 170 bp and 80 bp after digestion indicating a Col-like mtDNA configuration An undigested fragment of 250 bp stands for a C24-like mtDNA configuration. M: DNA Marker. The accession No exhibited a C24 like mtDNA configuration (data not shown here).

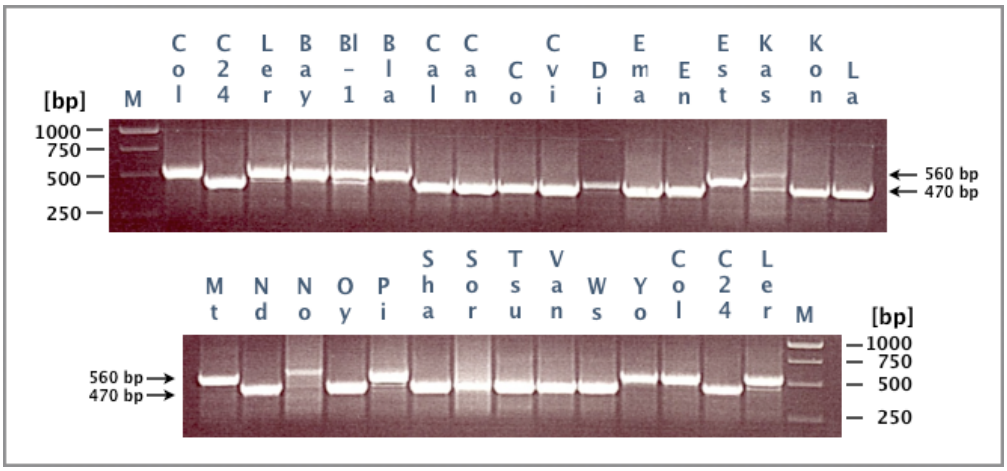


Figure 12: ccmC 5‘ ends in different accessions
CR-RT-PCR analysis of *ccmC* mRNA extremities separated alongside a DNA marker (M). PCR products were generated with primers Atccb3-Mega5‘.nah and Atccb3-Mega3‘.nah. Eight other accessions show a Col-like mRNA end phenotype while 16 other accessions show a C24-like mRNA end phenotype. In 2 accessions (Kas, No) an mRNA end phenotype deviating from Col and C24 was found.

A comparison of the observed genotypes and phenotypes (Table 2) shows a strong correlation of the *Kpn21* restriction site polymorphism and the transcript termini as detected by CR-RT-PCR analysis in all but 3 accessions. In Sorbo digestion revealed a Col-like mtDNA configuration but nevertheless a phenotype resembling the one from C24 was observed.

Accession	observed ccmC mRNA phenotype	ccmC upstream genotype as indicated by Kpn21 digestion
Col	Col	Col
C24	C24	C24
Ler	Col	Col
Bay-0	Col	Col
Bl-1	Col	Col
Bla-10	Col	Col
Cal	C24	C24
Can-0	C24	C24
Co	C24	C24
Cvi	C24	C24
Dijon	C24	Col
Ei	n.d.	C24
Ema-1	C24	C24
En-1	C24	C24
Est	Col	Col
Kas-1	different	C24
La-1	C24	C24
Mt	Col	Col
Nd	C24	C24
No-0	different	C24
Oy-1	C24	C24
Pi-2	Col	Col
Shakdara	C24	C24
Sorbo	C24	Col
Tsu-1	C24	C24
Van-0	C24	C24
Ws	C24	C24
Yo-0	Col	Col

Table 2: Correlation of *ccmC* mRNA phenotypes determined by CR-RT-PCR and *CcmC* upstream mtDNA configurations (as indicated by a *Kpn21* restriction polymorphism) in different accessions.

In the accessions No-0 and Kas-1 CR-RT-PCR products deviated from those found in Col, C24 and all other accessions while the *ccmC* upstream mtDNA configuration of No-0 and Kas-1 was that of C24. In No-0, weak products between 250 bp and 650 bp were amplified while in Kas-1 two weak products of around 480 bp and 650 bp were generated. Sequencing of the 650 bp PCR products revealed a 5‘ end for the No-0 and Kas-1 mRNA at around position -542 relative to the ATG (Figure 13) whereas no clear 5‘ mRNA was discernable after direct sequencing of the 480 bp product in Kas-1 (data not shown).

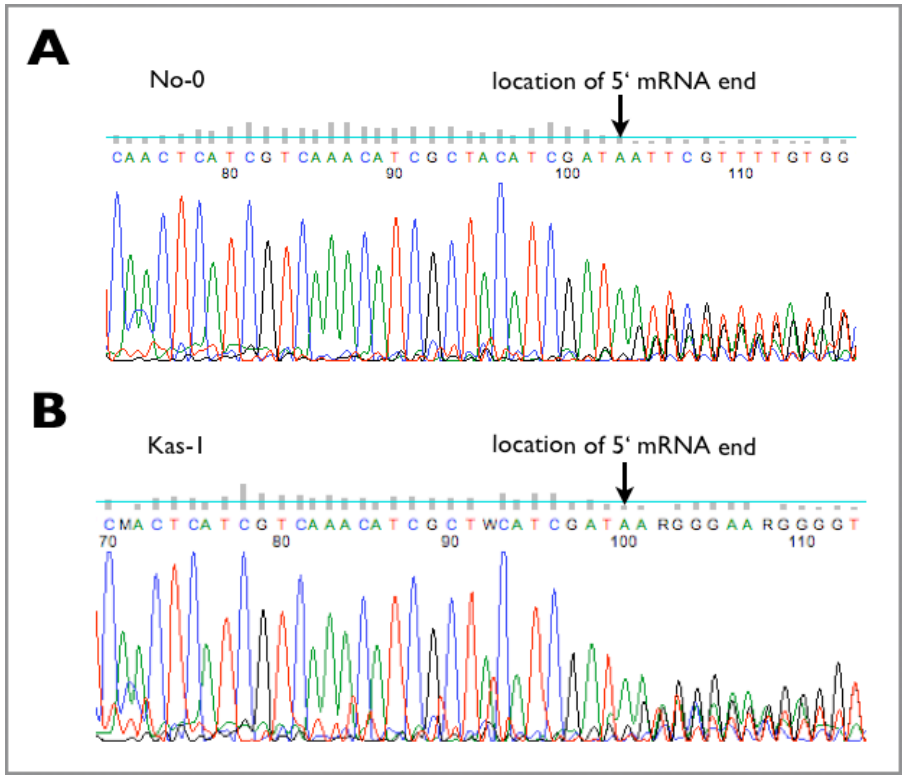


Figure 13: Chromatograms from sequencing of CR-RT-PCR products. A The 650 bp PCR product amplified with primers Atcbb3-Mega3'.nah and Atccb3-Mega5'.nah in No-0 contains a 5' mRNA end at position -542 in relation to the *ccmC* ATG. B The 650 bp PCR product amplified with primers Atcbb3-Mega3'.nah and Atccb3-Mega5'.nah in Kas-1 indicates the same 5' mRNA end.

To obtain detailed sequence information on the *ccmC* upstream configurations and to substantiate the correlation of the *Kpn21* restriction site polymorphism and the *ccmC* transcript phenotypes the corresponding mtDNA regions were sequenced from eight accessions (Figure 14). Bay-0, Est, L*er* and Yo-0 (group 1) represent the Col-like *Kpn21* restriction pattern while No-0, Kas-1, Shakdara, Sorbo and Wassilewsija (group 2) group with the C24-like restriction pattern (Table 2). Figure 14 shows the putative *ccmC cis*

```
Col         GTTACCATATTCCGGATGAACGGGATATAGAACGAGCAATAAATGTTGTGATGACTGAACATGAAA -513
Ler         GTTACCATATTCCGGATGAACGGGATATAGAACGAGCAATAAATGTTGTGATGACTGAACATGAAA -513
Yo-0        GTTACCATATTCCGGATGAACGGGATATAGAACGAGCAATAAATGTTGTGATGACTGAACATGAAA -513
Bay-0       GTTACCATATTCCGGATGAACGGGATATAGAACGAGCAATAAATGTTGTGATGACTGAACATGAAA -513
Est         GTTACCATATTCCGGATGAAGGGGATATAGAACGAGCAATAAATGTTGTGATGACTGAACATGAAA -513
Sorbo       GTTACCATATTCCGGATGAACGGGATATAGAACGAGCAATAAATGTTGTGATGACTGAACATGAAA -513
C24         -ATACAAAATTCCAGATAAAAACGATATAAAAAGAATTATCGATGTAGCGATGTTTGACGATGAGT -513
No-0        GATACAAAATTCCAGATAAAAACGATATAAAAAGAATTATCGATGTAGCGATGTTTGACGATGAGT -513
Kas-1       GATACAAAATTCCAGATAAAAACGATATAAAAAGAATTATCGATGTAGCGATGTTTGACGATGAGT -513
Shakdara    GATACAAAATTCCAGATAAAAACGATATAAAAAGAATTATCGATGTAGCGATGTTTGACGATGAGT -513
Ws          GATACAAAATTCCAGATAAAAACGATATAAAAAGAATTATCGATGTAGCGATGTTTGACGATGAGT -513
```

Figure 14: Clustal W allignment of the mitochondrial *CcmC* upstream sequences from 241334 to 241270 (-578 to -513 in relation to ATG) from Col, L*er*, Yo-0, Bay-0, Est, Sorbo, C24, No-0, Kas-1, Shakdara and Ws. The *Kpn21* restriction site in Col is underlined. Sequences from accessions with a Col-like mtDNA configuration (group 1) are in black while sequences from accessions with a C24-like mtDNA configuration (group 2) are in blue. Mismatches within members of a group are in red.

element from those accessions and in Supplemental Figure S4 the whole alignments of the -670 to -158 region of groups 1 and 2 can be seen.

Apart from one SNP in the accessions Est at position -557 and a single nucleotide deletion in C24 at position -578 the upstream mtDNA sequences in the -578 to -513 region are 100% conserved in all the members of group 1 and group 2 respectively further supporting the assumption that this region contains a *cis* element involved in *ccmC* mRNA processing. Surprisingly though, the corresponding mtDNA sequence in Sorbo, which shows a *ccmC* mRNA phenotype similar to C24, is completely identical to the Col sequence. For that reason the Sorbo *ccmC* mRNA phenotype was reinvestigated by CR-RT-PCR with two sets of primers (Figure 15).

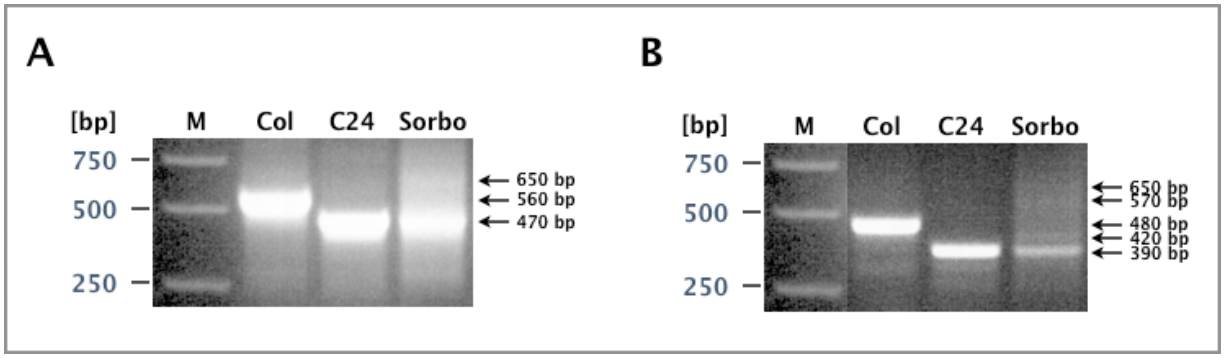


Figure 15: ccmC 5‘ ends in Col, C24 and Sorbo
CR-RT-PCR analysis of *ccmC* mRNA extremities separated on an agarose gel alongside a DNA marker (M). A: PCR products were generated with primers Atccb3-Mega5‘.nah and Atccb3-Mega3‘.nah. In Col and C24 560 and 470 bp products correspond to mature *ccmC* transcripts. In Sorbo a product pattern slightly differing from C24 was generated. Here a major product of 470 bp and a weak product of 650 bp were amplified.
B: PCR was performed with primers Atccb3-Mega5‘.nah and Atccb3-3. In Col and C24 the mature *ccmC* transcripts are represented by cDNA products of 490 and 400 bp. In Sorbo weak products of 390, 420, 570 and 650 bp are visible. For the location of the primers see Figure 6C.

This analysis clearly shows that the C24 and Sorbo *ccmC* phenotypes are similar but not identical. Although the major products seem to have identical sizes in C24 and Sorbo, additional larger products were repeatedly amplified in Sorbo. These results indicate that processing is somewhat impaired in this accession. Sorbo seems to have a different setup of nuclear processing factors taking part in the generation of the *ccmC* transcript since the mtDNA configuration is the same as in Col yet still different 5‘ ends are formed.

3.3.4 Introduction of Col RPF3 does not restore the *ccmC* phenotype in No-0

The *ccmC* mRNA phenotypes observed in Kas-1 and No-0 are similar to the one found in the *rpf3-1* mutant. Hardly any mature *ccmC* transcript is detectable by CR-RT-PCR or

Northern blot analysis (data not shown). This leads to the assumption that RPF3 might be defective in these accessions. To test this hypothesis and to check whether the Col allele does have any effect in the generation of mature *ccmC* mRNA in No-0 where a C24 mtDNA configuration is present the RPF3 allele from Col was introduced into No-0. The same construct has been used for complementation of the *rpf3-1* mutant (Figure 16, Materials & Methods RPF3 manuscript).

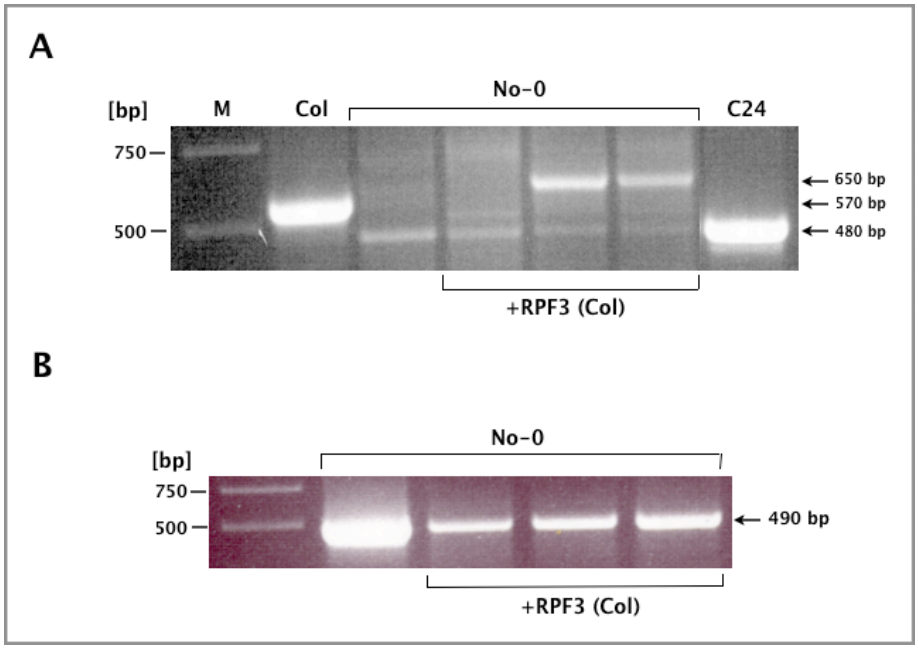


Figure 16: CR-RT-PCR analysis of *ccmC* mRNA extremities. The PCR products were separated alongside a DNA marker (M). A: PCR products were generated with primers Atccb3-Mega5‘.nah and Atccb3-Mega3‘.nah. In Col and C24 distinct products of 570 bp and 480 bp indicate the presence of mature *ccmC* mRNAs. Hardly any difference is visible between products amplified from No-0 and complemented No-0 plants (+RPF3 (Col)). In both cases only weak products could be generated (also compare to Figure 12).
B: To test the integrity of the No-0 cDNAs a control PCR was performed with primers Atatp9-Mega.3‘ and Atatp9-Mega.5‘ to amplify an atp9 cDNA of 490 bp.

After the introduction of the gene still only very weak CR-RT-PCR products could be amplified in No-0 WT and different transformants. It is also notable that there is some variability in the CR-RT-PCR product patterns generated in different experiments. Both No-0 WT plants and the transformed No-0 plants sometimes exhibit a more abundant product of around 650 bp and sometimes they do not (Figure 12 and Figure 16). What is more important though, is that neither the Col-like *ccmC* transcript (560 bp) with 5‘ ends around -484 nor the C24-like *ccmC* mRNA (470 bp) with 5‘ ends around -391 accumulates to a level comparable to Col or C24. The latter transcript seems to be present in trace amounts though. This indicates that the introduction of the Col allele of RPF3 into No-0 is not sufficient to enhance the generation of mature *ccmC* transcripts. Again the reason for this might be the presence of the C24-like mtDNA configuration in No-0. Additionally, in a

northern blot analysis of C24 (♀) x *rpf3-1 (♂)* F_2 hybrids lacking a functional RPF3 the mature *ccmC* mRNA was still detectable in considerable amounts (Figure 17). Since the mitochondrial genome is maternally inherited these hybrids possess a C24 mtDNA configuration, a chimeric C24/Col nuclear genome and a homozygous *rpf3-1* T-DNA allele. As the mature *ccmC* transcript is still detectable, it seems that *ccmC* transcript processing is not as dependant on RPF3 when a C24-like mtDNA configuration is present.

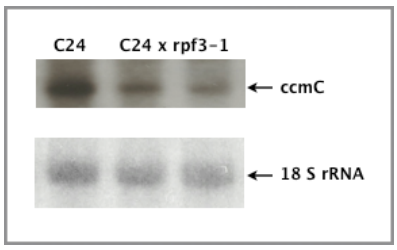


Figure 17: Northern blot analysis of gRNAs from C24 WT and two C24 x *rpf3-1* F_2 hybrids. A probe spanning the *ccmC* CDS from 208 to 607 with respect to the ATG still detects mature *ccmC* RNA in the hybrid plants. Loading of the gel was checked by hybridization with a 18S rRNA specific probe.

As a first step to getting more insight into the interaction of the different RPF3 proteins with the accesssion specific *cis* elements they should be overexpressed in *E.coli.*

3.3.5 Overproduction of RPF3 in E.*coli*

As described in the previous section at least two different *ccmC* upstream configurations are present in different accessions. In addition there seem to be accession-specific functions of RPF3. These specific functions might be attibuted to distinct interactions between the variants of RPF3 and the *ccmC* upstream region. To test the binding properties of the RPF3 protein from Col it should be overexpressed in *E.coli.* Col RPF3 is predicted to contain 67.4% α-helical secondary structures (predictprotein.org) which is in agreement with the current model of PPR protein structure and supports its role as an RNA binding protein. An EMSA using recombinant protein and *in vitro* transcribed *ccmC* RNAs from Col-0 and C24 could reveal further information on the *cis* elements required for the formation of mitochondrial *ccmC* mRNA termini. To this end overproduction of recombinant RPF3 protein was performed in *E.coli*. To optimize the conditions overproduction was carried out at 4°C, 20°C, 30°C and 37°C. The amount of soluble protein was then checked by an immunodetection analysis (Figure 18A) with a Coomassie stained gel serving as a loading control (Figure 18B). The maximum amount of soluble RPF3 protein was obtained when overexpression was carried out at 30°C.

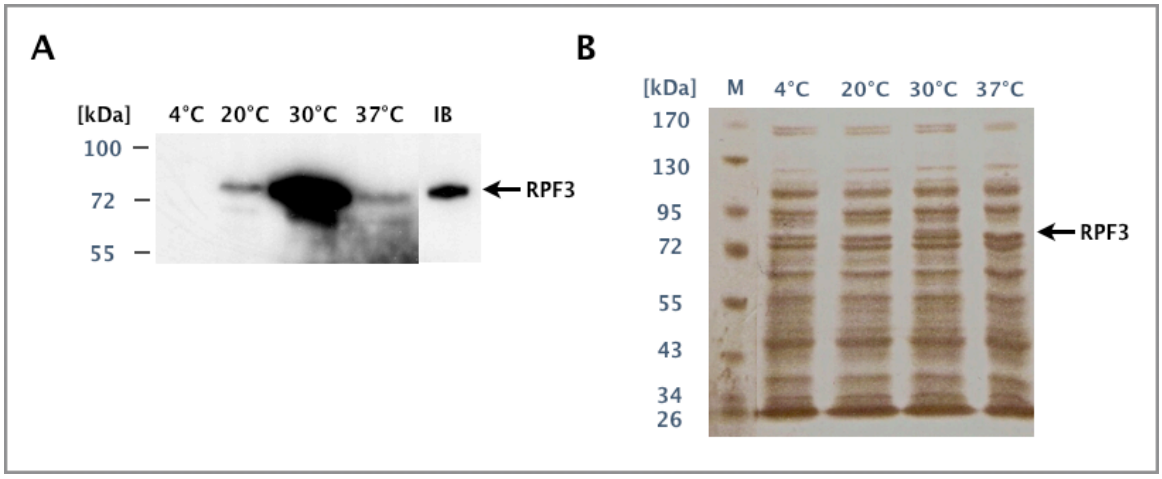


Figure 18: Overproduction of RPF3 in E.*coli* strain BL21AI. A: Immunodetection analysis of recombinant RPF3 in an inclusion body fraction (IB) and soluble fractions from E.*coli* strain BL21AI incubated at different temperatures. Overexpression was carried out for 3h. Binding of S-protein to the S-Tag of the recombinant protein leads to signals at around 80 kDa which is in agreement with the calculated mass (86kDa).
B: Loading of the gel was checked by Coomassie blue staining.

4. Discussion

4.1 RF-like PPR proteins in *Arabidopsis thaliana*

Over the last years it has become evident that PPR proteins play an important role in mitochondrial gene expression and biogenesis. They take part in virtually every step of the organellar RNA metabolism. Apart from the mitochondrial splicing factor OTP43 (de Longevialle et al. 2007) and several mitochondrial editing factors (Zehrmann et al. 2009; Verbitskiy et al. 2010; Takenaka et al. 2010; Bentolila et al. 2010) all of the characterized mitochondrial PPR proteins encode for RESTORERS OF FERTILITY (Bentolila et al. 2002; Brown et al. 2003; Desloire et al. 2003; Koizuka et al. 2003; Hanson and Bentolila. 2004; Wang et al. 2006; Chase. 2007). Those are important in the CMS restoration system. CMS is a maternally inherited trait where plants are no longer able to produce or release fertile pollen. It is caused by a chimeric mitochondrial peptide and can be restored by RF proteins which lead to posttranscriptional reduction of the CMS-associated transcripts. For most of the RFs it is presently unclear how this decrease is achieved in particular. The sterility-fertility systems are of great commercial interest as they can be used in agriculture to exploit the heterosis effect (Laser and Lersten 1972). They also provide an interesting basis for the research of nuclear-cytoplasmic interactions. In the autogamous plant species *Arabidopsis thaliana* CMS has not been reported yet. Surprisingly a comparison of the *RfO* from radish (Desloire et al. 2003) and the petunia RF (Bentolila et al. 2002) with protein databases revealed that about 20 proteins with high similarity to these RF proteins exist in *Arabidopsis* (Desloire et al. 2003). In this model species RF-like PPR genes are located in two clusters around 4.3 megabases and 23 megabases on chromosome 1 and all of them belong to the P class of PPR proteins. When compared to other P class PPR proteins outside of the two clusters it becomes even more evident that the RF-like PPR genes form a subgroup of their own within P class PPR proteins. On the amino acid level RF-like PPR proteins share about 50% identity and 70% similarity with each other while they are only about 25% identical and 50% similar to other P class PPR proteins. Thus, it is probable that the RF-like and the true RF genes share a common progenitor and evolved separately from the other PPR genes in *Arabidopsis* (Geddy and Brown 2007). In recent years several RF genes have been mapped. However, the exact mechanisms of fertility restoration are often unclear. Among the better characterized RF systems is the Boro II CMS from rice (*Oryza sativa*). Here RF1A induces endonucleolytic cleavage of the dicistronic CMS-related transcript while RF1B seems to

enhance its degradation by an as of yet unknown process (Wang et al., 2006; Kazama et al., 2008). The here presented analyses now assigned functions to two RF-like genes, *RPF2* and *RPF3*, in *Arabidopsis*. Both of them take part in the mitochondrial RNA metabolism. RPF2 and RPF3 are necessary for efficient generation of mature mRNAs of the nad9 and cox3 as well as of the ccmC gene. This has been unambiguously demonstrated by T-DNA insertion lines that showed defects in the formation of the *nad9* and *cox3* and *ccmC* transcripts and the restoration of the mutant phenotypes by complementation (Jonietz et al. 2010, Figure 3 and 4, RPF3 manuscript Figure 3). Although the mechanistic aspect of RPF2 and RPF3 function still needs further experimental clarification, it seems plausible that both factors mediate endonucleolytic processing of their respective RNA targets, similar to the above mentioned RF1A from rice. This assumption is supported by several observations. First, the knockout of RPF2 leads to an increase in *cox3* precursor RNAs while less of the mature transcripts are detectable (Figure 1) Second, an analogous observation was made in the *rpf3-1* knockout plants in which the mature *ccmC* mRNA is barely detectable while precursor RNAs accumulate to higher levels (Figure 6). Third, the affected *nad9* and *ccmC* RNAs derive from processing events. This has been shown by primer extension analysis of RNA treated with terminator exonuclease (Supplemental Figure S3; Figure 7). Fourth, the preferential amplification of a *ccmC* 5‘ leader molecules in Col WT shows that RPF3 is needed for endonucleolytic cleavage of the *ccmC* precursor molecules (Figure 8). Taken together these results make a function of RPF2 and RPF3 as either transcription factors or stability factors highly unlikely. Consequently, these results strongly suggest that RPF2 and RPF3 are RNA processing factors needed for endonucleolytic cleavage of their respective RNA targets. So apart from their striking sequence similarities some of the RF and RF-like proteins also have analogous functions.

Apart from RPF2 and RPF3 another factor taking part in mitochondrial mRNA processing, termed RPF1, has been identified (Hoelzle et al. 2010). RPF1 is necessary to generate mature *nad4* transcript and is also encoded in one of the clusters of RF-like PPR genes on chromosome 1. Since all RF-like proteins characterized so far are involved in mRNA 5‘ processing it is tempting to speculate that most, if not all of these PPR proteins take part in mRNA processing in *Arabidopsis*.

However, this function does not seem to be restricted to RF-homologous PPR proteins. Recently a non RF-like P class PPR protein that has been termed RPF5 has been identified. This factor is required for efficient generation of distinct mRNA ends of the *nad6, atp9* and *26S rRNA* transcripts (Hauler 2010).

4.2 The physiological relevance of 5‘ mRNA processing

In many cases the knockout of genes taking part in the mitochondrial RNA metabolism has severe detrimental effects on plant development and plant fitness. Embryos of homozygous T-DNA insertion lines with insertions in PPR genes often abort at early developmental stages (Tzafrir et al. 2003, Lurin et al. 2004). Other phenotypes include male sterility (Budar et al. 2003) or deeply retarded growth like dwarfism (Lurin et al. 2004). Concerning processing of mitochondrial RNAs, knockout of the mitochondrial 3‘-5‘ exonuclease PNPase leads to embryolethality (Holec et al. 2006) and loss of RNR1 function results in severely impeded growth (Bollenbach et al. 2005). In this context it is notable that no 3‘ end mRNA polymorphisms have been found yet (Forner et al. 2008). So either those are very rare compared to the 5‘ end polymorphisms or would have fatal consequences for the plant. These findings clearly support the importance of mitochondrial RNA 3‘ processing. What about RNA 5‘ processing though? The *rpf2-1* and *rpf3-1* mutants do not exhibit any macroscopic phenotypes under normal growth conditions and show normal fertility. Thus knockout of single 5‘ processing factors and the consequent loss of mRNAs with distinct 5‘ ends does not seem to have detrimental effects on plant fitness. This observation might be explained with the complexity of gene expression in *Arabidopsis* mitochondria, especially when transcription and 5‘ RNA processing are concerned.

Mitochondrial genes are transcribed from multiple promoters which are sometimes found several kilobases upstream of translation start codons (Kühn et al. 2005, 2009). This might be due to the recombinatorial nature of the mitochondrial genome (Shedge et al. 2007; Kubo and Newton 2008) where frequent recombinations could lead to novel and sometimes very long precursor RNAs. 5‘ RNA processing might be neccessary to shorten those long precursors to ensure efficient translation. Hereby a wide range in the lengths of the 5‘ UTRs seems to be tolerated by mitochondrial ribosomes. Mature mitochondrial mRNAs exhibit 5‘ UTRs between 0 and 450 nucleotides in length. Often several 5‘ ends in close vicinity to each other are generated per transcript (Forner et al. 2007) and consequently the loss of one of these transcripts does not have grave consequences. An example illustrating this is the expression of the *nad9* transcript. In Col two 5‘ ends around -243 and -202 are generated while in C24 a genome recombination within the RPF2 reading frame leads to an almost complete loss of the -202 transcript (Jonietz et al. 2010, Figure 1). However, since the -243 transcript is still present in C24 this does not lead to any physiological consequences for the plant. Another example is the expression of the *ccmC* transcript. Several transcription initiation sites have been identified upstream of the

ccmC gene ranging from -1834 to -45 in relation to the ATG (Kühn et al. 2009). Consequently, multiple precursor transcripts should exist for this transcript. Indeed, in our own Northern blot and primer extension analyses *ccmC* precursor RNAs of various sizes have been detected (RPF3 manuscript Figure 1, Figure 7). Upon knockout of RPF3 only trace amounts of mature *ccmC* transcript are detectable anymore while precursor transcripts accumulate to a higher level (Figure 6, RPF3 manuscript Figure 1). Accordingly, hardly any ccmC protein could be found in immunodetection experiments (RPF3 manuscript Figure 5). These results suggest that unprocessed precursor RNAs might indeed only be translated very inefficiently or not at all. So it is possible that 5‘ processing is needed to counterbalance the frequent genome recombinations and relaxed transcription found in *Arabidopsis* mitochondria by generating translatable mRNAs.

In this respect it is quite surprising that the loss of mature *ccmC* transcript in the *rpf3-1* mutant remains without consequences for the plant. In *Arabidopsis* a model for cytochrome c maturation similar to the one from *E.coli* has been proposed (Kranz et al. 2009). Here CcmC along with other cytochrome maturation proteins is required to deliver heme from the cytoplasm into the periplasm and is thus needed for the formation of holocytochrome c. In *Arabidopsis* it is assumed that heme gets translocated from the mitochondrial matrix into the inter membrane space by an analogous mechanism to form holocytochrome c (Giegé et al. 2008). As holocytochrome c is needed to shuttle electrons from the cytochrome c reductase (complex III) to the cytochrome c oxidase (complex IV) of the respiratory chain a severe phenotype would have been expected for the *rpf3-1* mutant. Since no growth phenotype is visible either CcmC is not required for the biogenesis of cytochrome c in *Arabidopsis* or the residual amount of CcmC left in the *rpf3-1* mutant is still enough to fulfill the function.

4.3 *Cis* elements of mitochondrial mRNA end processing in *Arabidopsis*

Little is known about the *cis* elements involved in mitochondrial mRNA processing in plants. In the cases of the $ccmF_c$ and *rps3* transcripts tRNA genes have been found located upstream of the coding sequences (Forner et al. 2007). This results in the precursor RNAs forming characteristic secondary structures that get recognized by RNase Z which leads to cleavage and thus defines the 5‘ RNA ends of these transcripts. Similarly, reduced versions of tRNA genes, so called t-elements, where only the sequence of the pseudo-uridine arm of the tRNA is retained, also get recognized by RNase Z and PRORP1 and serve as *cis* elements in the formation of both 5‘ and 3‘ RNA ends (Forner et al. 2007;

Gobert et al. 2009). However, for most of the mitochondrial genes no such secondary structures have been found. So, either very complex higher order structures are formed here or primary sequence elements aid in the recognition of processing sites. As almost no consensus sequences near the remaining known 5‘ or 3‘ mRNA ends are discernable it might be possible that most processing sites are defined by individual primary sequence elements which get recognized by individual processing factors.

Cis elements like that were observed in the RNA metabolism of chloroplasts of different species. In maize, sequence specific binding of PPR5 serves to stabilize the *trnG*-UCC precursor RNA by blocking the access to an endonuclease sensitive site (Beick et al. 2008). Another maize protein named PPR10 also binds a *cis* element defined by primary sequence, thereby determines the 5‘ and 3‘ ends of *atpH* and *psaJ* mRNAs and stabilizes them by acting as a border to exonucleolytic attack (Pfalz et al. 2009). In the green alga *Chlamydomonas reinhardtii* the chloroplast *petA* and *rbcL* transcripts are stabilized through binding of the PPR proteins MCA1 and MRL1 to the 5‘ UTR of the RNA (Loiselay et al. 2008, Johnson et al. 2010).

The only mitochondrial genes that share a common primary sequence at the 5‘ mRNA ends are *atp9, nad6* and the 26S rRNA gene. Here a stretch of 24 nucleotides directly adjacent to the 5‘ ends is found conserved. Interestingly one nuclear *trans* factor, the above mentioned RPF5, is mediating endonucleolytic processing of all of these RNA ends (Hauler 2010). It seems feasible that this 24 nucleotide primary sequence acts as a *cis* element that gets recognized by RPF5 which in turn recruits the processing enzyme. This could either be the tRNA processing enzymes RNase Z and PRORP1 or an as of yet unidentified endonuclease. Thus, while primary structure can also act as a *cis* element in mitochondrial mRNA end generation the mechanisms by which the ends are generated seem to be different in both organelles. In chloroplasts exonucleases seem to be responsible for shortening of precursor RNAs generating the 5‘ and 3‘ ends of mature mRNAs in the process. In mitochondria, where 5‘-3‘ exonuclease activity has not been described yet, the 5‘ ends of transcripts seem to be generated by endonucleases alone while 3‘ ends might be defined by both endo- and exonucleases.

For the remaining mitochondrial processing factors characterized so far the situation is less clear. In the case of RPF3 and the *ccmC* transcript processing a ~65 nucleotide stretch about 60 nucleotides upstream of the mature 5‘ ends is present in two different configurations (Figure 10). These configurations strongly correlate with distinct mRNA 5‘ ends being generated in different *Arabidopsis* accessions (Figure 11 and Figure 12)

making it very probable that the ~65 nucleotide region contains a *cis* element. It is unclear however, if the primary structure alone is sufficient for binding of RPF3 or if a secondary structure is formed. Even less is known about the recognition sites of RPF1 and RPF2. RPF1 only seems to be neccessary for processing of one 5‘ end of the *nad4* transcript (Hölzle et al. 2010). Also no binding of recombinant RPF1 to a synthetic *nad4* RNA was observed *in vitro* making it impossible to discern a binding site for this RNA processing factor (Hölzle et al. 2010). RPF2 is needed for processing of *nad9* and *cox3* mRNAs and has been shown to bind a synthetic *nad9* RNA *in vitro* (Jonietz et al. 2010). However, the regions upstream of the *nad9* and *cox3* 5‘ ends do not show any similarities which speaks in favor of RPF2 recognizing a certain secondary structure rather than primary structure.
Hence, both primary and secondary structures seem to be able to act as *cis* elements of mitochondrial mRNA end processing in *Arabidopsis*.

4.4 RPF3 might fulfill different roles in different *Arabidopsis* accessions

Previous studies of the *ccmC* transcript polymorphism between Col and C24 have connected this polymorphism to a differing region between -576 to -510 upstream of the mature mRNA ends since the remaining sequence surrounding the ends is almost identical in Col and C24 (Forner et al. 2008, Figure 10). These analyses have now been extended by phenotyping and genotyping the corresponding transcript and mtDNA region from 26 *Arabidopsis* accessions (Figure 12). In all but three (Kas-1, No-0 and Sorbo) a correlation of the mtDNA configuration and the transcript phenotype has been observed, further strenghtening the assumption that this DNA region harbors a *cis* element that affects *ccmC* RNA processing. What remains puzzling though, is the nature of this *cis* element and what exact function RPF3 and/or other nuclear factors fulfill in the *ccmC* processing events in the different *Arabidopsis* accessions.
The easiest scenario seems to be the one in Col. While other factors beside the processing endonuclease and RPF3 could also be contributing, processing at the -484/-482 sites seems to ultimately depend on the activity of RPF3. In the *rpf3-1* line RPF3 function is abolished by the presence of the T-DNA and this leads to an almost complete loss of the mature mRNA (Figure 6). This phenotype can be reverted by the introduction of a functional RPF3 allele into the mutant (RPF3 manuscript Figure 3). RPF3 could fulfill two different roles here. It either directly recruits the processing machinery to the cleavage site or is neccessary for the formation of a secondary structure which then gets recognized by other processing factors.

In C24 the scenario gets more complicated. The C24 mtDNA configuration gives rise to a *ccmC* mRNA phenotype differing from the one found in Col. In C24 and C24 (♀) x Col (♂) F_1 hybrids a mature *ccmC* mRNA with a 5‘ end at -391/390 is found. In homozygous C24 (♀) x *rpf3-1* (♂) F_2 plants the C24 allele of RPF3 is replaced by the T-DNA allele from the *rpf3-1* mutant which results in a reduction of the mature *ccmC* mRNA of about 50% in comparison to C24 wildtype (Figure 17). Thus it seems that the generation of the mature *ccmC* transcript is more or less independant from RPF3 when a C24 mtDNA configuration is present. Because of that it is very unlikely that RPF3 directly recruits the processing enzyme. If this was the case processing would have to be abolished completely upon loss of RPF3 similar to what is observed in *rpf3-1* with the Col background. The best explanation would be that another factor binds the *ccmC* RNA in C24 and that RPF3 only functions as an adapter that strenghtens the interaction of the first factor with the processing enzyme. Alternatively, if a secondary structure is formed as processing signal C24 RPF3 is only assisting in the formation and not necessary like Col RPF3. Until further analyses have been carried out all of these models are very speculative though as the nuclear genome of the C24 (♀) x *rpf3-1*(♂) F_2 is mixed up of Col and C24 sequences which might also be affecting *ccmC* RNA processing.

Since in No-0 hardly any mature *ccmC* transcript is detectable by CR-RT-PCR (Figure 12) or northern blot analysis (data not shown) at least one essential nuclear factor required for the generation of the processed *ccmC* transcript seems to be inactive here. No-0 and C24 have identical mtDNA configurations so the same processing factors might be involved in both accessions. Reciprocal crossing of No-0 with Col, C24 and the *rpf3-1* mutant as well as the introduction of the RPF3 alleles from Col and C24 into No-0 could be performed as first steps to gain more insight into the function of RPF3 and other factors taking part in *ccmC* mRNA processing in No-0.

The situation found in Sorbo illustrates again how flexible the system of gene expression and processing seems to be. Here, a Col-like mtDNA configuration is present and still a *ccmC* RNA phenotype resembling, yet different from the one in C24 was detected (Figure 12 and 15). This implicates that in Sorbo the generation of the mature *ccmC* transcript is goverened by a different (set of) nuclear factor(s). Again additional analyses are required to get insight into how the *ccmC* transcripts are formed here.

Judging from the data collected so far, it seems that optimized combinations of mtDNA configurations and processing factors might have evolved in the different *Arabidopsis* accessions. If this really is the case, it would be the first example of a nucleo-cytoplasm

co-adaptation that was recently proposed (Moison et al. 2010). However, as seen in the examples of No-0, Kas-1 and Sorbo the optimized state might not always be reached which leads to inefficient processing in these cases. With the occurence of different alleles of nuclear *trans* factors and distinct mitochondrial DNA configurations the *ccmC* transcript end processing in different accessions provides a promising tool to get more insight into-cyto-nuclear co-adaptation in *Arabidopsis.*

5. Summary

In higher plants mitochondrial gene expression is a very complex process where precursor RNAs undergo numerous modifications until mature mRNAs are formed. One of the least understood maturation steps is the posttranscriptional generation of 5‘ ends. In this process the 5‘ regions of the large precursor transcripts, which can be up to several kBs long get shortened down to a maximum of a few hundred nucleotides. Among the different accessions of *Arabidopsis thaliana* variations in the lengths of the 5‘ untranslated regions exist for some of the genes. These are linked to natural genetic variations within either the mitochondrial or the nuclear genomes of those ecotypes. Up to now, very few proteins taking part in 5‘ RNA processing have been characterized.

This work describes the identification of two nuclear encoded factors needed for the generation of mature mitochondrial mRNAs. Two different approaches were carried out. Based on natural genetic variation a forward genetic approach was used to identify the nuclear-encoded RNA Processing Factor 2 (RPF2), which is required for the generation of *nad9* mRNAs with 5‘ ends around -202 in respect to the translation start codon as well as the generation of a *cox3* mRNA with a 5‘ end around -380 in the accessions Columbia and C24. Another factor which was found through a reverse genetic approach by screening the mitochondrial mRNA end phenotypes in different T-DNA insertion lines and was named RNA Processing Factor 3 (RPF3). This protein plays an important role in the formation of mature *ccmC* transcripts with 5‘ ends around -484/-482 in the ecotype Columbia and affects the steady state level of *ccmC* mRNAs with a 5‘ end around -391 in the accession C24. In both cases accessions specific *cis* elements seem to have an influence on the 5‘ processing events.

Both proteins belong to a subfamily of the P class of pentatricopeptide repeat proteins with high similarity to RESTORER OF FERTILITY genes from different plant species. They are both imported into mitochondria where they presumably bind the respective targets RNAs which ultimately leads to endonucleolytic cleavage and the formation of mature, translatable mRNAs.

5. Zusammenfassung

In höheren Pflanzen ist die mitochondriale Genexpression ein sehr komplexer Prozess bei dem Vorläufer RNAs zahlreiche Modifikationen durchlaufen, bis eine reife mRNA gebildet wird. Einer der am wenigsten verstandenen Reifungsschritte ist die posttranskriptionelle Entstehung von 5‘-Enden. Dabei werden die 5‘-Regionen von großen Vorläufertranskripten, die mehrere Kilobasen lang sein können, auf ein Maximum von wenigen hundert Nukleotiden gekürzt. Betrachtet man verschiedene Ökotypen von *Arabidopsis thaliana* können bei manchen Genen Längenunterschiede der 5‘ untranslatierten Bereiche auftreten. Die Ursache dieser können natürliche genetische Variation innerhalb der mitochondrialen- oder der Kerngenome dieser Ökotypen sein. Bis jetzt sind nur wenige Proteine bekannt, die bei der 5‘-Prozessierung von RNAs beteiligt sind.

Diese Arbeit beschreibt die Identifikation von zwei kernkodierten Faktoren, die zur Generierung von reifen, mitochondrialen mRNAs benötigt werden. Es wurden zwei verschiedene Ansätze zur Identifikation durchgeführt. Basierend auf den natürlichen genetischen Variationen wurde ein vorwärts-genetischer Ansatz vollzogen. Hierduch wurde der RNA Prozessierungs Faktor 2 (RPF2), welcher in den Ökotypen Col und C24 für die Entstehung von *nad9* mRNAs mit 5‘-Enden um -202 in Relation zum Translationsstartcodon und von *cox3* mRNAs mit einem 5‘ Ende um -380 benötigt wird, kartiert. Ein weiterer Faktor wurde durch einen rückwärts-genetischen Ansatz, mit Hilfe eines Screenings der mitochondrialen mRNA Enden in verschiedenen T-DNA Insertionslinien gefunden und wurde RNA Prozessierungs Faktor 3 (RPF3) genannt. Dieses Protein spielt eine wichtige Rolle bei der Generierung von reifen *ccmC* Transkripten mit 5‘-Enden um -484/-482 im Ökotyp Columbia. Desweiteren hat er einen EInfluss auf die steady-state Menge von *ccmC* mRNAs mit 5‘-Enden um -391 im Ökotyp C24. In beiden Fällen scheinen ökotyp-spezifische *cis* Elemente ebenfalls einen Einfluss auf die Prozessierung zu haben.

Beide Proteine gehören zu einer Subfamilie der P-Klasse pentatricopeptide repeat Proteine, die eine hohe Ähnlichkeit zu RESTORER OF FERTILITY Genen von verschiedenen Pflanzenspezies aufweist. Beide werden in die Mitochondrien importiert, wo sie höchstwahrscheinlich ihre jeweiligen Ziel-RNAs binden, was letztendlich zur endonukleolytischen Prozessierung dieser führt. Hierdurch werden reife, translatierbare mRNAs erzeugt.

6. References

Binder S, Hölzle A, Jonietz C. (2010) RNA processing and RNA stability in plant mitochondria, in *Plant mitochondria,* ed. Kempken F., Springer Press New York Inc., pp. 107-130

Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Mol Cell Biol **28:** 5337-5347

Bentolila, S., Alfonso, A.A., and Hanson, M.R. (2002). A pentatrico- peptide repeat-containing gene restores fertility to cytoplasmic male sterile plants. Proc. Natl. Acad. Sci. USA 99: 10887–10892.

Bentolila S, Knight W, Hanson M. (2010) Natural Variation in Arabidopsis Leads to the Identification of REME1, a Pentatricopeptide Repeat-DYW Protein Controlling the Editing of Mitochondrial Transcripts. Plant Physiol;154(4):1966-82. Epub 2010 Oct 25.

Bollenbach, T.J., Lange, H., Gutierrez, R., Erhardt, M., Stern, D.B., and Gagliardi, D. (2005). RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. Nucleic Acids Res. 33, 2751-2763.

Brown, G.G., Formanova, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W.Y., and Landry, B.S. (2003). The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. 35: 262–272.

Canino, G., Bocian, E., Barbezier, N., Echeverria, M., Forner, J., Binder, S., and Marchfelder, A. (2009). Arabidopsis encodes four tRNase Z enzymes. Plant Physiol. 150: 1494–1502.

Chase, C.D. (2007). Cytoplasmic male sterility: A window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23: 81–90.

de Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D. (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana. Plant Cell 19: 3256–3265.

Desloire, S., et al. (2003). Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep. 4: 588–594.

Forner, J., Hölzle, A., Jonietz, C., Thuss, S., Weber, B., Schwarzländer, M., Meyer, R.C., and Binder, S. (2008). Mitochondrial mRNA polymorphisms in different Arabidopsis thaliana accessions. Plant Physiol. 148: 1106–1116.

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.

Forner, J., Weber, B., Wietholter, C., Meyer, R.C., and Binder, S. (2005). Distant sequences determine 5‘ end formation of cox3 transcripts in Arabidopsis thaliana ecotype C24. Nucleic Acids Res. 33: 4673–4682.

Gagliardi, D., and Binder, S. (2007). Expression of the plant mitochondrial genome. In Plant Mitochondria, D. Logan, ed (Ames, IA: Blackwell Publishing), pp. 50–95.

Geddy, R., and Brown, G.G. (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: 130.

Giege JM, Grienenberger JM, Bonnard G (2007) Cytochrome c biogenesis in mitochondria. Mitochondrion 8**:** 61-73

Gobert A, Gutmann B, Taschner A, Gössringer M, Holzmann J, Hartmann RK, Rossmanith W, Giegé P. (2010) A single Arabidopsis organellar protein has RNase P activity. Nat Struct Mol Biol. 2010 Jun;17(6): 740-4. Epub 2010 May 16.

Hanson, M.R., and Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16 (suppl.): S154–S169.
Hauler, A. (2010) Charakterisierung des mitochondrialen RNA Prozessierungsfaktors 5 (RPF5) in Arabidopsis thaliana, Diploma thesis, University of Ulm

Hedtke, B., Börner, T., and Weihe, A. (2000). One RNA polymerase serving two genomes. EMBO Rep. 1: 435–440.

Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T., and Gagliardi, D. (2006). Relaxed transcription in *Arabidopsis* mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide

Johnson X, Wostrikoff K, Finazzi G, et al. (2010) 2. MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. Plant Cell **22:** 234-248

Jonietz, C. (2007) Kartierung eines nad9-spezifischen 5‘ Prozessierungsfaktors in Arabidopsis thaliana, Diploma thesis, University of Ulm

Kazama, T., Nakamura, T., Watanabe, M., Sugita, M., and Toriyama, K. (2008). Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J. 55: 619–628.

Kennedy, E. P., and Lehninger, *A. L.*, *J. Biol. Chem.,* 1949, 179, 957.

Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S., and Imamura, J. (2003). Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male sterile Kosena radish. Plant J. 34: 407–415.

Kranz RG, Richard-Fogal C, Taylor JS, Frawley ER (2009) Cytochrome c biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control. Microbiol Mol Biol Rev 73**:** 510-528

Kubo, T., and Newton, K.J. (2008). Angiosperm mitochondrial genomes and mutations. Mitochondrion 8: 5–14.

Kühn, K., Weihe, A., and Börner, T. (2005). Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. Nucleic Acids Res. 33: 337–346.

Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., Falcon de Longevialle, A., O’Toole, N., Börner, T., Millar, A.H., Small, I.D., and Whelan, J. (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of Arabidopsis thaliana. Plant Cell 21: 2762–2779.

Laser KD and Lersten NR, (1972). Anatomy and cytology of microsporo- genesis in cytoplasmic male sterile angiosperms. Bot Rev 38: 425–454.

Loiselay C, Gumpel NJ, Girard-Bascou J, Watson AT, Purton S, Wollman FA, Choquet Y (2008) Molecular identification and function of cis- and trans-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol Cell Biol **28:** 5529-5542

Lurin, C., et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16: 2089–2103.

Matthes, A., Schmidt-Gattung, S., Köhler, D., Forner, J., Wildum, S., Raabe, M., Urlaub, H., and Binder, S. (2007). Two DEAD-box proteins may be part of RNA-dependent high-molecular-mass protein complexes in Arabidopsis mitochondria. Plant Physiol. 145: 1637– 1646.

McGinnis, S., and Madden, T.L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32: W20–25.

References

Moison, M., Roux, F., Quadrado, M., Duval, R., Ekovich, M., Le, D.H., Verzaux, M. and Budar, F. (2010). Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in Arabidopsis thaliana. The Plant Journal 63: 728–738

Nakamura T., G. Schuster, M. Sugiura and M. Sugita (2004) Chloroplast RNA-binding and pentatricopeptide repeat proteins Biochemical Society Transactions 32, (571–574)

Perrin, R., Lange, H., Grienenberger, J.M., and Gagliardi, D. (2004a). AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria. Nucleic Acids Res. 32: 5174–5182.

Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M., and Gagliardi, D. (2004b). Two exoribonucleases act sequentially to process mature 3‘-ends of atp9 mRNAs in Arabidopsis mitochondria. J. Biol. Chem. 279: 25440–25446.

Pfalz J, Bayraktar OA, Prikryl J, Barkan A (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. EMBO J **28:** 2042-2052

Sambrook, J., and Russel, D.W. (2000). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

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

Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. Plant Cell;19(4):1251-64. Epub 2007 Apr 27.

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.

Takenaka, M. (2010). MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the *nad7* transcript in mitochondria of Arabidopsis. Plant Physiol. **152,** 939-947.

Tracy RL, Stern DB. (1995) Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr Genet. 1995 Aug;28(3):205-16.

Tzafrir I, Dickerman A, Brazhnik O, Nguyen Q, McElver J, Frye C, Patton D, Meinke D. (2003) The Arabidopsis SeedGenes Project. Nucleic Acids Res. 1;31(1):90-3.

Unseld, M., Marienfeld, J.R., Brandt, P. & Brennicke, A. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366.924 nucleotides. Nat Genet 15(1):57-61

Verbitskiy, D., Zehrmann, A., Brennicke, A., and Takenaka, M. (2010). A truncated MEF11 protein shows site-specific effects on mitochondrial RNA editing. Plant Signal Behav. **Epub ahead of print May 31**.

Walter, M., Killian, J., Kudla, J. (2002) PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. The EMBO Journal 21, 6905 - 6914

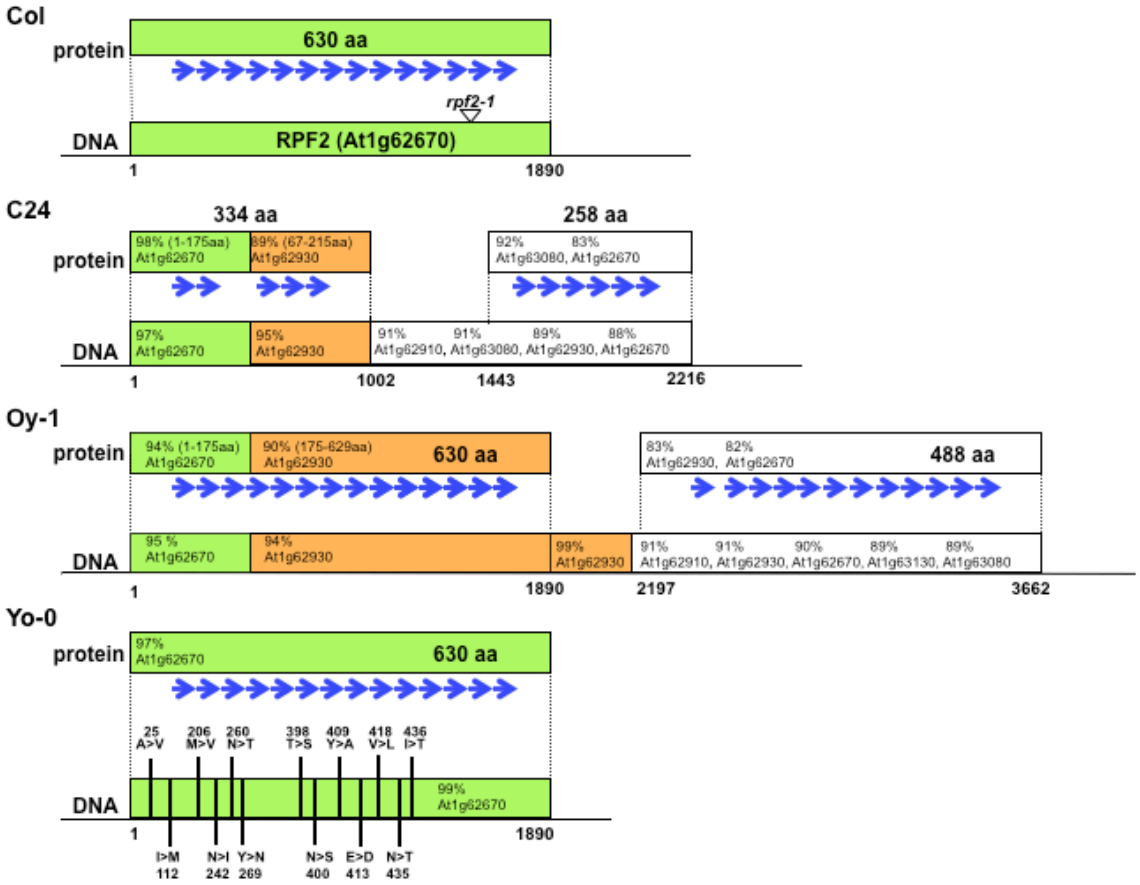
Wang, Z., et al. (2006). Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18: 676–687.

Weigel, D., and Glazebrook, J. (2002). Arabidopsis: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

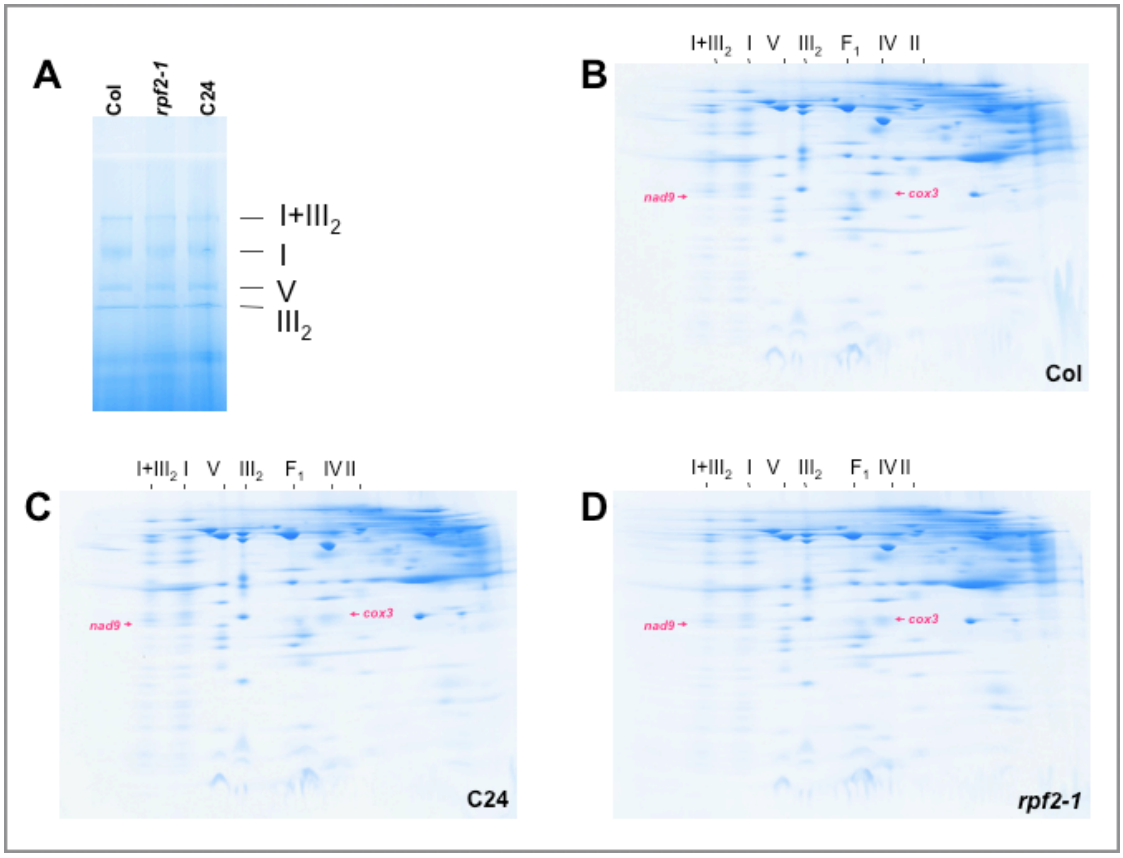
Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. Plant Cell 21: 558–567.

7. Appendix

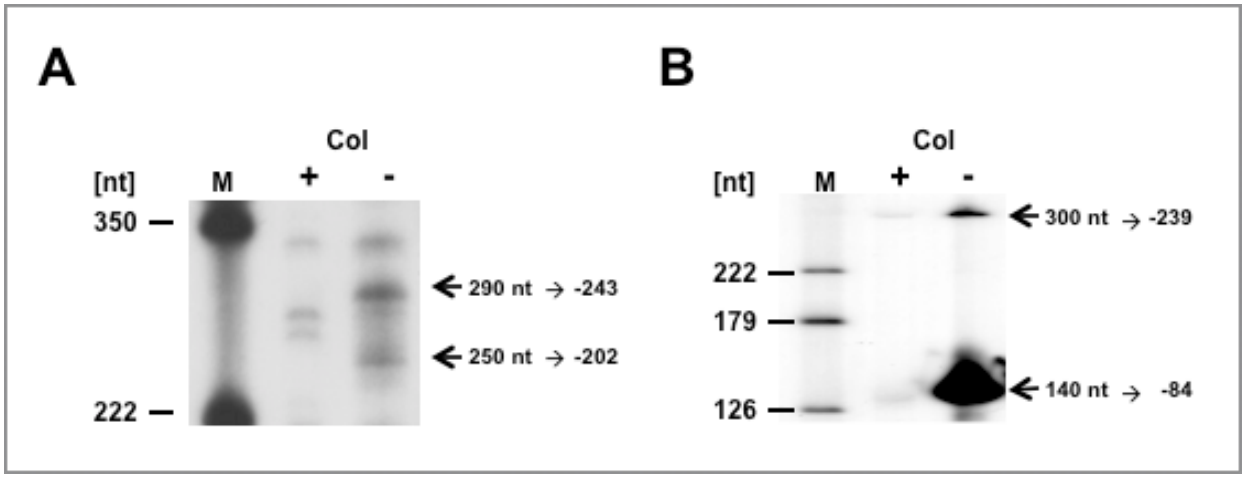
7.1 Supplemental Figures



Supplemental Figure S1: *RPF2* (At1g62670) alleles in accessions Col, C24, Oy-1 and Yo-0. Nucleotide and amino acid sequence identities between *RPF2* from these accessions and different RF-like PPR genes from Col are given. The C-terminal parts of the *RPF2* alleles from C24 and Oy-1 correspond to different parts of At1g62930 (C24: 67-215; Oy-1: 175-630). Green boxes indicate highest similarity to RPF2 (At1g62670), orange boxes represent regions with highest similarity to At1g62930 and white boxes stand for highest similarities to various RF-like PPR genes. Amino acid identities that are different between the *RFP2* alleles from Col and Yo-0 are indicated (Col > Yo-0), Their approximate positions are indicated by black vertical lines. Lengths of the proteins are given in amino acids (aa). Numbering refers to the first and last nucleotide position of the reading frame (without stop codon) and is given with respect to the ATG (A = +1) of the first reading frame. The T-DNA insertion in *rpf2-1* is indicated by a triangle. PPR repeats are given as blue arrows.



Supplemental Figure S2: Blue native (BN)/SDS polyacrylamide gel electrophoresis (PAGE) of mitochondrial protein from *Arabidopsis*. (A) About 1.0 mg of mitochondrial protein isolated from cell suspension cultures from accession Columbia (Col), from *rpf2-1* and from C24 were separated by blue native PAGE in the first dimension. (B to D) Lanes from the first dimension were cut from the gel and proteins subsequently separated by SDS PAGE in the second dimension. A comparison between the protein patterns from the different lines (B to D) did not reveal any differences in the abundance of *nad9* and *cox3* polypeptides (indicated by red arrow heads). Thus altered 5' processing of the corresponding mRNAs has no influence on the steady levels of these polypeptides. This result is consistent with the immunodetection analysis of the *nad9* protein (Fig. 5). Positions of different respiratory chain complexes are indicated.



Supplemental Figure S3: Primer extension analysis of *nad9* (A) and *atp9* (B) transcripts from untreated (-) and Terminator™ exonuclease-treated (+) mitochondrial RNA. *nad9* mRNAs with -243 and -202 5' ends were completely degraded by the Terminator™ exonuclease. This demonstrates that these 5' termini derived from posttranscriptional processing. Two extension products of about 270 and 280 nucleotides are barely or not detectable before the treatment of the RNA and are not related to the -243 and -202 5' termini of *nad9* transcripts. In a control experiment a low abundant *atp9* 5' end originating from transcription initiation (-239, Kühn et al., 2005) was still detectable after Terminator™ exonuclease treatment, while the highly abundant mRNA with the processed 5' extremity at position -84 was almost completely digested. Sizes of the primer extension products are given in nucleotides (nt).

A

```
Bay-0       TCGGGG-AGCAAGTAAAAGCAGAGAGTGATAAATGCGAAAAAATCAAGGCAAAAATAATA -610
Yo-0        TCGGGG-AGCAAGTAAAAGCAGAGAGTGATAAATGCGAAAAAATCAAGGCAAAAATAATA -610
Est         TCGGGG-AGCAAGTAAAAGCAGAGAGTGATAAATGCGAAAAAATCAAGGCAAAAATAATA -610
Ler         TCGGGG-AGCAAGTAAAAGCAGAGAGTGATAAATGCGAAAAAATCAAGGCAAAAATAATA -610
Col         TCGGGGAAGCAAGTAAAAGCAGAGAGTGATAAATGCGAAAAAATCAAGGCAAAAATAATA -610
            ****** **************************************************

Bay-0       GCAAAAACACATGAACTCTTAGTATCCGAAGGTTACCATATTCCGGATGAACGGGATATA -550
Yo-0        GCAAAAACACATGAACTCTTAGTATCCGAAGGTTACCATATTCCGGATGAACGGGATATA -550
Est         GCAAAAACACATGAACTCTTAGTATCCGAAGGTTACCATATTCCGGATGAAGGGGATATA -550
Ler         GCAAAAACACATGAACTCTTAGTATCCGAAGGTTACCATATTCCGGATGAACGGGATATA -550
Col         GCAAAAACACATGAACTCTTAGTATCCGAAGGTTACCATATTCCGGATGAACGGGATATA -550
            ************************************************* ********

Bay-0       GAACGAGCAATAAATGTTGTGATGACTGAACATGAAATGATCGATATCGATCGGCGTAAG -490
Yo-0        GAACGAGCAATAAATGTTGTGATGACTGAACATGAAATGATCGATATCGATCGGCGTAAG -490
Est         GAACGAGCAATAAATGTTGTGATGACTGAACATGAAATGATCGATATCGATCGGCGTAAG -490
Ler         GAACGAGCAATAAATGTTGTGATGACTGAACATGAAATGATCGATATCGATCGGCGTAAG -490
Col         GAACGAGCAATAAATGTTGTGATGACTGAACATGAAATGATCGATATCGATCGGCGTAAG -490
            ************************************************************

Bay-0       AAAAGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGGTGGAG -430
Yo-0        AAAAGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGGTGGAG -430
Est         AAAAGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGGTGGAG -430
Ler         AAAAGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGGTGGAG -430
Col         AAAAGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGGTGGAG -430
            ************************************************************

Bay-0       CTACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCC -370
Yo-0        CTACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCC -370
Est         CTACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCC -370
Ler         CTACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCC -370
Col         CTACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCC -370
            ************************************************************

Bay-0       ATTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTT -310
Yo-0        ATTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTT -310
Est         ATTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTT -310
Ler         ATTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTT -310
Col         ATTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTT -310
            ************************************************************

Bay-0       CATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATA -250
Yo-0        CATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATA -250
Est         CATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATA -250
Ler         CATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATA -250
Col         CATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATA -250
            ************************************************************

Bay-0       GCGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATC -190
Yo-0        GCGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATC -190
Est         GCGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATC -190
Ler         GCGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATC -190
Col         GCGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATC -190
            ************************************************************

Bay-0       TTTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Yo-0        TTTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Est         TTTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Ler         TTTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Col         GTTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
             ******************************
```

B

```
C24         TCGGGGAGCAAGTAAAAGCAGAGAGTGGTAAATGCGAAAAAATCAAGGCAAAAATAATAG -610
No-0        TCGGGGAGCAAGTAAAAGCAGAGAGTGGTAAATGCGAAAAAATCAAGGCAAAAATAATAG -610
Kas-1       TCGGGGAGCAAGTAAAAGCAGAGAGTGGTAAATGCGAAAAAATCAAGGCAAAAATAATAG -610
Ws          TCGGGGAGCAAGTAAAAGCAGAGAGTGGTAAATGCGAAAAAATCAAGGCAAAAATAATAG -610
Shakdara    TCGGGGAGCAAGTAAAAGCAGAGAGTGGTAAATGCGAAAAAATCAAGGCAAAAATAATAG -610
            ************************************************************

C24         CAAAAACACATGAACTCTTAGTATCCGAAG-ATACAAAATTCCAGATAAAAACGATATAA -550
No-0        CAAAAACACATGAACTCTTAGTATCCGAAGGATACAAAATTCCAGATAAAAACGATATAA -550
Kas-1       CAAAAACACATGAACTCTTAGTATCCGAAGGATACAAAATTCCAGATAAAAACGATATAA -550
Ws          CAAAAACACATGAACTCTTAGTATCCGAAGGATACAAAATTCCAGATAAAAACGATATAA -550
Shakdara    CAAAAACACATGAACTCTTAGTATCCGAAGGATACAAAATTCCAGATAAAAACGATATAA -550
            ***************************** ****************************

C24         AAAGAATTATCGATGTAGCGATGTTTGACGATGAGTTGATCGATATCGATCGGCGTAAGA -490
No-0        AAAGAATTATCGATGTAGCGATGTTTGACGATGAGTTGATCGATATCGATCGGCGTAAGA -490
Kas-1       AAAGAATTATCGATGTAGCGATGTTTGACGATGAGTTGATCGATATCGATCGGCGTAAGA -490
Ws          AAAGAATTATCGATGTAGCGATGTTTGACGATGAGTTGATCGATATCGATCGGCGTAAGA -490
Shakdara    AAAGAATTATCGATGTAGCGATGTTTGACGATGAGTTGATCGATATCGATCGGCGTAAGA -490
            ************************************************************

C24         AACGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGATGGAGC -430
No-0        AACGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGATGGAGC -430
Kas-1       AACGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGATGGAGC -430
Ws          AACGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGATGGAGC -430
Shakdara    AACGATTCTACTATCTTTATTCCCGGTTAGGAAAAAACGGAGACAAGTTTTGGATGGAGC -430
            ************************************************************

C24         TACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCCA -370
No-0        TACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCCA -370
Kas-1       TACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCCA -370
Ws          TACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCCA -370
Shakdara    TACTGGAATTGCTGTCTGACTACGACATAAATAAAAGCGATTCCGACGACTAAACCGCCA -370
            ************************************************************

C24         TTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTTC -310
No-0        TTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTTC -310
Kas-1       TTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTTC -310
Ws          TTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTTC -310
Shakdara    TTTCAGGTTTACGTCTTTTGTTAGGTCTCGACTTCTTCGCTATGATCTTCCCGGTAGTTC -310
            ************************************************************

C24         ATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATAG -250
No-0        ATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATAG -250
Kas-1       ATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATAG -250
Ws          ATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATAG -250
Shakdara    ATATAGGAGCTATAGCCGTTTCATTCCTTTTCGTTGTTATGATGTTCCATATTCAAATAG -250
            ************************************************************

C24         CGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATCT -190
No-0        CGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATCT -190
Kas-1       CGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATCT -190
Ws          CGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATCT -190
Shakdara    CGGAGATTCACGAAGAAGTATTGCGCTATTTACCAGTGAGTGGGATTATTGGACTGATCT -190
            ************************************************************

C24         TTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
No-0        TTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Kas-1       TTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Ws          TTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
Shakdara    TTTGGTGGTGAGTTTCCTTTTTTCTTTTTTT -158
            ******************************
```

Supplemental Figure S4: Clustal W allignment of the mitochondrial *CcmC* upstream sequences from 240952 to 241419 (-670 to -158 in relation to ATG) in different accessions. Stars mark sequence identity.
A: Accessions showing a Col mtDNA configuration (group 1).
B: Accessions showing a C24 mtDNA configuration (group 2).

7.2 List of Publications

Publications:

Forner J, Hölzle A, Jonietz C, Thuss S, Weber B, Schwarzländer M, Meyer RC, Binder S (2008) Mitochondrial mRNA polymorphisms in different Arabidopsis thaliana accessions. Plant Physiol **148:** 1106-1116

Jonietz C, Forner J, Holzle 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:** 443-453

Hölzle A, Jonietz C, Törjek O, Altmann T, Binder S, Forner J (2010) A RESTORER OF FERTILITY-like PPR gene is required for 5' processing of the *nad4* mRNA in mitochondria of *Arabidopsis thaliana*. manuscript accepted

Reviews:

Binder S, Hölzle A, Jonietz C. (2010) RNA processing and RNA stability in plant mitochondria, in *Plant mitochondria,* ed. Kempken F., Springer Press New York Inc., pp. 107-130

Poster:

2009 Poster at the „International Conference for Plant MItochondrial Biology“ in Lake Tahoe / California (USA): „Identification of RPF2, an *Arabidopsis thaliana nad9* and *cox3* mRNA processing factor

Oral presentations:

2009 Talk at the „International Conference for Plant MItochondrial Biology“ in Lake Tahoe / California (USA): „Identification of RPF2, an *Arabidopsis thaliana nad9* and *cox3* mRNA processing factor

2010 Talk at the „Plant Regio Meeting“ in Tübingen: „Natural genetic variation affects mitochondrial mRNA processing in *Arabidopsis thaliana*“

7.2 Publications/Manuscripts

The Plant Cell, Vol. 22: 443–453, February 2010, www.plantcell.org © 2010 American Society of Plant Biologists

RNA PROCESSING FACTOR2 Is Required for 5′ End Processing of *nad9* and *cox3* mRNAs in Mitochondria of *Arabidopsis thaliana* [W]

Christian Jonietz, Joachim Forner, Angela Hölzle, Sabine Thuss, and Stefan Binder[1]

Institut Molekulare Botanik, Universität Ulm, D-89069 Ulm, Germany

In mitochondria of higher plants, the majority of 5′ termini of mature mRNAs are generated posttranscriptionally. To gain insight into this process, we analyzed a natural 5′ end polymorphism in the species *Arabidopsis thaliana*. This genetic approach identified the nuclear gene At1g62670, encoding a pentatricopeptide repeat protein. The functional importance of this mitochondrial restorer of fertility-like protein, designated RNA PROCESSING FACTOR2 (RPF2), is confirmed by the analysis of a respective T-DNA knockout mutant and its functional restoration by in vivo complementation. RPF2 fulfills two functions: it is required for the generation of a distinct 5′ terminus of transcripts of subunit 9 of the NADH DEHYDROGENASE complex (*nad9*) and it determines the efficiency of 5′ end formation of the mRNAs for subunit 3 of the CYTOCHROME C OXIDASE (*cox3*), the latter also being influenced by mitochondrial DNA sequences. Accordingly, recombinant RPF2 protein directly binds to a *nad9* mRNA fragment in vitro. Two-dimensional gel electrophoresis and immunodetection analyses reveal that altered 5′ processing does not influence accumulation of the *nad9* and *cox3* polypeptides. In accessions C24, Oystese-1, and Yosemite-0, different inactive *RPF2* alleles exist, demonstrating the variability of this gene in *Arabidopsis*. The identification of *RPF2* is a major step toward the characterization of 5′ mRNA processing in mitochondria of higher plants.

INTRODUCTION

Plant mitochondrial genomes encode ~60 genes (Kubo and Newton, 2008). This essential genetic information is expressed via a multistep process in which posttranscriptional events contribute to the complexity of the plant mitochondrial gene expression system. Alterations of the structure and primary sequence of transcripts are necessary for the formation of translatable mitochondrial mRNAs. Posttranscriptional processes include C-to-U RNA editing, splicing of *cis* and *trans* introns, and the formation of mature 5′ and 3′ ends (Gagliardi and Binder, 2007; Bonen, 2008; Takenaka et al., 2008).

The consequences of splicing and RNA editing are quite evident as they directly affect the sequence of the encoded protein, while the biological impact of posttranscriptional processing of mRNA extremities is less clear. Functional studies of mRNA 3′ maturation indicated the importance of this process. Two 3′ to 5′ exoribonucleases, the mitochondrial polynucleotide phosphorylase (PNPase) and the RNase R homolog 1, have crucial functions in the removal of extra sequences at the 3′ termini of mitochondrial RNA. The knockout or knockdown of these proteins has severe consequences for plant fitness, demonstrating the importance of mitochondrial mRNA 3′ processing (Perrin et al., 2004b). Trimming of mRNA 3′ ends is linked to mRNA stability, which is influenced also by polyadenylation. As in bacteria and chloroplasts, 3′ poly(A) tails enhance degradation of mRNAs in plant mitochondria.

In contrast to mRNA 3′ maturation, virtually nothing is known about the posttranscriptional generation of 5′ termini in plant mitochondria. A systematic analysis of major extremities of mitochondrial mRNAs in *Arabidopsis thaliana* showed that most 5′ termini of mature mRNAs are generated posttranscriptionally. Only a few mature 5′ transcript ends directly originate from transcription initiation (Forner et al., 2007). The prevalence of processed 5′ ends indicates that their generation is an important process, which might play an unidentified regulatory role in plant mitochondrial gene expression.

The inspection of mitochondrial mRNA extremities in *Arabidopsis* also revealed the frequent occurrence of multiple, abundant 5′ ends per gene. These can be processed ends or termini generated by transcription initiation as observed in other species. Thus, high variability of 5′ ends is a common feature of plant mitochondrial mRNAs. In addition, variation in 5′ mitochondrial transcript ends is seen between different *Arabidopsis* accessions. Comparative mapping of mRNA extremities in the three accessions Columbia (Col), C24, and Landsberg *erecta* (L*er*) identified several 5′ end polymorphisms, while no difference was found in the 3′ ends of mitochondrial transcripts (Forner et al., 2008). This further demonstrates the variation and flexibility of 5′ ends of plant mitochondrial mRNAs, opposing the clearly defined nature of 3′ ends. The 5′ variation in different *Arabidopsis* accessions can be associated with differences in the mitochondrial DNA or with distinct factors encoded in the nucleus (Forner

[1] Address correspondence to stefan.binder@uni-ulm.de.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Stefan Binder (stefan.binder@uni-ulm.de).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.109.066944

et al., 2005, 2008). The 5′ transcript end polymorphisms can now be used to identify the proteins that are required for posttranscriptional generation of 5′ termini of mitochondrial transcripts.

To date, only a few proteins implicated in posttranscriptional RNA metabolism have been identified and/or characterized in plant mitochondria. These include the two above-mentioned exoribonucleases (Perrin et al., 2004a, 2004b), two DEAD-box proteins (Matthes et al., 2007), and a small number of pentatricopeptide repeat (PPR) proteins. These PPR proteins almost exclusively represent restorers of male fertility (encoded by *RF* genes) in cytoplasmic male sterility (CMS) systems in different plant species (Andrés et al., 2007; Schmitz-Linneweber and Small, 2008). The RF proteins have particular roles in translation, RNA stabilization, and RNA cleavage of the CMS-specific mitochondrial transcripts, while their intrinsic normal functions are almost exclusively unknown. Apart from the *RF* genes, two other PPR proteins, ORGANELLAR TRANSCRIPT PROCESSING DEFECT43, involved in *trans*-splicing of the intron 1 of the mRNA for subunit 1 of the NADH DEHYDROGENASE (*nad1*) in mitochondria of *Arabidopsis* (de Longevialle et al., 2007), and MITOCHONDRIAL RNA EDITING FACTOR1, have been characterized (Zehrmann et al., 2009). However, no protein involved in 5′ end processing of plant mitochondrial mRNA has yet been described.

Here, we report the identification of a nuclear gene, *RNA PROCESSING FACTOR2* (*RPF2*), required for the formation of 5′ ends 202 bp upstream of the *nad9* reading frame in *Arabidopsis* mitochondria. The *RPF2* locus is identified by map-based cloning, analysis of a respective T-DNA insertion mutant, and in vivo complementation studies. *RPF2* encodes a PPR protein that is similar to the PPR proteins encoded by the restorer of fertility genes (Budar and Pelletier, 2001; Bentolila et al., 2002; Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003; Wang et al., 2006). Phenotyping of mitochondrial transcript extremities in several PPR knockout mutants reveals that RPF2 is also involved in the formation of distinct *cox3* 5′ transcript ends. In C24, *cox3* 5′ end formation is influenced both by RPF2 and a *cis*-element in the mitochondrial DNA.

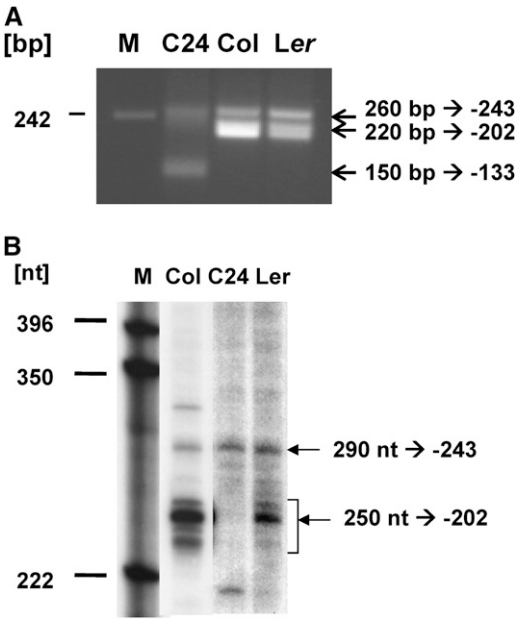


Figure 1. *nad9* 5′ Ends in Different *Arabidopsis* Accessions.

(A) CR-RT-PCR analysis (at least three biological and technical replicates) of *nad9* mRNAs. PCR products of 260, 220, and 150 bp contain 5′ termini at positions −243, −202, and −133 relative to the ATG (A = +1).
(B) Primer extension analysis of *nad9* mRNAs. Extension products of 290 nucleotides confirm the presence of the *nad9* −243 5′ end in all accessions investigated. The cDNA strands of ~250 nucleotides (nt) correspond to *nad9* 5′ ends around position −202 in Col und L*er*. These results are consistent with the CR-RT-PCR analysis. Accession-specific products of 330 nucleotides (Col) and 210 nucleotides in C24 were not detected in respective CR-RT-PCR experiments.

RESULTS

The *nad9* 5′ Transcript End Polymorphism in Different *Arabidopsis* Accessions

In *Arabidopsis*, the mitochondrial *nad9* gene is transcribed into mRNAs of 0.85 and 0.9 kb in the accessions Col and L*er* with distinct 5′ ends. In these accessions, circularized RNA (CR)-RT-PCR analysis identified two major 5′ ends at positions −202 and −243 with respect to the translation start codon (NATG, $n = -1$) (Forner et al., 2007). In C24, the latter end is present at levels comparable to Col and L*er*, while the −202 5′ ends are hardly detectable (Figure 1A) (Forner et al., 2008). A reexamination of the *nad9* mRNAs by primer extension confirmed the presence of the −243 end in these three accessions. In Col and L*er*, a group of additional 5′ termini were found around position −202, in agreement with the CR-RT-PCR product pattern, which showed at least two products from this cluster of ends (Figure 1B). This analysis confirmed the absence of substantial amounts of the 5′ ends at −202 in C24. However, a second CR-RT-PCR approach amplified a faint cDNA product, indicating that minute amounts of −202 ends still exist in C24 (see Supplemental Figure 1 online). Also, in C24, a unique *nad9* 5′ transcript terminus was found 133 bp upstream of the ATG (Figure 1A).

The investigation of the *nad9* −243 and −202 termini by primer extension analysis of terminator exonuclease-treated mtRNA confirmed these ends to be derived from posttranscriptional processing (see Supplemental Figure 2A online). This exonuclease specifically digests mRNAs containing 5′ monophosphate groups typically generated by posttranscriptional processing. By contrast, primary mRNAs with 5′ ends directly generated by transcription initiation contain 5′ tri- or diphosphate groups and are not degraded by this enzyme. Apart from our experimental data, a recent analysis of transcription initiation sites in *Arabidopsis* mitochondria identified four *nad9* promoters more than 1 kb upstream of the −202 5′ ends (Kühn et al., 2009).

Delimiting the Genomic Region Containing the *RPF2* Locus

Reciprocal crosses between Col and C24 have previously shown that the *nad9* 5′ end polymorphism is attributed to a nuclear

locus, which we now designate *RPF2* (Forner et al., 2008). In order to map this locus, an F2 mapping population was established from Col/C24 F1 hybrids. Phenotyping of *nad9* transcript ends in 200 F2 plants revealed the C24 *nad9* 5′ end phenotype (i.e., the almost complete absence of the −202 ends) in 48 F2 individuals (see Supplemental Figure 3 online; Table 1). These plants correspond to 24% of the population, consistent with a 1:3 ratio of the C24 *nad9* 5′ end phenotype versus the *nad9* mRNA 5′ end phenotype of Col. Therefore, a single genetic locus is responsible for this *nad9* 5′ end polymorphism, with the almost complete absence of the *nad9* −202 mRNA end being a recessive trait. In addition, this analysis revealed that the nearly complete lack of the *nad9* −202 5′ termini in C24 is independent from the occurrence of the *nad9* −133 end (see Supplemental Figure 3 online; see below), demonstrating that another genetic locus is responsible for the generation of this terminus in C24.

Using insertion/deletion markers distributed over the nuclear genome (see Supplemental Table 1 online), 47 plants with the C24 *nad9* 5′ end phenotype were genotyped (see Supplemental Figure 2 online). For all markers located on chromosomes 2, 3, 4, and 5, an approximately equal distribution of the Col and C24 genotypes was found. By contrast, the C24 genotype of marker CER 479928 on chromosome 1 is strongly linked to the recessive C24 *nad9* 5′ end phenotype. Therefore, further markers on the lower arm of chromosome 1 were tested to delimit the genetic interval containing the *RPF2* locus (see Supplemental Figure 4 online). This mapping approach defined a genetic interval containing the *RPF2* locus between 22.3 to 24.0 Mbp on the lower arm of chromosome 1. This region encodes 442 protein-coding genes, from At1g60470 to At1g64620 (Arabidopsis Genome Initiative, 2000).

A T-DNA Insertion in At1g62670 Promotes the Absence of the *nad9* −202 5′ Ends

RPF2 is expected to encode a protein targeted to mitochondria with a predicted function in nucleic acid metabolism. These criteria reduced the number of candidate genes in the defined interval to 23 genes encoding PPR proteins (Lurin et al., 2004). Thirteen of these PPR genes are very similar to restorers of fertility (*RF*) genes implicated in cytoplasmic male sterility systems in various plant species (Budar and Pelletier, 2001; Bentolila et al., 2002; Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003; Wang et al., 2006). Since fertility restoration correlates with altered expression of the CMS-related mitochondrial gene, it was investigated whether one of these genes is linked to the *nad9* 5′ end polymorphism. To this end, homozygous T-DNA insertion lines were established for nine candidate PPR genes and the 5′ termini of the *nad9* mRNAs were inspected by CR-RT-PCR. No *nad9* 5′ end difference was observed between wild-type Col and the SALK lines carrying T-DNA insertions in At1g63070, At1g62680, At1g64580, At1g63080, At1g62590, At1g64100, At1g63230, and At1g63130 (Figure 2; see Supplemental Figure 5 online). However, in line SALK 146198, which harbors a T-DNA insertion in At1g62670, the *nad9* −202 5′ ends were undetectable in a single round of CR-RT-PCR as in C24. An RT-PCR analysis of this mutant with primers flanking the T-DNA insertion demonstrated that the T-DNA causes a knockout of At1g62670 (Figure 2B). This correlates the inactivation of this gene with the nearly complete absence of the *nad9* −202 5′ ends. Like C24, this knockout mutant contains minimal amounts of the 5′ ends around −202 (see Supplemental Figure 1 online). In addition, this CR-RT-PCR analysis revealed that the C24-specific *nad9* −133 5′ end is not present in this T-DNA insertion line (Figure 2A), further confirming the abundance of the −202 5′ ends to be independent from the accumulation of *nad9* transcripts with the −133 end.

Table 1. CR-RT-PCR Analysis of Col/C24 F2 Hybrid Plants

	♀ C24 × Col ♂	♀ Col × C24 ♂	Total
Number of plants			
Without *nad9* −202	25	23	48
With *nad9* −202	75	77	152
Percentage without *nad9* −202	25%	23%	24%
Ratio without:with *nad9* −202	1:3.0	1:3.4	1:3.2

The Wild-Type Col Allele of At1g62670 Restores the Formation of Abundant *nad9* -202 5′ Ends in Its T-DNA Insertion Line and in C24

The barely detectable levels of the *nad9* -202 5′ ends in the T-DNA knockout line of At1g62670 are a strong indication that this gene is required for the formation of these termini. However, we could not exclude that another T-DNA insertion in line SALK 146198 might cause this phenotype. Thus, the At1g62670 gene from the Col wild type, including the regions ~1.6 kb upstream and 1.3 kb downstream, was cloned into the pMDC123 vector and transformed into SALK 146198 and C24 plants. The *nad9* transcript ends were then investigated by CR-RT-PCR analysis in several T1 plants. In all transformants (SALK 146198 and C24), the presence of the At1g62670 wild-type allele from Col restores the formation of the *nad9* −202 5′ ends (Figure 3A). This result was confirmed by RNA gel blot analysis of total RNA from the different lines. In SALK 146198 and C24, the smaller *nad9* transcript with the −202 5′ termini was only detected after introduction of the At1g62670 gene from Col (Figure 3B). In addition, both the CR-RT-PCR and the RNA gel blot analysis consistently showed that the presence of this gene in C24 leads to increased amounts of the *nad9* mRNAs with the −202 5′ ends, while the transcripts with the −243 5′ ends were reduced in comparison to the C24 wild type. In C24, another mRNA of ~740 nucleotides is detected. This *nad9* mRNA species contains the −133 5′ end (Figure 1, panel C24). In summary, these data unambiguously demonstrate that At1g62670 encodes RPF2 required for the formation of considerable quantities of *nad9* mRNAs with 5′ ends around nucleotide −202. The T-DNA insertion allele in SALK 146198 is accordingly designated *rpf2-1*.

To further explore the role of RPF2 in mRNA processing, we investigated whether this protein can bind to *nad9* RNAs. To this

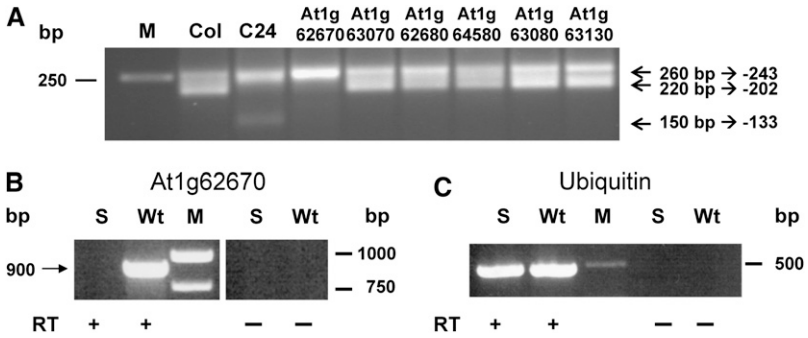


Figure 2. Phenotypes of *nad9* 5′ Termini in Different *Arabidopsis* Mutant Lines.

(A) Nad9 mRNA extremities were analyzed by CR-RT-PCR analysis as described in Methods. While the T-DNA insertions in At1g63070 (SALK 128068), At1g62680 (SALK 139736), At1g64580 (SALK 087235), At1g63080 (SALK 020638), and At1g63130 (SALK 020230) did not alter the 5′ end pattern of *nad9* transcripts in comparison to Col wild type, an insertion in At1g62670 (SALK 146198) abolished the detection of the *nad9* 5′ ends at −202 as seen in the C24 wild type. In this mutant, the *nad9* −133 5′ end is not detectable. CR-RT-PCR analysis of SALK 146198 was performed with three independent biological samples in at least three technical replicates. Lines with CR-RT-PCR product patterns identical to the wild type were investigated with single tests.

(B) RT-PCR analysis of SALK 146198 (S) mutant and Col wild type. An At1g62670 cDNA product of the expected size of 900 bp was amplified from wild-type RNA, while no such fragment is detectable in the mutant (RT +). No product is obtained without reverse transcriptase (RT −), demonstrating that the product derives from cDNA and not from DNA contaminations.

(C) The integrity and the use of equal amounts of the RNA and cDNA are indicated by the amplification of a 500-bp product from the ubiquitin 10 mRNA.

end, electrophoretic mobility shift assays (EMSAs) were performed with recombinant RPF2 and a synthetic *nad9* transcript covering the −202 5′ ends from –296 to –124 with respect to the translation initiation codon (NATG, $n = -1$). This experiment revealed that RPF2 indeed has the potential to bind to this RNA, further confirming its importance in 5′ end processing of *nad9* transcripts (see Supplemental Figure 6 online).

Knockout of *RPF2* Affects the 5′ Ends of *cox3* Transcripts

To investigate if the 5′ termini of other mitochondrial mRNAs are affected by the knockout of different *Arabidopsis* PPR genes with similarity to *RF* genes, we analyzed the 5′ and 3′ ends of all mitochondrial mRNAs in Col, L*er*, C24, *rpf2-1*, and several additional lines with T-DNA insertions in *RF*-like PPR genes (Figure 4; see Supplemental Figure 5 online). In Col and L*er*, a CR-RT-PCR product of 230 bp indicated a single *cox3* 5′

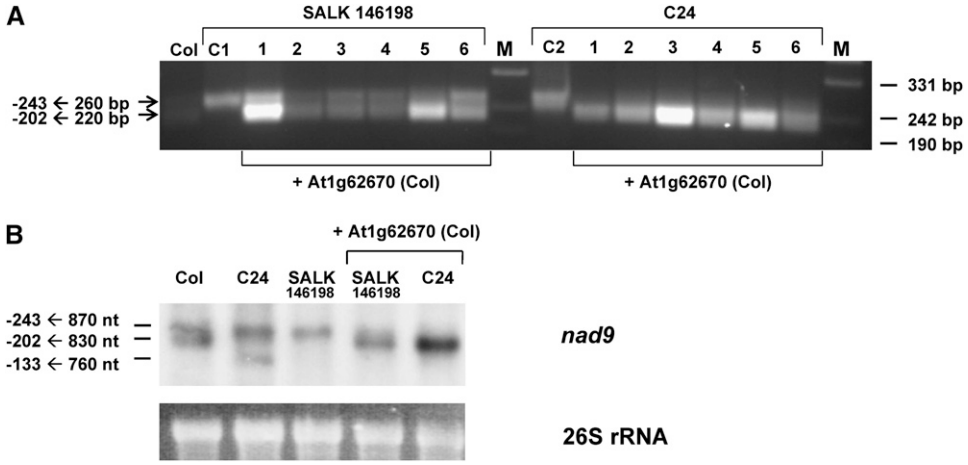


Figure 3. The At1g62670 Gene from Col Restores the Efficient Formation of the *nad9* −202 5′ Ends in SALK Line 146198 and in C24.

(A) A 4.75-kb construct carrying the At1g62670 from Col was transformed into At1g62670 knockout line SALK 146198 and into the C24 wild type (+At1g62670). *Nad9* mRNA of six individual transformants of each type was inspected by single CR-RT-PCRs. C1 (SALK 146198) and C2 (C24) are untransformed control plants.

(B) RNA gel blot analysis of total RNA from Col, C24, and SALK 146198, as well as SALK 146198 and C24 complemented with the At1g62670 gene from Col (+At1g62670). Loading of the gel was regulated by visualization of 26S rRNA.

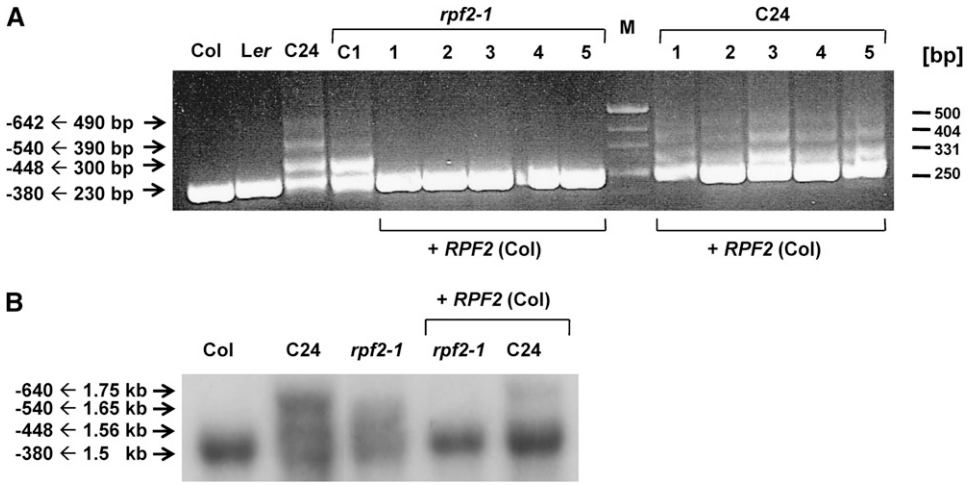


Figure 4. RPF2 and a C24-Specific Mitochondrial *cis*-Element Influence *cox3* mRNA 5′ End Formation.

(A) The 5′ termini of the *cox3* transcripts were determined by single CR-RT-PCRs in wild-type plants from accession Col, L*er*, C24, and *rpf2-1* (C2), as well as five *rpf2-1* and five C24 plants transformed with the *RPF2* complementation construct (+*RPF2* Col).
(B) RNA gel blot analysis of total RNA from Col, C24, and *rpf2-1*, as well as C24 and *rpf2-1* containing the *RPF2* gene from Col (+*RPF2* Col) using a *cox3* probe.

transcript end at position −380 relative to the ATG, while in C24, larger products (300, 390, and 490 bp) showed the presence of additional termini located upstream of the mature 5′ end at −448 (Figure 4A). These results are consistent with previous observations (Forner et al., 2005, 2008). The knockouts of the other *RF*-like PPR genes investigated did not exhibit any altered 5′ termini of mitochondrial transcripts (see Supplemental Figure 5 online); however, we found that *cox3* mRNA ends are affected in the *rpf2-1* mutant (Figure 4A). In addition to the single major *cox3* mRNA 5′ end seen in Col wild type (−380), an additional CR-RT-PCR product was amplified from the *rpf2-1* mutant. Sequence analysis of this product identified a *cox3* 5′ end at position −448 (Figure 4A; see Supplemental Figure 7 online). This observation suggested that RPF2 is also involved in 5′ end maturation of *cox3* transcripts. To substantiate this result, we also inspected *cox3* transcripts in *rpf2-1* and C24 plants containing the Col *RPF2* gene, which restored the *nad9* −202 5′ end formation. In complemented *rpf2-1* plants, a single major 5′ end of *cox3* mRNA is detected by CR-RT-PCR analysis as in Col wild type (Figure 4A). By contrast, the complemented C24 plants still have additional *cox3* 5′ transcript ends due to a specific *cis*-element in the mitochondrial DNA of this accession (Forner et al., 2005, 2008). This result was confirmed by a RNA gel blot analysis of complemented *rpf2-1* and C24 plants (Figure 4B). In both lines, a *cox3* probe detected mature *cox3* mRNAs of ~1.5 kb and two additional *cox3* RNA species of 1.56 and 1.65 kb. An additional *cox3* RNA species of 1.75 kb was found in C24 consistent with previously mapped ends (Forner et al., 2005). These *cox3* RNA patterns changed substantially after the introduction of the *RPF2* gene into both lines. In the *rpf2-1* knockout line, the efficient generation of the 1.5-kb mature *cox3* mRNA was fully restored by the presence of RPF2. Likewise, the presence of the Col *RPF2* gene in C24 plants enhanced the formation of the mature *cox3* mRNA; however, in these plants, *cox3* RNA species of 1.75 and 1.56 kb were still detectable.

Taken together, these results confirm that in both Col and C24, 5′ end formation of *cox3* transcripts is influenced by *RPF2*. In C24, 5′ end maturation of *cox3* mRNAs is additionally affected by a mitochondrially encoded *cis*-element (Forner et al., 2005).

Impaired 5′ Processing Does Not Interfere with Protein Accumulation

The defective *RPF2* gene results in the loss of one of the *nad9* mRNA species and a substantial reduction of mature *cox3* mRNA (Figures 3 and 4). A potential impact of these deficiencies on accumulation of the *nad9* polypeptide was investigated by an immunodetection analysis with an antiserum against the *nad9* polypeptide from wheat (*Triticum aestivum*; Lamattina et al., 1993). In mitochondrial lysates from cell suspension cultures established from the Col wild-type and the *rpf2-1* knockout lines, respectively, proteins with an apparent molecular mass of ~21 kD were detected, which is consistent with the size expected for the *nad9* polypeptide (Figure 5). In both lines, the *nad9* subunits were present in identical amounts, demonstrating that the almost complete lack of the *nad9* mRNA with the −202 ends in the *rpf2-1* mutant has no influence on accumulation of this protein. This result was confirmed by two-dimensional blue native (BN)/SDS-PAGE of mitochondrial protein of the above-mentioned lines and a C24 wild-type culture. In addition, *cox3* polypeptides were found in identical amounts in all lines investigated, and the total amount of protein complexes and their composition was indistinguishable between these lines (see Supplemental Figure 8 online). These results demonstrate that the altered *nad9* and *cox3* mRNA patterns in C24 and *rpf2-1* do not interfere with accumulation of the respective polypeptides.

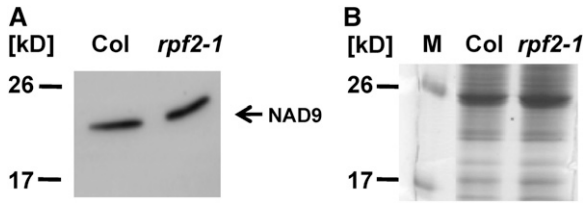


Figure 5. Immunodetection Analysis of Mitochondrial Protein from Col Wild Type or *rpf2-1*.

(A) An antiserum against the *nad9* polypeptide from wheat binds to a protein with an apparent size of ~21 kD, consistent with the calculated size of *nad9* in *Arabidopsis* (22.7 kD). No difference was observed between the amounts of the *nad9* polypeptides in wild-type and *rpf2-1* mutant plants.

(B) Loading of the gel was controlled by Coomassie blue staining.

Defective *RPF2* Alleles in Different *Arabidopsis* Accessions

In C24, impaired *nad9* and *cox3* 5′ end processing strongly suggests that this accession contains a defective *RPF2* allele. To resolve the structure of the *RPF2* gene in this and other accessions deficient in *nad9* and *cox3* 5′ end processing (Oy-1 and Yo-0), the gene was amplified with primer pair At1g62670.H/At1g62670.R (see Supplemental Table 2 online) annealing to sequences at the 5′ and 3′ extremities of the *RPF2* reading frame. L*er* and Col were included as controls. In Col, L*er*, and Yo-0, products of 2.0 kb were amplified from total DNA, as expected from the genomic sequence from Col-0 (see Supplemental Figure 9 online) (*Arabidopsis* Genome Initiative, 2000). By contrast, a 2.4-kb product was obtained in C24 while in Oy-1 two fragments of 3.7 and 0.4 kb were generated. These results suggest the presence of different types of *RPF2* alleles defective in *nad9* and *cox3* 5′ end processing. Two of these alleles seem to have alternative structures, as indicated by the sizes of the PCR products that differ from the expected size in Col (C24 and Oy-1), while another one has a size identical with Col (Yo-0). Sequencing of the PCR products obtained from C24 and Oy-1 confirmed the aberrant *RPF2* gene structures and revealed that these alleles are chimeras composed of *RPF2* and At1g62930 sequences (see Supplemental Figures 10 and 11 online). In contrast with these chimeric alleles, the *RPF2* gene in Yo-0 contains 13 single nucleotide polymorphisms, which alter the amino acid sequence of RPF2 in comparison to Col.

To obtain information about the function of the *RPF2* alleles from Oy-1 and Yo-0, these alleles were tested for the potential to complement impaired *nad9* and *cox3* 5′ end processing in the *rpf2-1* mutant. We found that both alleles of *RPF2* were incapable of restoring efficient processing in *rpf2-1* (data not shown). On the contrary, introduction of the Col *RPF2* gene into these accessions led to an enhanced generation of a *nad9* mRNA with −202 5′ ends as well as efficient and complete *cox3* 5′ end processing. These results demonstrate that the *RPF2* alleles of Oy-1 and Yo-0 were nonfunctional with respect to their contribution to 5′ end processing of *nad9* and *cox3* mRNAs (see Supplemental Figure 12 online). In summary, these data indicate that the *RPF2* gene in *Arabidopsis* does not encode a vital protein and that both changes in gene structure or alteration of amino acid identities can lead to the loss of function in 5′ end processing of *nad9* and *cox3* transcripts.

RPF2 Is a Mitochondrial Protein

As outlined above, RPF2 is a nuclear-encoded factor expected to be imported into mitochondria. To investigate the subcellular localization of RPF2, DNA fragments representing the full-length *RPF2* reading frame or 80 amino acids of its N terminus were fused in frame upstream of the gene encoding the green fluorescent protein (GFP; Davis and Vierstra, 1998). These constructs were transiently transformed into protoplasts prepared from transgenic tobacco (*Nicotiana tabacum*) stably expressing the isovaleryl-CoA dehydrogenase:red fluorescent fusion protein (eqFP611) as a mitochondrial marker (Forner and Binder, 2007). Inspection of transformed protoplasts by microscopy identified the GFP fluorescence exclusively in mitochondria, as indicated by the almost identical pattern of the green and the red fluorescence (Figure 6; see Supplemental Figure 13 online). This result suggests that the native RPF2 protein is imported into mitochondria in vivo.

DISCUSSION

RPF2 Is Involved in 5′ End Formation of the *nad9* and *cox3* mRNAs

We recently found several nuclear encoded natural polymorphisms affecting the 5′ termini of mitochondrial transcripts in distinct *Arabidopsis* accessions (Forner et al., 2008). Taking advantage of this natural genetic variation, we now identified At1g62670 as *RPF2*, a gene encoding a PPR protein. Several lines of evidence unambiguously demonstrate the function of the identified gene in 5′ end formation of mitochondrial *nad9* and *cox3* mRNAs. First, the knockout of At1g62670 in Col results in the absence of *nad9* 5′ ends around nucleotide −202 (Figures 2A and 3B) and in the accumulation of additional *cox3* 5′ transcript termini (Figure 4). Second, these phenotypes could be restored in vivo by introducing the *RPF2* wild-type allele from Col into the *rpf2-1* knockout line or into C24 (Figures 3 and 4). Third, C24, Oy-1, and Yo-0 contain aberrant *RPF2* alleles. Accordingly, these accessions lack the *nad9* mRNAs with −202 5′ ends and accumulate additional *cox3* RNAs with 5′ termini located further upstream than in accessions with functional *RPF2* alleles.

RPF2, an Efficiency Factor for the Maturation of *cox3* mRNAs

As shown by our analysis, RPF2 has at least two functions. It is required for the generation of substantial amounts of *nad9* mRNA with 5′ ends around −202, and it is involved in the formation of 5′ ends of the *cox3* mRNAs. In the latter process, the inactivation of *RPF2* provokes the accumulation of additional abundant *cox3* RNAs with 5′ ends at position −448 (1.56-kb RNA species) and −540 (1.65-kb RNA species), upstream of the major *cox3* 5′ end (−380; 1.5-kb mature mRNA) (Figure 4). The relative amounts of these transcripts indicate that the occurrence of the larger *cox3* mRNA species is accompanied by a decrease of the mature *cox3*

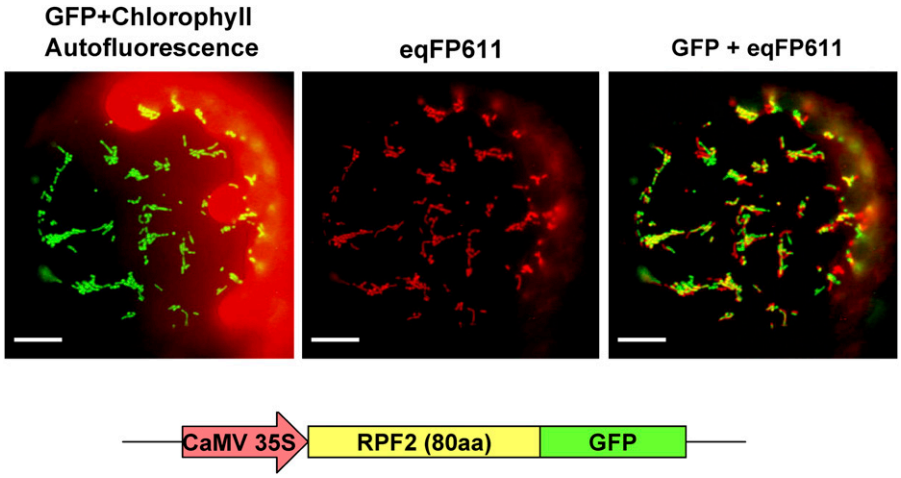


Figure 6. RPF2 Is a Mitochondrial Protein.

A fusion protein consisting of the RPF2 mitochondrial targeting sequence (80 amino acids) and GFP was transiently expressed in transgenic tobacco protoplasts. Fluorescence of GFP and chlorophyll and the red fluorescent protein from the sea anemone *Entacmaea quadricolor* (eqFP611) were visualized through different filter sets. eqFP611 C-terminally fused to the N-terminal part of the mitochondrial isovaleryl-CoA dehydrogenase has previously been established as a mitochondrial marker in transgenic tobacco (Forner and Binder, 2007). A merger (GFP + eqFP611) of both images demonstrates the match of both patterns (yellow represents colocalization). Small differences were due to the rapid movement of mitochondria in the protoplast. Bars = 10 μm.

mRNA. This indicates that the 5′ extended RNA molecules are intermediates occurring during the generation of the mature *cox3* mRNA, suggesting that the inactivation of RPF2 reduces the efficiency of this process. Thus, RPF2 is probably an efficiency factor for the generation of mature *cox3* mRNA and is potentially required for different steps of this process, as indicated by the different 5′ extended *cox3* precursor molecules. The accumulation of the *cox3* processing intermediates argues against a function of this protein in modulating mRNA stability or a role of RPF2 as *cox3* transcription factor. Furthermore, no difference in the efficiency in RNA editing of *nad9* and *cox3* mRNAs has been observed between *rpf2-1* and wild-type plants, indicating that an involvement of RPF2 in RNA editing is highly unlikely.

The dual function of RPF2 in 5′ processing of *nad9* and *cox3* transcripts is reminiscent of the MMT (modifier of mitochondrial transcripts) locus in *Brassica*. The MMT locus also influences transcripts of two mitochondrial genes (*ccmF* and *nad4*); however, the changes of the transcript patterns of these genes might be linked to different genes within this nuclear locus (Singh et al., 1996).

Maturation of *nad9* Transcripts

In contrast with the role of RPF2 in *cox3* mRNA processing, the function of this protein in *nad9* mRNA formation is less obvious. In C24 and *rpf2-1* plants, the substantial reduction in the abundance of the short *nad9* mRNAs with the 5′ ends around −202 is not accompanied by an apparent increase of the larger mRNA with the −243 5′ end or any other *nad9* precursor molecules (Figure 3B; see Supplemental Figure 14 online). This observation suggests that the larger *nad9* RNA might not serve as an intermediate or precursor for the formation of the short *nad9* mRNA. On the other hand, the introduction of the intact *RPF2* allele from Col into *rpf2-1* knockout plants results in the generation of the *nad9* −202 5′ ends, while the *nad9* mRNA species with the −243 end is reduced. This effect is particularly strong when the Col *RPF2* gene is brought into C24. In these complemented plants, the short *nad9* mRNA accumulates in relatively high amounts, while the larger *nad9* transcript almost completely disappears (Figure 3B). This result is confirmed by the absence of the 260-bp CR-RT-PCR product containing the −243 5′ end (Figure 3A). These observations suggest that the larger *nad9* mRNA can indeed serve as precursor for the generation of the short *nad9* transcript with −202 5′ termini. The particularly strong accumulation of the short *nad9* mRNA in the complemented C24 plants might be explained with specific features of the nuclear background, but this phenomenon needs further investigation.

The apparent lack of accumulation of *nad9* precursors in the *rpf2-1* knockout line and in C24 might be explained by very moderate increases of precursor RNAs of different sizes originating from various upstream located promoters (Kühn et al., 2009).

Taken together, our data suggest that RPF2 is a factor that enhances 5′ processing of *cox3* mRNAs and also of *nad9* transcripts.

Mode of 5′ End Maturation of *nad9* and *cox3* mRNAs

The exact molecular function of RPF2 in 5′ processing of *nad9* and *cox3* transcripts is presently unclear. We could not detect a 5′ leader molecule with a 3′ end located directly upstream of the −202 5′ mRNA termini that would be indicative of endonucleolytic processing. Nevertheless, a 5′ to 3′ exonucleolytic generation of the different 5′ ends seems to be rather unlikely, since no

such enzymatic activity has ever been reported in plant mitochondria (Gagliardi and Binder, 2007). Moreover, such an exonucleolytic formation of the *nad9* −202 5′ ends would mean that RPF2 has to protect the −202 end from degradation, similar to what has been observed for the plastidal PPR10, which binds in a sequence-specific manner to different RNAs, protecting them from 5′ as well as 3′ exonucleolytic degradation (Pfalz et al., 2009). However, such a potential function of RPF2 in mRNA protection would be consistent with a function as stability factor, which is unlikely, as indicated by the increase of *cox3* precursor RNAs in the *rpf2-1* mutant, as mentioned above.

Considering all aspects, RPF2 is most likely a factor that enhances the endonucleolytic generation of the *nad9* −202 and *cox3* −380 5′ ends. A direct function as endonuclease is very unlikely, since the *cox3* −380 5′ terminus and the *nad9* −202 5′ ends are generated in the absence of RPF2. An endonucleolytic activity has been assigned to several DYW domain PPR proteins (Okuda et al., 2009), one of which (At2g02980) is targeted to mitochondria (Nakamura and Sugita, 2008). RPF2 does not contain a DYW domain required for endonucleolytic cleavage, supporting that RPF2 might interact with such PPR proteins or other endonucleases. In this scenario, RPF2 might interact with mitochondrial RNaseP and RNaseZ, enzymes normally involved in endonucleolytic tRNA processing. However, the latter has recently been found to cleave also some mitochondrial tRNA-like structures, so-called t-elements, which retained some features of real tRNAs but lost their original function (Canino et al., 2009). In plant mitochondria, such t-elements are present at 5′ or 3′ ends of several mRNAs, indicating that both RNaseP an RNaseZ are also involved in posttranscriptional generation of mRNA extremities. It seems conceivable that RPF2 might direct these endonucleases to higher-order structures that determine processing sites otherwise not or barely cleaved by these endonucleolytic enzymes. This model, which favors definition of the cleavage sites by secondary structure, is also compatible with the observation that the primary sequence at the *nad9* −202 and *cox3* −380 5′ ends is not conserved. The involvement of secondary structures is also implicated by the C24-specific mitochondrial DNA influencing the *cox3* 5′ end maturation process (Forner et al., 2005). Here, sequences located at least 200 bp upstream of −380 provoke the generation of alternative *cox3* 5′ termini. Such an effect across a long distance is best explained by the formation of alternative secondary structures, which interfere with 5′ processing of *cox3* mRNAs. Further studies will be necessary to resolve potential processing complexes and to unambiguously define *cis*-elements.

RPF2 Is a Variable Restorer of Fertility-Like PPR Protein

In recent years, a number of PPR proteins have been identified as *RF* genes in CMS systems from various species (Bentolila et al., 2002; Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003; Hanson and Bentolila, 2004; Wang et al., 2006; Chase, 2007). CMS is a maternally inherited deficiency to produce or release viable pollen (Budar and Pelletier, 2001; Chase, 2007). This trait, which is linked to chimeric, CMS-specific mitochondrial genes, can be used to generate F1 hybrids in many crops. Here, crosses between male sterile lines with fertile lines containing nuclear-encoded *RF* genes produce hybrid seeds. The presence of the *RF* genes restores male fertility in the F1 plants, most often by altering the expression of the CMS-specific mitochondrial genes at the posttranscriptional level. Most proteins encoded by the *RF* genes are highly similar to each other and to a group of PPR proteins from the autogamous plant species *Arabidopsis*, where CMS has not been found (Desloire et al., 2003; Schmitz-Linneweber and Small, 2008). RPF2, comprising 15 PPR repeats, groups with *RF*-like PPR proteins that form a small cluster of ~20 genes within the huge family of 442 PPR genes identified in *Arabidopsis* (Lurin et al., 2004; O'Toole et al., 2008). Among the restorers of fertility, RPF2 is very similar to RF1B (32% identical amino acids) and RF1A (31%) from rice (*Oryza sativa*) as well as to RF592 (35%) from petunia (*Petunia* × *hybrida*; Bentolila et al., 2002; Wang et al., 2006). While the molecular function of RF592 in restoration of *Petunia* CMS is unclear, the two RF proteins from rice exhibit distinct fertility restoration functions in the Boro II CMS system. RF1A promotes the endonucleolytic cleavage of the discistronic *B-atp6/orf79* transcript, which is consistent with the role of RPF2 in the *nad9* and *cox3* mRNA processing. Cleavage by RF1A generates two different mRNAs with the *B-atp6* and the *orf79* reading frame, respectively. This cleavage restores pollen fertility since the *orf79* mRNA is inefficiently translated. RF1B enhances the degradation of the discistronic *B-atp6/orf79* mRNA by an as yet unknown mechanism. The concerted action of RF1A and RF1B substantially reduces levels of the toxic *orf79* protein, which allows the formation of fertile pollen (Wang et al., 2006; Kazama et al., 2008).

Interestingly, *Arabidopsis RPF2* is a relatively variable gene with three distinct alleles being inactive in *nad9* and *cox3* 5′ end processing in different accessions. It presently cannot be excluded that these alleles retained or even gained other yet unknown functions; however, their divergent structures suggest that another function common to these different alleles is rather unlikely. The diversity between these alleles seems only possible since the loss of the short *nad9* mRNA and the impaired *cox3* 5′ end processing have no influence on protein accumulation and, thus, no detrimental effects on plant survival (Figure 5; see Supplemental Figure 4 online).

The sequences of the *RPF2* loci from C24 and Oy-1 originate from different *RF*-like PPR genes (like At1g62930, At1g63130, or At1g62910), suggesting that chimeric gene structures frequently arise from recombinations between these highly similar genes. In line with this assumption, the *RPF2* alleles from C24 and Oy-1 probably do not originate from a single recombination event, but are likely the outcome of independent recombinations, as indicated by the different parts of At1g62930, which are linked to the N-terminal portion of At1g62670 (see Supplemental Figures 6 and 7 online). This generation of new PPR genes is related to the Death-and-Birth model, which suggests either nonconservative or conservative transposition followed by rapid loss of a gene. This leads to a so-called nomadic character of PPR genes with highly variable positions in otherwise colinear segments of closely related genomes (Geddy and Brown, 2007).

In summary, our analysis reveals basic information about the intrinsic function of the *RF*-like PPR genes in *Arabidopsis* and allows insights into the evolution of such genes in this model species.

METHODS

Plants and Plant Cultivation

Arabidopsis thaliana and *Nicotiana tabacum* petit Havana plants and *Arabidopsis* cell suspension cultures were established and cultivated as described previously (Forner et al., 2007, 2008).

Analysis of RNA

Phenotypes (i.e., the 5′ ends of mitochondrial mRNAs) were determined by CR-RT-PCR analysis as described (Kuhn and Binder, 2002; Forner et al., 2007, 2008). Briefly, total RNA (up to 5 μg) was self-ligated by RNA ligase (10 units) in a bulk reaction (total volume 100 μL). After desalting on Microcon YM-10 microconcentrators, ligated RNA was used as template for first-strand cDNA synthesis using 200 units of M-MLV reverse transcriptase under conditions recommended by the manufacturer. One-tenth of the cDNA was used as template for a PCR across the ligated 3′ and 5′ ends of a given RNA species. PCR products were then separated by agarose gel electrophoresis, visualized by ethidium bromide staining, and recovered from the gel using the GFX Gel Band purification kit according to instructions given by the manufacturer (GE Healthcare). RNA extremities were determined by sequence analyses of the PCR products.

To determine *nad9* 5′ends, cDNA synthesis was initiated with oligonucleotide Atnad9-1 followed by PCR with the primer pair Atnad9-3/Atnad9-7. For the 5′ end mapping of *cox3* mRNAs, cDNA synthesis was initiated from oligonucleotide Atcox3-3′PS followed by PCR with primers Atcox3-Mega3′ and Atcox3-Mega5′. Total RNA and mtRNA were isolated using an RNeasy plant mini kit (Qiagen). To discriminate between primary ends derived from transcription initiation (carrying 5′ triphosphate groups) and secondary ends generated by posttranscriptional processing (having 5′ monophosphate groups), 8 μg of mitochondrial RNA was digested with 1 unit of Terminator phosphate-dependant exonuclease according to the instructions given by the manufacturer (Epicentre Biotechnologies). Both exonuclease-treated and untreated RNAs were subsequently analyzed by primer extension analysis according to standard protocols (Sambrook and Russel, 2000). Oligonucleotides used in the primer extension experiments were nad9-2 and atp9-2. All primer sequences are given in Supplemental Table 2 online. RNA gel blot hybridization was performed with Duralon UV membranes according to instructions given by the manufacturer (Stratagene). Hybridization probes were generated by PCR with primer pairs cox3-5′ Primersonde/cox3-3′ Primersonde (*cox*3) and Atnad9-NS.H/Atnad9-NS.R (*nad9*).

Investigation of Nuclear and Mitochondrial DNAs

Crude and fine mapping of the genomic region containing *RPF2* was done using insertion/deletion and single nucleotide polymorphism markers (Jander et al., 2002; Schmid et al., 2003). PCRs were performed with genomic DNA isolated with the DNeasy plant mini kit (Qiagen) from 14- to 18-d-old plants. A complete list of all markers used is given in Supplemental Table 1 online.

The *RPF2* gene was amplified with oligonucleotide pair At1g62670.H/At1g62670.R from ~200 ng total DNA from 14-d-old seedlings of the different accessions. The *cox3* upstream DNA fragments were amplified with primers Atcox3-2 and Atcox3-23 for the Col/C24 configurations and primers Le-cox3-HindIII.H and Le-cox3-HindIII.R for the L*er* configuration. These PCRs were performed with total DNA isolated as described above using Go-Taq DNA polymerase according to the instructions given by the manufacturer (Promega).

In Vivo Complementation

Complementation constructs were established in pMDC123 using standard procedures (Curtis and Grossniklaus, 2003). The *RPF2* gene (At1g62670), including the 1.6-kb upstream and 1.3-kb downstream sequence, was amplified with primers At1g62670-Kompl.H and At1g62670-Kompl.R, digested with *Asc*I and *Pac*I, and cloned into the respective sites in pMDC123. Since we did not sequence the cloned PCR fragments, we transformed a mixture of 20 different clones per construct to avoid potential effects of PCR errors. Plant transformation was done by floral dip (Clough and Bent, 1998). Selection of transformed plants was done with Basta (120 mg/L; Bayer CropScience). Selected plants were analyzed as described above.

Miscellaneous Methods

Two-dimensional analysis of 1 mg of total mitochondrial protein obtained from Percoll purified organelles was done by BN/SDS-PAGE as described previously (Wittig et al., 2006; Matthes et al., 2007). Immunodetection assays were performed with protein extracted from mitochondria isolated and purified as described before (Matthes et al., 2007). The antiserum raised against *nad9* from wheat (*Triticum aestivum*) was used at a 1/10,000 dilution.

EMSAs used gel-purified RNA directly transcribed from PCR products (amplified with primers Atnad9-EMSA.H + Atnad9-EMSA.R; see Supplemental Table 2 online) using a T7 RNA polymerase. The synthetic *nad9* transcript corresponds to sequences in the 5′-untranslated region ranging from position −296 to −124 with respect to the translation initiation codon (NATG, $n = -1$), which includes the −202 and −243 5′ ends. The binding reactions were essentially performed as described previously with slight modifications (Williams-Carrier et al., 2008). Two different buffer systems were used. Reactions performed with buffer B1 contained 20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.2 mM EDTA, 4 mM DTT, 0.4 μg/μL BSA, 5% (v/v) glycerol, 20 units Ribolock (Fermentas), 150 pM RNA, and 1 μg of purified RPF2 or control lysate (see below) in a total volume of 10 μL. Assays performed in buffer B2 contained 50 mM Na phosphate, pH 8.0, 12.5 mg/mL heparin, 20 mM NaCl, 0.2 mM EDTA, 4 mM DTT, 0.4 μg/mL BSA, 20 units Ribolock (Fermentas), 150 pM RNA, and 150 to 300 ng of purified protein or control protein extract.

To generate recombinant protein, the *RPF2* reading frame was amplified without the putative mitochondrial targeting sequence (24 amino acids), with primers At1g62670UEx.H2 and At1g62670UEx.R. The PCR product was cloned into the *Bam*HI/*Sac*I sites in the expression vector pET32a (Novagen). Expression and affinity purification using S-protein agarose was performed according to the instructions of the manufacturer (Novagen), except that bacteria (*Escherichia coli*) were first grown at 37°C to an OD_{600} of 0.4 and then kept for 20 min at 20°C and, after induction with L-arabinose (0.2% [w/v]) and isopropyl β-D-1-thiogalactopyranoside (1.0 mM), grown for another 2 h at 20°C. A control lysate was generated under identical experimental conditions, except that an empty pET32a vector was used.

An RPF2 fragment corresponding to the first 80 amino acids was amplified with oligonucleotides At1g62670-Kompl.H3 and At1g62670-Kompl.R5 and cloned into the *Sma*I site of the psmGFP4 vector. The RPF2 full-length reading frame was amplified with primers At1g62670-Kompl.H3 and At1g62670-Kompl.R3 and analogously cloned into the psmGFP4 vector. Protoplast transformation and fluorescence microscopy were done as described previously (Matthes et al., 2007).

DNA sequencing was commercially performed (4baselab). All other basic methods of molecular biology and plant analysis were done following standard protocols (Weigel and Glazebrook, 2002). In silico sequence analyses were done at the National Center for Biotechnology Information server, applying various BLAST tools (McGinnis and Madden, 2004), and amino acid sequence alignments were done with ClustalW2 (Larkin et al., 2007).

Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession

numbers: *RPF2*, At1g62670; *RPF2* allele in Oy-1, FN665655; *RPF2* allele in Yo-0, FN665658; and *RPF2* allele in C24, FM211647.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Detection of the −202 5′ End of nad9 mRNA in C24 and *rpf2-1*.

Supplemental Figure 2. The *nad9* −243 and −202 Transcript Ends Are Generated by Posttranscriptional Processing.

Supplemental Figure 3. Phenotyping of *nad9* mRNA Extremities in an F2 Mapping Population.

Supplemental Figure 4. Rough Mapping of the *RPF2* Locus in Chromosome 1.

Supplemental Figure 5. CR-RT-PCR Analyses of Mitochondrial mRNA Extremities in PPR T-DNA Insertion Mutants.

Supplemental Figure 6. RPF2 Binds to a *nad9* Transcript.

Supplemental Figure 7. Mapping of the *cox3* −448 5′ Terminus.

Supplemental Figure 8. Accumulation of *nad9* and *cox3* Polypeptides Is Not Affected in the *rpf2-1* Mutant.

Supplemental Figure 9. PCR Analysis of *RPF2* Alleles in Different Accessions.

Supplemental Figure 10. RPF2 Alleles in Accessions Col, C24, Oy-1, and Yo-0.

Supplemental Figure 11. Alignment of Amino Acid Sequence Deduced from the *RPF2* Alleles (62670) from Accessions Col, Yo-0, Oy-1, and C24 and from the At1g62930 Gene (62930) from Col.

Supplemental Figure 12. Complementation Studies with RPF2 Alleles from Oy-1 and Yo-0.

Supplemental Figure 13. RPF2 Is a Mitochondrial Protein.

Supplemental Figure 14. RNA Gel Blot Analysis of *nad9* Transcripts in Col, C24, and rpf2-1 Plants.

Supplemental Table 1. List of Genetic Markers Used for Mapping of the *RPF2* Locus.

Supplemental Table 2. Oligonucleotide Sequences.

ACKNOWLEDGMENTS

We thank Gonny Guha and Uli Tengler Ulm University for excellent technical assistance and Anna Rapp Ulm University for her help in the mutant genotyping. We also thank Dagmar Lewejohann and Hans-Peter Braun from the Leibniz University in Hannover for their help with BN/SDS-PAGE analyses, the plant mitochondria group from the Institut de Biologie Moléculaire des Plantes in Strasbourg for the kind gift of the *nad9* antiserum, and Solomon Stonebloom (Berkeley) for his help with the *nad9* immunodetection assay and his valuable comments on the manuscript. J.F. was a fellow of the Studienstiftung des Deutschen Volkes. This work was supported by the Deutsche Forschungsgemeinschaft (Bi 590/10-1).

Received March 11, 2009; revised January 20, 2010; accepted February 4, 2010; published February 26, 2010.

REFERENCES

Andrés, C., Lurin, C., and Small, I. (2007). The multifarious roles of PPR proteins in plant mitochondrial gene expression. Physiol. Plant. **129:** 14–22.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408:** 796–815.

Bentolila, S., Alfonso, A.A., and Hanson, M.R. (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc. Natl. Acad. Sci. USA **99:** 10887–10892.

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

Brown, G.G., Formanova, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W.Y., and Landry, B.S. (2003). The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. **35:** 262–272.

Budar, F., and Pelletier, G. (2001). Male sterility in plants: Occurrence, determinism, significance and use. C. R. Acad. Sci. III **324:** 543–550.

Canino, G., Bocian, E., Barbezier, N., Echeverria, M., Forner, J., Binder, S., and Marchfelder, A. (2009). Arabidopsis encodes four tRNase Z enzymes. Plant Physiol. **150:** 1494–1502.

Chase, C.D. (2007). Cytoplasmic male sterility: A window to the world of plant mitochondrial-nuclear interactions. Trends Genet. **23:** 81–90.

Clough, S.J., and Bent, A.F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana.* Plant J. **16:** 735–742.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. **133:** 462–469.

Davis, S.J., and Vierstra, R.D. (1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol. Biol. **36:** 521–528.

de Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D. (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in *Arabidopsis thaliana.* Plant Cell **19:** 3256–3265.

Desloire, S., et al. (2003). Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep. **4:** 588–594.

Forner, J., and Binder, S. (2007). The red fluorescent protein eqFP611: Application in subcellular localization studies in higher plants. BMC Plant Biol. **7:** 28.

Forner, J., Hölzle, A., Jonietz, C., Thuss, S., Weber, B., Schwarzländer, M., Meyer, R.C., and Binder, S. (2008). Mitochondrial mRNA polymorphisms in different *Arabidopsis thaliana* accessions. Plant Physiol. **148:** 1106–1116.

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.

Forner, J., Weber, B., Wietholter, C., Meyer, R.C., and Binder, S. (2005). Distant sequences determine 5′ end formation of *cox3* transcripts in *Arabidopsis thaliana* ecotype C24. Nucleic Acids Res. **33:** 4673–4682.

Gagliardi, D., and Binder, S. (2007). Expression of the plant mitochondrial genome. In Plant Mitochondria, D. Logan, ed (Ames, IA: Blackwell Publishing), pp. 50–95.

Geddy, R., and Brown, G.G. (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:** 130.

Hanson, M.R., and Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell **16** (suppl.): S154–S169.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., and Last, R.L. (2002). *Arabidopsis* map-based cloning in the post-genome era. Plant Physiol. **129:** 440–450.

Kazama, T., Nakamura, T., Watanabe, M., Sugita, M., and Toriyama, K. (2008). Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J. **55:** 619–628.

Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S., and Imamura, J. (2003). Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J. **34:** 407–415.

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

Kuhn, J., and Binder, S. (2002). RT-PCR analysis of 5′ to 3′-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. Nucleic Acids Res. **30:** 439–446.

Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., Falcon de Longevialle, A., O'Toole, N., Börner, T., Millar, A.H., Small, I.D., and Whelan, J. (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of *Arabidopsis thaliana.* Plant Cell **21:** 2762–2779.

Lamattina, L., Gonzalez, D., Gualberto, J., and Grienenberger, J.M. (1993). Higher plant mitochondria encode an homologue of the nuclear-encoded 30-kDa subunit of bovine mitochondrial complex I. Eur. J. Biochem. **217:** 831–838.

Larkin, M.A., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947–2948.

Lurin, C., et al. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16:** 2089–2103.

Matthes, A., Schmidt-Gattung, S., Köhler, D., Forner, J., Wildum, S., Raabe, M., Urlaub, H., and Binder, S. (2007). Two DEAD-box proteins may be part of RNA-dependent high-molecular-mass protein complexes in Arabidopsis mitochondria. Plant Physiol. **145:** 1637–1646.

McGinnis, S., and Madden, T.L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. **32:** W20–25.

Nakamura, T., and Sugita, M. (2008). A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. FEBS Lett. **582:** 4163–4168.

Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. Plant Cell **21:** 146–156.

O'Toole, N., Hattori, M., Andres, 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.

Perrin, R., Lange, H., Grienenberger, J.M., and Gagliardi, D. (2004a). AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in *Arabidopsis thaliana* mitochondria. Nucleic Acids Res. **32:** 5174–5182.

Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M., and Gagliardi, D. (2004b). Two exoribonucleases act sequentially to process mature 3′-ends of atp9 mRNAs in *Arabidopsis* mitochondria. J. Biol. Chem. **279:** 25440–25446.

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:** 2042–2052.

Sambrook, J., and Russel, D.W. (2000). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schmid, K.J., Sörensen, T.R., Stracke, R., Törjek, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B. (2003). Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana.* Genome Res. **13:** 1250–1257.

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

Singh, M., Hamel, N., Menassa, R., Li, X.Q., Young, B., Jean, M., Landry, B.S., and Brown, G.G. (1996). Nuclear genes associated with a single *Brassica* CMS restorer locus influence transcripts of three different mitochondrial gene regions. Genetics **143:** 505–516.

Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., and Brennicke, A. (2008). The process of RNA editing in plant mitochondria. Mitochondrion **8:** 35–46.

Wang, Z., et al. (2006). Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell **18:** 676–687.

Weigel, D., and Glazebrook, J. (2002). Arabidopsis: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Williams-Carrier, R., Kroeger, T., and Barkan, A. (2008). Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. RNA **14:** 1930–1941.

Wittig, I., Braun, H.P., and Schagger, H. (2006). Blue native PAGE. Nat. Protoc. **1:** 418–428.

Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana.* Plant Cell **21:** 558–567.

RNA PROCESSING FACTOR2 Is Required for 5' End Processing of nad9 and cox3 mRNAs in Mitochondria of Arabidopsis thaliana
Christian Jonietz, Joachim Forner, Angela Hölzle, Sabine Thuss and Stefan Binder
PLANT CELL 2010;22;443-453; originally published online Feb 26, 2010;
DOI: 10.1105/tpc.109.066944

This information is current as of December 21, 2010

Supplemental Data	http://www.plantcell.org/cgi/content/full/tpc.109.066944/DC1
References	This article cites 44 articles, 23 of which you can access for free at: http://www.plantcell.org/cgi/content/full/22/2/443#BIBL
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs for *THE PLANT CELL* at: http://www.plantcell.org/subscriptions/etoc.shtml
CiteTrack Alerts	Sign up for CiteTrack Alerts for *Plant Cell* at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists

ADVANCING THE SCIENCE OF PLANT BIOLOGY

RNA PROCESSING FACTOR 3 is crucial for the accumulation of mature *ccmC* transcripts in mitochondria of *Arabidopsis thaliana* accession Columbia

Christian Jonietz, Joachim Forner, and Stefan Binder

Institut Molekulare Botanik, Universität Ulm, D-89069 Ulm, Germany

This work was supported by the Deutsche Forschungsgemeinschaft (Bi 590/10-1).

Corresponding author:

Stefan Binder

E-mail: stefan.binder@uni-ulm.de

Fax: ++49 731 502 2626

ABSTRACT

RNA PROCESSING FACTORS 1 and 2 are pentatricopeptide repeat (PPR) proteins involved in 5' processing of different mitochondrial mRNAs in *Arabidopsis thaliana*. Both factors are highly similar to RESTORERS OF FERTILITY (RF), which are part of cytoplasmic male sterility/restoration systems in various plant species. These findings suggest a predominant role of RF-like PPR proteins in posttranscriptional 5' processing. To further explore the functions of this group of proteins we examined a number of T-DNA lines carrying insertions in the corresponding PPR genes. This screening identified a nearly complete absence of mature *ccmC* transcripts in an At1g62930 T-DNA insertion line, a phenotype that could be restored by the introduction of the intact At1g62930 gene into the mutant. The insertion in this nuclear gene, which we now call RNA PROCESSING FACTOR 3, also leads to a severe reduction of the *CcmC* protein in mitochondria. The analysis of C24/Col F_2 hybrids lacking functional RPF3 genes revealed that this gene has less influence on the generation of mature *ccmC* 5' transcript end derived from a distinct *ccmC* 5' upstream configuration found in mitochondrial DNAs from C24 and other accessions. Our new results further substantiate the fundamental role of RF-like PPR proteins in the posttranscriptional generation of plant mitochondrial 5' transcript termini. The data further show that particular a function of an RF-like protein is required only in connection with a distinct mtDNA configuration.

INTRODUCTION

Plant mitochondria contain a complex genetic framework to realize the genetic information encoded in their DNA (Kubo and Newton, 2008). These systems include many different proteins required for the posttranscriptional processing of the organellar transcripts. Among these factors, pentatricopeptide repeat (PPR) proteins play a crucial role (Delannoy et al., 2007; Schmitz-Linneweber and Small, 2008). In higher plants, PPR proteins, which are characterized by degenerate 35-amino acid repeats arranged in tandem, form one of the largest protein families (Small and Peeters, 2000; Lurin et al., 2004), comprising more than 400 members in *Arabidopsis thaliana*, now referred to as *Arabidopsis*. PPR proteins predominantly localize to chloroplasts or/ and mitochondria and take part in virtually all processes affecting RNA. This includes processes that link the formation of RNA extremities to RNA stability. In chloroplasts, this is achieved by specific binding of the PPR proteins to the target RNA (Loiselay et al., 2008; Pfalz et al., 2009; Johnson et al., 2010). The binding leads to the formation of so-called "caps", which define the RNA ends by building barriers to 5' or/and 3' exonucleases. This relatively simple mechanism, which has been found for PPR 10 from maize and similarly for two other proteins from *Chlamydomonas*, might account for the formation and stabilization of the majority of chloroplast RNAs. But PPR proteins influence also endonucleolytic processing reactions. For instance binding

of PPR5 protects a distinct $tRNA^{Gly}$ precursor transcript from endonucleolytic cleavage in maize chloroplasts (Beick et al., 2008).

In mitochondria, a number of nuclear encoded PPR proteins also influence processing, accumulation and stability of RNAs. These are primarily proteins that function as RESTORERS OF FERTILITY (RF) in various cytoplasmic male sterility/restoration systems described in different species. Most RF proteins reconstitute male fertility by altering transcript levels and/or patterns of CMS-related mitochondrial genes, which reduces expression of the corresponding CMS-associated proteins (Budar and Pelletier, 2001; Bentolila et al., 2002; Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003; Wang et al., 2006; Schmitz-Linneweber and Small, 2008). Up to now, the exact mechanisms influencing the steady state levels of the CMS-specific transcripts are mostly unknown. The sole exception is the Boro II CMS/restoration system in rice. Here RF1A stimulates the cleavage of the CMS-related *B-atp6/orf79* dicistronic transcript probably by directing an unknown endonuclease to a specific cleavage site (Wang et al., 2006; Kazama et al., 2008). Analogous functions have been suggested for RNA PROCESSING FACTORS 1 and 2. These PPR proteins, which are highly similar to RFs, participate in 5' processing of the major transcripts of *nad4*, *nad9* and *cox3* genes in mitochondria of the autogamous species *Arabidopsis*, for which no CMS lines have been described (Hölzle et al., 2010; Jonietz et al., 2010).

Here we report the identification of RNA PROCESSING FACTOR 3 (RPF3), which is required for formation of the mature *ccmC* mRNA in mitochondria of the *Arabidopsis* accession Columbia (Col). The gene was identified by a screening of mutants with T-DNA insertions in *RF*-like PPR genes. The presence of the T-DNA in the *RPF3* gene provokes a nearly complete absence of the mature *ccmC* transcripts, accompanied by an increase of corresponding precursor RNAs. The recovery of *ccmC* transcript accumulation by the introduction of the intact *RPF3* gene into the mutant unambiguously demonstrates its function in 5' processing of *ccmC* transcripts. In contrast to Col, RPF3 is less important for the formation of a mature *ccmC* 5' transcript end, which is generated from a specific mtDNA configuration found in C24 and other accessions (Forner et al., 2008). The extremely reduced amount of the *CcmC* protein in mitochondria of the *rpf3-1* mutant has no negative influence on the activity of the respiratory chain or on plant fitness. These observations suggest that unlike in bacteria, the plant mitochondrial *CcmC* protein does not play an important role in cytochrome *c* biogenesis.

RESULTS

At1g62930 is required for the accumulation of *ccmC* transcripts

We have recently identified two genes involved in the generation of 5' ends of *nad4*, *nad9* and *cox3* mRNAs in mitochondria of *Arabidopsis*, (Hölzle et al., 2010; Jonietz et al., 2010). Both genes, *RNA PROCESSING FACTOR 1* (*RPF1*, At1g12700) and *RNA PROCESSING FACTOR 2* (*RPF2*, At1g62670) encode pentatricopeptide repeat (PPR) proteins with high similarity to RESTORER OF FERTILITY (RF) active in cytoplasmic male sterility (CMS)-restoration systems in various plant species (Budar and Pelletier, 2001; Bentolila et al., 2002; Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003; Wang et al., 2006). This suggests an important role of RF-like PPR proteins in 5' end formation of mitochondrial mRNAs in the autogamous species *Arabidopsis*. Therefore, a CR-RT-PCR approach was used to analyze major mRNA extremities in various lines with T-DNA insertions in *RF*-like PPR genes (data not shown). In almost all mutants, no aberrant PCR product patterns were found when compared to wild type. This indicates that these PPR genes do not play an important role in the generation of main 5' and 3' ends of mitochondrial transcripts. However, in the SAIL line 18E04, we could not amplify a CR-RT-PCR product that would be consistent with the previous results obtained for the *ccmC* mRNAs in accessions Col, L*er* and C24 (560 and 470 bp, respectively) (Forner et al., 2008). Instead, comparatively weak products of about 700 and 760 bp were generated in the mutant (Fig. 1A). This result suggested the absence of the mature *ccmC* mRNA and a slightly elevated accumulation of alternative *ccmC* transcripts in this mutant.

To examine the *ccmC* transcripts by an alternative approach, a northern blot analysis was performed with a probe covering the *ccmC* reading frame (from position 208 to 607 with respect to the ATG). In Col, L*er* and C24 total RNA, *ccmC* transcripts of about 1200 (Col and L*er*) and 1100 (C24) nucleotides were detected, which is consistent with previous mapping results (Fig. 1B) (Forner et al., 2007; Forner et al., 2008). In addition, weak signals corresponding to RNAs with sizes of about 1.4, 1.8 and 2.6 kb were seen in all RNA samples. In contrast, almost no mature *ccmC* mRNA was detected in the SAIL 18E04 mutant. Instead, the approximately 2.6 kb RNA was found to be increased. In addition, approximately 1.35 and 2.15 kb *ccmC* RNAs were detected, which were not seen in the wild-type RNAs of the different accessions (Fig. 1B). The elevated accumulation of the mutant-specific larger *ccmC* RNAs was confirmed by a probe corresponding to regions upstream of the *ccmC* reading frame (-513 to –734 with respect to the ATG). This hybridization experiment identifies the 1.35, 2.15 and 2.6 kb transcripts as *ccmC* precursor RNAs. To the contrary, no increase was observed for the signals corresponding to the 1.4 and 1.8 kb RNAs.

In addition, a primer extension analysis of total mitochondrial RNA isolated from cell suspension cultures established from Col, C24 and the SAIL 18E04 line was performed (Fig. 1C). In Col wild

type, strong extension products of about 370 and 390 nucleotides and a relatively weak 350-nucleotide cDNA correspond to 5' ends at -500, -484/482 and -465, all of which correlate with mature *ccmC* transcripts of approximately 1.200 nucleotides. In the mutant and C24, the 390- and 370-nucleotide extension products are barely detectable, which is consistent with the almost complete absence of the mature 1200-nucleotide *ccmC* transcripts in these lines (Forner et al., 2008). Only the 350-nucleotide extension product is detectable in the mutant or even slightly increased in C24. In this accession, a very strong 270-nucleotide cDNA corresponds to the –391/390 5' termini of the mature *ccmC* mRNA of about 1.100 nucleotides. These ends are also present in Col and the mutant, but only in trace amounts. Apart from that extension products of 440 and 650 nucleotides are present in similar amounts in all investigated lines or are overrepresented in the mutant (550, 1300 and 1900 nucleotides).

Since the primer extension experiment indicated a complex 5' end pattern for the mature *ccmC* transcript in Col, we reexamined this end by sequencing of ten clones obtained from the 570-bp CR-RT-PCR product. This revealed two clusters of ends (Fig. 2). One group of termini found in five clones surrounds the previously described ends at -484/-482, which had been determined by direct sequencing of a PCR product (Forner et al., 2007). The second cluster comprising four clones is found at positions -501 to -503. Both clusters correspond to the primer extension products of 370 and 390 bp, respectively, suggesting that the mature *ccmC* transcript from Col has indeed two 5’ ends located about 20 nucleotides apart from each other. An additional 5' end at position -476 was found represented only in single clone.

At1g62930 rescues the formation of mature *ccmC* mRNAs in the SAIL 18E04 mutant

To test whether the formation of mature *ccmC* mRNAs is indeed linked to the T-DNA insertion in At1g62930, a 2.6 kb DNA fragment containing the At1g62930 gene and 0.2 kb of the 5’ and 0.5 kb of the 3’ flanking sequences from Col wild type was stably introduced into the SAIL 18E04 mutant. After selection of transformants, the presence of the transgene was confirmed by PCR with T-DNA-specific primers. The accumulation of the mature *ccmC* mRNA with the –484/482 5' ends was then examined by CR-RT-PCR and northern blot analyses (Fig. 3).

In five independent transformants, a CR-RT-PCR analysis showed that the presence of At1g62930 rescues the formation of the mature *ccmC* mRNA. Using total cDNA as template, 560 bp PCR products were amplified in the transformed mutant as in Col and L*er* wild-type plants. No such product is seen in the untransformed mutant (Fig. 3A). Additionally, a northern blot analysis of two complemented plants confirmed that mature *ccmC* mRNAs with sizes identical to wild type

accumulate only in the presence of At1g62930, while no such transcript was seen in the untransformed mutant (1.200 nucleotides, Fig. 3B). These results unambiguously demonstrate that an intact At1g62930 gene is required for the formation of the mature *ccmC* mRNA in *Arabidopsis* accession Col. We designate this gene *RNA PROCESSING FACTOR 3* (*RPF3*) and the corresponding mutant *rpf3-1*.

RPF3 is a factor involved in the posttranscriptional generation of *ccmC* transcripts

The failure to generate mature *ccmC* mRNAs in the *rpf3-1* mutant can be attributed to a defect in transcription initiation or to a disturbance of posttranscriptional processing. To discriminate between both possibilities, we investigated whether the mature *ccmC* mRNAs carry 5' triphosphate or 5' monophosphate groups by primer extension analysis of Terminator exonuclease-treated mitochondrial RNA (mtRNA). This exonuclease specifically degrades RNA containing 5' monophosphate groups, which were generated by posttranscriptional cleavage. Primary RNAs having 5’ triphosphate groups, which derive from transcription initiation, are not degraded by this enzyme. As shown in Fig. 4A, strong extension products of approximately 270 and 370/390 nucleotides were obtained from untreated mtRNA preparations. These indicate the mature –391/390 in C24 and -484/482 and -501 to -503 5' termini in Col. Again further signals indicated 5' termini of precursor molecules, consistent with the results shown in Fig. 1. In contrast, no signals corresponding to the mature *ccmC* 5’ ends have been detected with Terminator exonuclease-treated mtRNA from both accessions. Only the less abundant extension products were seen after Terminator exonuclease-treatment. In both accessions, the strongest signals correspond to extension products of about 650 nucleotides. Additionally, relatively weak extension products of about 1 kb and more have been observed. These results strongly suggest that the -484/482 and -501 to -503 5' termini in Col are generated posttranscriptionally, which makes a function of RPF3 in transcription initiation unlikely. Similarly, the mature 5' ends of the *ccmC* transcripts in C24 are also formed posttranscriptionally.

To support the posttranscriptional origin of the major 5' termini in Col and to obtain information whether these ends could be derived from endonucleolytic processing, total mitochondrial RNA was searched for the presence of *ccmC* 5’ leader RNAs which might result from a cleavage at positions -484/482 or -501 to -503. To this end, the mitochondrial 18S rRNA was used as “anchor”, which was linked to the 3' end of such potential leader molecules in a bulk RNA ligation reaction. After cDNA synthesis initiated from an oligonucleotide annealing to the 18S rRNA the applied PCR strategy selected for *ccmC* 5’ leader-18S rRNA ligation products (Fig. 4B). After a single PCR, a

product with a size of about 410 nucleotides was generated from Col mtRNA (Fig. 4C), while a much weaker product of the same size was amplified from *rpf3-1* mtRNA. A half nested approach using an alternative primer annealing to 18S rRNA amplified a strong product of approximately 210 nucleotides. Almost no cDNA product of this size was seen in the mutant. The 210-bp product was inserted into pCR 2.1-TOPO vectors and 9 individual clones were sequenced. These sequences identified 5' termini at positions -492 to -497 (Fig. 3D). Taken together these results demonstrate that *ccmC* 5' leader molecules exist in *Arabidopsis* mitochondria. The 3' ends of these molecules are located 10 to 14 nucleotides upstream of the -484/482 and around eight nucleotides downstream of the -503 to -501 5' ends of the mature *ccmC* mRNA. However, the preferential amplification of these leader molecules from wild-type mtRNA demonstrates that their accumulation depends on a functional *RPF3* gene. This correlation strongly suggests a function of RPF3 in endonucleolytic cleavage.

A *CcmC* protein is detectable in the *rpf3-1* mutant

The above described northern blot hybridization revealed (Fig. 1) that the mature *ccmC* transcript is almost completely absent in the mutant, which might be at least partially compensated by increased amounts of *ccmC* precursor molecules. To determine the abundance of transcripts containing the coding sequence (CDS) in the mutant, Real-Time quantitative RT-PCR was performed. Amplification of a product spanning the *ccmC* CDS from positions 273 to 368 relative to the ATG (A = +1) revealed a relative amount of *ccmC* RNAs containing the reading frame of about 84 % in *rpf3-1* compared to Col wild type. In contrast, an amplification reaction covering *ccmC* precursor molecules from –510 to –607 revealed about six times more *ccmC* precursor RNA species in the mutant, confirming a defect in *ccmC* 5' processing in the *rpf3-1* mutant. These data suggested that a *CcmC* protein might still be translated in the mutant. To test this assumption, an immunodetection analysis was performed with total mitochondrial protein from *rpf3-1* and Col wild-type cell suspension cultures, respectively. Using an antiserum against the *CcmC* polypeptide, a strong signal corresponding to a protein with the apparent molecular mass of about 34 kDa detected the *CcmC* protein in Col wild type (Fig. 5). To the contrary, a much weaker signal shows a substantially reduced accumulation of the *CcmC* polypepetide in *rpf3-1* but confirmed that still minor amounts of this protein can be synthesized in this mutant.

RPF3 is a mitochondrial protein

RPF3 is nuclear-encoded protein involved in a processing event that occurs in mitochondria. Accordingly this protein is expected to be imported into theses organelles. To test this assumption, a DNA fragment encoding 80 amino acids at the N-terminus of RPF3 was fused in frame upstream of the gene encoding the green fluorescence protein (GFP)(Davis and Vierstra, 1998). After transient transformation of this construct into transgenic tobacco protoplasts, which express a isovaleryl-CoA dehydrogenase:red fluorescence fusion protein as mitochondrial marker (Forner and Binder, 2007), the subcellular sorting of the fusion proteins was inspected by fluorescence microscopy. Green fluorescence was exclusively observed in mitochondria, as demonstrated by the nearly identical pattern of the green and red fluorescence (Fig. 6). This result strongly suggests that RPF3 is also targeted to mitochondria *in vivo*.

RPF3 has a minor influence on the generation of the mature *ccmC* transcript in C24

In a previous analysis of mitochondrial transcript extremities in different *Arabidopsis* accessions we identified a *ccmC* 5' end polymorphism between accessions Col and C24 (Forner et al., 2008). The analysis of reciprocal Col/C24 F_1 hybrids clearly demonstrated that this polymorphism is linked to differences in the mitochondrial DNA sequences, most likely to a small segment located upstream of both accession-specific 5' ends (Supplemental Fig. S2). Since RPF3 is crucial for the generation of the mature 5' end of the *ccmC* transcript in Col, we wanted to know whether this protein might also be required for 5' processing of the mature *ccmC* transcript derived from the C24 mtDNA configuration. Thus C24 (egg cell donor) was crossed with the *rpf3-1* mutant (pollen donor). Since in *Arabidopsis* mtDNA is maternally inherited, the offspring of this cross contains the C24 *ccmC* upstream configuration, which was confirmed by PCR. In the F_2 generation, plants homozygous for the *rpf3-1* T-DNA allele were selected by PCR and investigated for their ability to generate mature *ccmC* transcripts. As shown in Fig 7, CR-RT-PCR products with sizes identical to the C24 wild-type control could be amplified from two F_2 plants. This result was confirmed by a northern blot hybridization, which revealed mature *ccmC* mRNAs with identical sizes in these F_2 and C24 wild-type plants. These results demonstrate that the function of the *RPF3* gene is per se not required for the generation of substantial amounts of the mature C24-type *ccmC* mRNA in these F_2 hybrids and probably also in C24. This might indicate that a distinct mtDNA configuration, i.e. in the *ccmC* upstream region, might be associated with a particular version of the *RPF3* reading frame. Thus we examined the *RRF3* coding region in C24. A PCR analysis of the CDS sequence generated products of identical sizes in Col and C24 (data not shown). Sequencing of the product from C24 identified

35 amino acid polymorphisms in comparison to Col (Supplemental Tab. S1). None of the nucleotide exchanges interrupts the coding region. These results suggest that C24 retained a functional *RPF3* reading frame.

Taken together these results demonstrated that in both C24 and Col RPF3 is involved in *ccmC* transcript processing, but that the gene is of different importance for the generation of the mature 5' ends of *ccmC* mRNAs in the different accessions.

DISCUSSION

RPF3 is required for the generation of mature 5' ends of *ccmC* mRNAs in *Arabidopsis* accession Col

In *Arabidopsis* mitochondria, RPF1 and RPF2, two PPR proteins with high similarity to RESTORERS OF FERTILITY, are involved in the generation of the mature 5’ ends of *nad9*, *cox3* and *nad4* transcripts. This observation suggests that RF-like PPR proteins play a predominant role in 5' processing. In accordance with this hypothesis, several lines of evidence now demonstrate an analogous function of the RF-like PPR protein RPF3 in the formation of mature 5' end of *ccmC* mRNAs in *Arabidopsis* accession Col.

First, a T-DNA insertion in the *RPF3* gene prevents the accumulation of the mature *ccmC* transcript in the accession Col. Second, the generation of this mRNA could be restored by the introduction of the functional *RPF3* gene into *rpf3-1*. Third, as demonstrated by Real-Time quantitative RT-PCR, northern blot hybridization and primer extension analysis, severely impaired generation of the mature *ccmC* mRNA is accompanied by an increase of different 5’-extended precursor molecules strongly suggesting that RPF3 functions in posttranscriptional generation of the mature 5’end. Fourth, this function is corroborated by the analysis of Terminator exonuclease-treated RNA, which shows the investigated mature *ccmC* 5' ends in Col but also the mature 5' termini in C24 to be derived from processing rather than from transcription initiation, although the 5’-CRTA-3’ promoter motif if found in close vicinity to these transcript ends (Kühn et al., 2005; Forner et al., 2007). Fifth, the preferential detection of *ccmC* 5' leader molecules in Col wild type indicates that the accumulation of these by products depends on an intact *RPF3* gene. This observation strongly suggests a function of RPF3 in endonucleolytic cleavage, although the identified 3’ ends of these leader molecules slightly deviate from the detected 5' ends. This pattern of ends indicates that after the endonucleolytic cuts, a few nucleotides are removed most likely by 3’ exonucleases.

Alternatively, the few nucleotides might be lost by two endonucleolytic cuts in close vicinity to each other (e.g. -484/-482 mRNA 5' ends and -492 to -497 3' leader ends).

In summary, our data unambiguously demonstrate a function of RPF3 in the generation of the mature 5' end of the mitochondrial *ccmC* transcript in the accession Col. Thus three RF-like PPR proteins are now shown to be required for the posttranscriptional generation of 5' transcript termini corroborating the predominant role of this group of proteins for this process (Hölzle et al., 2010; Jonietz et al., 2010). In addition, RPF1 to RPF3 as well as RF1A from rice (Wang et al., 2006; Kazama et al., 2008) are involved in endonucleolytic cleavage while no evidence has been found that would support a function of these proteins as barriers to exonucleolytic degradation as it has been seen for some PPR proteins from chloroplasts (Loiselay et al., 2008; Pfalz et al., 2009; Johnson et al., 2010).

Function of RPF3 in mitochondria of *Arabidopsis* accession C24

In Col, the function of RPF3 is very important for the generation of the mature 5' ends since almost no mature *ccmC* transcripts are detectable in the corresponding *rpf3-1* mutant. In contrast, the analysis of C24 X Col F_2 plants without functional *RPF3* indicated, that this gene is *per se* not necessary for the generation of the -391/-390 5' termini of *ccmC* RNAs derived from the mtDNA configuration present in C24 and other accessions (Fig. 7)(Forner et al., 2008). In these plants still substantial amounts of mature *ccmC* mRNAs are detectable by northern blot analysis. However, the level of these transcripts is reduced to about 50 % (Fig. 7). This indicates that the RPF3 has a function in RNA processing of *ccmC* transcripts originating from two different upstream configurations. While in Col the absence of an intact RPF3 almost completely prevents accumulation of mature *ccmC* mRNAs, in C24 RPF3 has only a moderate stimulatory effect on the processing event that results in the -391/-390 5' termini.

The RPF3 protein from C24 shows 35 changes of amino acid identities in comparison to Col, most of them located in the N-terminal half. But all 15 repeats retain the characteristic degenerative motif, even repeat 6, where 12 amino acid exchanges were seen (Supplemental Table S1). These observations suggest that RPF3 might have a slightly different function in C24. However, we cannot exclude that the reduction of *ccmC* mRNA level in the F_2 plants is due to other factors in the nuclear genome, which is combined and mixed up from C24 and Col. Thus further analyses are required to elucidate the exact function of the *RPF3* gene in C24.

The mature *ccmC* mRNA is almost completely absent in *rpf3-1*

In the previously analyzed *rpf2-1* mutant still substantial amounts of mature *nad9* and *cox3* mRNA accumulate and accordingly no changes were seen in the steady-state levels of the corresponding proteins (Jonietz et al., 2010). Likewise, L*er* plants lacking a functional *RPF1* gene accumulate an approximately 170-nucleotide longer precursor RNA to amounts comparable to the normal mature *nad4* transcript found in Col with the intact *RPF1* (Hölzle et al., 2010). In contrast to these observations, the northern analysis shows that the presence of the T-DNA insertion in *RPF3* results in a nearly complete loss of the mature *ccmC* mRNA and accordingly only small amounts of this protein are detectable in the *rpf3-1* mutant. On the one hand there might be still enough mature *ccmC* mRNA to be translated. On the other hand it is also possible that one or the other species of the precursor molecules can be translated. It is rather unlikely though that all of the *ccmC* pre RNA species are used for protein biosynthesis. In this case more *CcmC* protein might be expected, since quantitative RT-PCR results indicated only a 15 % reduction of transcripts containing the *ccmC* reading frame (Supplemental Fig S1).

In bacteria, the *CcmC* protein is an essential component of the cytochrome *c* biogenesis pathway where it is involved in the translocation of heme *c* from the cytoplasm into the periplasm (Kranz et al., 2009). Likewise, plant *ccmC* might be involved in the delivery of heme *c* from the mitochondrial matrix into the inter membrane space (Giege et al., 2007). If mitochondrial *CcmC* also has a crucial function in plants, one might expect a reduction of the respiratory activity in the *rpf3-1* mutant. However, the activities of complex I, III and IV have been found to be identical in the mutant and in wild-type mitochondria (Hildebrandt, Jonietz, Binder and Braun, unpublished results). In addition, no changes were seen in the macroscopic phenotype or in the development of the mutant under normal growth conditions (Jonietz, and Binder, unpublished results). Thus mitochondria of *rpf3-1* seem to have retained normal activity, which can be explained by two scenarios. First, the residual amount of *CcmC* protein in the mutant still supports heme *c* translocation without negative consequences for the respiratory potential of mitochondria. But this explanation seems to be rather unlikely considering the strong reduction of the protein in *rpf3-1*. Second, the plant mitochondrial *CcmC* protein has no essential function in cytochrome *c* biogenesis in plants. This scenario is consistent with the fact that the enzymes required for protoporphyrin IX biosynthesis localize to chloroplasts. Moreover, the mitochondrial localization of ferrochelatase, which catalyzes the insertion of iron into the protoporphyrin IX, has not yet been unambiguously demonstrated. Thus it is speculated that heme *c* is delivered from the “outside” directly to the mitochondrial inter membrane space so that there is no need for a transport of heme *c* across the

inner mitochondrial membrane (Giege et al., 2007). In this scenario, *CcmC* might fulfill an alternative function, which is not essential for cytochrome *c* biogenesis and for the function of the respiratory chain. Thus the *rpf3-1* mutant is a valuable tool for further studies of *CcmC* and its role in cytochrome *c* biogenesis in mitochondria of higher plants.

MATERIAL AND METHODS

Plants and plant cultivation

Arabidopsis plants were grown and checked for correct genotypes as described (Forner *et al.*, 2008). Cell suspension cultures were established and cultivated as outlined previously (Köhler et al., 2010). The *Arabidopsis* SAIL line 18E 04 was obtained from NASC (http://arabidopsis.info/). Crosses to generate F_1 and F_2 hybrid plants were done following standard procedures (Weigel and Glazebrook, 2002).

Analysis of RNA

The 5' ends of mitochondrial *ccmC* mRNA were determined by CR (circularized RNA)-RT-PCR analysis (Kuhn and Binder, 2002; Forner *et al.*, 2007; Forner *et al.*, 2008; Jonietz *et al.*, 2010) using primers Atccb3Mega5‘.nah and Atccb3Mega3‘.nah. The 3’ termini of 5’ leader RNA molecules were determined by an established alternative 3’ RACE approach (Forner *et al.*, 2007). First strand cDNA synthesis was initiated from primer At18S-Adapter.R3 followed by the first PCR with primer pair Atccb3QPCR.3/At18S-Adapter.R3 and a second half-nested PCR with oligonucleotides Atccb3QPCR.3/At18S-Adapter.R4. PCR products were directly sequenced and cloned into pCR 2.1-TOPO (Invitrogen), respectively. *ccmC* 5' ends were determined by sequence analysis of ten clones. Total RNA and total mtRNA were isolated using RNeasy plant mini kits (Qiagen, http://www1.qiagen.com). Northern blot hybridization, Terminator exonuclease treatment and primer extension analyses were done as described formerly (Jonietz et al., 2010). Extension was initiated at oligonucleotide Atccb3Mega5‘.nah, the probes for the northern analyses were amplified using primer pairs Atccb3-21/Atccb3-22 (CDS) and Atccb3-Mega5’-fern/Atccb3-12 (upstream probe). PCR products were labeled using a random primed DNA labeling kit and [α^{32}P]dCTP following the recommendation given by the manufacturer (GE Healthcare).

Real-Time quantitative RT-PCR was done using Transcriptor High Fidelity cDNA Synthesis and Light-Cycler 480 SYBR Green I Master Kits on a Light-Cycler 480 Real-Time PCR System according to instructions given by the manufacturer (Roche). Data were evaluated using the Light-Cycler 480 software (1.5). Relative transcript levels were measured in respect to the validated

reference genes encoding ubiquitin conjugating enzyme 9 (UBC9, At4g27960) and protein phosphatase 2A subunit 3 (PP2A, At1g13320) with established primer pairs (Czechowski et al., 2005). Oligonucleotides for quantitative *ccmC* transcript analysis (Atccb3QPCR1 to 4) were selected using the primer design tool at the NCBI home page (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHomeAd).

Complementation assays

A complementation construct containing the *RPF3* gene as well as 184 and 480 bp upstream and downstream regions (amplified with primers At1g62930-Kompl.H and At1g62930-Kompl.R) was established in pMDC123 using standard procedures (Sambrook and Russel, 2000; Curtis and Grossniklaus, 2003). To circumvent potential effects by PCR errors a mixture of 20 different clones was used for each construct. Plant transformation was done by floral dip (Clough and Bent, 1998). The presence of the transgene was tested with a PCR with primers KAN1 and KAN2.

Miscellaneous methods

DNA sequencing was commercially performed (4baselab, http://www.4base-lab.de/index_4.html). Immunodetection analysis of mitochondrial protein was performed as described (Jonietz et al., 2010). Basic methods of molecular biology followed standard protocols (Sambrook and Russel, 2001). Protoplast transformation and fluorescence microscopy were done following an established protocol (Matthes et al., 2007). *In silico* sequence analyses were done at the NCBI server applying various blast tools (McGinnis and Madden, 2004). Oligonucleotide sequences are listed in Supplemental Table S2.

ACKNOWLEDGEMENTS

We are very grateful for the excellent technical assistance from Conny Guha and Uli Tengler. We thank Geraldine Bonard and Philippe Giegé from the IBMP in Strasbourg as well as Kazuyoshi Kitazak and Tomohiko Kubo from the Hokkaido University in Japan for the kind gift of *Nad9* and *CcmC* antisera. The efforts of T. Hildebrandt and H.P. Braun for testing respiratory chain complexes are deeply acknowledged. We also thank Katrin and Birgit Stoll for their technical support.

FIGURES

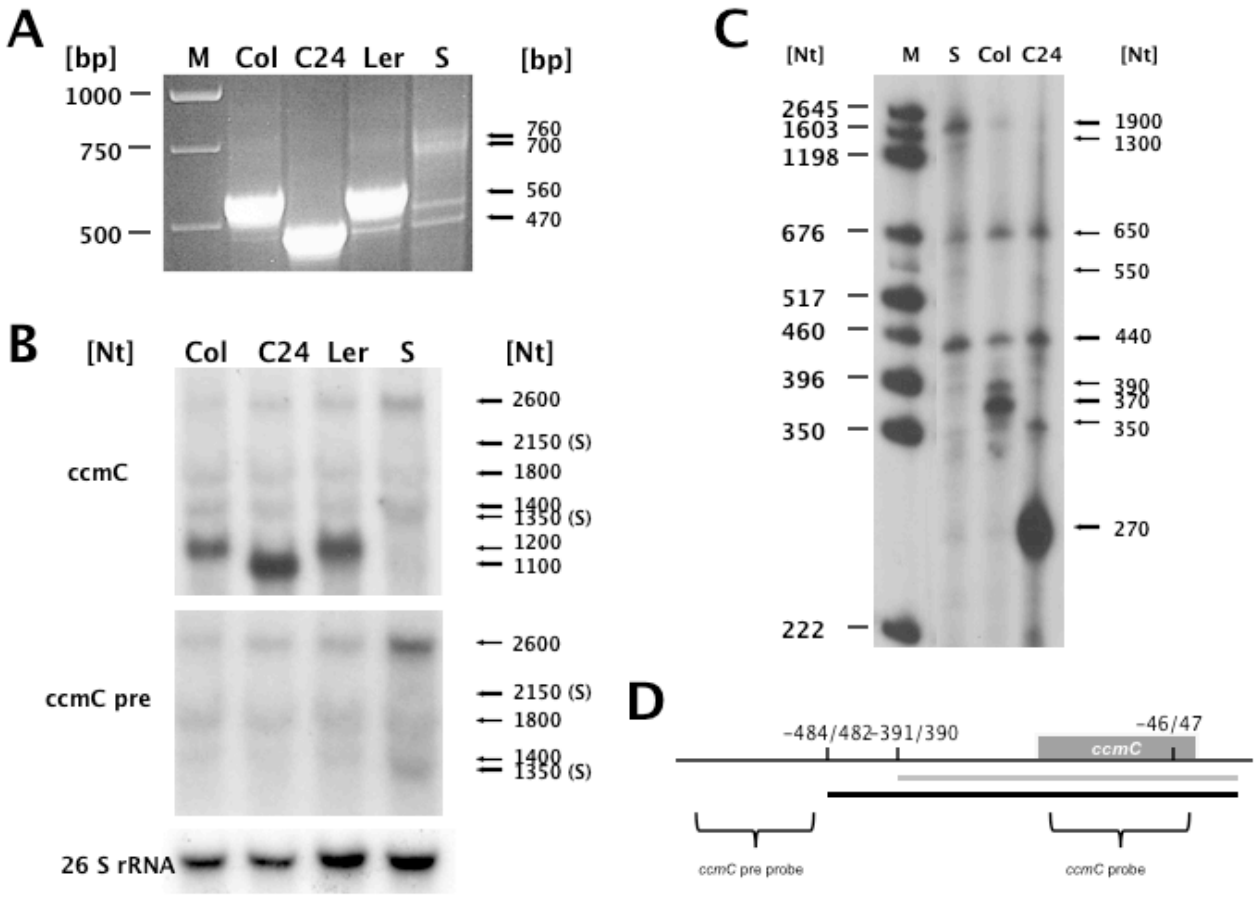


Figure. 1: Analysis of *ccmC* transcripts. RNAs derived from the *ccmC* gene in accession Col, C24 and L*er* and the mutant line SAIL 18E04 (S) were analyzed by CR-RT-PCR (A), northern blot hybridization (B) and primer extension analysis (C). Lengths of size markers are given on the left hand sides of the images. Relevant products or signals are indicated by small arrows and corresponding sizes on the right hand sides. Asterisks indicate mRNA species predominantly detected in the mutant (D) Schema of the *ccmC* gene (dark grey box) and the major transcripts (grey (C24) and black (Col) bold lines). Mature 5' (relative to the translation start codon NATG, N = -1) and 3’ ends (relative to the translation stop codon, TAAN, N = +1) are indicated. Probes used for the northern analysis are depicted in the bottom part of the schema and indicated on the left hand side of the images in (B).



Figure. 2: RNA extremities of *ccmC* RNAs. The 5' termini found in ten clones obtained after cloning of CR-RT-PCR products of Col *ccmC* transcripts (shown in Fig. 1A) are given as bent arrows above the sequence. Previously mapped 5' end are indicated by grey boxes (Forner et al., 2007). The 3' ends of *ccmC* leader RNAs determined by sequence analysis of nine individual clones derived from the 3’ RACE analysis of the *ccmC* 5’ leader molecules (see also Fig. 4B) are indicted by vertical arrows below the sequence.

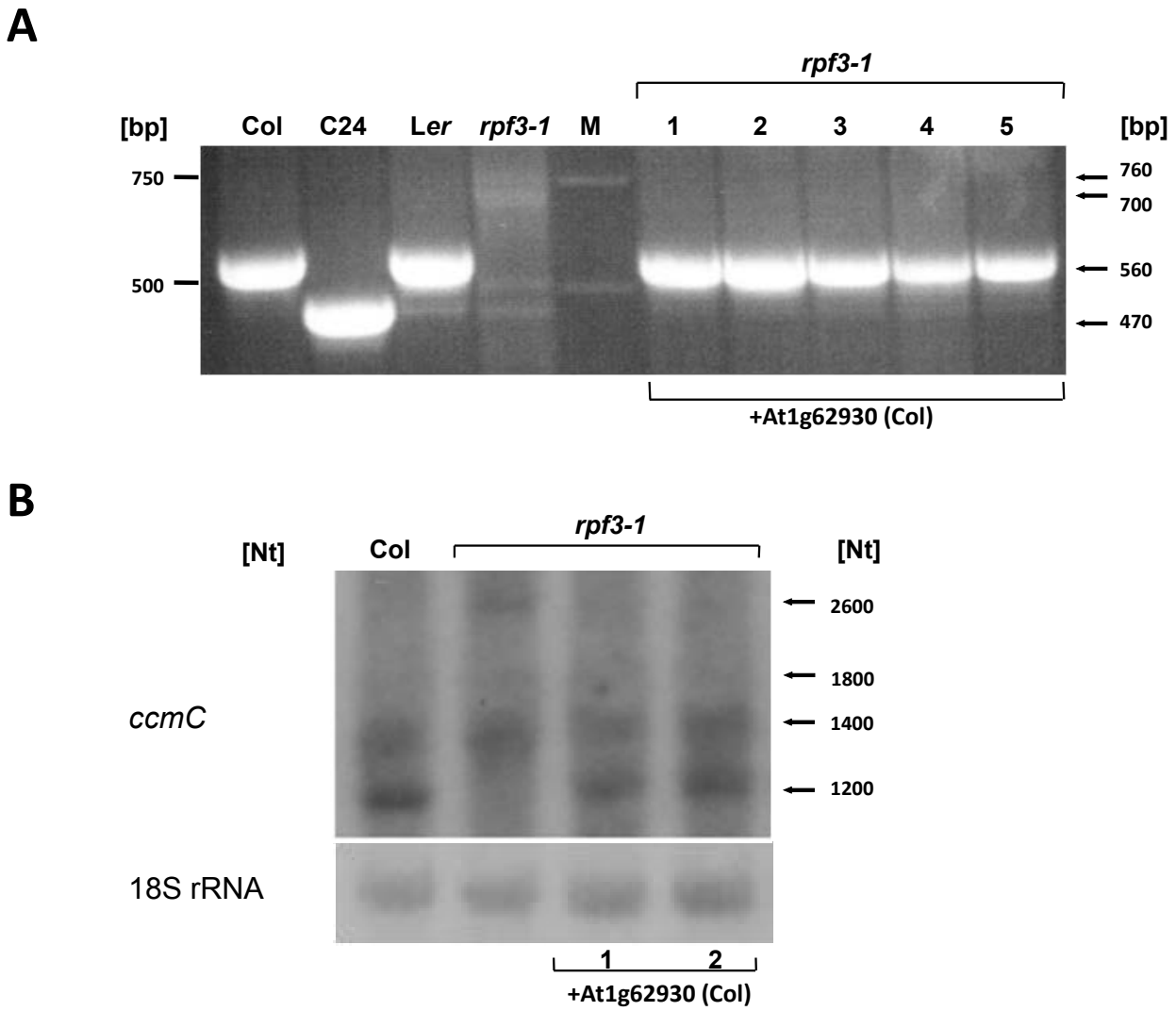


Figure. 3: Complementation of *rpf3-1* with At1g62930 rescues the formation of mature *ccmC* mRNA. (A) CR-RT-PCR analysis of five *rpf3-1* plants containing the wild –type allele of At1g62930. 560 bp products were amplified from total RNA from all of these plants and from Col and L*er* wild-type plants, while no such product was amplified from the un-transformed *rpf3-1* mutant. A 470 bp product was generated in C24. (B) A northern analysis confirmed the generation of the 1.200 nucleotide *ccmC* in two complemented plants. The size of the *ccmC* transcripts is identical in the complemented plants and in Col wild type. No mature *ccmC* mRNA was detected in the un-transformed mutant. Lengths of size markers are given in the left margin, relevant products are indicated in the right margin.

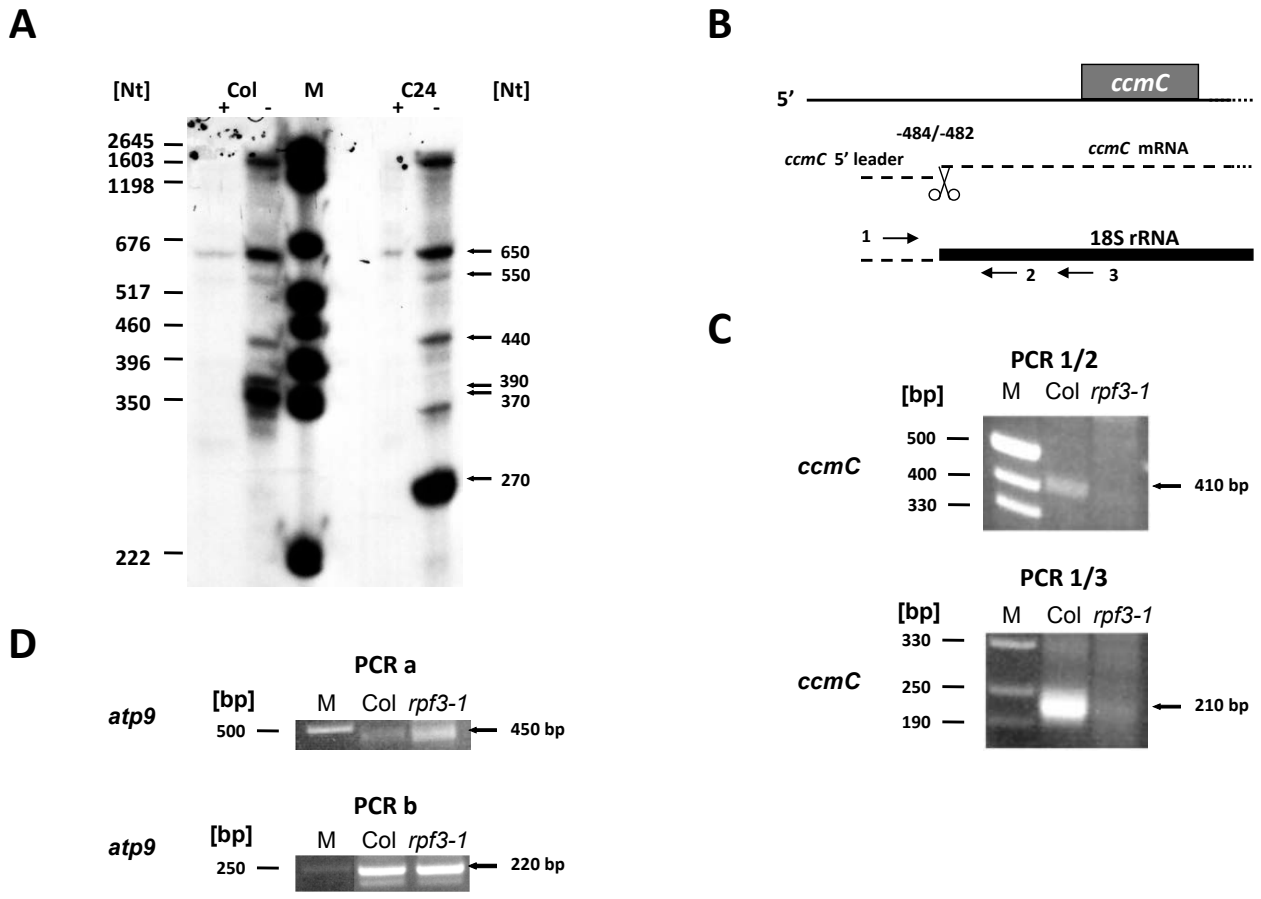


Figure. 4: The mature *ccmC* 5' ends are generated by posttranscriptional processing. (A) Primer extension analysis of untreated (-) and Terminator exonuclease-treated (+) mtRNA from Col and C24. (B) Schema explaining the analysis of 3' ends of potential *ccmC* 5' leader molecules. The *ccmC* reading frame is indicated by a grey box, *ccmC* RNAs are given as dashed lines. The 18S rRNA, which is used as anchor molecule, is depicted as a black bold line. Primers (horizontal arrows) used for this analysis are Atccb3QPCR.3 (1), At18S-Adapter.R4 (2) and At18S-Adapter.R3 (3). (C) PCR products amplified with primer pair 1/3 (upper panel) and 1/2 (lower panel) from Col wild-type and from *rpf3-1* total mtRNA. Lengths of size markers are given on the left, relevant products are indicated on the right hand side of the images (A and C). (D) Control reactions performed with primer pairs Atatp9Mega-Endo.A/ At18S-Adapter.R3 (PCR a) and Atatp9-11/ At18S-Adapter.R4 (PCR b) amplified *atp9* leader RNA molecules as described previously (Forner et al., 2007). These reactions were performed with identical volumes of the same cDNA preparation used in the PCRs shown in (C). The results of these amplifications demonstrate that in *rpf3-1* the reduced amount of PCR product derived of *ccmC* leader molecules is not due to the use of lower amounts or insufficient quality of the cDNA.

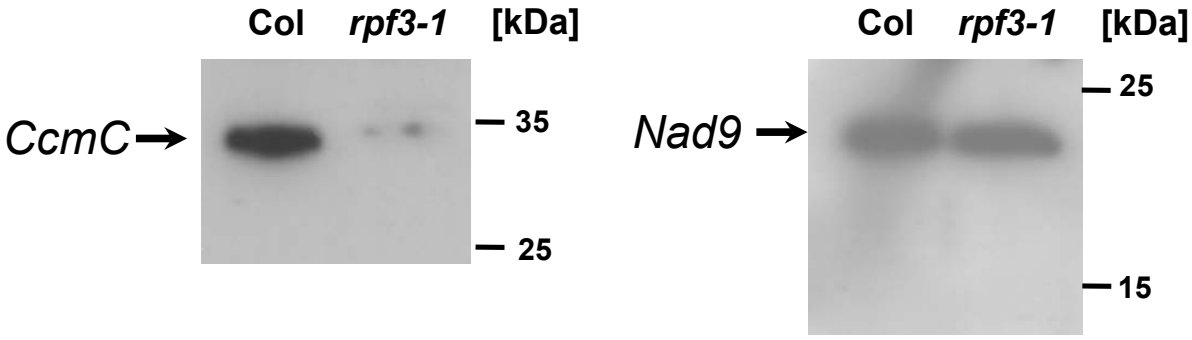


Figure. 5: The steady state level of the *CcmC* polypeptide is substantially reduced in the *rpf3-1* mutant. Immunodetection analyses of total mitochondrial protein isolated from cell suspension cultures from Col wild type (Col) and the *rpf3-1* mutant (*rpf3-1*). While in both lines a *Nad9* antiserum detects the target protein in similar amounts, an antiserum raised against an oligopeptide representing the core region of the *CcmC* protein (Kitazaki et al., 2009) revealed a substantially reduced level of *CcmC* polypeptide in the mutant. Sizes of marker proteins are indicated on the right hand side of the images.

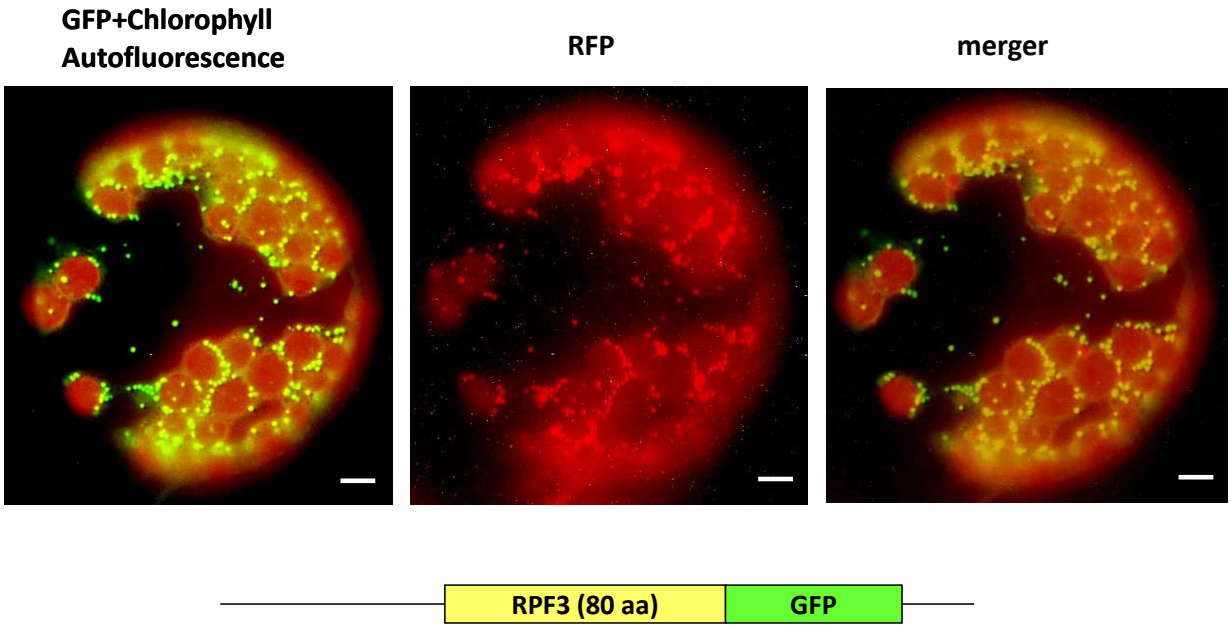


Figure. 6: Subcellular targeting on RPF3. A hybrid protein consisting of the RPF3 mitochondrial targeting sequence (80 amino acids) and the green fluorescent protein (GFP) was transiently expressed in transgenic tobacco protoplasts. Fluorescence of GFP and chlorophyll and the red fluorescent protein from the sea anemone *Entacmaea quadricolor* (eqFP611) were visualized through different filter sets. eqFP611, which is C-terminally fused to the N-terminal part of the mitochondrial isovaleryl-CoA dehydrogenase had been established as mitochondrial marker in transgenic tobacco (Forner and Binder, 2007). A merger (GFP + eqFP611) of both images demonstrated the nearly perfect match of both patterns. Small differences were due to the rapid movement of mitochondria in the protoplast. Size bars correspond to 10 µm.

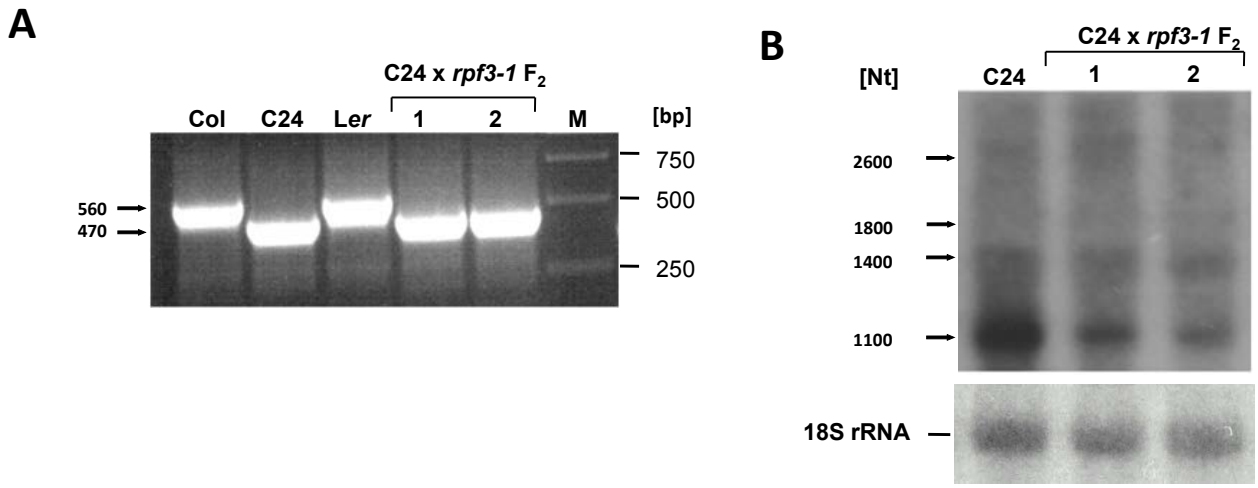
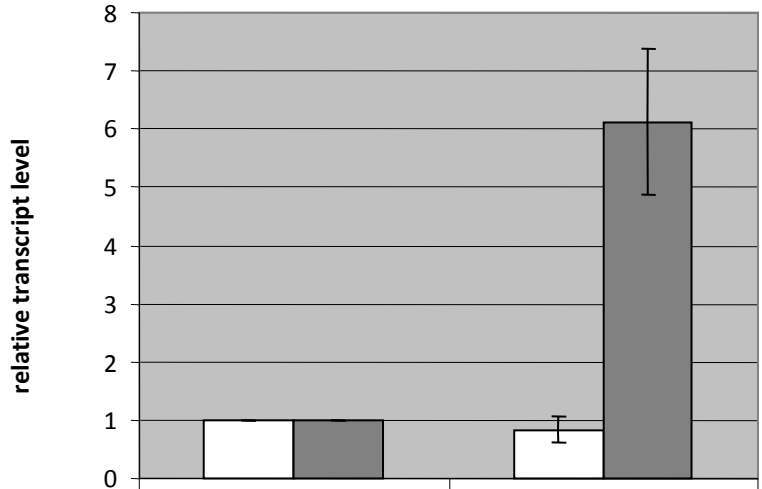


Figure. 7: RPF3 is less important for the generation of the –391/390 5' termini in C24. (A) CR-RT-PCR analysis of total RNA from two F_2 hybrids obtained from a cross of C24 (♀) with the *rpf3-1* mutant (♂) and from Col, C24 and L*er* wild type. Products of 470 bp were generated from total RNA both F_2 plants lacking an intact *RPF3* gene and from C24. Products of 560 bp were amplified from Col and L*er*. (B) F_2 hybrids were also analyzed by northern blot hybridization. Relevant signals are indicted in the left, the position of the cytosolic 18S rRNA (about 1.800 nucleotides) is given in the right hand side. The northern probe against the CDS is shown in Fig. 1D. A hybridization with an oligonucleotide complementary to the cytosolic 18S rRNA showed equal loading of the RNA gel.

SUPPLEMENTAL MATERIAL

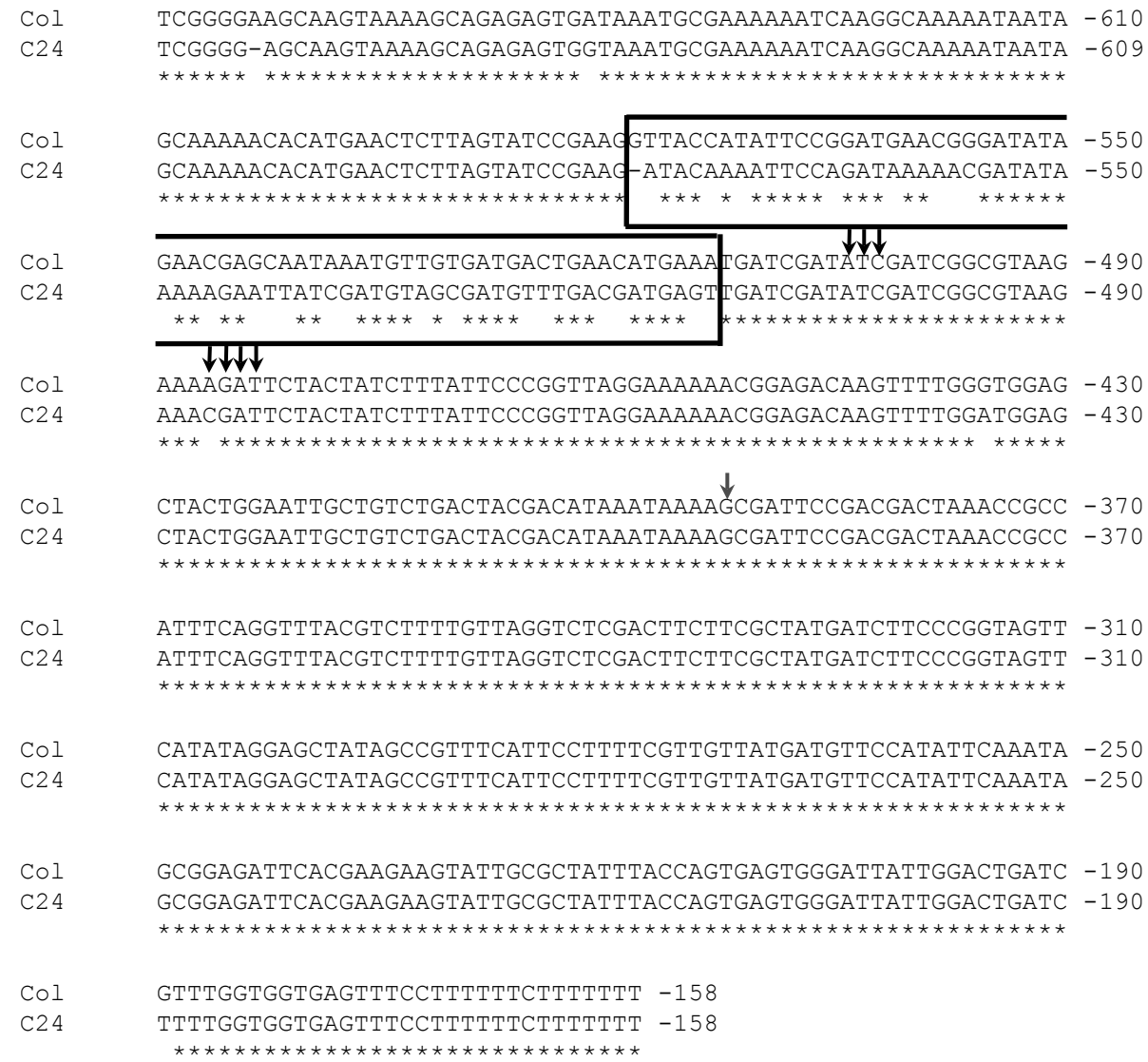
Supplemental Figure S1: Real-Time quantitative RT-PCR of *ccmC* transcripts in *rpf3-1*.



	Col	*rpf3-1*
CDS	1	0,8409
precursor mRNAs	1	6,121

Real-Time quantitative RT-PCR of *ccmC* transcripts in Col wild type and in the rpf3-1. Two PCR approaches cover either the coding sequence (CDS, white bars, primer pair: Atccb3QPCR.1/ Atccb3QPCR.2) from 273 to 368 relative to the ATG (A = +1) or the 5' upstream region (precursor RNA, grey bars, primer pair: Atccb3QPCR3/ Atccb3QPCR4) from –510 to –607 (NATG (N = -1). Transcript levels are given as x-fold change (right column) relative to the wild-type control (left column), which was arbitrarily set to 1.

Supplemental Figure S2: Sequences of the *ccmC* upstream regions in the mitochondrial DNAs from *Arabidopsis* accessions Col and C24.



Sequence alignment of the regions upstream of the ccmC reading frames in accessions Col and C24. Vertical arrows indicate the 5' ends of the mature ccmC mRNAs in Col (black arrows) or C24 (grey arrow). The box indicates the segment with highest sequence diversity between these accessions. Asterisks indicate identical nucleotide identities.

Supplemental Table S1: Amino acid exchanges between RPF3 proteins from Col and C24

Position	Col --> C24	Repeat
55	V --> A	
58	D --> N	
65	V --> I	
66	D --> G	
94	N --> K	1
163	G --> S	3
168	E --> D	3
177	F --> V	3
178	V --> E	3
180	E --> G	3
182	Q --> K	3
184	N --> D	4
208	I --> V	4
210	R --> Q	4
213	A --> Q	4
221	F --> V	5
225	T --> A	5
259	T --> N	6
264	A --> G	6
267	N --> K	6
270	N --> H	6
271	V --> I	6
272	N --> D	6
275	L --> F	6
276	N --> A	6
279	T --> N	6
280	E --> K	6
282	D --> E	6
283	N --> T	6
289	N --> D	7
291	V --> F	7
298	R --> S	7
315	D --> N	7
569	N --> D	15
572	T --> C	15

Supplemental Table S2: Oligonucleotide sequences.

Primer	Sequence [5'-3']
Atccb3-Mega5'.nah	cttgtgcgtggcactccttgct
Atccb3-Mega3'.nah	tgggagcattagccgatttggtacatc
Atccb3-Mega5'.fern	tcttacgccgatcgatatcgatcaactc
Atccb3-12	gtacaaggaaagtggctcaaaacgcc
Atccb3Sonde5'	tcgctatgcgccgctccg
Atccb3Sonde3'	aaagagagcgagtgaactagg
At26S-Adapter.R2	ttgccctttaggaatcttcagtcaaggtacg
Atccb3QPCR.1	atcccacacccaaaaggtc
Atccb3QPCR.2	aaccggtatagaaatgggtgc
Atccb3QPCR.3	aaacacatgaactcttagtatccgaag
Atccb3QPCR.4	catttcatgttcagtcatcacaac
At1g13320.1	taacgtggccaaaatgatgc
At1g13320.2	gttctccacaaccgcttggt
At4g27960.1	tcacaatttccaaggtgctgc
At4g27960.2	tcatctgggtttggatccgt
At18S-Adapter.R3	cattggcttggtcagcc
At18S-Adapter.R4	cctagctccccgagaacaacgttcg
At1g62930Kompl.H	atgcagtcgagcggcgcgcccctgaacaaaagagacattaactgaac
At1g62930Kompl.R	gacttcgagtttaattaattcgaaccgacaaattgtttgtc
At1g62930UEx.H	atgcagtcgagcggatccatgacaagttgtgttcatttagggattg
At1g62930GFP.R2	gacttagagtttcccgggtttattgaattcaacaatggaaggtaaagg
Atatp9Mega-Endo.A	gaagacgaagacggatcaaattgaataatcga
Atatp9-11	agatgggaccctagctac
KAN1	gaacaagatggattgcac
KAN2	gcgatagaaggcgatg

LITERATURE CITED

Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Mol Cell Biol **28:** 5337-5347

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A **99:** 10887-10892

Brown GG, Formanova N, Jin H, et al. (2003) The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J **35:** 262-272

Budar F, Pelletier G (2001) Male sterility in plants: occurrence, determinism, significance and use. C R Acad Sci III **324:** 543-550

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735-742

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133:** 462-469

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol **139:** 5-17

Davis SJ, Vierstra RD (1998) Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol Biol **36:** 521-528

Delannoy E, Stanley WA, Bond CS, Small ID (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochem Soc Trans **35:** 1643-1647

Desloire S, Gherbi H, Laloui W, et al. (2003) Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep **4:** 588-594

Forner J, Binder S (2007) The red fluorescent protein eqFP611: application in subcellular localization studies in higher plants. BMC Plant Biol **7:** 28

Forner J, Hölzle A, Jonietz C, Thuss S, Weber B, Schwarzländer M, Meyer RC, Binder S (2008) Mitochondrial mRNA polymorphisms in different Arabidopsis thaliana accessions. Plant Physiol **148:** 1106-1116

Forner J, Weber B, Thuss S, Wildum S, 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

Giege JM, Grienenberger JM, Bonnard G (2007) Cytochrome c biogenesis in mitochondria. Mitochondrion **8:** 61-73

Hölzle A, Jonietz C, Törjek O, Altmann T, Binder S, Forner J (2010) A RESTORER OF FERTILITY-like PPR gene is required for 5' processing of the *nad4* mRNA in mitochondria of *Arabidopsis thaliana*. Plant J **accepted for publication**

Johnson X, Wostrikoff K, Finazzi G, et al. (2010) 2.

MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. Plant Cell **22:** 234-248

Jonietz C, Forner J, Holzle 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:** 443-453

Kazama T, Nakamura T, Watanabe M, Sugita M, Toriyama K (2008) Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J **55:** 619-628

Kitazaki K, Nomoto Y, Aoshima A, Mikami T, Kubo T (2009) A mitochondrial gene involved in cytochrome c maturation (ccmC) is expressed as a precursor with a long NH2-terminal extension in sugar beet. J Plant Physiol **166:** 775-780

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

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J **34:** 407-415

Kranz RG, Richard-Fogal C, Taylor JS, Frawley ER (2009) Cytochrome c biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control. Microbiol Mol Biol Rev **73:** 510-528

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

Kuhn J, Binder S (2002) RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. Nucleic Acids Res **30:** 439-446

Kühn K, Weihe A, Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. Nucleic Acids Res **33:** 337-346

Loiselay C, Gumpel NJ, Girard-Bascou J, Watson AT, Purton S, Wollman FA, Choquet Y (2008) Molecular identification and function of cis- and trans-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol Cell Biol **28:** 5529-5542

Lurin C, Andres C, Aubourg S, et al. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16:** 2089-2103

Matthes A, Schmidt-Gattung S, Köhler D, Forner J, Wildum S, Raabe M, Urlaub H, Binder S (2007) Two DEAD-box proteins may be part of RNA-dependent high-molecular-mass protein complexes in Arabidopsis mitochondria. Plant Physiol **145:** 1637-1646

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res **32:** W20-25

Pfalz J, Bayraktar OA, Prikryl J, Barkan A (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. EMBO J **28:** 2042-2052

Sambrook J, Russel DW (2000) Molecular Cloning: A Laboratory Manual, Ed Ed 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sambrook J, Russel DW (2001) Molecular Cloning: A Laboratory Manual. *In*, Ed Ed 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

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

Small ID, Peeters N (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci **25:** 46-47

Wang Z, Zou Y, Li X, et al. (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell **18:** 676-687

Weigel D, Glazebrook J (2002) Arabidopsis, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Habor, New York

7.3 Danksagung

Zuerst möchte ich mich sehr herzlich bei Prof. Dr. Stefan Binder für die hervorragende Betreuung dieser Doktorarbeit bedanken. Vielen Dank für die Unterstützung, Aufmunterung und das tolle Arbeitsklima in deiner Gruppe.

Desweiteren gilt mein Dank Prof. Dr. Anita Marchfelder, die sich bereit erklärt hat als Zweitkorrektor für diese Arbeit zu fungieren und Prof. Dr. Axel Brennicke für die Möglichkeit im Institut Molekulare Botanik zu arbeiten.

Besten Dank an Janet Imhof (Captain Janet), Conny Guha, Ulrike Tengler, Aron Hauler (Paprikäron), Angela Hölzle und Joachim Forner für die Hilfe, die super Arbeitsatmosphäre und eine lustige, produktive Zeit in der Arbeitsgruppe Binder.

Vielen Dank an Prof. Dr. Hans Peter Braun, Dagmar Lewejohann und Tatjana Hildebrandt für die Untersuchungen der mitochondrialen Proteine mittels 2D Gelelektrophorese und der Aktvitätsbestimmung der Atmungskettenkomplexe.

Bei Bianca Wolf und Matthias Kugelmann möchte ich mich für den gemeinsamen Kampf gegen die morgendliche Müdigkeit, sowie unzählige spannende Ball- und Kartenspiele und kleine Gefallen bedanken.

Herzlichen Dank auch an Mizuki Takenaka, Daniil Verbitskiy und Friedrich Ossenbühl für interessante Diskussionen an Tagen wo es einmal später wurde.

Außerdem möchte ich mich bei der Arbeitsgruppe Marchfelder, vor allem bei Ruth Heyer und Britta Stoll, für diverse wohlschmeckende Backspezialitäten und die gute Nachbarschaft bedanken.

Zuletzt auch noch ein Dankeschön für eure Freundlichkeit und Hilfsbereitschaft an alle nicht namentlich erwähnten Mitglieder des Instituts Molekulare Botanik.

7.4 Curriculum vitae

Persönliche Daten

Name: Christian Jonietz

Geburtsdatum: 20.05.1981

Geburtsort: Bobingen/Deutschland

Nationalität: Deutsch

Familienstand: Ledig

Ausbildung

1987-1991 Grundschule Bobingen SIedlung

1991-2000 Leonhard-Wagner-Gymnasium Schwabmünchen
Abschluss: Abitur

2000-2001 Wehrdienst in der Sanitätsstaffel der Welfenkaserne Landsberg am Lech

2001-2007 Studium der Biologie an der Universität Ulm
Diplomarbeit „Kartierung eines *nad9*-spezifischen 5‘ Prozessierungs-faktors in *Arabidopsis thaliana*“ im Institut Molekulare Botanik der Universität Ulm
Abschluss: Diplom

2007-2010 Dissertation „Identification of mitochondrial RNA processing factors in *Arabidopsis thaliana*“ unter Betreuung von Prof. Dr. Stefan Binder, Institut Molekulare Botanik, Universität Ulm

7.5 Erklärung über die in Anspruch genommenen Hilfen

Ich erkläre hiermit, dass ich diese Arbeit selbständig angefertigt, keine anderen als die angegebenen Hilfsmittel und Quellen benutzt sowie wörtlich und inhaltlich übernommene Stellen als solche erkenntlich gemacht habe.

Ulm, den