Analysis of mRNA termini in mitochondria of *Arabidopsis thaliana*

DISSERTATION

zur Erlangung des Doktorgrades Dr. rer. nat.
der Fakultät für Naturwissenschaften
der Universität Ulm

vorgelegt von

JOACHIM FORNER

aus Bopfingen

2007

Amtierender Dekan der Fakultät für Naturwissenschaften:

Prof. Dr. Klaus-Dieter Spindler

Erstgutachter:

Prof. Dr. Stefan Binder, Universität Ulm

Zweitgutachter:

Prof. Dr. Axel Brennicke, Universität Ulm

Drittgutachter:

Prof. Dr. Hans-Peter Braun, Universität Hannover

Tag der Promotion:

13.04.2007

MEINEN ELTERN

Table of Contents

1. Introduction 5

2. Results 8

2.1 Comprehensive mapping of mRNA ends in mitochondria of *A. thaliana* 8

2.2 Investigation of the *cox3* mRNA 5' end polymorphism 11

2.3 Establishment of the red fluorescent protein eqFP611 as mitochondrial marker in plants 13

3. Discussion 16

3.1 Transcript ends and their generation 16

3.2 Application of eqFP611 in plant cells 21

4. Summary 22

5. References 24

6. Deutschsprachige Zusammenfassung 29

7. Appendix 33

7.1 Own contribution 33

7.2 Manuscripts 34

7.2.1 Manuscript 1: "Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5' and 3' end formation" 34

7.2.2 Manuscript 2: "Distant sequences determine 5' end formation of *cox3* transcripts in *Arabidopsis thaliana* ecotype C24" 68

7.2.3 Manuscript 3: "The red fluorescent protein eqFP611: application in subcellular localization studies in higher plants" 82

7.3 Acknowledgements 111

7.4 Curriculum vitae 112

7.5 List of publications 113

7.6 Erklärung über die in Anspruch genommenen Hilfen 114

1. Introduction

Although higher plants grow photoautotrophically, they depend on functional mitochondria to complete their life cycle successfully. Apart from hosting a multiplicity of biochemical reactions, the main function of theses organelles is to provide energy by oxidative phosphorylation (Logan, 2006). Not only heterotrophic tissues like roots entirely rely on this way of ATP production, but also the complete plant during early stages of embryo development and germination. A partial dysfunction of plant mitochondria often manifests itself as male sterility, i.e. the inability to produce functional pollen (Chase, 2006). This phenomenon is attributed to insufficient mitochondrial ATP production since this process requires a lot of energy and occurs in chloroplast-free tissues (Warmke & Lee, 1978).

Essential components of the respiratory chain are encoded within the mitochondria themselves. As descendants of formerly free living bacteria, these organelles still possess an own genome, although most genes have been transferred to the nucleus or have been lost completely (Kutschera & Niklas, 2005). In *Arabidopsis thaliana* for instance, only 57 genes are present in the mitochondrial genome (Unseld et al., 1997). All of the respective gene products are involved directly or indirectly in the assembly of the respiratory chain. The 32 encoded proteins are mainly either direct components of complexes I to V or are involved in the biogenesis of cytochrome c. The residual proteins – as well as the mitochondrially encoded 22 tRNAs and 3 rRNAs – are part of the ribosomes, which in turn are necessary to translate the mRNAs of those former proteins. All other proteins present in mitochondria are encoded in the nucleus and imported posttranslationally into this intracellular compartment. These include e.g. the remaining components of the respiratory chain and of the ribosomes as well as all the proteins required for maintenance and transcription of the mitochondrial DNA and for all steps of post-transcriptional RNA modification.

The 57 genes encoded in *A. thaliana* mitochondria represent just a small part of the mitochondrial DNA, which comprises 367 kb in total. Known coding sequences account only for 38 kb (Unseld et al., 1997). The genes are either grouped as little clusters or are found solitarily. Some of the genes are interrupted by group II introns and for three of them, the single exons are encoded at different loci, either individually or in small groups. The clusters and single genes are separated by long stretches of intergenic sequences and are thus probably transcribed individually. This requires an extra promoter for each of these transcription units, and in many cases even multiple promoter motifs are found upstream of them (Kühn et al., 2005). Thus, a multitude of primary transcripts is generated in *A. thaliana* mitochondria, which undergo a series of posttranscriptional modifications. These include *cis* and *trans*

splicing, hundreds of editing events, additional base modifications for tRNAs and rRNAs and processing by exo- or endonucleases (Gagliardi & Binder, 2007).
Of all these posttranscriptional processes, especially the generation of the secondary transcript termini by exo- or endonucleolytic cleavage is poorly understood. For most protein-coding genes, not even descriptive data on the existence, number and location of secondary transcript ends is available. Except for the mature 3' ends of the *atp9* and *atp8* mRNAs, which are generated by exonucleolytic trimming catalyzed by PNPase and RNaseII (Perrin et al., 2004b), it is unclear whether the secondary mRNA termini are created exo- or endonucleolytically. Furthermore, it is unknown how the position of the final end is determined. This can either be done by *cis* elements being part of the pre-mRNA or by the binding of *trans* factors to a specific site on the immature transcript. In case of exonucleolytic trimming, the nucleotide at the mature transcript end must be inaccessible for the processing exonuclease and thus, the terminal nucleotide is either protected by the secondary structure of the transcript or by a protein bound to it. If the processing enzyme is an endonuclease, it either directly recognizes the cleavage site on the pre-mRNA or in concerted action with a site-specific *trans* factor. The recognition signals for both putative *trans* factors and endonucleases could be either certain primary sequences or given secondary structures of the pre-mRNA. *Trans* factors could be encoded mitochondrially or nuclearly and be proteins or RNAs. Actually, several reports on mitochondrial RNA processing suggest pentatricopeptide repeat proteins (PPRs) encoded in the nucleus to be important factors in various reactions (Gagliardi & Binder, 2007).
Additionally, it is not known whether the generation of secondary mRNA termini is of any functional importance in terms of translation efficiency or RNA stability.

The aim of the studies presented in this thesis was to collect data about transcript ends and their generation in mitochondria of *A. thaliana* and thus to help to understand how plant mitochondria express their genetic information.
The results of this work are summarized in three manuscripts which are discussed and presented below.

The first manuscript contains a complete list of the 5' and 3' mRNA ends derived from mapping the trancript termini of all protein-coding genes in mitochondria of *A. thaliana*. Furthermore, by analyzing the (pre-mRNA) sequences surrounding these mRNA termini for

similarities, tRNA-like structures, so-called t-elements, have been found upstream or downstream of some of the mature transcript ends.

A polymorphism in the mitochondrial DNA between different ecotypes of *A. thaliana* and a correlated ecotype-specific mRNA 5' end are investigated in the next publication. An examination of reciprocal F_1 hybrids indicates a *cis* element to determine the formation of the respective 5' terminus.

To characterize putative nuclearly encoded *trans* factors involved in mitochondrial mRNA processing, their import into mitochondria has to be investigated. This is often done by expressing fluorescent proteins fused to the protein under investigation. The last manuscript describes the characterization of a new red fluorescent protein as reporter in plant cells, with special emphasis on its use as mitochondrial marker.

2. Results

2.1 Comprehensive mapping of mRNA ends in mitochondria of *A. thaliana*

As a basic prerequisite for the analysis of the mitochondrial mRNA 5' and 3' end formation in *Arabidopsis thaliana*, these ends have to be mapped. Up to date, precisely determined mRNA ends for mitochondrial genes of this species have been reported only sporadically (Kühn et al., 2005; Perrin et al., 2004b; Raczynska et al., 2006). Therefore, a complete and systematic survey was necessary. It was based mainly on the CR-RT-PCR (circularized RNA-reverse transcription-polymerase chain reaction) analysis (Kuhn & Binder, 2002). This is a method to map simultaneously the 5' and 3' ends of a given gene. An indispensible requirement for this experimental approach is that the genes of interest have been sequenced, which is the case for the mitochondrial genome of *A. thaliana*. Briefly, isolated RNA is circularized by RNA ligase and reverse transcribed starting from a gene-specific primer annealing in the reading frame. The 5' and 3' extremities are then amplified by PCR using primers annealing close to the ends of the reading frame. Products of this reaction are either sequenced directly or cloned before sequencing. Comparison of the sequences obtained with the respective genomic sequence identifies the ligation site and thus the 5' and 3' ends of the original mRNA.

Such assays were carried out for all protein-coding genes in the mitochondrial genome of *A. thaliana*. For each gene or dicistronic gene pair except *matR* at least one 5' and one 3' end was obtained. The untranslated regions generally cover 15 to several hundred nucleotides, but in some cases the transcript ends are found within the reading frame. If several PCR products per gene were detected and analyzed, the different PCR products were attributable to varying 5' termini in nearly all cases, while generally only one 3' end was found per gene. The sequences surrounding the observed ends were compared to search for common nucleotide motifs. Promoter motifs (Dombrowski et al., 1999; Kühn et al., 2005) were found at a group of seven major and two minor 5' ends that are therefore probably generated by transcription initiation. All other 5' termini are most likely secondary ends derived from posttranscriptional processing. Around the secondary 5' termini and the 3' ends no common sequence motif was found when searching for primary sequence similarities.

Some of the ends observed are found immediately adjacent to tRNAs and are thus most likely created during tRNA maturation. The 3' ends of *atp6-1* and *atp6*-2 are located immediately upstream of a $tRNA^{Ser}$, and the 5' ends of *ccb6c* and *rps3* are located in close vicinity to the 3' ends of $tRNA^{Gly}$ and $tRNA^{Lys}$, respectively.

Other ends are closely associated with tRNA-like structures. These so called t-elements have been identified upstream of the 5' ends of *cox1*, *rps4* and *ccb6n1* and downstream of the 3' termini of *ccb3* and *nad6* by searching for the presence of the GTTCRANYC motif indicative for the T-arm of tRNAs (Hanic-Joyce et al., 1990). In addition, the sequences upstream of the 5' ends of the mature *rpl5* and *atp6-2* mRNAs could fold into a stem-loop structure that in regard to the length of the paired region and the single unpaired nucleotide at the 3' end closely resembles the acceptor stem of a tRNA. These stem-loops and the t-elements upstream of the mature mRNAs could possibly be substrates for a tRNase Z (Vogel et al., 2005). A putative stem-loop is also found immediately downstream the mature 5' end of the *nad7* transcripts. This 5' end as well as the mRNA 3' ends upstream of t-elements could therefore be generated by an RNase P activity (Frank & Pace, 1998).
No further common possible secondary structures around the mapped ends were discernible. Especially almost no double or single stem-loops are present upstream of the mature 3' ends. In previous *in vitro* studies, such structures have been shown to be involved in the formation of the 3' termini of certain mitochondrial transcripts (Dombrowski et al., 1997). They both induce 3' to 5' exonucleolytic degradation of the downstream sequences and protect the mature 3' termini against further processing. Thus, the 3' ends of most mRNAs must be generated and/or stabilized by another mechanism.
In three cases, ends of two or three indepent genes were found at identical positions within duplicated sequences, suggesting that they are formed by the same mechanism. These groups consist of the 3' ends of *nad1* und *atp9*, the 3' ends of *atp6-1* und *atp6-2* and the 5' ends of *atp9*, *nad6* and of the 26S rRNA (Binder et al., 1994).

The ends detected in this survey could principally be primary or secondary termini. Primary ends are created directly by transcription initiation or termination, while secondary ends are generated post-transcriptionally by exo- or endonucleolytic cleavage. The presumably secondary 5' transcript termini of *cox1* and *atp9* were exemplarily investigated if they are generated by endonucleolytic cleavage. To this end, a modified CR-RT-PCR analysis was performed that allows the detection of small molecules. If the 5' end of a mature transcript is formed by an endonucleolytic cut of the precursor mRNA, a 5' leader is released as a by-product. The 5S rRNA, one of the most abundant transcript species, was used as an internal anchor. cDNAs derived from ligation of the 5' leader to the 5S rRNA anchor were then amplified. These were cloned and several individual clones were sequenced. In case of *cox1*, the 3' ends of nearly all 5' leader molecules analyzed were found exactly at the position

upstream of the 5' end of the mature *cox1* mRNA, clearly showing that this transcript terminus is generated by endonucleolytic cleavage. The tRNA-like generation of the *cox1* 5' end is further substantiated by the non-encoded C and A nucleotides found at 65 % of the individual 5' trailer molecules, reminding of the CCA triplet added to tRNAs. For *atp9*, several of the molecules analyzed ended immediately upstream of the mature *atp9* 5' end, as expected for the hypothetic 5' trailer. All other ends were found a variable number of nucleotides further upstream, indicative of a 3' to 5' exonucleolytic degradation of the 5' trailer. Thus, the 5'-terminus of the mature *atp9* transcript is also created by endonucleolytic cleavage.

For some genes, primer extension and northern blot experiments were carried out to control the results of the CR-RT-PCR analyses by independent methods. In general, the results of all three experimental approaches are consistent. For example, the primer extension analyses of *cox1* indicate the presence of a single main transcript end at position –239 to –241 which is identical with the 5' end determined by CR-RT-PCR. A single transcript species was detected in a corresponding northern blot analysis, its length of approximately 1,800 nucleotides fitting perfectly to the major 5'- and 3'-transcript termini determined by CR-RT-PCR. However, in some cases, there are some smaller discrepancies. The primer extension analysis of *nad1* for instance indicates that there is only a single main 5' end, namely the one at position -645, whereas the CR-RT-PCR maps the major end equally to positions -645, -355 and -149.

2.2 Investigation of the *cox3* mRNA 5' end polymorphism

To create the distinct mRNA ends observed in mitochondria of *A. thaliana*, the processing enzyme(s) must be able to recognize the nucleotides at the final transcript termini exactly. In principal, the location of the later mRNA end could be specified by both *cis* elements and *trans* factors, i.e. specific sequences on the precursor mRNA and independent molecules acting in *trans*, respectively. Most likely, the *trans* acting factors bind to *cis* elements and thus both jointly specify the location of the mRNA end. However, differences between the transcript ends of a certain gene in different ecotypes of *A. thaliana* offer the possibility to dissect and identify the components responsible for the specificity of the mRNA end formation.

In the course of mapping all major 5' and 3' mRNA ends in mitochondria of *A. thaliana*, a polymorphism between the *cox3* transcripts in the ecotypes Columbia (Col), Landsberg *erecta* (L*er*) and C24 was discovered. CR-RT-PCR analyses of *cox3* mRNAs in these three ecotypes revealed a common 3' end at position +314 relative to the stop codon of the *cox3* reading frame. Furthermore, *cox3* mRNAs in all three accessions share an identical 5'-terminus at position –379/-378 upstream of the translation start codon. But an additional 5' end at -444/-437 is present exclusively in C24. A northern blot analysis confirmed the results of the CR-RT-PCR. Whereas only one major *cox3* transcript of approximately 1,500 nucleotides is found in ecotypes Col and L*er*, C24 possesses an additional transcript of about 1,560 nucleotides. This mRNA species is present in similar amounts as the smaller one. The transcript sizes fit to the mRNA end positions mapped by CR-RT-PCR. Furthermore, since the abundance of the shorter and of the longer *cox3* transcript in C24 is approximately equal, the additional mRNA species seems to be of physiological relevance.

To determine whether the presence of the additional transcript in C24 is due to differing *cis* elements or *trans* factors, *cox3* mRNAs were investigated in reciprocal F_1 hybrids of Col and C24. In *A. thaliana* the nuclear genome is inherited biparentally, while the mitochondrial DNA is exclusively transmitted through the egg cell (Nagata et al., 1999). Thus, reciprocal F_1 hybrids do not differ in regard to the origin and composition of the nuclear genome and potential *trans* factors, but are different in terms of the origin of the mitochondrial DNA and *cis* elements encoded therein. Both a primer extension and a northern blot analysis clearly showed that the C24-specific *cox3* mRNA species is present only in plants with a C24 female parent, i.e. with a C24-type mitochondrial genome. Thus, a C24-specific *cis* element must be responsible for the generation of the additional mRNA 5' end, excluding a nuclear encoded *trans* factor to be responsible for the observed polymorphism.

If a specific *cis* element is present in C24, differences of the mitochondrial DNA in the *cox3* region must exist between C24 and the two other ecotypes. Indeed, Southern blot experiments as well as PCR analyses followed by restriction digestion and sequencing of a corresponding Col mtDNA HindIII clone revealed different genomic configurations upstream of the *cox3* gene in the three ecotypes. The *cox3* reading frame and the region downstream of it are completely identical as well as the first 592 nucleotides upstream of the ATG codon. Beyond this point, a 1,799 bp insertion is present in Col as compared to C24, followed again by identical sequence in both accessions. In L*er*, at least a part of the same insertion as in Col is present, but obviously a recombination event has disrupted this 1.8 kb insertion, so that in this ecotype the *atp9* gene is situated upstream of the *cox3* region. This L*er*-type genomic configuration is also present in a cell suspension culture of ecotype Col from which the RNA for the mapping of the mRNA ends of all mitochondrial genes has been isolated. Interestingly, the C24-specific 5' end is located in a region where the sequence is identical in the three ecotypes, approximately 150 nucleotides downstream of the the polymorphic region. Thus, the *cis* element involved in the formation of the C24-specific 5' end is separated by at least 150 nucleotides from this terminus.

2.3 Establishment of the red fluorescent protein eqFP611 as mitochondrial marker in plants

The coding sequence of eqFP611, a red fluorescent protein from the sea anemone *Entacmaea quadricolor*, has recently been cloned and expressed in bacterial and mammalian cells (Wiedenmann et al., 2002), but eqFP611 has not yet been tested for application in plants. Therefore, it was investigated whether this polypeptide could be expressed in plant cells from its original cDNA. After transient transformation of tobacco protoplasts with a plasmid carrying the eqFP611 reading frame sequence under control of a CaMV 35S promoter, red fluorescence was clearly visible in an epifluorescence microscope using an appropriate filter set. The protein accumulated in the nucleus and in the cytosol of the transformed cells. By contrast, it was clearly excluded from the chloroplasts and thus probably also from any other organelle. No aggregate formation was observed and the viability of protoplasts expressing eqFP611 was not adversely affected.

To test if the subcellular localization of eqFP611 can be efficiently controlled by N-terminal targeting sequences and to check if the fusion of an additional polypeptide would prevent fluorescence, the mitochondrial presequence of isovaleryl-CoA-dehydrogenase (IVD) was added to eqFP611. Tobacco protoplasts were transiently transformed with the respective plasmid and analyzed for red fluorescence. The transformed cells showed fluorescence in many small rod-shaped structures of about 1 to 2 μm in length, a pattern typical for mitochondria (Van Gestel & Verbelen, 2002). No red signal was visible outside these structures.

To verify that the rod-shaped structures were indeed mitochondria and to test if eqFP611 and GFP can be co-expressed in the same cell, tobacco protoplasts were co-transformed with two constructs simultaneously. In addition to the IVD-eqFP611 construct, a plasmid encoding a IVD-smGFP fusion protein was used. This latter fusion protein has previously been shown to be targeted exclusively to mitochondria (Däschner et al., 2001). The red and the green fluorescence in doubly transformed protoplasts were analyzed with two different filter sets. In both channels, fluorescence was clearly visible and the spatial patterns of the red and the green signals were completely identical, confirming the mitochondrial localization of the IVD-eqFP611 fusion protein.

To exclude that the subcellular targeting of eqFP611 and smGFP possibly interferes, another protoplast co-transformation experiment with two different constructs was carried out. This time, IVD-smGFP was used together with KAT2-eqFP611. The latter fusion protein consists of the N-terminal part of the peroxisomal 3-keto-acyl-CoA thiolase 2 (KAT2) which includes

a peroxisomal targeting signal 2 (PTS2) and of eqFP611 (Germain et al., 2001; Kato et al., 2000). Doubly transformed cells showed again both clear red and green fluorescence. While the green fluorescence was again found exclusively in the rod-shaped mitochondria, the red fluorescence now was visible in roundish structures of about 1.5 μm in diameter. No overlap of the red and green signals was observed. The roundish structures were shown to be indeed peroxisomes in protoplasts transformed in parallel with plasmids encoding either KAT2-eqFP611 or GFP-APX chimeric proteins, where the spatial distribution of the red and the green fluorescence was identical. The APX moiety consists of the C-terminal 36 amino acids of the cotton ascorbate peroxidase, targeting the fusion protein to the peroxisomal membrane (Bunkelmann & Trelease, 1996).

For all dual-labeling experiments described above, two different constructs were used, requiring at least two transformation events per cell. But in some cases, only the uptake of one construct is seen. Thus, to ensure that every transformed protoplast expresses both fluorescent proteins and that the ratio of RFP and GFP is equal in all cells, a single plasmid encoding both eqFP611 and smGFP was constructed and used to transform tobacco protoplasts. As expected, every transformed cell displayed both clear red and green fluorescence in mitochondria.

To create a convenient source of tobacco protoplasts with endogenous red fluorescent mitochondria, tobacco plants were stably transformed with the vector pBI121 containing the IVD-eqFP611 expression cassette. T_0 plants were regenerated from transgenic calli and grown to maturity. The presence of red fluorescent mitochondria in the T_1 plants was verified by epifluorescence microscopy. For further experiments, only descendants of the T_1 plant with the strongest and most stable eqFP611 fluorescence were used. The red fluorescence was stably transmitted to the following generations and no detrimental effects on plant growth and development were observed.

Protoplasts derived from these plants were finally used to investigate the subcellular localization of MTERF HOMOLOG 1 (MTH1). This protein is one of the *A. thaliana* homologs of the human mitochondrial protein mTERF, which promotes both transcription initiation and transcription termination at the rDNA region in the human mitochondrial DNA (Martin et al., 2005). MTH1 was *in silico* predicted to be targeted to mitochondria via a mitochondrial presequence at its N-terminus by different algorithms (Altschul et al., 1997; Claros & Vincens 1996; Emanuelsson et al., 2000). Indeed, when a construct encoding a fusion protein of the first 120 amino acids of MTH1 and smRSGFP was transiently expressed in tobacco protoplast derived from the transgenic plants mentioned above, the patterns of the

red and of the green fluorescence were completely identical, indicating a mitochondrial localization of MTH1 in *A. thaliana*.

Taken together, these findings clearly show that eqFP611 is a suitable marker in plants.

3. Discussion

3.1 Transcript ends and their generation

The most striking feature of the mRNA termini mapped in this study is the diversity of the surrounding sequences, in contrast to e.g. the conserved dodecamer motif at the 3' termini of mitochondrial mRNAs in yeast (Hofmann et al., 1993). Common primary sequence motifs or secondary structures cannot be found around the vast majority of the *A. thaliana* transcript ends. Therefore, individual regulatory sequences for the formation of each transcript terminus are likely to exist.

However, some of the mRNA termini do share common features. A group of 5' ends detected in this study is located downstream of sequences resembling known promoter motifs (Dombrowski et al., 1999; Kühn et al., 2005), suggesting these mRNA ends to be primary termini derived directly from transcription initiation. The exact ends of these transcripts are sometimes detected a few nucleotides downstream of the predicted transcription start sites. This is expected since intact primary transcripts cannot be self-ligated by T4 RNA ligase due to their 5'-terminal triphosphate groups. Thus, only those molecules that are partially degraded at their 5' ends - most likely during the RNA isolation procedure - can be amplified by CR-RT-PCR (Kuhn & Binder, 2002).

All other detected ends are probably generated post-transcriptionally, although this has been demonstrated experimentally only for a few termini.

Up to date no evidence for a 5' to 3' exonucleolytic activity has been observed in plant mitochondria (Gagliardi & Binder, 2007). This suggests all secondary 5' ends to be generated by an endonuclease, as has been indirectly shown for the 5' ends of the 18S and the 5S rRNAs (Perrin et al, 2004a). Although little or nothing is known about an endonuclease involved in rRNA or mRNA processing, the tRNA processing endonucleases RNase P and tRNase Z, which have been detected and described in plant mitochondria (Marchfelder & Brennicke, 1994), might be good candidates for the cleavage of other RNAs. For instance, the 5' mRNA termini of the *ccb6c* and *rps3* genes map immediately adjacent to tRNAs, strongly suggesting these mRNA termini to be created during tRNA processing by tRNase Z. Since t-elements, secondary structures similar to tRNAs, are also processed by the same plant mitochondrial extracts like tRNAs *in vitro* (Hanic-Joyce et al., 1990), the pre-mRNAs of *cox1*, *rps4* and *ccb6n* containing such secondary structures are probably also cut by this enzyme immediately upstream of the mature 5' mRNA end.

In *A. thaliana*, four genes encoding putative tRNases Z are present, two for the short version and two for the long version of this enzyme (Vogel et al., 2005). The long versions of tRNase Z from human, pig and yeast have been shown to accept relatively short RNAs as substrates, provided that they include a paired region resembling the acceptor stem of a tRNA (Takaku et al., 2004). In *A. thaliana*, both long tRNase Z proteins are predicted to be targeted to mitochondria (Vogel et al., 2005). Thus, these proteins are promising candidates for creating the mRNA 5' ends found immediately downstream of t-elements and even for the 5' termini downstream of acceptor stem-like secondary structures, as has been found in case of *rpl5* and *atp6-2*. 5' ends upstream of such structures could be generated by an RNase P activity, like for instance the end at position -375 of *nad7*.
These processing events strongly resemble the mRNA maturation process in mammalian mitochondria, where the mature mRNAs are released from the long precursor transcripts during tRNA processing (Ojala et al., 1981).

However, most 5' ends are not flanked by sequences with the potential to form tRNA-like or acceptor stem-like structures. These ends are thus either generated by a hitherto unknown alternative endoribonuclease or they require binding of one or more additional *trans* factors, which then in turn direct(s) cleavage by tRNase Z or RNase P. Considering the variety of the structures around the different 5' ends, probably an individual *trans* factor is required for each processing site. A member of the pentatricopeptide repeat (PPR) protein family has recently been shown to promote endonucleolytic processing of the CMS-associated B-*atp6-orf79* transcript in rice mitochondria (Wang et al., 2006). Since there are more than 400 of these proteins in *A. thaliana* most of which are predicted to be targeted to mitochondria (Lurin et al., 2004), members of this familiy are the most promising candidates for nuclear encoded *trans* acting specificity factors involved in endonucleolytic mitochondrial mRNA processing. For groups of transcript termini like the 5' ends of the 26S rRNA, *atp9* and *nad6* which are found within duplicated sequences, only a single specifity factor would be necessary.

Surprisingly, almost non of the 3' ends detected in this study lie immediately downstream of obvious single or double stem loop structures. In previous reports, such elements have often been found upstream of 3' transcript termini in mitochondria of higher plants (Bellaoui et al., 1997; Schuster et al., 1986). Additionally, *in vitro* experiments have shown these structures to determine the location of the mature 3' ends of certain transcripts. They both initially recruite and finally stop the 3' to 5' exonucleolytic activity present in plant mitochondria

(Dombrowski et al., 1997). This latter function also protects the respective mature mRNAs against exonucleolytic degradation and such stabilizes the transcripts (Kuhn et al., 2001). *In vivo*, two 3' to 5' exoribonucleases have been shown to be involved in RNA degradation and in 3' end formation of at least some transcripts in *A. thaliana* mitochondria (Holec et al., 2006; Perrin et al., 2004a; Perrin et al., 2004b). PNPase degrades large RNA molecules after they have been tagged by oligo-adenylation, while RNaseII removes small unstructured 3' extensions including oligo(A)-tails but is stopped by secondary structures. These characteristics of RNaseII nicely explain the correlation between stem-loop structures and stable 3' ends immediately downstream of them, since both unpaired nucleotides in the precursor RNA and oligo(A)-stretches necessary for degradation by PNPase are removed by RNaseII.

However, the formation and especially the stability of 3' transcript termini without preceeding stem loops must be achieved differently. The regions immediately upstream of the mature 3' ends might pair with complementary sequences located far away within the same transcript or they might be hidden by the tertiary structure of the RNA and therefore be unaccessible for the exonucleases. Alternatively, they could be protected by proteins. If such stabilizing *trans* factors exist, one would postulate an individual protein factor for each mRNA. Unspecific general RNA binding proteins would equally protect the sequences downstream of the mature 3' ends and there are no evident conserved primary sequences or secondary structure motifs upstream or downstream of almost all mature mRNA 3' termini to serve as common binding sites.

Similar to the 5' transcript termini, the 3' ends detected in this study could either be generated by transcription termination or by post-transcriptional processes.

Since they are found immediately upstream of the mature $tRNA^{Ser}$ 5' end, the mRNA 3' ends of the *atp6-1* and *atp6-2* genes are most likely generated post-transcriptionally by the endonucleolytic processing of RNase P during tRNA maturation. This enzyme could also be involved in the formation of the mature 3' transcript ends of *ccb3* and *nad6*, as these termini are located upstream of putative t-elements. The detection of the additional 3' transcript terminus at the downstream end of the t-element indicates that transcription proceeds beyond the mature *ccb3* mRNA 3' end and thus further substantiates that this latter terminus is generated post-transcriptionally.

In the CR-RT-PCR analyses, 3' ends downstream of the major 3' mRNA termini were rarely detected and only found for some genes, e.g. in 3 clones out of 38 for *atp1*, 1 out of 32 clones

for *cox1* and a few clones for *ccb2*. These longer ends could be either generated erraneously due to inefficient transcription termination or they could present the regular primary transcript ends. But most likely they are unstable processing intermediates.
The mature 3' transcript ends of *atp9*, *atp8*, 18 S rRNA and 5S rRNA have been shown to be generated post-transcriptionally from longer precursors by the exonucleolytic activity of PNPase and RnaseII (Perrin et al., 2004a; Perrin et al., 2004b). Furthermore, large regions downstream of the rRNA and tRNA genes are regularly transcribed (Finnegan & Brown, 1990; Holec et al., 2006), implying that this occurs downstream of protein-coding genes as well. Thus, probably all the 3' ends detected in this study are secondary termini generated from longer precursor RNAs, but there is no experimental data yet.

The presence of an ecotype specific transcript end offers the possibility to further characterize the regulatory elements involved in its formation. In case of the C24-specific *cox3* mRNA 5' terminus, these specificity factors have been shown to be mitochondrially encoded. Thus, they are most likely *cis* elements being part of the precursor RNA. However, the sequences in C24 and in the other two ecotypes investigated are completely identical around the position of the C24-specific transcript 5' terminus up to 150 nucleotides upstream, showing that the *cis* element and the corresponding processing site are located relatively far away from each other. This long distance action could be explained by base-pairing of complementary RNA sequences on the same molecule separated by hundred of nucleotides forming a secondary structure required for processing. The differing upstream sequences could also be involved in the formation of a certain tertiary structure of the pre-mRNA necessary to specify the processing site or these sequences could provide the binding site for a *trans* factor without interacting with other parts of the RNA precursor.
No investigation of a *cis* element involved in the generation of secondary mRNA 5' ends in plant mitochondria has been reported before.
Alternatively, the maternal mode of inheritance of the C24-specific *cox3* 5' transcript end could be due to mitochondrially encoded *trans* factors. Since the number of proteins encoded in the mitochondrial genome is evidently not sufficient to provide a specific *trans* factor for each transcript end, these *trans* acting molecules would most likely be RNAs. Indeed, several non-coding regions have been found to be transcribed in *A. thaliana* mitochondria including small RNAs, but it is still unclear whether these transcripts fulfill a biological function at all (Holec et al., 2006; Marker et al, 2002).

A comprehensive screen for transcript end polymorphisms between different ecotypes of *A. thaliana* will probably reveal ecotype-specific mRNA termini that are due to polymorphic nuclear encoded *trans* factors. Map-based cloning of the corresponding nuclear genes will then identify the postulated proteins involved in mitochondrial mRNA processing and gene expression.

3.2 Application of eqFP611 in plant cells

Apart from a number of other applications (Lippincott-Schwartz & Patterson, 2003), fluorescent proteins are a powerful tool to examine the subcellular localization of a protein. For this purpose, the intra-cellular spatial distribution of a fusion protein composed of the protein under investigation and the fluorescent reporter is analyzed by fluorescence microscopy. However, the compartments to which the fusion protein is targeted have to be unambigously identified. This can be done for instance by the parallel expression of a spectrally different fluorescent protein fused to a well characterized targeting sequence. The most frequently used fluorescent protein in molecular biology is GFP, the first fluorescent protein ever cloned (Prasher et al, 1992). Red fluorescent proteins are suitable partners for GFP in dual-labeling experiments, since they are spectrally different and can be detected separately with appropriate filters. Up to date almost exclusively DsRED and derivatives thereof are used as RFPs in molecular biology (Shaner et al., 2004). To broaden the palette of RFPs available for plant cellular biology, the red fluorescent protein eqFP611 was characterized for its application in plants.

The studies performed with this protein in plants showed that problems like splicing of a cryptic intron or formation of aggregates which were observed with other fluorescent proteins (Haseloff et al., 1997; Yanushevich et al., 2002) have not been seen with eqFP611. Furthermore, fusion of additional sequences to a fluorescent protein can interfere with the development of fluorescence (Shaner et al., 2004), but eqFP611 seems to tolerate at least the short N-terminal extensions used in this study. Thus, eqFP611 can readily be utilized in plants without need for prior optimization.

The classical way to visualize mitochondria is staining with MitoTracker® Red CM-H2Xros (Molecular Probes, Eugene, OR), a dye interacting with the respiratory chain. Expression of mitochondrially targeted eqFP611 eliminates the need for the staining procedure. Conveniently, the filter set optimized for visualizing MitoTrackerRed can also be used for eqFP611 visualization. In terms of stability and brightness, the fluorescence of this mitochondrially targeted eqFP611 and MitoTrackerRed are comparable, so that eqFP611 offers a quick and cheap replacement of this dye. For the investigation of putative nuclear encoded components of the RNA processing apparatus of mitochondria, especially the transgenic tobacco line established in this work will faciliate the experimental procedure. No staining with MitoTracker dye or parallel introduction of a second construct is required to localize the mitochondria in protoplasts derived from such plants when expressing GFP chimeras of these candidate genes.

4. Summary

In this thesis mitochondrial mRNA ends and their generation were analyzed in *Arabidopsis thaliana*.

5' and 3' ends of all mRNAs were mapped mainly by circularized RNA-reverse transcription-polymerase chain reaction (CR-RT-PCR). At least one 5' and one 3' end were obtained for each transcript except *matR*, a maturase-like pseudogene.

The untranslated regions of most transcripts range from 15 to several hundred nucleotides. In case of *rps4*, *ccb3*, *nad6* and *orfX* the major transcript terminus was found within the reading frame, suggesting these mRNAs to be possibly non-functional or to code for shorter proteins than expected from the genomic sequence.

Nine 5' transcript ends are located in close vicinity to known promoter motifs, indicating that these termini originate directly from transcription initiation. No other common primary sequence motif was discernible around the remaining 5' and the 3' ends.

Two 5' and two 3' termini were found directly adjacent to tRNAs. Thus, these mRNA ends are most likely generated during tRNA processing by tRNase Z and RNase P, respectively. Furthermore, tRNA-like structures, so-called t-elements, are present immediately upstream of the *cox1*, *rps4* and *ccb6n1* 5' ends as well as downstream of the *ccb3* and *nad6* 3' termini. Additionally, secondary structures resembling the acceptor-stems of tRNAs are found upstream of the 5' ends of *rpl5* and *atp6-1* and downstream of the *nad7* 5' terminus. The association of these structures with the mRNA ends suggests these to be likewise generated by tRNase Z and RNase P.

Single or double stem-loop structures directly preceeding the 3' mRNA termini were found only for *atp9*, *nad1* and *cox2*. In previous studies, such secondary structures were shown *in vitro* to be involved in the exonucleolytic generation of certain mitochondrial RNA 3' ends and in the stabilization of the respective transcripts. Thus, most of the 3' termini detected in this study must be generated and stabilized by another mechanism.

Since most of the transcript ends are not associated with any obvious common sequence or structural motif, individual regulatory elements are probably involved in the generation of each mRNA terminus.

The *cox3* transcript 5' end was investigated in detail in different ecotypes of *A. thaliana*. In ecotype C24, an mRNA end is present at position -444/-437 upstream of the start codon, while no such terminus was found in ecotypes Columbia (Col) and Landsberg *erecta* (Ler). In reciprocal F_1 hybrids derived from crosses between C24 and Col, the C24-specific end was

only detected in plants with C24 as female parent. The regulatory element required for the formation of the C24-specific transcript terminus must therefore be encoded in the mitochondrial genome, since in this species the mitochondria are transmitted exclusively through the egg cell. Indeed, differences between the three ecotypes in the mitochondrial DNA at the *cox3* locus were detected. In Col and L*er*, the genomic arrangement upstream of position -592 differs from that in observed in C24 due to an 1.8 kb insertion.
However, the C24-specific 5' terminus at position -444/-437 is located in a region where the sequences of the different ecotypes are identical, indicating that this mRNA end and the *cis* element necessary for its formation are separated by at least 150 nucleotides.

Finally, the red fluorescent protein from the sea anemone *Entacmaea quadricolor* was introduced as marker protein in plants. A tobacco line stably expressing eqFP611 targeted to mitochondria was established to create a convenient source of cells with an "endogenous" mitochondrial marker. Protoplasts from these plants can be used in future experiments to determine the subcellular localization of putative components of the mitochondrial RNA processing apparatus.

5. References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res. 25: 3389 - 3402

Bellaoui M, Pelletier G and Budar F (1997) *The steady-state level of mRNA from the Ogura cytoplasmic male sterility locus in Brassica cybrids is determined post-transcriptionally by its 3′ region.* EMBO J. 16: 5057 - 5068

Binder S, Thalheim C and Brennicke A (1994) *Transcription of potato mitochondrial 26S rRNA is initiated at its mature 5′ end.* Curr Genet. 26: 519 - 523

Bunkelmann JR and Trelease RN (1996) *Ascorbate peroxidase. A prominent membrane protein in oilseed glyoxysomes.* Plant Physiol. 110: 589 - 598

Chase CD (2006) *Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions.* Trends Genet. doi:10.1016/j.tig.2006.12.004

Claros MG and Vincens P (1996): *Computational method to predict mitochondrially imported proteins and their targeting sequences.* Eur. J. Biochem. 241: 779 - 786

Däschner K, Couée I and Binder S (2001) *The mitochondrial isovaleryl-coenzyme a dehydrogenase of arabidopsis oxidizes intermediates of leucine and valine catabolism.* Plant Physiol. 126: 601 - 612

Dombrowski S, Brennicke A and Binder S (1997) *3′-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators.* EMBO J. 16: 5069 - 5076

Dombrowski S, Hoffmann M, Guha C and Binder S (1999) *Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria.* J Biol Chem. 274: 10094 - 10099

Emanuelsson O, Nielsen H, Brunak S and von Heijne G (2000) *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence*. J Mol Biol. 300: 1005 - 1016

Finnegan PM and Brown GG (1990) *Transcriptional and Post-Transcriptional Regulation of RNA Levels in Maize Mitochondria.* Plant Cell. 2: 71 - 83

Frank DN and Pace NR (1998) *Ribonuclease P: unity and diversity in a tRNA processing ribozyme.* Annu Rev Biochem. 67: 153 - 180

Gagliardi D and Binder S (2007) *Expression of the plant mitochondrial genome.* In Logan D (ed.), Plant mitochondria. Blackwell Publishing, Ames, IA, USA. In press

Germain V, Rylott EL, Larson TR, Sherson SM, Bechtold N, Carde JP, Bryce JH, Graham IA and Smith SM (2001) *Requirement for 3-ketoacyl-CoA thiolase-2 in peroxisome development, fatty acid beta-oxidation and breakdown of triacylglycerol in lipid bodies of Arabidopsis seedlings.* Plant J. 28: 1-12

Hanic-Joyce PJ, Spencer DF and Gray MW (1990) *In vitro processing of transcripts containing novel tRNA-like sequences ('t-elements') encoded by wheat mitochondrial DNA.* Plant Mol Biol. 15: 551 - 559

Haseloff J, Siemering KR, Prasher DC and Hodge S (1997) *Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly.* Proc Natl Acad Sci U S A. 94: 2122 - 2127

Hofmann TJ, Min J and Zassenhaus HP (1993) *Formation of the 3′ end of yeast mitochondrial mRNAs occurs by site-specific cleavage two bases downstream of a conserved dodecamer sequence.* Yeast. 9: 1319 - 1330

Holec S, Lange H, Kühn K, Alioua M, Börner T and Gagliardi D (2006) *Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase.* Mol Cell Biol. 26: 2869 - 2876

Kato A, Hayashi M, Kondo M, Nishimura M (2000) *Transport of peroxisomal proteins synthesized as large precursors in plants*. Cell Biochem Biophys. 32: 269 - 275

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 - 46

Kuhn J, Tengler U and Binder S (2001) *Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria.* Mol Cell Biol. 2001 21: 731 - 742

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

Kutschera U and Niklas KJ (2005) *Endosymbiosis, cell evolution, and speciation.* Theory Biosci. 124: 1 - 24

Lippincott-Schwartz J and Patterson GH (2003) *Development and use of fluorescent protein markers in living cells.* Science. 300: 87 - 91

Logan DC (2006) *The mitochondrial compartment.* J Exp Bot. 57: 1225 - 1243

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

Marchfelder A and Brennicke A (1994) *Characterization and partial purification of tRNA processing activities from potato mitochondria.* Plant Physiol. 105: 1247 - 1254

Marker C, Zemann A, Terhorst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie JP, Brosius J and Hüttenhofer A (2002) *Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana.* Curr Biol. 12: 2002 - 2013

Martin M, Cho J, Cesare AJ, Griffith JD and Attardi G (2005) *Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis.* Cell. 123: 1227-1240

Nagata N, Saito C, Sakai A, Kuroiwa H and Kuroiwa T (1999) *The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance.* Planta. 209: 53 - 65

Ojala D, Montoya J and Attardi G (1981) *tRNA punctuation model of RNA processing in human mitochondria.* Nature. 290: 470 - 474

Perrin R, Lange H, Grienenberger JM 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 EH, Zaepfel M, Kim YJ, Mache R, Grienenberger JM, Gualberto JM 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

Prasher DC, Eckenrode VK, Ward WW, Prendergast FG and Cormier MJ (1992) *Primary structure of the Aequorea victoria green-fluorescent protein.* Gene. 111: 229 - 233

Raczynska KD, Le Ret M, Rurek M, Bonnard G, Augustyniak H and Gualberto JM (2006) *Plant mitochondrial genes can be expressed from mRNAs lacking stop codons*. FEBS Lett. 580: 5641 - 5646.

Schuster W, Hiesel R, Isaac PG, Leaver CJ, Brennicke A (1986) *Transcript termini of messenger RNAs in higher plant mitochondria.* Nucleic Acids Res. 14: 5943 - 5954

Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE and Tsien RY (2004) *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nat Biotechnol. 22: 1567-1572

Takaku H, Minagawa A, Takagi M and Nashimoto M (2004) *The N-terminal half-domain of the long form of tRNase Z is required for the RNase 65 activity.* Nucleic Acids Res. 32: 4429 - 4438

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

Van Gestel K and Verbelen JP (2002) *Giant mitochondria are a response to low oxygen pressure in cells of tobacco (Nicotiana tabacum L.).* J Exp Bot. 53: 1215 - 1218

Vogel A, Schilling O, Späth B and Marchfelder A (2005) *The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties.* Biol Chem. 386: 1253 - 64

Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y and Liu YG (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 - 87

Warmke HE and Lee S-LJ (1978) *Pollen abortion in T cytoplasmic male-sterile corn (Zea mays): a suggested mechanism.* Science. 200: 561 – 563

Wiedenmann J, Schenk A, Röcker C, Girod A, Spindler KD and Nienhaus GU (2002) *A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria).* Proc Natl Acad Sci U S A. 99: 11646 - 11651

Yanushevich YG, Staroverov DB, Savitsky AP, Fradkov AF, Gurskaya NG, Bulina ME, Lukyanov KA and Lukyanov SA (2002) *A strategy for the generation of non-aggregating mutants of Anthozoa fluorescent proteins.* FEBS Lett. 511: 11 - 14

6. Deutschsprachige Zusammenfassung

In der vorliegenden Arbeit wurden Untersuchungen zu mRNA-Enden in Mitochondrien von *Arabidopsis thaliana* durchgeführt.

Da bisher nur Transkriptenden vereinzelter Gene bekannt waren, wurden zunächst systematisch in einem standardisierten Ansatz mittels CR-RT-PCR (circularized RNA-reverse transcription-polymerase chain reaction) 5'- und 3'-RNA-Enden aller Protein-kodierenden Gene im mitochondrialen Genom von *A. thaliana* erfasst. Daneben wurden für einige Transkripte Primer-Extension- und Northern-Blot-Experimente durchgeführt, um die Ergebnisse der CR-RT-PCR-Analysen zu untermauern

In drei Fällen (*ccb3, nad6* und *orfX*) liegt das 3'-Ende des Haupttranskripts noch innerhalb des Leserasters, während bei *rps4* das 5'-Ende innerhalb des Start-Codons liegt. Es ist somit nicht sicher, ob es sich bei diesen Transkripten um funktionale RNAs handelt. Bei allen anderen mRNAs liegt die Ausdehnung der 5'- und 3'- nicht translatierten Bereiche zwischen 15 und mehreren hundert Nukleotiden.

Die Sequenzen um die so erhaltenen Enden wurden auf Ähnlichkeiten hinsichtlich der Nukleotidabfolge und potentieller Sekundärstrukturen untersucht. Es gibt kein erkennbares gemeinsames Motiv innerhalb der Primärsequenz in der Umgebung der verschiedenen Transkriptenden. Allerdings handelt es sich bei einigen der gefundenen 5'-Termini wahrscheinlich um primäre Enden, und etwas stromaufwärts dieser Enden finden sich mehr oder weniger gut konservierte Promotormotive. Zudem liegen einige Transkripttermini jeweils an derselben Stelle innerhalb eines mehr oder weniger großen duplizierten Bereichs. Dabei handelt es sich um die 3'-Enden von *nad1* und *atp9* einerseits sowie von *atp6-1* und *atp6-2* andererseits beziehungsweise die 5'-Enden von *atp9*, *nad6* und der 26S rRNA.

Was Gemeinsamkeiten in der Sekundärstruktur anbelangt, so liegen die 3'-Enden von *atp6-1* und *atp6*-2 jeweils unmittelbar stromaufwärts der $tRNA^{Ser}$, ebenso wie sich die 5'-Enden von *ccb6c* und *rps3* stromabwärts der $tRNA^{Gly}$ bzw. der $tRNA^{Lys}$ finden. Die 5'-Enden von *cox1*, *rps4* und *ccb6n1* liegen hingegen unmittelbar stromabwärts von so genannten t-Elementen, d.h. tRNA-ähnlichen Strukturen. Die 3'-Enden von *ccb3* und *nad6* finden sich unmittelbar stromaufwärts solcher Strukturen. Die 5'-Enden der Transkripte von *rpl5* und *atp6-2* liegen stromabwärts von potentiellen Haarnadelschleifen-Strukturen, die von der Länge der gepaarten Region und dem ungepaarten 3'-terminalen Nukleotid her Ähnlichkeit mit dem Akzeptor-Stamm einer tRNA aufweisen. Das 5'-Ende von *nad7* wiederum liegt unmittelbar stromaufwärts einer Haarnadelschleife. Die mit tRNAs, t-Elementen oder solchen Haarnadelschleifen assoziierten Transkriptenden könnten somit alle durch direkte

endonukleolytische Prozessierung der prä-mRNAs durch eine tRNase Z bzw. RNase P erzeugt werden. Für die meisten 5'- und fast alle 3'-Enden lassen sich aber keine dieser Sekundärstrukturen finden, so dass an deren Erzeugung zusätzliche Hilfsfaktoren oder andere Ribonukleasen beteiligt sein müssen.
Auffällig ist, dass sich von *atp9*, *nad1* und *cox2* abgesehen unmittelbar stromaufwärts der 3'-Enden keine doppelten oder einfachen Haarnadelschleifen finden. Frühere *in vitro*-Studien haben gezeigt, dass bei einigen mitochondrialen mRNAs der 3'-5'-exonukleolytische Abbau der Vorläufertranskripte an solchen Strukturen zum Stillstand kommt, wodurch sowohl die Position des reifen 3'-Endes festgelegt als auch die RNA gegen weiteren Abbau stabilisiert wird. Dies muss bei den meisten Transkripten somit auf andere Art und Weise erreicht werden.

Bei den beobachteten Transkriptermini kann es sich entweder um primäre Enden handeln, die direkt auf die Initiation bzw. die Termination der Transkription zurückgehen, oder aber um sekundäre Enden, die posttranskriptional durch endo- oder exonukleolytische Prozessierung erzeugt werden. In weiter gehenden Untersuchungen wurde beispielhaft gezeigt, dass die 5'-Enden von *cox1* und *atp9* durch endonukleolytische Prozessierung entstehen. In modifizierten CR-RT-PCR-Ansätzen, die auf der Ligation der hypothetischen 5'-leader-Moleküle an 5S rRNA beruhen, wurde die Existenz von RNA-Molekülen nachgewiesen, deren 3'-Enden unmittelbar stromaufwärts der 5'-Enden der zugehörigen reifen mRNAs liegen, wie bei einem endonukleolytischen Schnitt an dieser Stelle zu erwarten.

Die notwendige Information für die Festlegung eines bestimmten Transkriptendes kann entweder in *cis*-Elementen enthalten sein oder durch *trans*-Faktoren vermittelt werden. *cis*-Elemente sind in der mitochondrialen DNA kodiert. Auf DNA-Ebene dienen sie als Bindestellen für Transkripitonsinitiations- oder terminationsfaktoren. In einer prä-mRNA können sie durch Ausbildung einer bestimmten Sekundärstruktur oder aufgrund ihrer Primärstruktur als Bindestellen für an der Prozessierung beteiligte Proteine wirken. Bei *trans*-Faktoren handelt es sich sehr wahrscheinlich um Kern-kodierte Proteine, die entweder Transkriptionsinitiations- oder –terminationsfaktoren darstellen oder an die prä-mRNA binden und dort eine nukleolytische Prozessierung entweder ermöglichen, ausführen oder verhindern. Es wäre auch denkbar, dass es sich bei den *trans*-Faktoren um RNAs handelt bzw. dass sie mitochondrial kodiert werden.

Ein Ansatzpunkt diese Determinanten zu identifizieren bietet sich, wenn einzelne Transkriptenden nur in bestimmten Ökotypen von *A. thaliana* auftreten. Im Fall des *cox3*-Gens wurde im Ökotyp C24 ein spezifischer 5'-Transkriptterminus gefunden, der in den Ökotypen Columbia (Col) und Landsberg *erecta* (L*er*) nicht gebildet wird. Die CR-RT-PCR-Analysen zeigten, dass alle drei Ökotypen dasselbe 3'-Ende an Position +314 aufweisen sowie ein gemeinsames 5'-Ende an Position –378/-379 teilen. Das 5'-Ende bei –444/-437 findet sich hingegen ausschließlich bei C24. In einem zugehörigen Northern-Blot-Experiment traten bei C24 zwei Hauptsignale auf, bei Col und L*er* dagegen nur eines. Die Größen der so detektierten RNAs (ca. 1,50 bzw. 1,56 kb) stehen in Einklang mit den bereits ermittelten Transkriptenden bei –378/-379 und +314 bzw. –444/-437 und +314. Somit ist das C24-spezifische Transkriptende auch von physiologischer Relevanz, da das zugehörige Transkript stabil und mengenmäßig mit dem kürzeren, in allen drei Ökotypen auftretenden Transkript vergleichbar ist.

In reziproken F_1-Hybriden zwischen C24 und Col tritt das C24-spezifische *cox3*-5'-Ende nur dann auf, wenn C24 als Mutter in der Kreuzung fungiert, wie die Ergebnisse sowohl einer Primer-Extension- als auch einer Northern-Blot-Analyse übereinstimmend zeigen. Das C24-spezifische 5'-Ende wird somit rein maternal und damit nur über die in der Eizelle weiter gegebene mitochondriale DNA vererbt. Schließt man mitochondrial kodierte *trans*-Faktoren aus - im vollständig sequenzierten mitochondrialen Genom von C24 findet man keine Hinweise auf die Existenz solcher *trans*-Faktoren -, muss somit ein *cis*-Element für das Auftreten des C24-spezifischen Transkriptendes verantwortlich sein. Tatsächlich unterscheiden sich die DNA-Sequenzen im Bereich des *cox3*-Gens zwischen den drei verschiedenen Ökotypen, wie aus den Ergebnissen mehrerer Southern-Blot-Experimente, PCRs mit anschließendem Restriktionsverdau und der Sequenzierung eines entsprechenden klonierten HindIII-Fragments aus Col abgeleitet werden kann. Demnach sind der Bereich stromabwärts des *cox3*-Leserasters, das *cox3*-Leseraster selbst sowie die ersten 592 Nukleotide stromaufwärts des Start-Codons bei allen drei untersuchten Ökotypen identisch. Stromaufwärts davon liegt in Col relativ zu C24 eine Insertion von 1.799 bp vor, während L*er* diese Insertion zwar auch aufweist, es aber zusätzlich innerhalb dieser Insertion relativ zu Col zu einer Rekombination gekommen ist.

Das C24-spezifische 5'-Ende liegt interessanterweise in einem Bereich, der in allen drei Ökotypen identisch ist, die polymorphe Region beginnt erst etwa 150 Nukleotide weiter stromaufwärts. Somit findet sich das für das C24-spezifische 5'-Ende verantwortliche *cis*-Element mindestens 150 Nukleotide von seinem Wirkort entfernt.

Sollten mögliche Kern-kodierte *trans*-Faktoren gefunden werden, so besteht der erste Schritt ihrer genaueren Charakterisierung darin, zu zeigen dass es sich um mitochondriale Proteine handelt. Dies geschieht oft mit Hilfe von Fusionsproteinen aus dem zu untersuchenden Protein und einem fluoreszierenden Protein, in der Regel dem grün fluoreszierenden Protein (GFP). Dabei wird davon ausgegangen, dass das fluoreszierende Fusionsprotein die gleiche subzelluläre Lokalisation aufweist wie das zu untersuchende Protein. Allerdings muss das zelluläre Kompartiment, in dem sich das Fusionsprotein ansammelt, eindeutig identifiziert werden, was mittels eines zweiten, spektral verschiedenen fluoreszierenden Protein, an das eine bekannte Signalsequenz angehängt ist, erfolgen kann. Zu diesem Zweck wurde das rot fluoreszierende Protein eqFP611 aus der Anemone *Entacmaea quadricolor* auf seine Einsetzbarkeit in Pflanzen und seine Kompatibilität mit GFP getestet. Durch transiente Transformation von Tabakprotoplasten wurde gezeigt, dass eqFP611 ohne Änderung der kodierenden Sequenz in Pflanzenzellen stark genug exprimiert werden kann, um durch Epifluoreszenzmikroskopie problemlos detektierbar zu sein. Seine subzelluläre Lokalisation lässt sich durch Anhängen N-terminaler Signalsequenzen effektiv steuern, ohne dass dies mit einem Verlust der Fluoreszenz einhergeht, wie anhand einer mitochondrialen Präsequenz und eines peroxisomalen Targetingsignals demonstriert wurde. Wird eqFP611 zusammen mit GFP in denselben Zellen eingesetzt, so können beide Fluoreszenzen durch entsprechende Filter klar voneinander getrennt werden. Es kommt weder zu einer wechselseitigen Beeinträchtigung der Fluoreszenzentwicklung noch zu gegenseitigen Beeinflussungen beim intrazellulären Transport. Zudem wurde eine transgene Tabaklinie etabliert, in der eqFP611 mit vorangehender mitochondrialer Präsequenz stabil exprimiert wird. Cytotoxische Effekte von eqFP611 wurden in keinem Fall beobachtet.

Somit stehen jetzt Tabakpflanzen zur Verfügung, aus denen sich bequem Protoplasten gewinnen lassen, deren Mitochondrien rot fluoreszieren. Wenn nachgewiesen werden soll, dass es sich bei einem putativen *trans*-Faktor tatsächlich um ein mitochondriales Protein handelt, so müssen nur noch ein Fusionsprotein aus dem zu untersuchenden Gen und GFP in diesen Zellen exprimiert und die räumlichen Muster der roten und der grünen Fluoreszenz verglichen werden.

7. Appendix

7.1 Own contribution

7.1.1 Manuscript 1

I have performed most of the CR-RT-PCR analyses of *cox1*, *nad4*, *ccb3* and *atp1* each and carried out the primer extension analyses of these four genes, *nad1* and *rpl5* and the northern blot of *ccb3*. Except for parts of the *ccb2*, *atp8* and *atp9* analyses, I have investigated the other transcripts in collaboration with a technician or diploma students. I have developed the strategy for those CR-RT-PCR analyses in which separate primers were used for cDNA synthesis and amplification. Finally, I have collected and analyzed all the data and results.

7.1.2 Manuscript 2

I have performed the complete Southern blot experiment shown in Fig. 2, the northern blot experiments in Fig. 4 and 6.A and the primer extension experiment in Fig. 6.B. I have conceived and carried out the PCRs presented in Suppl. Fig. S1 and S2 as well as the genotyping of the cell suspension cultures and the F_1 plants of Col/C24 crossings. I have supervised the *cox3* CR-RT-PCR analyses of mtRNA and planned and performed the analogous investigation of total RNA preparations. I have designed Fig. 1, 5, 7 and all supplementary figures and assisted writing the manuscript.

7.1.3 Manuscript 3

I have conceived all the plasmids constructed in this study, designed all primers, carried out all the respective PCRs, performed all cloning steps with one exception, prepared the plasmids for protoplast transformation, carried out protoplast transformations and analyzed the transformed protoplasts. In addition, I have grown and selected the transgenic tobacco lines and prepared the draft version of the manuscript.

7.2 Manuscripts

7.2.1 Manuscript 1: "Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5' and 3' end formation"

As preparatory work to investigate the generation and function of mitochondrial mRNA termini in *A. thaliana*, these ends have to be identified. A standardized comprehensive study of transcript ends has been performed for all mitochondrial protein-coding genes in *A. thaliana*. The results are presented in the manuscript below. Additionally, the sequences surrounding the ends obtained in this survey have been searched for similarities of primary sequences or secondary structures *in silico*. Secondary structures were found to be associated with transcript ends for at least a subset of mitochondrial mRNA. Furthermore, evidence is provided for the endonucleolytic generation of the main *cox1* and *atp9* 5' termini.

[A revised version of this manuscript has been accepted for publication in Nucleic Acids Research in April 2007.]

Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5' and 3' end formation

Joachim Forner, Bärbel Weber, Sabine Thuss, Steffen Wildum and Stefan Binder
Institut Molekulare Botanik, Universität Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany

ABSTRACT

Using CR-RT-PCR as basic method we have mapped 5' and 3' ends of all mitochondrial mRNAs in *Arabidopsis thaliana*. Almost all transcripts (31 of 32) have single major 3' termini, while for several genes multiple 5' ends have been found. Of all identified 5' ends some have been mapped within promoter motifs suggesting these ends to be derived from transcription initiation. But the majority of the 5' termini seems to be generated posttranscriptionally. Mapping of the extremities of 5' leader RNAs revealed clear evidence for an endonucleolytic generation of the major *cox1* and *atp9* 5' mRNA ends. We identified tRNA-like structures, so-called t-elements, to be associated with both 5' and 3' termini of several mRNAs. These secondary structures most likely function as *cis* signals for the endonucleolytic cleavage most likely by RNaseZ and/or RNaseP. Since no consensus motif is evident at posttranscriptionally derived ends we suggest higher order structures to be predominant *cis* elements. This analysis provides novel insights into 5' end formation of mRNAs. In addition the complete transcript map is a substantial and important basis for future studies of gene expression in mitochondria of higher plants.

INTRODUCTION

Seed plants contain the largest mitochondrial genomes investigated so far. Although their sizes reach up to about half of the *E. coli* genome, the mitochondrial genomes more or less encode only the same small set of approximately 60 genes. This gene collection is found with minor variations in all complete mitochondrial genome sequences from Arabidopsis, sugar beet, rice, rape seed, tobacco, wheat and rice (1-8). The transcription units, mono- as well as polycistronic, are distributed over the complete genomes separated by large spacer sequences without any obvious function. Transcription of mitochondrial genes frequently starts from multiple promoters of various types, generating precursor RNAs that have to pass through

various processing steps such as RNA editing, splicing of group II introns, 3' end trimming as well as formation of secondary 5' termini (9). Many or all of these steps are required to generate a mature translatable mRNA or functional transfer as well as ribosomal RNAs and can potentially contribute to the regulation of mitochondrial gene expression. But up to date it is still unknown whether or to which extend these processes control or even regulate the realization of the mitochondrial genetic information.

Transcription initiation is one of the most important levels to regulate gene expression in bacteria, archaea as well as nuclei from eukaryotes. This process has also been intensively examined in mitochondria of various plant species. Functional studies with corresponding in vitro transcription systems from mono- and dicot species defined promoter sequences with similar structures but differing sequence motifs (10-15). The plant mitochondrial promoters are single entities each containing a core motif surrounding the transcription initiation site and an upstream element. Both types of elements contain important positions indispensable for promoter function but also positions, where changes in the nucleotide identities are accepted but modulate promoter strength. In *Arabidopsis thaliana* at least two conserved promoter motifs have been found, but also a number of additional non-conserved transcription initiation sites are present (16-18). The situation might be even more complex since not all mitochondrial transcription units have been analyzed in this respect. The presence of a variety of different promoter sequences in a single plant species suggests that there might be several different proteins binding to these *cis* elements. This is supported by the constitution of the mitochondrial transcription machineries in yeast and animal combining the action of phage-type single subunit RNA polymerases with two types of transcription factors (19). While the same type of RNA polymerases has been found in plant mitochondria (20,21), no auxiliary proteins have been yet identified in these organelles (22). Still unclear is also whether plant mitochondrial gene expression is regulated or controlled at the transcriptional level. Clearly, run on transcription studies showed that mitochondrial genes are transcribed at different rates, most likely determined by differing promoter strength (23,24). But so far there is no clear evidence that expression of individual genes is actively regulated during transcription initiation events.

Several reports provided convincing evidence that also posttranscriptional processes influence plant mitochondrial mRNA steady levels (25,26). This was demonstrated by comparing the transcriptional rates with the steady state RNA levels. The observed discrepancies for several genes were interpreted to originate from posttranscriptional processes influencing RNA stability. Of course there must be *cis* elements as well as *trans* factors, which probably

determine RNA stability in a concerted action. Stem loop structures, either single or double, are good candidates for such *cis* acting processing signals. Such structures at or near the 3' ends of several plant mitochondrial RNAs have been found to influence the mRNA stability in vivo and in vitro (27-30). It is assumed that they prevent 3’ to 5’ exonucleolytic degradation of the RNA being steric barriers for exoribonucleases such as a mitochondrial PNPase and a RNaseII-like enzyme (31). Another important (*cis*) factor is the polyadenylation state on an RNA. Short oligo(A) tails have been detected predominantly at the mature 3’ ends of many plant mitochondrial RNAs (28,32,33). They have been found to destabilize RNA both in vivo and in vitro in mitochondrial protein extracts.

The 5' ends of mRNA can be generated directly by transcription initiation or by succeeding 5’ processing events. Support for the latter has been obtained by mapping of such ends of various genes in different plans species, however, it is still unclear how these ends are generated. Up to now no evidence has been found for a 5' to 3' exonucleotlytic activity. Consequently the creation of secondary 5' termini has been attributed to (a) endoribonuclease(s). So far two different endonucleolytic activities have been described in plant mitochondria both being involved in maturation of tRNAs. An RNase P-like activity has been found to cut precisely 5’ of mature tRNA 5’ end, while tRNase Z cleaves directly or one nucleotide downstream of the discriminator nucleotide (34-36). The prerequisite for the cleavage of precursor molecules by these activities is the formation of the tRNA secondary structure (37,38). Since the cleavage reaction also occurs at incomplete cloverleaf structures (Ref), tRNA-like elements (t-elements) are also substrates for these enzymes at least in vitro. Such t-elements have been detected in wheat mtDNA and could potentially be involved in cleavage of long transcripts generating secondary 5’ ends (39). But so far no evidence has been found for the function of such t-elements in vivo.

To gain more information about 5' and 3' ends of mitochondrial mRNAs in plant mitochondria, we have analyzed the extremities of all protein-coding genes annotated in the mitochondrial genome of *Arabidopsis thaliana*. This was done in ecotype Col both from mitochondrial and total RNA isolated from cell suspension culture. Single 3’ ends are found for almost all transcripts, while in several cases multiple 5’ ends are found. Some of the 5’ ends are derived from transcription initiation, while most seem to be generated posttranscriptionally. Analysis of sequences surrounding the 5’ and 3’ termini revealed that many of the ends coincide with termini of t-elements or stem-loop structures suggesting that predominantly tRNase Z but in some instances also RNase P are involved in mRNA 5’ and 3’ end processing.

MATERIALS AND METHODS

Preparation of RNA from Arabidopsis

An *Arabidopsis thaliana* ecotype Columbia cell suspension culture was cultivated on a shaker (120 rpm) in dark at 23°C. The ecotype of this culture was recently confirmed by analyzing corresponding informative genetic markers (40). Mitochondria were isolated from cultures 6 days after the transfer to new media according to a method described previously (41). RNA from these organelles (mtRNA) was extracted following previously established protocols. Alternatively mtRNA was extracted from 100 mg frozen mitochondria (fresh weight), which were disrupted in a mortar. The organelles were then suspended in the lysis buffer of a RNeasy Plant Mini kit. RNA was then isolated following the instructions of the manufacturer (Qiagen).

Total cellular RNA was isolated from 100 mg suspension cells. To this end cells were harvested 24 h after transfer to fresh medium and grinded in liquid nitrogen. From this material total RNA was purified using RNeasy Plant Mini kit as mentioned above.

Analysis of RNA

CR-RT-PCR analyses were either performed as described before (Fig. 1A) (42) and/or carried out a using modified protocol (Fig. 1B). Briefly, large scale self ligation was performed with up to 50 µg RNA in a total volume of 100 µl. After ligation samples were desalted using Microcon YM-10 or -30 micro concentrators (Millipore) and stored as aliquots of 15 µl at -20°C. First strand cDNA synthesis was done with 5 µg of total RNA and 2 µg of mitochondrial RNA, respectively, and 200 U M-MLV RNase H Minus (point mutant, Promega) reverse transcriptase under conditions recommended by the manufacturer. The RNA template was then degraded by adding 1/5 volume of 1 M NaOH and an incubation of 10 minutes at room temperature. The sample was subsequently neutralized with an equal amount of 1 M HCl and the cDNA was purified with the GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare). This purification step also removed the primer used for cDNA synthesis. This oligonucleotide could cause the amplification of multiple PCR-products when it is present in the first amplification reaction containing another primer with the same orientation. One fifth of the cDNA sample was used as template in a single RT-PCR. In some cases amplification was done in a 2-step PCR using primers with melting temperature of about 73 °C. This allows annealing and synthesis to be performed at 68 °C.

The extremities of the *nad6* t-element were determined by using mitochondrial 5S rRNA as linker molecules. These molecules were linked to the target RNA by self ligation of 5 µg of mitochondrial RNA as described above. To map the 5' end of the t-elements cDNA synthesis was initiated from primer Atnad6-6 (40 to 23). cDNA across the ligation site was amplified in a PCR with primer pair At5S-Mega.H (-2 to 22)/Atnad6-Endo-3'.R (37 to 11). The 3' terminus was determined after cDNA synthesis with oligonucleotide At5S-5 (to)and amplification with primer pair Atnad6-Ende-3'. H (-15 to +19)/At5S-Mega.R (107 to 82).Ana analogous approach was applied to map the 3' end of the cox1 t-elements. Details are given in Results.

For mapping of the 3' end of *ccb3* t-element a conventional 3' RACE analysis was done using an oligo(dT)17-anchor primer (DTXSC) for cDNA first strand synthesis followed by two amplfication reaction done with primer pairs XSC/Atccb3-8 (729 to 748) and XSC/Atccb3-9 (767 to 787). All PCR products were cloned and individual clones sequenced.

For northern blot experiments 3 to 10 µg of total and 1 to 3 µg of mitochondrial RNA, respectively, were size fractionated on 1 % (w/v) agarose gels using glyoxal as denaturing agent. Gel electrophoresis, blotting and hybridization were done as described previously (43). Nucleic acids were blotted onto Duralon UV (Stratagene) or Hybond-XL membranes (GE Healthcare) and hybridized with radioactively labeled probes as outlined in the manufacturers' guidelines. Primer extension analyses were performed according to standard protocols (43).

Miscellaneous methods

Sequencing of PCR products was commercially obtained (MWG Biotech and 4base lab). Individual cDNA clones were sequenced using Thermo Sequenase Primer Sequencing kit as recommended by the manufacturer (GE Healthcare) and an ALF sequence system (GE Healthcare). In silico sequence analyses were done at the NCBI server using various blast tools (44). Secondary structure prediction were done the with program "stemloops" from Wisconsin GCG software package version 10.2, with program searching for palindromes (`http://bioweb.pasteur.fr/seqanal/interfaces/palindrome.html`) and a web interface RNA-fold program (`http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi`). Searches for conserved sequence motifs were done with the MEME search tools (http://meme.nbcr.net/meme/meme.html).

RESULTS

The experimental approach to map major mRNA termini in *A. thaliana* mitochondria

The *Arabidopsis thaliana* mitochondrial genome encodes 32 protein-coding genes. To gain more information about the major termini of the transcripts of these genes in this model plant species we analyzed these RNAs by CR-RT-PCR. This experimental approach was used since it allows a simultaneous determination of 5' and 3' termini, the detection of non-encoded nucleotides and delivers unambiguous information by sequencing the amplified cDNA products. In addition this method allows a fine mapping of the ends on the nucleotides level.

Generally two different approaches have been used, both starting from gene-specific cDNA first strands synthesized on self ligated RNA (Fig. 1). In the first approach, which was mainly used in the initial phase of the project, products were amplified from mitochondrial RNA isolated from *A. thaliana* ecotype Columbia (Col) cell suspension culture. A first PCR was performed with the primer used for cDNA synthesis, which anneals in the 5' terminal region of the reading frame and a forward primer complementary to sequences located in the 3' terminal part of the gene. Products of this reaction were inspected on agarose gels and most prominent fragments selected for further amplification in (a) nested PCR(s). The resulting fragment(s) was (were) then cloned and a set of about 20 different cDNA clones per PCR product were sequenced (Fig. 1A).

In the alternative approach gene-specific cDNA synthesis was performed on both mitochondrial and total RNA from Col cell suspension culture. After cDNA synthesis only single PCR was performed with oligonucleotides annealing to 5' and 3' terminal sequences of the reading frame (Fig. 1B). In this PCR the primer annealing to the 5' part of the gene is different form the oligonucleotide used for cDNA synthesis. In some cases long primers were used to increase specificity in these reactions. Products, which appear identical in the PCR analyses of both RNAs were selected for sequencing in the majority of cases from both directions using the PCR primers. This delivers unambiguous sequences up to the ligation site. Downstream of this site several sequences are superposed with multiple peaks at each nucleotide position, which in many cases results in non-readable sequence. Usually the product obtained from mtRNA was used in the sequence reaction.

All protein-coding genes have been analyzed using the second approach. For those genes investigated with both approaches identical results were obtained. To verify the reliability of the CR-RT-PCR results primer extension as well northern blot analyses were performed for some genes.

Identification of the *cox1* transcript ends

The analysis of *cox*1 gene is here described as an example. This gene is flanked by 10 kb 5' and 17 kb 3' non-coding sequences and this isolated location suggests a monocistronic *cox*1 mRNA transcribed from its own promoter (Fig 2A). A transcription initiation site has indeed been found 355 nucleotides upstream of the ATG (18).

The *cox1* transcripts have been analyzed by both CR-RT-PCR approaches. In the first approach cDNA synthesis was initiated from oligonucleotide Atcox1-1 complementary to sequences from position +156 to +138 relative to the ATG (+1). The same primer and oligonucleotide Atcox1-3 (+1473 to 1493) were used in the first amplification reaction yielding three products with sizes of about 400, 500 and 650 nucleotides, respectively (Fig. 2A and B). These were subsequently used as DNA templates in three different reactions (second PCR) each performed with primer pair Atcox1-2 (+118 to +98)/Atcox1-4 (+1501 to +1520). While no product was obtained in the reaction with the smallest DNA template, identical fragments of 500 and about 300 nucleotides were generated in reactions with the 650 and 500 bp templates, respectively. The 500 bp product from the second PCR was then cloned, which yielded cDNA clones with sizes between 270 and 500 bp (data not shown). In the sequence of 32 clones of various sizes, 3' ends were found 23 times 47 nucleotides (position 349,783 in complete sequence nc_001284.2), six times 46 (349,784) and one time 45 (349,785) nucleotides downstream of the stop codon, respectively. In addition a single cDNAs were found with a 3' end in the reading frame at position 349,871, 40 nucleotides upstream of the stop codon and 64 nucleotides downstream of the reading frame, respectively. This strongly suggests a single, slightly scattering 3' ends around position 349,783 located 47 bp downstream of the reading frame. No non-coding adenosines or other homopolymeric sequences were found in the *cox1* mRNAs, however, two clones contained four and 29 bp between the 5' and 3' ends, the larger of which was identified as mitochondrial 18S rRNA fragments.

In contrast to the homogenous 3' terminus the 5' ends found in the different clones vary substantially. The 5' termini in the largest cDNA inserts were found five times at position 351,654, two times at 351,653 and three times 351,652, 241 to 239 nucleotides upstream of the start codon. All other clones revealed ends between -216 to -20 with two minor clusters between 187 and 179 and -48 and -43. Taken together this suggests that a major 5' terminus of the steady state *cox1* mRNA is located at position 351,654 241 nucleotides upstream of the ATG.

In the second approach cDNA synthesis was again started from primer Atcox1-1 followed by a single amplification reaction with primer pair Atcox1-2/Atcox1-4. Beside some very weak products of about 350 bp, a strong 500 bp cDNA was amplified from both RNAs (Fig. 2C). The latter was thus directly sequenced with oligonucleotide Atcox1-2 directed to the 5' end (Fig. 2D). This revealed a sequence across the ligation sites of the *cox1* steady state mRNA pool. The chromatogram shows a clear sequence up to the adenosine at position 351,654 which is followed by a sequence with continuous minor signals. This transition marks the ligation site and determines the transcript termini. From this it can be concluded that the majority of the *cox1* transcripts start with a thymidine at position 351,654 241 nucleotides upstream of the ATG and ends with another thymidine at position 349,783 47 nucleotides downstream of the reading frame. The fact that the sequence can be clearly followed across the ligation site indicates that the vast majority of the *cox1* mRNA molecules have exactly identical 5' and 3' ends. The appearance of minor signals beyond the ligation site also shows the presence of a minor fraction of *cox*1 RNA molecules with slightly different ends. Both approaches consistently identify the *cox1* mRNA 5' end at position 351,654 located 241 upstream of the ATG and the 3' terminus at position 349,783 corresponding to a 47 nucleotide 3' UTR.

To check the mapping results by independent methods primer extension reactions were carried out with primers Atcox1-2 and Atcox1-5 (-129 to -148) using mtRNA as template. Atcox1-2 is elongated to a strong product of 360 nucleotides corresponding to the above detected 5' end at -241 and some weak products indicating additional minor ends up- and downstream of the main terminus (Fig. 3A). A higher resolution is obtained after separation of Atcox1-5 extension products along with sequencing reaction products obtained with the same primer. Again major products were found at position -241 to -239, interestingly with same quantitative distribution as seen in the CR-RT-PCR (Fig. 3B). In addition smaller and larger minor products were found, one of which could correspond to an end at position -332. For further analysis by an independent method a northern blot experiment was performed. This detects a single transcript of about 1900 nucleotides in each RNA preparation consistent with the size calculated on the basis of the mapping data (i.e. 1872 nucleotides) (Fig. 3C).

In summary these results clearly show that in *A. thaliana* the major *cox1* transcripts range from nucleotide position 351,654 located 241 upstream of the ATG to position 349,783 47 bp downstream of the stop codon consistent with a mRNA of 1872 nucleotides. Thus the major 5' terminus is located 114 bp downstream of the previously identified promoter 355 bp

upstream of the ATG (18). In addition these data document again the reliability of the CR-RT-PCR analysis, particularly of the approach with direct sequencing of PCR products.

Single 3' ends are detected for almost all Arabidopsis mitochondrial mRNAs

Using the experimental approaches shown in Fig. 1 we analyzed all transcripts of the mitochondrially encoded proteins in *A. thaliana*. It can be assumed that both CR-RT-PCR strategies detect ends that are the most abundant ones and which are not too far away from the primers used. This has been confirmed for many transcripts, either by northern or primer extension analysis for instance for *cox1* and *atp9*.

All 3' termini detected are given in Table 1. For almost all mRNAs single slightly scattering 3' ends are found. The only exception is the *ccb6c* transcript for which two ends have been detected without a specific search for additional ends. One of the *ccb6c* ends (around +115) shows a remarkably strong scattering over several nucleotides (Table 1).

The identified 3’ ends are usually located downstream of the reading frame with 3’ UTRs up to 498 nucleotides as for instance found for *rps4*. But the major 3' termini of the *ccb3* and *nad6* genes were found within the reading frame (*ccb3*: -46 and *nad6*: -17) confirming the results published recently (45). Consistent with this previous analysis CR-RT-PCRs with primers located downstream of these 3' ends did not detect further downstream located ends. However, an alternative approach applying a 3’ RACE analysis to detect polyadenylated ccb3 mRNA 3’ ends identified such termini downstream of the reading frame. The same holds true for *nad6*. To map additional 3’ mRNA ends downstream of this gene a CR-RT-PCR with 5S rRNA as anchor molecule was performed. Likewise this revealed 3' termini downstream of the translation stop codon (details see below).

Two other 3' termini were identified within pseudo genes (*orfX* and *sdh4*) further corroborating that these truncated reading are non-functional (46).

A MEME analysis of sequences flanking the 3’ ends (-/+ 20 bp???) did not reveal any conserved sequence motif. However it is remarkable that transcripts ending with a guanidine are under-represented. The mRNAs ending with a G include the 3’ ends of both *atp6* transcripts, which are generated endonucleotlytically by RNase P cutting at the 5’ end of the downstream encoded $tRNA^{Ser}$ (UGA). Each ten transcripts end with cytidines and uridines, respectively, while the mRNAs of seven genes end with adenosines, which is about 25 % which would be expected from a random distribution.

The sequences flanking the 3' termini were also screened for the presence of inverted repeats encoding stem loop structures (SL). The most striking structure is a double stem loop found exactly upstream of the 3' end of the *atp9* mRNA (Fig.4). This has identically been found in a previous study, which suggested this structure to have a function in the exonucleolytic maturation of this end (31). Interestingly parts of this double inverted repeat are duplicated in the 3' flanking region of the *nad1* exon e with the downstream located SL being completely identical. The 3' ends of the mature *nad1* transcripts are like those of the *atp9* mRNAs indicating a function of the identical downstream located SL in 3' end formation. A single stem loop is found directly upstream of the *cox2* 3' terminus, also suggesting a function of this SL in the generation of this end (Fig. 4). All other inverted repeats form single SLs and are positioned several nucleotides upstream of the ends (data not shown). This includes inverted repeats in the 3' UTRs of the *cox1*, *nad2* and *rps7* genes. Two SLs separated only by ten nucleotides are present in the *atp8* 3' UTR. Interestingly the 3' terminus of the latter transcript maps one bp downstream of the 3' end of a short stable RNA (Ath-377) identified in a general RNomics analysis in *A. thaliana* (47). Likewise the 3' end of *nad4L-orf25* mRNA is identical with the terminus of Ath-290.

5' ends of mitochondrial transcripts in *A. thaliana*

The 5' termini identified in our analysis are listed in Table 2. In principle the ends could originate directly from transcription initiation (referred to as primary ends) or could be formed by posttranscriptional processing (known as secondary ends). Several major 5' ends detected in our experiments have been found in a previous search for primary ends in *A. thaliana* mitochondria (18). Exactly in agreement with results of this study we identified 5' termini for *atp6*-1 (-200) *atp8* (-157) and *atp8* (-228). In the latter case we found an additional minor end at position -224 instead of -226 as detected before. Likewise a difference is seen for a *ccb2* primary end, which we found five bp further downstream at position -205. Five other major 5' ends could as well be of primary origin with the strongest candidate being an end of the *cox2* mRNA at position -140, which maps within a perfectly conserved promoter motif (18). In addition two minor ends of *atp1* RNAs were found at the predicted site within the CNM1 promoter motif. These have only been observed in CR-RT-PCR experiments specifically designed to identify these predicted ends.

Considering the variety of different promoter sequences found in *A. thaliana* mitochondria a clear assignment of some of the ends identified in this study is difficult. Nevertheless it is

reasonable to assume that most of the 5’ ends detected are the result of posttranscriptional processes since the surrounding sequences do not show any similarity to previously characterized promoter motifs (14-16,18,48,49). Moreover the mechanism of RNA ligation allow only the connection of 5’ monophosphate ends and excludes 5’ termini with two or three phosphates as one would expect it at primary ends derived from transcription initiation (50,51).

Of the 32 major ends potentially derived from posttranscriptional processing only four transcripts start with a cytidine, six with guanidine, ten with adenosine and eleven with uridine. As for the 3’ ends no conserved general sequence motif emerged at or within the sequence flanking the secondary ends, while the MEME analysis identified the promoter motif when the primary ends were analyzed (data not shown). However we noticed a striking match between the primary structures enclosing identical 5’ terminal nucleotides of the 26S rRNA, the *nad6* as well as *atp*9 mRNAs (Fig. 5) Here 24 nucleotides are perfectly conserved between sequences around the 5’ ends of 26S and *nad6*, while discrepancies were seen in two positions of the *atp9* sequence. This might suggest that the primary sequences itself might be important for the generation of these ends or that they are a portion of an at least partially conserved secondary structure element, which is necessary for processing. The sizes of the 5' UTR are found to be up to 645 nucleotides (*nad1*). We have identified the 5’ end of *atp1* transcripts at -1898 but this end was detected in a particular search for the atp1 primary transcript ends at a predicted promoter motif. Curiously the mRNA of *rps4* has no 5' UTR at all. Here the major 5' terminus is found at position +2 within the ATG. It is thus unclear whether this gene is functional or not.

Potential dicistronic transcripts

The A. thaliana mitochondrial genome encodes a number of genes that are potentially transcribed within dicistronic transcripts. This includes *rpl5-cob*, *rpl2-orfX*, *nad3-rps12*, and *rps3-rpl16*. All these genes were examined by CR-RT-PCR considering them as momocistronic mRNA or as dicistronic transcripts. Clear PCR products and ends were found for potential monocistronic mRNAs of *rps12* and *rps3,* while no distinct ends of have been detected for such transcripts of all other genes. However, when we examined the potential monocistronic transcripts of the downstream located genes of *nad3-rps12* and *nad4L-orf25* arrangements we detected the 5’ ends upstream of the 5’ located gene, which strongly suggests a co-transcription of these genes in the case.

All putative dicistronic genes were also investigated with CR-RT-PCR with one primer annaleaing to 5' terminal region of the upstream reading frame and a primer annealing to 3' terminal part of the upstream located gene. In all cases these primer combinations yielded clear PCR products and ends. This together with the lack of clear ends of potential monocistronic RNAs suggests the existence dicistronic transcripts of these genes. In the case of *rps12* and *rps3* additional monocistronic mRNAs seem to be present in the steady state RNA pool.

Endonucleolytic cleavage generates the 5' end of the mature cox1 mRNA

As indicated in the previous section most 5' ends detected in our transcript end analysis are most likely generated posttranscriptionally. Since in plant mitochondria no 5' exonucleolytic activity has so far been detected these ends are most likely derived from endonucleolytic cleavage (9). To obtain experimental evidence for an endonucleolytic 5' processing reaction we used an experimental approach in which the mitochondrial 5S RNA was used as an "anchor" molecule. This rRNA was ligated to the 3' end of a potential 5' cleavage product (details see Materials and Methods and Fig. 6A).

For a potential cox1 5' cleavage product a 195 bp cDNA fragment is expected in an amplification reaction with primers Atcox1-lm.H and At5S-mega.R on a cDNA template synthesized from oligonucleotide At5S-5. Products of this reaction were separated on an agarose gel and the predominant cDNA fragment obtained in this reaction is indeed consistent with the expected size (Fig. 6B). This product was thus cloned and sequenced. In a total number of 20 clones 19 cDNA inserts contained 3' ends corresponding to nucleotide position -242. This is the position exactly upstream of the major mature 5' end of the cox1 mRNA at -241 (Fig. 2D and 3, Tables 2 and 3). In addition two ends were found at -243 and -239 for which the corresponding 5' ends of the *cox*1 mRNA at -242 and -238 have not been detected, indicating that the ends might vary slightly more than seen in the above described analysis of the *cox1* 5' mRNA extremities (Fig. 2 and 3). Thus the results from this indirect approach strongly suggest that the major 5' end of the *cox1* transcript around position -241 is generated by endonucleotlytic cleavage. Interestingly 15 cDNA clones representing *cox1* 5' leader RNAs contained non-encoded cytidines and adenosines at the 3' termini. These nucleotides might be added by a terminal tRNA nucleotidyltransferase and indeed perfectly conserved pseudo uridine arm as it is usually found in tRNAs is found upstream of this 3' end. Moreover the 87 nucleotides of the cox1 5' leader molecule can be folded into a tRNA-like structure

(Fig. 6C). This clearly indicates that the 5' end of the *cox1* mRNA at position -241 is generated by a cleavage reaction most likely catalyzed by RNase Z. This result raised the hypothesis that other termini of mitochondrial mRNA might be generated by RNase Z or by RNase P.

Analogous experiments performed to detect potential processed *atp9* 5' leader molecules revealed that also the major 5' end at position –83 is derived from an endonucleolytic cut (data not shown).

t-elements contribute to the formation of 5' and 3' ends

To check whether other mitochondrial 5' and 3' mRNA ends could be derived from cleavage by RNase Z or RNase P we screened the complete sequence of the *A. thaliana* mitochondrial DNA for the presence of the highly conserved 5'-GGTTCRANYCC-3' motif present in the pseudo uridine arm of tRNAs. Apart from the canonical tRNA genes we found such structures upstream of the 5' ends of *rps4* and *ccb6n1* as well as downstream of the gene-internal 3' ends of *ccb3* and *nad6*, indicating that they could be involved in the generation of 5' ends as well as 3' termini. Folding of these sequences revealed that these conserved motifs are parts of t-elements with ends coinciding with the experimentally mapped transcript termini (Fig. 7).

To gain more experimental evidence for the importance of these t-elements in RNA processing, we searched for 3' ends mapping downstream of the major *ccb3* 3' termini, which were located within the reading frame. Such ends would be expected if the t-elements would be cleaved off from respective precursor RNAs as postulated. To this ends we used a 3' RACE approach to identify 3' ends of *ccb3* t-elements. This analysis revealed a ccb3 3' end consistent with the t-element model presented in Fig. 7.

This strongly suggests that this end might be also generated by RNase Z cleavage. Moreover these results demonstrate that the t-element provoke the cleavage of *ccb3* mRNAs within the readings frame, reactions that are most likely performed by RNaseP. (Fig. 7).

In summary the results presented in the study show, that both tRNA processing enzymes generate 5' and 3' ends of plant mitochondrial mRNAs. This might not only occur at canonical tRNAs such as $tRNA^{Gly}$ and $tRNA^{Lys}$ located upstream of *ccb6c* and *rps3* and $tRNA^{Ser}$ encoded downstream of the *atp6* genes, but also at t-elements. RNase Z and RNase P are thus enzymes that are responsible for endonucleolytic processing of at least a subset of plant mitochondrial mRNAs.

DISCUSSION

General features on plant mitochondrial mRNAs

In our study we have analyzed the 5’ and 3’ ends of all mRNAs of the annotated mitochondrial genes in *A. thaliana*. This comprehensive study reveals several general features of these organellar transcripts in plants. First, almost all transcripts have a single 3’ end that in the vast majority of mRNAs scatters only over a few nucleotides. The only real exception is the *ccb6c* mRNA, which has two 3’ ends 105 nucleotides apart from each other. Second, for several genes we find mRNAs with different 5’ ends. Thus the 5' termini seem to be generally more variable than the 3’ ends. Third, with the exception of the primary 5’ ends within promoter sequences, there is no evident conserved sequence motif at the 5’ termini. This holds also true for the 3’ ends. Thus other determinants govern the posttranscriptional generation of these ends.

Fourth, the majority of the 5’ ends are derived from processing. There are several arguments that support this assumption. Most of the ends identified do not resemble any known promoter motif neither the conserved one nor those that have been found only once (18). In addition upstream of many of the 5' ends found in our study promoter sequences have been found for instance in the case of *atp9*, *atp1*, *cox1*, *cox2*, *rps3* (18) and potential conserved promoters can be predicted upstream of *cox3* (-640 in ecotype C24, here 5’ ends have been found at this position), *nad3* (-1320 and 1230), *nad1* (-1400) and *ccb6n1* (-890). Moreover our analysis shows that the 5’ ends of the *cox1* and the *atp9* mRNAs are generated by endonucleolytic cleavage. This might likewise occur for other secondary 5’ ends and might be even postulated for certain 3’ ends (detailed discussion see below).

Our analysis also revealed non-encoded nucleotides most likely attached to the 3' ends. Mostly single adenosines are found and with less frequency cytidines, but in some rare cases also thymidines and guanosines. The longer extensions can be grouped into two categories as observed previously. They are either oligohomopolymeric adenosine stretches (up to 24 As) or short extensions of adenosines and cytosines. However the vast majority of the mRNAs do not have non-encoded nucleotides.

Our analyses also showed that RNA editing in the 5’ and 3’ UTRs is a rare event. We detected such sites only in the 5’ UTR of *ccb3*. Here three C to U conversion were found at positions 241,044 (-286), 241,066 (-308) and 241,074 (-316). We also found a new editing site in *atp1* reading frame at position 302,265, however this site is only edited at a rate of about 35%. In contrary there are some postulated editing sites that have never been seen

edited in our analysis. This includes a cytidines at positions 219,244 in the peudo *sdh4* gene and at position 260,938 in *nad3*.

Taken together this comprehensive investigation of all mitochondrial mRNA ends in *A. thaliana* provides new generals insights about the nature and the generation of these ends. It is thus represents a broad basis for the further detailed studies of individual mRNA ends.

Endonucleolytic generation of both 5' and 3' ends

As mentioned above most or at least a large portion of the 5' termini are derived from posttranscriptional processing. Theoretically both endonucleolytic and exonucleolytic processes could contribute to the formation of these ends, but so far no evidence has been found for the presence of a 5' to 3' exonuclease in plant mitochondria (9). In contrast there is clear experimental evidence for an endonucleolytic generation of 5' mRNA ends. As mentioned in the previous section the major 5' end of the *cox1* mRNA arises from an endonucleolytic cut. The processing is directed by a t-element, which strongly suggests RNase Z to generate this 5' end.

In addition this *cox1* t-element might be cleaved at the 5' end, most likely generated by RNase P. This cleavage site might be indicated by the larger Atcox1-5 extension product, which would be in agreement with the proposed t-element secondary structure (Fig. 3B, indicated by arrow a and Fig. 6C), however, further experiments are required to unambiguously determine this additional cleavage site. In none of our *cox1* CR-RT-PCR or primer extension analyses we detected a 5' end at or close to the position of the primary end identified recently (18) suggesting that the endonucleolytic processing reaction generating the 5' end at -241 proceeds very rapidly.

The other t-elements, whose ends map to 5' or 3' ends of mRNAs, suggest this type of processing not to be restricted to *cox1*. At least the 5' ends of *rps4*, which is located within the ATG, and of *ccb6n1* to be generated by RNase Z, while the major 3' termini of *nad3* and *ccb3*, interestingly both located within in the reading frame, are most likely the results of endonucleolytic cuts performed by RNase P. This has been similarly found during the investigation of Ogura cytoplasmic male sterility (CMS) in Brassica cybrids. Here a t-element was identified downstream of the orf138, which is responsible for the CMS phenotype. Most likely cleavage by RNase P generates the 3' end of orf138 mRNAs. In addition also 3' cleavage of the t-element is observed giving rise to stable t-element RNA (52). For *ccb3* the presence of further downstream located ends indicate that the complete t-elements is removed from a *ccb3* precursor RNA. The same scenario can be expected for *nad6*.

In line with the hypothesis of an important role of RNase Z and RNase P in mRNA processing it can be assumed that even simpler secondary structures such as simple stem loops can be recognized as processing signal by tRNA processing enzymes. It has been shown for the homologous enzymes from other organisms that they indeed cleave at such simple structures such as stem loops (37,38). Three such structures might direct RNase Z to generate two *rpl5* 5' termini (-459 and -406) and *atp6*-2 5’ end (-268), while a simple SL downstream of *nad7* (-375) 5' terminus suggests RNase P to generate this end (Fig 8). Even the 3’ ends found downstream of the *cox2* SL might similarly direct an endonucleolytic cleavage at the 3' terminus of this mRNA (Fig. 4).

Contrary it is also possible that some RNAs form complex secondary structure that cannot be predicted, but which are still substrates for these tRNA processing enzymes. Thus it this reasonable to assume that both enzymes have an additional more general function beyond tRNA processing one of which is the generation of 5’ as well 3’ ends of mRNAs.

Function of secondary 5’ ends

The detection of many secondary 5’ ends raises the question why some mRNAs retain their original primary 5’ ends while most others are processed. A size reduction of the 5’ UTRs would be a simple explanation, but even those mRNAs that have originally relatively short 5’ UTR (for instance *atp9*) are processed, while some of the processed ends (cox3; rpl5) define still larger 5’ UTRs than in some instances where primary ends are present.

Of course the nature of the 5’ UTR could influence translation and it would be possible that some mRNAs are only accessible for ribosomes after 5’ processing introducing another level of expression control. Here two scenarios are possible. Either a potential ribosome entry sequence similar to the Shine-Dalgarno sequence is blocked in the non-processed RNA or a binding of specific translation factor, like those that have been described in yeast, is not possible prior to processing. The first scenario seems to be rather unlikely since no Shine-Dalgarno-like sequence has been identified in plant mitochondria. Although conserved sequence elements have been found in some mRNAs (53) the restriction of this motif to certain mRNAs excludes a general function. Moreover there is so far no experimental evidence for the importance of these conserved sequences. More likely seems the second scenario. mRNA-specific translation factors have indeed been found in mitochondria from *Saccharomyces cereviesiae* (54) and in chloroplasts from *Chlamydomonas reinhardtii* (55) and higher plants (56). In the latter case the CRP1 protein was found to specifically interact with 5’ UTRs of *petA* and psaC mRNAs and this protein activates the translation of these

RNAs. CRP1 belongs to PPR protein family of which many members are also transported to mitochondria. Here they could exhibit similar functions.

Stabilizing elements in plant mitochondrial mRNAs

As mentioned above almost no variability is found for the 3' ends of mitochondrial mRNA in A. thaliana. The only exception from this rule is the *ccb6c* mRNA where two different 3' termini were found located 105 nucleotides apart from each other. Nevertheless in contrast to the 5' ends the 3' termini seem to be determined very clear. Recently two 3' exonucleases have been identified as important trans factors necessary for the generation of both *atp9* and *atp8* mRNAs (31). Presently it is unclear whether additional proteins are required as it has been suggested in the case of chloroplasts (57). Also in this compartment the dual targeted RNaseII-like protein and the organelle-specific PNPase are present and previous reports suggested the participation of other proteins in 3' end maturation. In addition stem-loop structures have been found to be the major *cis* elements required for 3' processingin chloroplasts. In contrast to chloroplast the nature of such *cis* elements in mitochondria is rather unclear. While in some instances a stem-loop structure can be found at the 3' ends (27,58) in the majority of these cases no such secondary element can be found. Since also no conserved primary structure element can be found it seems likely that indeed secondary structures are required, but that these structure might be much more complex than easily detectable stem-loops. But also individual trans factors each binding to a distinct primary structure element can presently not be excluded.

However it is also possible that the 3' ends are generated by other mechanism. For instance some could directly be created by transcription termination and other could be the results of endonucleolytic cleavage. This is definitely the case for the two *atp6* mRNAs whose 3' ends are created by the cleavage of RNase P at the 5' end of $tRNA^{Ser}$ (UGA). But also the 3' ends of the truncated *nad6* and *ccb3* are produced by the same mechanism. Interestingly all these 3' ends are the direct result of the endonucleolytic cleavage reaction and no further exonucleolytic trimming is required. But as in all other cases no obvious secondary structure is evident that would prevent an exonucleolytic attack of these mRNAs. Thus the requirements for both the determination of most 3' ends as well as for the stabilization of the RNAs remain unknown.

ACKNOWLEDGEMENTS

This work was supported by the DFG (Bi 590/6-1 and 6-2) and by a fellowship of the Studienstiftung des deutschen Volkes to J.F.. We are very grateful to C. Guha for technical assistance and thank all co-workers and students who directly or indirectly contributed to this project.

REFERENCES

1. Clifton, S.W., Minx, P., Fauron, C.M., Gibson, M., Allen, J.O., Sun, H., Thompson, M., Barbazuk, W.B., Kanuganti, S., Tayloe, C. et al. (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol., 136, 3486-3503.
2. Handa, H. (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. Nucleic Acids Res., 31, 5907-5916.
3. Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A. and Mikami, T. (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res., 28, 2571-2576.
4. Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A. and Kadowaki, K. (2002) The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol. Genet. Genomics, 268, 434-445.
5. Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N. et al. (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res., 33, 6235-6250.
6. Satoh, M., Kubo, T., Nishizawa, S., Estiati, A., Itchoda, N. and Mikami, T. (2004) The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugar beet share the same complement of genes of known function but differ in the content of expressed ORFs. Mol. Genet. Genomics, 272, 247-256.
7. Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A. and Sugiura, M. (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol. Genet. Genomics, 272, 603-615.
8. Unseld, M., Marienfeld, J.R., Brandt, P. and Brennicke, A. (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat. Genet., 15, 57-61.
9. Gagliardi, D. and Binder, S. (2007) Expression of the plant mitochondrial genome.In Logan, D. (ed.), Plant mitochondria. Blackwell Publishing, Ames, IA, USA, pp. in press.

10. Hanic-Joyce, P.J. and Gray, M.W. (1991) Accurate transcription of a plant mitochondrial gene in vitro. Mol. Cell. Biol., 11, 2035-2039.
11. Rapp, W.D., Lupold, D.S., Mack, S. and Stern, D.B. (1993) Architecture of the maize mitochondrial atp1 promoter as determined by linker-scanning and point mutagenesis. Mol. Cell. Biol., 13, 7232-7238.
12. Rapp, W.D. and Stern, D.B. (1992) A conserved 11 nucleotide sequence contains an essential promoter element of the maize mitochondrial atp1 gene. EMBO J., 11, 1065-1073.
13. Dombrowski, S., Hoffmann, M., Guha, C. and Binder, S. (1999) Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria. J. Biol. Chem., 274, 10094-10099.
14. Binder, S., Hatzack, F. and Brennicke, A. (1995) A novel pea mitochondrial in vitro transcription system recognizes homologous and heterologous mRNA and tRNA promoters. J. Biol. Chem., 270, 22182-22189.
15. Hoffmann, M. and Binder, S. (2002) Functional importance of nucleotide identities within the pea atp9 mitochondrial promoter sequence. J. Mol. Biol., 320, 943-950.
16. Dombrowski, S., Hoffmann, M., Kuhn, J., Brennicke, A. and Binder, S. (1998) On mitochondrial promoters in Arabidopsis thaliana and other flowering plants.In Møller, I. M., Gardeström, P., Glimelius, K. and Glaser, E. (eds.), Plant Mitochondria: From Gene to Function. Backhuys Publications, Leiden Netherlands, pp. 165-170.
17. Giese, A., Thalheim, C., Brennicke, A. and Binder, S. (1996) Correlation of nonanucleotide motifs with transcript initiation of 18S rRNA genes in mitochondria of pea, potato and Arabidopsis. Mol. Gen. Genet., 252, 429-436.
18. Kühn, K., Weihe, A. and Borner, T. (2005) Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. Nucleic Acids Res., 33, 337-346.
19. Shadel, G.S. (2004) Coupling the mitochondrial transcription machinery to human disease. Trends Genet., 20, 513-519.
20. Hedtke, B., Borner, T. and Weihe, A. (1997) Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. Science, 277, 809-811.
21. Chang, C.C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S. and Stern, D.B. (1999) Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. Plant Cell, 11, 911-926.
22. Weihe, A. (2004) The transcription of plant organelle genomes.In Daniell, H. and Chase, C. (eds.), Molecular Biology and Biotechnolgy of Plant Organelles. Springer, Dordrecht, Netherlands, pp. 213-237.
23. Mulligan, R.M., Leon, P. and Walbot, V. (1991) Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression. Mol. Cell. Biol., 11, 533-543.
24. Finnegan, P.M. and Brown, G.G. (1990) Transcriptional and Post-Transcriptional Regulation of RNA Levels in Maize Mitochondria. Plant Cell, 2, 71-83.
25. Leino, M., Landgren, M. and Glimelius, K. (2005) Alloplasmic effects on mitochondrial transcriptional activity and RNA turnover result in accumulated transcripts of Arabidopsis orfs in cytoplasmic male-sterile Brassica napus. Plant J., 42, 469-480.
26. Giegé, P., Hoffmann, M., Binder, S. and Brennicke, A. (2000) RNA degradation buffers asymmetries of transcription in Arabidopsis mitochondria. EMBO Rep., 1, 164-170.
27. Dombrowski, S., Brennicke, A. and Binder, S. (1997) 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. EMBO J., 16, 5069-5076.

28. Kuhn, J., Tengler, U. and Binder, S. (2001) Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. Mol. Cell. Biol., 21, 731-742.
29. Kaleikau, E.K., Andre, C.P. and Walbot, V. (1992) Structure and expression of the rice mitochondrial apocytochrome b gene (cob-1) and pseudogene (cob-2). Curr. Genet., 22, 463-470.
30. Saalaoui, E., Litvak, S. and Araya, A. (1990) The apocytochrome from alloplasmic line wheat (T. aestivum, cytoplasm-T. timopheevi) exists in two differently expressed forms. Plant Sci., 66, 237-246.
31. Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M. and Gagliardi, D. (2004) Two exoribonucleases act sequentially to process mature 3'-ends of atp9 mRNAs in Arabidopsis mitochondria. J. Biol. Chem., 279, 25440-25446.
32. Gagliardi, D. and Leaver, C.J. (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. EMBO J., 18, 3757-3766.
33. Lupold, D.S., Caoile, A.G. and Stern, D.B. (1999) Polyadenylation occurs at multiple sites in maize mitochondrial cox2 mRNA and is independent of editing status. Plant Cell, 11, 1565-1578.
34. Kunzmann, A., Brennicke, A. and Marchfelder, A. (1998) 5' end maturation and RNA editing have to precede tRNA 3' processing in plant mitochondria. Proc. Natl. Acad. Sci. U S A, 95, 108-113.
35. Hanic-Joyce, P.J. and Gray, M.W. (1990) Processing of transfer RNA precursors in a wheat mitochondrial extract. J. Biol. Chem., 265, 13782-13791.
36. Marchfelder, A. (1995) Plant mitochondrial RNase P. Mol. Biol. Rep., 22, 151-156.
37. Vogel, A., Schilling, O., Spath, B. and Marchfelder, A. (2005) The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. Biol. Chem., 386, 1253-1264.
38. Frank, D.N. and Pace, N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. Annu. Rev. Biochem., 67, 153-180.
39. Hanic-Joyce, P.J., Spencer, D.F. and Gray, M.W. (1990) In vitro processing of transcripts containing novel tRNA-like sequences ('t-elements') encoded by wheat mitochondrial DNA. Plant Mol. Biol., 15, 551-559.
40. 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.
41. Klein, M., Binder, S. and Brennicke, A. (1998) Purification of mitochondria from Arabidopsis. Methods Mol. Biol., 82, 49-53.
42. 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.
43. Sambrook, J. and Russel, D.W. (2001) Molecular Cloning: A Laboratory Manual. Ed 3 ed. Cold Spring Habror Laboratory Press, Cold Spring Harbor, NY.
44. 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.
45. Raczynska, K.D., Le Ret, M., Rurek, M., Bonnard, G., Augustyniak, H. and Gualberto, J.M. (2006) Plant mitochondrial genes can be expressed from mRNAs lacking stop codons. FEBS Let., 580, 5641-5646.
46. Giege, P., Knoop, V. and Brennicke, A. (1998) Complex II subunit 4 (sdh4) homologous sequences in plant mitochondrial genomes. Curr. Genet., 34, 313-317.

47. Marker, C., Zemann, A., Terhorst, T., Kiefmann, M., Kastenmayer, J.P., Green, P., Bachellerie, J.P., Brosius, J. and Huttenhofer, A. (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana. Curr. Biol., 12, 2002-2013.
48. Binder, S. and Brennicke, A. (1993) Transcription initiation sites in mitochondria of Oenothera berteriana. J. Biol. Chem., 268, 7849-7855.
49. Brown, G.G., Auchincloss, A.H., Covello, P.S., Gray, M.W., Menassa, R. and Singh, M. (1991) Characterization of transcription initiation sites on the soybean mitochondrial genome allows identification of a transcription-associated sequence motif. Mol. Gen. Genet., 228, 345-355.
50. England, T.E., Gumport, R.I. and Uhlenbeck, O.C. (1977) Dinucleoside pyrophosphate are substrates for T4-induced RNA ligase. Proc. Natl. Acad. Sci. U S A, 74, 4839-4842.
51. Sugino, A., Snoper, T.J. and Cozzarelli, N.R. (1977) Bacteriophage T4 RNA ligase. Reaction intermediates and interaction of substrates. J. Biol. Chem., 252, 1732-1738.
52. Bellaoui, M., Pelletier, G. and Budar, F. (1997) The steady-state level of mRNA from the Ogura cytoplasmic male sterility locus in Brassica cybrids is determined post-transcriptionally by its 3' region. EMBO J., 16, 5057-5068.
53. Pring, D.R., Mullen, J.A. and Kempken, F. (1992) Conserved sequence blocks 5' to start codons of plant mitochondrial genes. Plant Mol. Biol., 19, 313-317.
54. Naithani, S., Saracco, S.A., Butler, C.A. and Fox, T.D. (2003) Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of Saccharomyces cerevisiae. Mol. Biol. Cell, 14, 324-333.
55. Auchincloss, A.H., Zerges, W., Perron, K., Girard-Bascou, J. and Rochaix, J.D. (2002) Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in Chlamydomonas reinhardtii. J. Cell Biol., 157, 953-962.
56. Schmitz-Linneweber, C., Williams-Carrier, R. and Barkan, A. (2005) RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. Plant Cell, 17, 2791-2804.
57. Marchfelder, A. and Binder, S. (2004) Plastid and plant mitochondrial RNA processing and RNA stability.In Daniell, H. and Chase, C. (eds.), Molecular Biology and Biotechnolgy of Plant Organelles. Springer, Dordrecht, Netherlands, pp. 261-294.
58. Morikami, A. and Nakamura, K. (1993) Transcript map of oppositely oriented pea mitochondrial genes encoding the alpha-subunit and the subunit 9 of F1F0-ATPase complex. Biosci. Biotechnol. Biochem., 57, 1530-1535.

TABLES

Table 1. 3' mRNA termini in *Arabidopsis* mitochondria

Major 3' ends

Gene	Position relative to the stop codon* (position in nc_001284)	Flanking sequences (5' to 3'), the most prominent end is underlined	Remarks
atp6-1	+45 (112.952)	TTATAATTGAACAAAAGCGA**G**GGATGGATGTCTGAGCGGTT	The G corresponds to the position one nucleotide upstream of $tRNA^{Ser}$ 5' end. 3' mRNA end generated by RNaseZ
atp6-2	+45 (296.775)	TTATAATTGAACAAAAGCGA**G**GGATGGATGTCTGAGCGGTT	The G corresponds to the position one nucleotide upstream of $tRNA^{Ser}$ 5' end. 3' mRNA end generated by RNaseZ
ccb3	-46 (240.033)	ATACCAAAACCTAGTTCAC**T**CGCTCTCTTTGCATCCATGGC	Major 3' end within the reading frame (45), most likely generated by RNaseP
nad6	-17 (76.658)	AATGAGGAGGACGACTGACC**C**ACTCACGATCTACTAAAGGG	Major 3' end within the reading frame (45), most likely generated by RNaseP
nad4L-orf25	+72 (188.012)	GGATCCAATACCAAGACTAC**T**TCTTTCTCAGGAAGTGCAGC	3' end corresponds to the 3' terminus of Ath-290 (47)
atp8	+121 (130.506)	CTTTCTTTTCAGGCTTGACT**C**ATTCGCTAGCTTCTCCTCCC	3' end one nucleotide downstream of Ath-377 (47)
nad1	+88 (143.131)	TTAATCAAGCAAGTTGGGGA**A**CAAAATCTTCCTTGTTAGTT	Identical with the 3' terminus of the *atp9* mRNA
atp9	+180 (279.332)	TAATCAAGCAAGTTGGGGAA**C**AAAATCTTCCTTGTTAGTTA	Identical with previously identified 3' terminus (31)
rpl2-orfX	-10 (158.342)	CCTAGAACCTGGCAAAGTAA**C**TATCAATGAATTCCAATAAA	
nad3-rps12	+15 (260.209)	ATATGAATGGAAGATGCCTC**T**GGAACTAGTATCGGTCAGTC	
nad4	+30 (169.704)	GCAAAGAAAAGAAAAACGGG**T**CAACATCTTAATGTGTATTT	
nad2	+35 (327.855)	AAGGTTCAAAGAAAGGGTAG**G**CCGTCGGTTTGAATCCAATA	
cox1	+47 (349.783)	TAAGAACCTAACAGAACTTT**T**AGAAATTTATGCATGGTTAT	
nad7	+52 (138.205)	TTATTGCGAGCCCAGAAGTC**T**CTCTTTTTTTCGGCCTTCAG	
nad9	+55 (23.608)	CAAATCAATAGGAAATGCTA**T**AGGAAATGCTATTTGCTTTG	
rpl5-cob	+58 (61.474)	CACCAAGAATTGACAAGCAG**A**TAAGTTTTCTAGTTTGCTAT	
rps3-rpl16	+67 (25.009)	AAGTAAGGACCAAGGATAAT**C**TTTTCGGAAAGGAGGAGTAG	
ccb2	+81 (31.164)	ATGTAAGCCATGTATCTAGG**A**GGAATTAGAAAGAAGGGCTT	
nad5	+86 (20.485)	TCTTTCTCTTCTCTCTTTTT**T**TAGTTTAGTGACAGCTCATC	
ccb6c	+115 (51.178)	AATAGAAATCGTATTCGTGA**A**TAAATCCCCTTTGTTTGAAT	
atp1	+160 (302.006)	TGGCATGCTCCCCAGGGCCC**T**CTTTTTTTTAGATAGGTAGG	
ccb6n2	+170 (256.695)	TTCGGAAGATGGTCAAGGTA**C**GAAGTGACAAGCCACTGCGC	
cox3	+314 (219.391)	TTGAAGAACCTAATGGATCG**A**ACTCATCCTACGAAAGATAT	
cox2	+316 (40.186)	CGGCCTACTTATAGCCTTTT**C**CCCTTTATTCGGGGACTTCA	
rps7	+351 (315.424)	CACTTCAAAGGGAGGGAAGG**C**TAGGAATCTCGTTTTTGCTC	
ccb6n1	+468 (231.426)	GGTCCAGTTCTTTGCTAGTA**A**GTGAGGAAAATATATCACAC	
rps4	+498 (81.530)	CAAAAGGACCAAGGACGATC**T**TTTCGGAAAGGAGGAGTAAG	
matR		No defined 3' end detectable	

Minor 3' ends

Gene	Position relative to the stop codon* (position in nc_001284)	Flanking sequences (5' to 3'), the most prominent end is underlined	Remarks
ccb6c	+10 (51.283)	TAATTGCATAAAAGGAGTCA**A**AGTAGTGGCTGCGGCGCGTC	
ccb3	+33 (239.955)/+37 (239.951)	TAGGTTCGATTCCTGCTGGA**T**GCA**C**TTGGAACGTTCAGTTCAATC	t-element, ends generated by RNaseZ

*the first nucleotides downstream of the stop codon corresponds to position +1

Table 2. 5' ends of mitochondrial mRNAs from *A. thaliana*

Major ends

Gene	Position relative to the start codon** (position in nc_001284)	Flanking sequences (5' to 3'), the most prominent end(s) is (are) underlined	Remarks
Real and potential primary ends			
atp8	-157 (129.752)	AAACTATCAATCT *CATAAGAGAAGAAATCTCTATGCCCCCT*	Primary end, CNM1, also detecetd in (18)
atp8	-228/4 (129.681/5)	TTCCATACCATAA *CATATATAGAATCGATTTTCTTTTCTGA*	Primary end , CNM2, also detecetd in (18)
atp6-1	-200 (111.550)	GCAGCCAATAATA *CGTATATAAAGAAGAGGACTGCTTACGGG*	Primary end , CNM2, also detecetd in (18)
ccb2	-140 (30.323)	AGTCTTTTAGTAAG *CGTATATAAAGCCAGCAGTTTAGTGTATA*	Primary end, CNM2
ccb3	-484/-482 (241.242/241.240)	TATCGATCGG *CGTAAGAAACGATTCTACTATCTTTATTCCC*	Putative primary end, CNM1
cox2	-151 (42.779)	AACTCAAAATAT *TGTAAGAAGAAGAAGAATCTTACGCCCCA*	Putative primary end, CNM1
cox2	-205 (42.833)	GTTGGTTT *CGTATATAAGAAGACTGCTGCTTTTAGGAGAGT*	*, Putative primary end , CNM2, also detecetd in (18)
5' mRNA ends generated by RNaseZ cleavage at the 3' ends of upstream located tRNAs			
ccb6c	-124 (53.735)	TTCAAGTCCCTCCTTCCGCTCCTGGTGTTCGAACTAGTCAT	The 5' terminal cytidine overlaps with the 3' terminal nucleotide of $tRNA^{Gly}$
rps3-rpl16	-163 (28.896)	TCAAGTCCTGCTATACCCAAACCTACCTTACCACACTATAT	The adenosine is nucleotides downstream of the $tRNA^{Lys}$ 3' terminus
5' mRNA ends most likely generated by RNaseZ cleavage at the 3' ends of upstream located t-elements			
rps4	+2 (83.115)	TTCGAGTCCCAGGGACGCAATGTGGCTGCTTAAAAAACTGA	t-element, RNase Z
ccb6n1	-66 (233.108)	TCGAATCCTGCCACCTTTCTTGTGGATCATCCTGTGGTTAC	t-element, RNaseZ
cox1	-241 (351.654)	TTCGAATCCAATGGTAGGCCTGGGCTCTCCTACGCAGGAGA	t-element, RNaseZ
5' ends directly adjacent to stem-loop structures			
rpl5-cob	-459 (57.315)	CTAACTAAATAGTTAGTTAAGAAGACGCGTGCTTTGTTTGG	SL, RNaseZ
rpl5-cob	-406 (57.368)	ATTTCACTATAAGATAAGATAAGAGTGAAGTGCTTTGTTTG	SL, RNaseZ
atp6-2	-268 (298.137)	GTTTTTGTTGGTCAACAACCAACCACAACTTTTTATAGTTG	SL, RNaseZ
nad7	-375 (131.696)	TAAAGCCCTCTCTT TATATAAAGAGGGCTTTGCGCCTTCG	SL, RNaseP
Miscellaneous transcript ends			
atp9	-84 (278.811)	CCGTGGGTATAAGATCGCAAAGAATGCATTCCAAGTGAGAT	Identical with the 5' terminus of the *nad6* mRNA
nad6	-179 (77.438)	CGTGGGTACAAGATCGAAAAGAATGCATTAAATGGATGCAT	Identical with the 5' terminus of the *atp9* mRNA
nad5	-21 (143.019)	TGTCTTGATTTTAGGTTTTTGATCACACTCGAAATTAAATT	
nad5	-72 (143.070)	GTTCTCCGGACGTCGAGAAATCTATCACTCAATCACTGGCC	*
nad2	-122 (81.419)	TTGAGCCGAATCACTATCATTATATTATATATTCATTGTTG	
ccb6n1	-149 (233.191)	AGCGCATCTGTTTCGGGTACAGAGGCGACGAGGGCGAAGCA	*
rpl2-orfX	-198 (154.546)	AAAATAATATAAATAATAATATATAGAATATAGAAAGAATT	actually -194 (four bp deletion in the 5' UTR of this gene was found in ecotype Col mtDNA)
nad9	-202 (24.437)	CCACGAATTTGATCATTCGCTTTCAATCAAATTAGATTGTC	
nad4	-228/6 (161.465/7)	TCGGATAAAGAATGCAAAAATAGAAAGAATATAAATAAAGG	
nad3-rps12	-229 (261.235)	GAGGACCACTAGTGGGGAAGAGTCCTTGGAGCGTCTTTATG	
ccb2	-347/6 (30.116/7)	GGCAAGGCTATAAAGTAAGTAAAAGTTGGAAGCCTAAGCCAA	
cox3	-378 (217.902)	GAGTGGAGGGGAAGGTTTAATACGGAATTAGCAGGGATAGA	
rps7	-448 (314.179)	ACTCATATATATATATATATGATGACTTATATCTAAATGGA	
nad1	-645 (319.035)	GAACCCTGGCCCATTACCCGAAATCTCGGGTTGGAGTCTTC	*
atp1	-124 (303.813)	AAGTTAAGATAGATGTAACAATTTGAACCAAAGGCCTTTTT	
atp1	-267 (303.956)	GCTCAACTCGACATATGGAGGCAGGGGCTTGATCAGGAATT	*
atp1	-361 (304.050)	CCGTGGGCGGGAACTAGCCGCCCGCATCGATTGGGAGGTGC	
matR		No clear 5' end detectable	

Table 2. 5' ends of mitochondrial mRNAs from *A. thaliana* (continued)

ccb6n2	-57 (257.533), -35 (257.511)	TTCGGCGAATCATGTGGGGCTCAGAAGAGTGCTCCCTACGACGGTAAATCGGGGCAAAAGA
nad4L-orf25	-318 (189.549) / -294 (189.525)	GGGGGGAACGGGGGAAATCACTAGATCCCTCATGGGATCAAAAACAACCTATCACCGAAGTCCCA

Minor ends

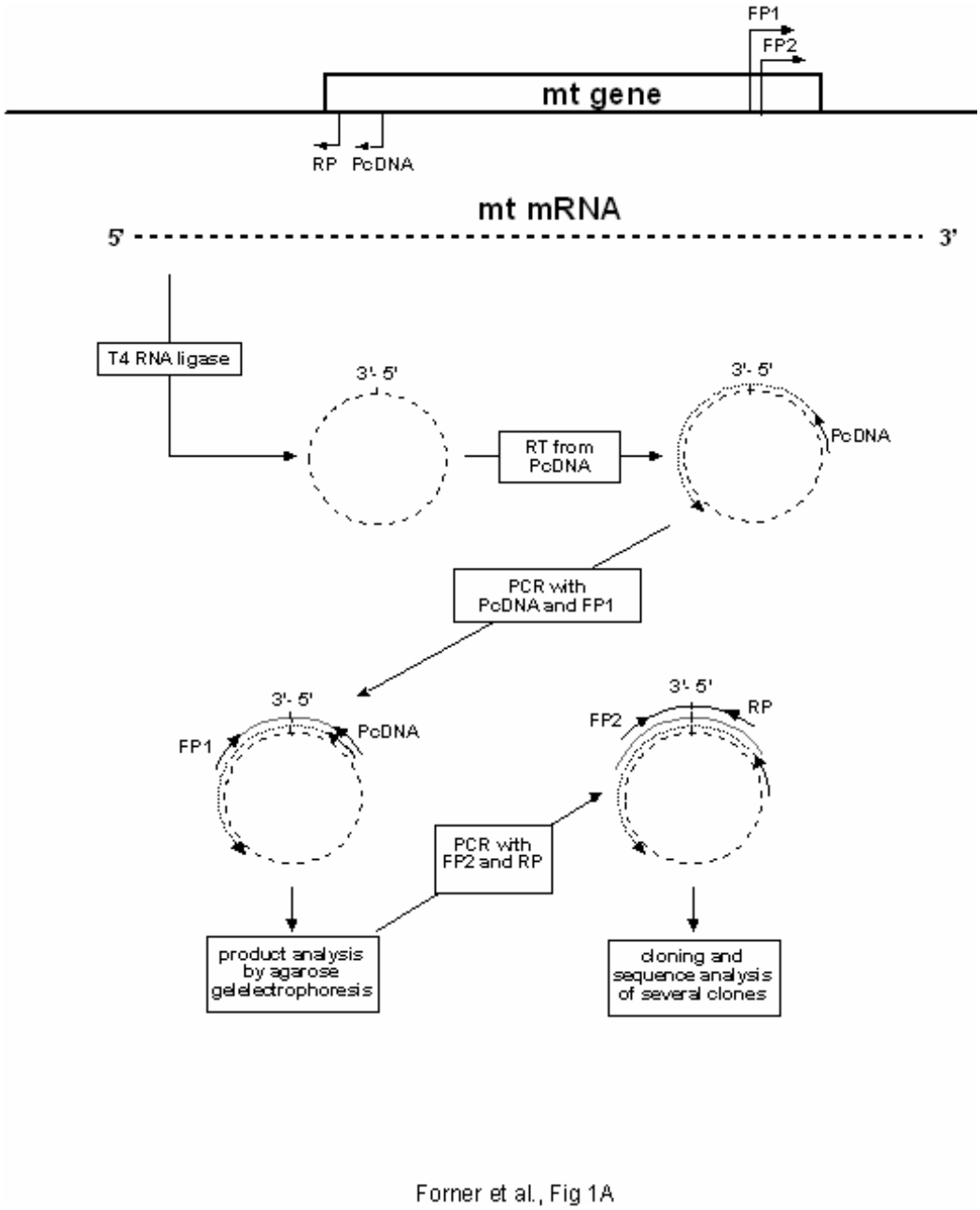
Gene	Position relative to the start codon** (position in nc_001284)	Flanking sequences (5' to 3'), the most prominent end(s) is (are) underlined	Remarks
atp1	-1.898 (305.587)	AAACTATCAATTT*CATAAGA*GAAGAAAGAAAGTAGCCTAGT	Primary end, CNM1
atp9	-239 (278.656)	AAACTATCAATTT*CATAAGA*GAAGACGAAGACGGATCAAAT	Primary end, CNM1
atp6-2	-63 (297.332)/-44 (297.913)	TATAAGTAGACAC*CATAATT*AACCACCCCTTCTAGTGTTCGGGTACAGTAGCTCTCGCAG	
rpl2-orfX	-119 (154.625)	CTACTAAAAGTAAAGTTGAGCACCCGGACCAGACAGAGGTGGCCG	
nad7	-68 (132.003)	GTCGTCGAAGGAGATGCATTTCTGGTACAAGTGGTATTGGA	*
nad7	-109 (131.962)	CAAGCAAGCGTAATAAGGGGAGGGGGACTAGGGGTGGAAGG	*
nad1	-149 (318.539)	CTTTATTTCGGGAGTAAAGTGATAGGGGGTTCGTTCCATTG	*
nad1	-355 (318.745)	TCAAGAGGCTTTTTTCGACTTCTTGAACGAGAGGCGGGACG	*

Table 3. 3' ends of *cox1* 5' leader RNA molecules

clone	3' end position in the mitochondrial genome	3' end position in respect to the ATG (+1)	Non-encoded nucleotides
27	351,656/351,655*	-243/-242*	
6	351,655	-242	
8	351,655	-242	
11	351,655	-242	
28	351,655	-242	
29	351,655	-242	
9	351,655	-242	A
18	351,655	-242	A
5	351,655	-242	C
7	351,655	-242	C
13	351,655	-242	C
20	351,655	-242	C
1	351,655	-242	CC
2	351,655	-242	CC
14	351,655	-242	CC
16	351,655	-242	CA
17	351,655	-242	CAAAA
22	351,655	-242	CAAAA
19	351,655	-242	CCAAA
10	351,652	-239	

* A single nucleotide at the ligation site could be either assigned to *cox1* or 5S RNA. Thus the *cox1* t-elements ends at position -243 or -242.

FIGURES



Forner et al., Fig 1A

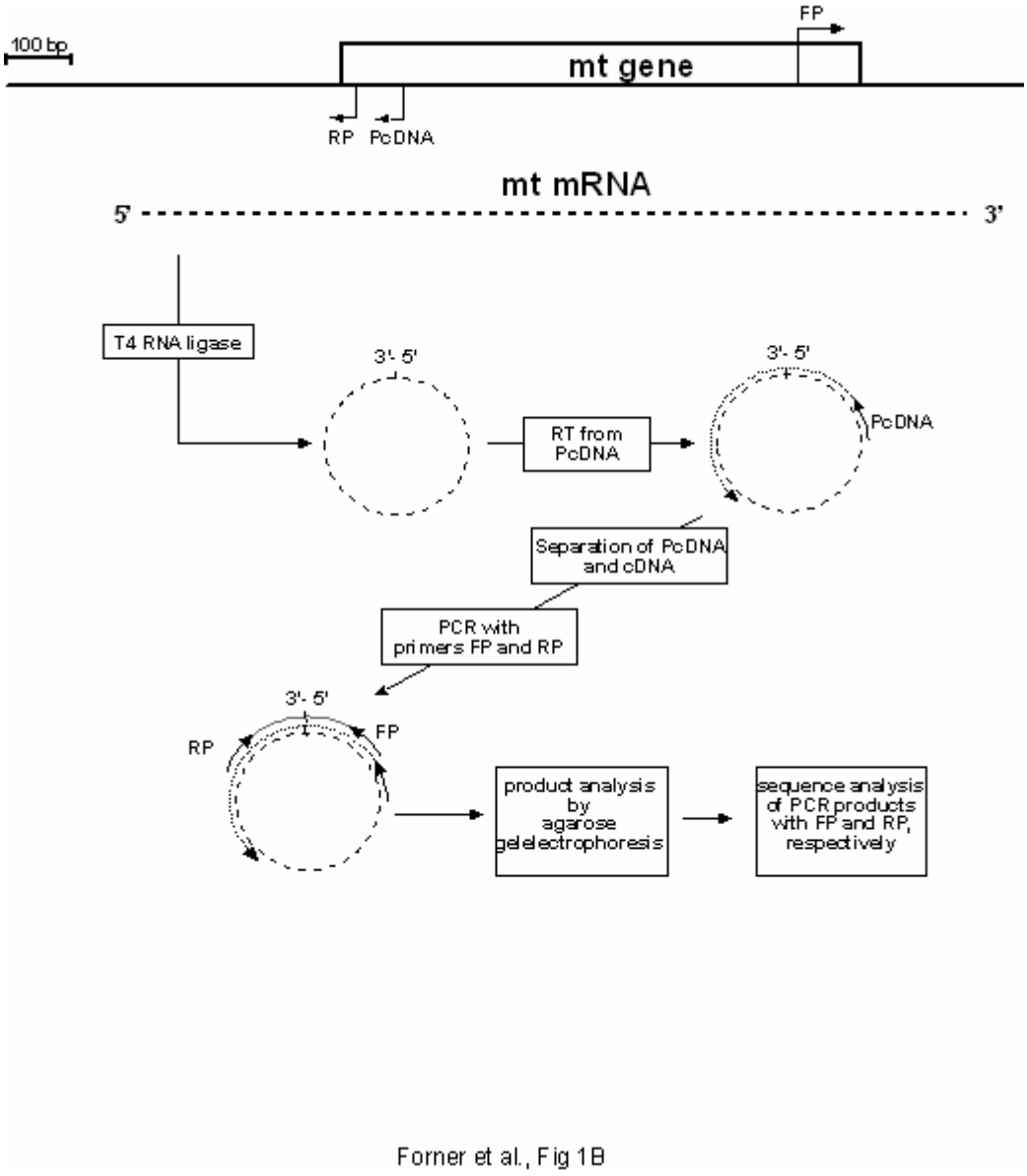
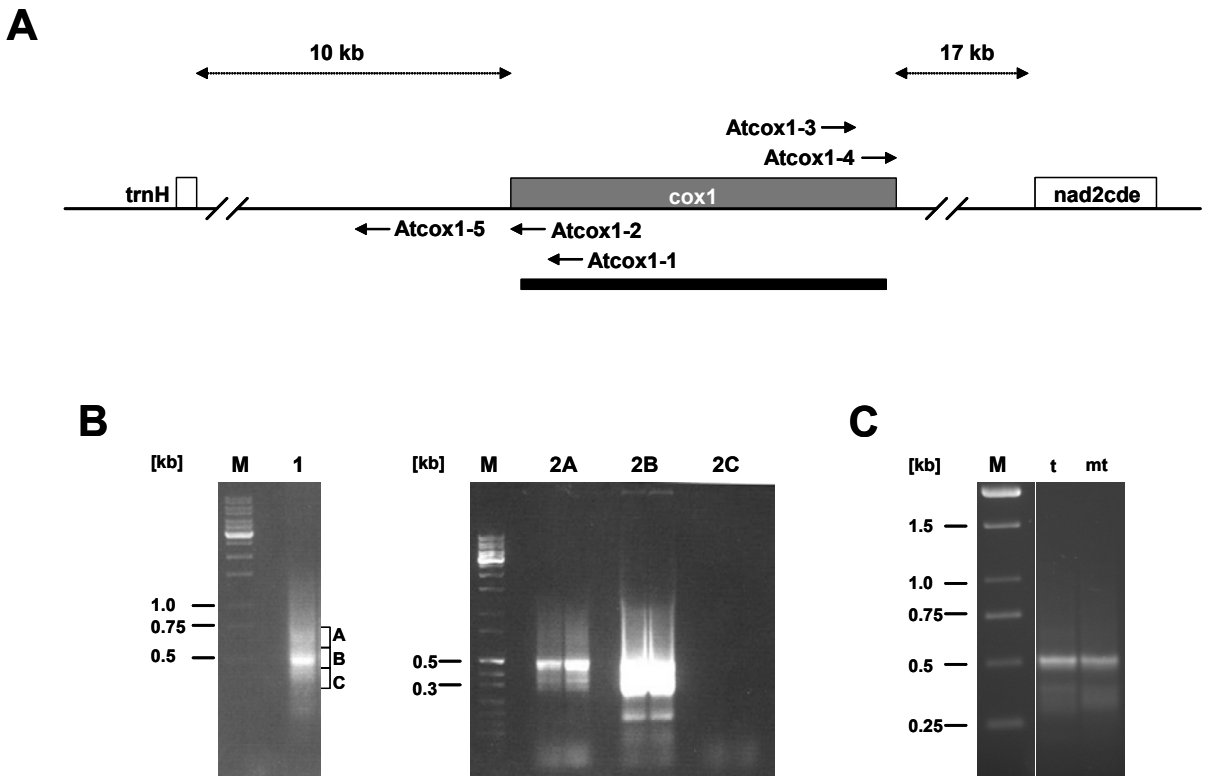


Figure 1. CT-RT-PCR analysis of mitochondrial mRNAs. Two different CR-RT-PCR strategies were followed in the transcript analysis of the 32 protein-coding genes encoded in the mitochondrial genome of A. thaliana. The main difference is that the approach outlined in (**A)** includes two PCRs and cloning with subsequent sequence analysis of individual clones while in the alternative approach **(B)** PCR products obtained by a single PCR step were directly sequenced. Further explanations see in the text.



Forner et al., Fig. 2A-C

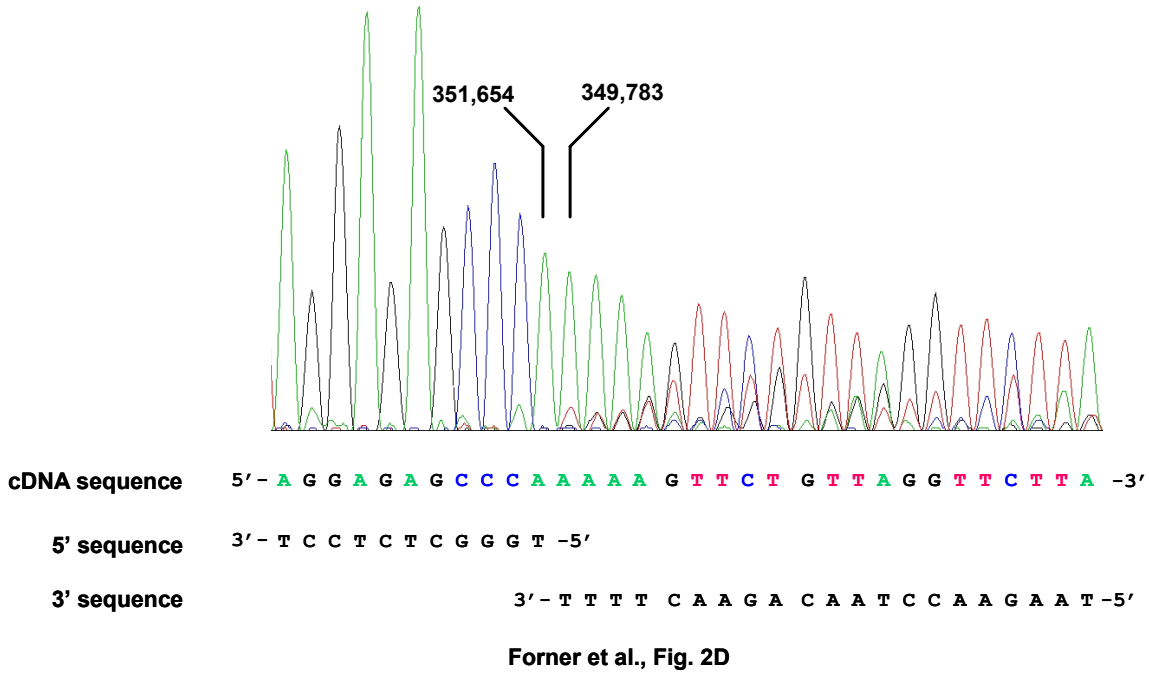


Figure 2. Mapping of *cox1* transcript ends. **A** The *cox1* reading frame (grey box) is flanked by 10 and 17 kb non-coding DNA (indicated by dotted arrows), respectively, from the next genes (white boxes). The location of primers (black arrows) is depicted in the scheme. The probe used in the northern analysis is given as a black bar. **B** *cox1* transcript analysis using the experimental strategy depicted in Fig. 1A. Products obtained in the first PCR performed with primer pair Atcox1-1/Atcox1-3 separated on a 1% (w/v) agarose gel (lane 1). cDNA fragments with sizes of about 650, 500 and 400 bp (corresponding to products A, B and C given at the right margin of lane 1) were used as DNA templates in the second PCR, which was carried out with primer pair Atcox1-2/Atcox1-4. The products were separated in lanes 2A to 2C (designation of the lanes corresponds to the products of the first PCR used as templates). **C** CR-RT-PCR analysis following the single PCR approach outlined in Fig. 1B using oligonucleotide Atcox1-1 for cDNA synthesis and primer pair Atcox1-2/Atcox1-4 for amplification. cDNA fragments were obtained from total RNA (t) and mitochondrial RNA (lane mt) from *A. thaliana* cell suspension culture of ecotype Col. Sizes of DNA marker fragments (lanes M) are given in kb. **D** Sequence chromatogram of a cDNA fragment representing the *cox1* steady state mRNA pool. The corresponding sequences up to the 5' as well as the 3' end of the cDNA are given. Further explanations are given in the text.

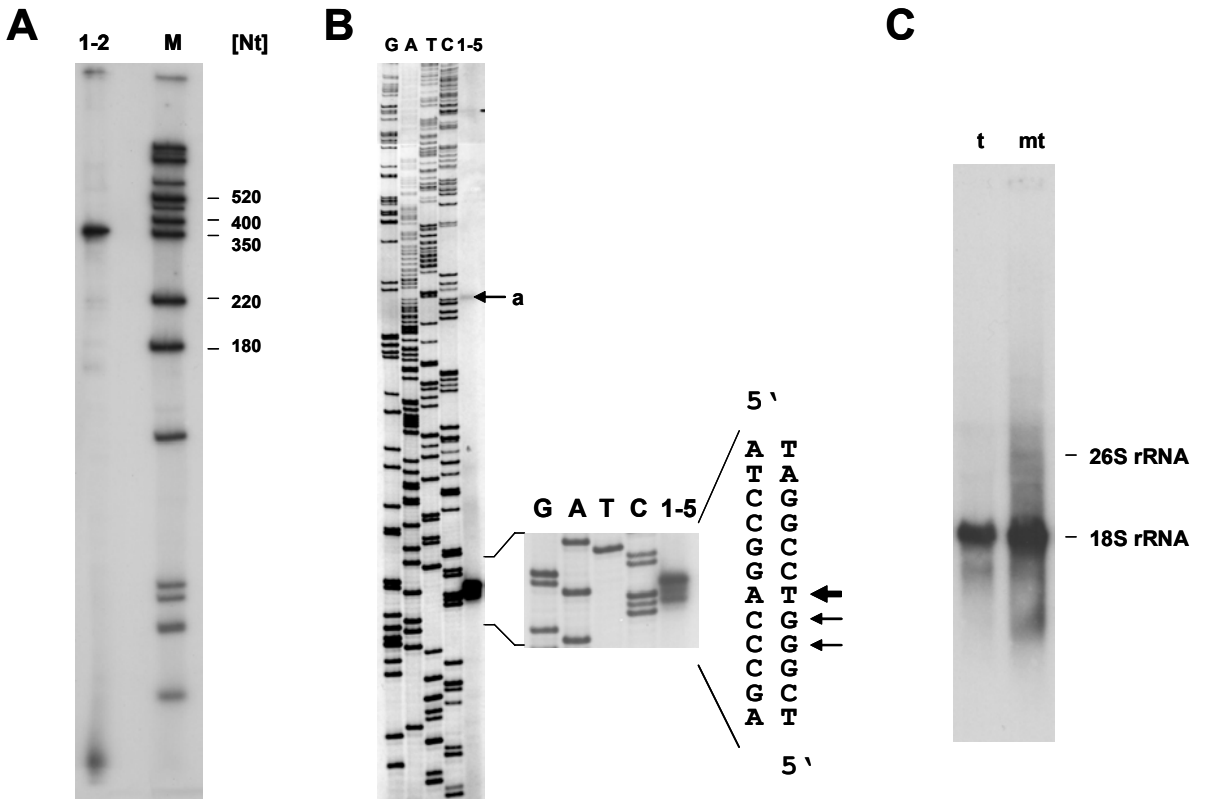


Figure 3. Primer extension and northern blot analysis of *cox1* transcripts. **A** Extension products generated in a reverse transcription reaction performed with oligonucleotide Atcox1-2 (see Fig. 2A) are separated along DNA marker fragments. A single major product (black arrow) detects a 5' end about 240 nucleotides upstream of the ATG, which is consistent with the major 5' terminus mapped in the CR-RT-PCR analysis. **B** The major end(s) is (are) also detected in the extension reaction with oligonucleotide Atcox1-5. The corresponding products (lane 1-5) were separated together with sequencing reaction products (lanes G, A, T and C) obtained with the same primer. These ends (black arrows) scatter over three nucleotides with the most prominent end found at -241 and minor termini at -240 and -239. The sequence is given in both orientations. In addition weaker products are observed, one of which indicates an end at approximately -330 (black arrow a). **C** Northern blot analysis performed with a probe covering almost the complete *cox1* reading frame as indicated in Fig. 2A. A single mRNA of about 1800 nucleotides is detected in total (t) and mitochondrial (mt) RNA preparations from *A. thaliana* cell suspension culture of ecotype Col.

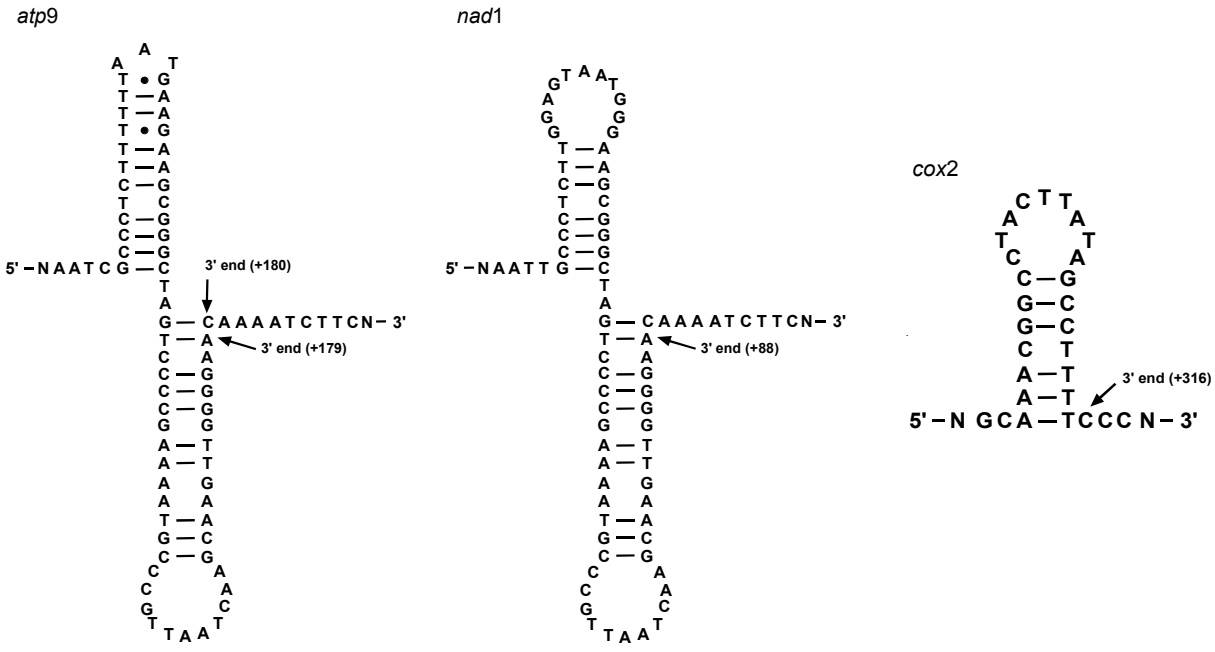


Figure 4. Stem-loop structures found directly upstream of the mapped 3' ends. The termini are indicated by arrows.

26S rRNA: GTACAAGATCGAAAAGAATGCATT

***nad6*: GTACAAGATCGAAAAGAATGCATT**

***atp9*: GTA*t*AAGATCG*c*AAAGAATGCATT**

Figure 5. Sequences found to be conserved at the 5' termini of the 26SrRNA as well as *atp9* and *nad6* mRNAs. The 5' terminal nucleotides are given as underlined letter. Differing nucleotide identities are indicated with lower case letters given in italics.

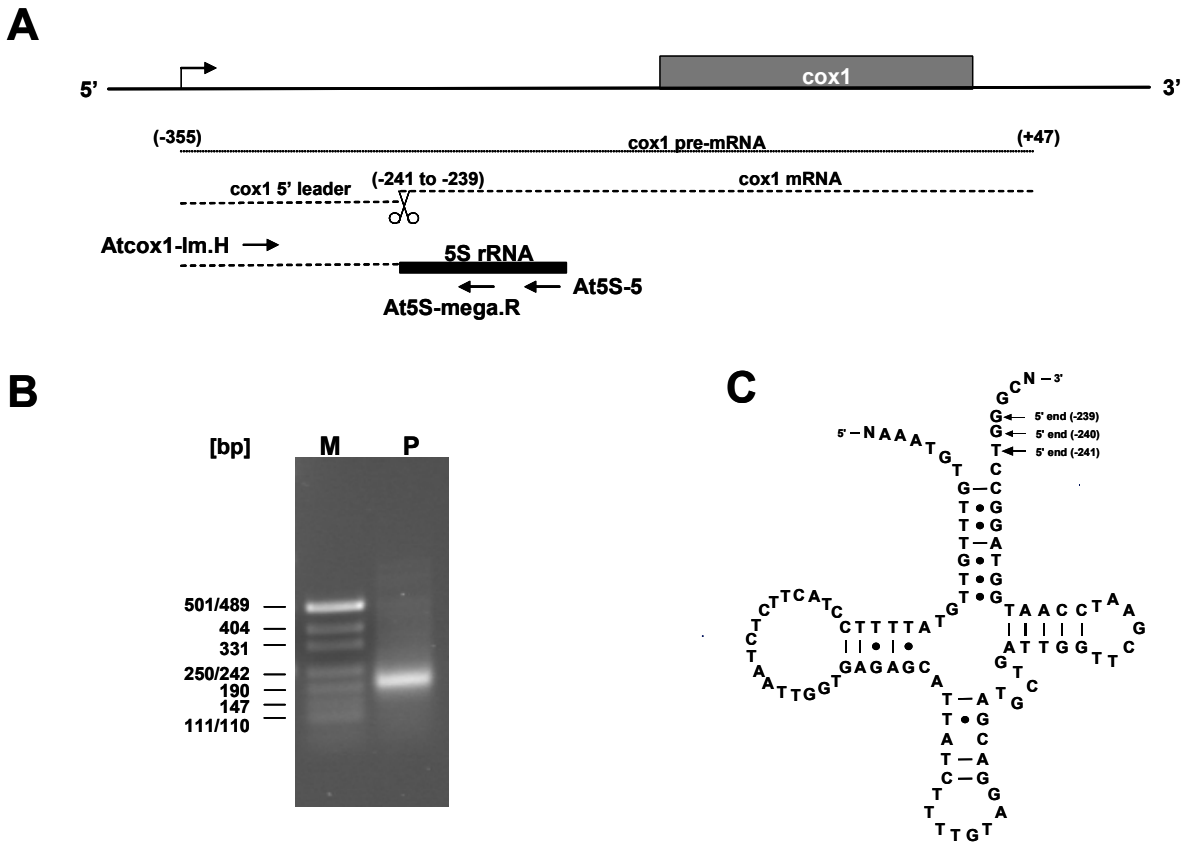


Figure 6. The 5' terminus of the cox1 steady state mRNA is generated by an endonucleolytic cleavage. **A** The *cox1* gene is transcribed from a promoter (bent arrow) located 355 nucleotides upstream of the ATG. The precursor RNA (dotted line) is processed by endonucleolytic cleavage (scissor) into mature cox1 mRNA and 5' leader RNA. In an RNA ligation reaction endogenous mitochondrial 5S rRNA (black bar) is ligated to the *cox1* 5' leader RNA and then used as anchor in the cDNA synthesis initiated at primer At5S-5 and a subsequent PCR with primer At5S-mega.R and Atcox1-lm.H. **B** PCR products obtained in the reaction described above are separated on an agarose gel (lane P). A single 200 bp product is obtained as expected from the size upon the ligation of 5S rRNA to the 3' end of the potential *cox1* 5' leader. **C** The leader sequence upstream of the mature *cox1* 5' end can be folded into a tRNA-like structure forming a perfectly conserved ψ-uridine arm.

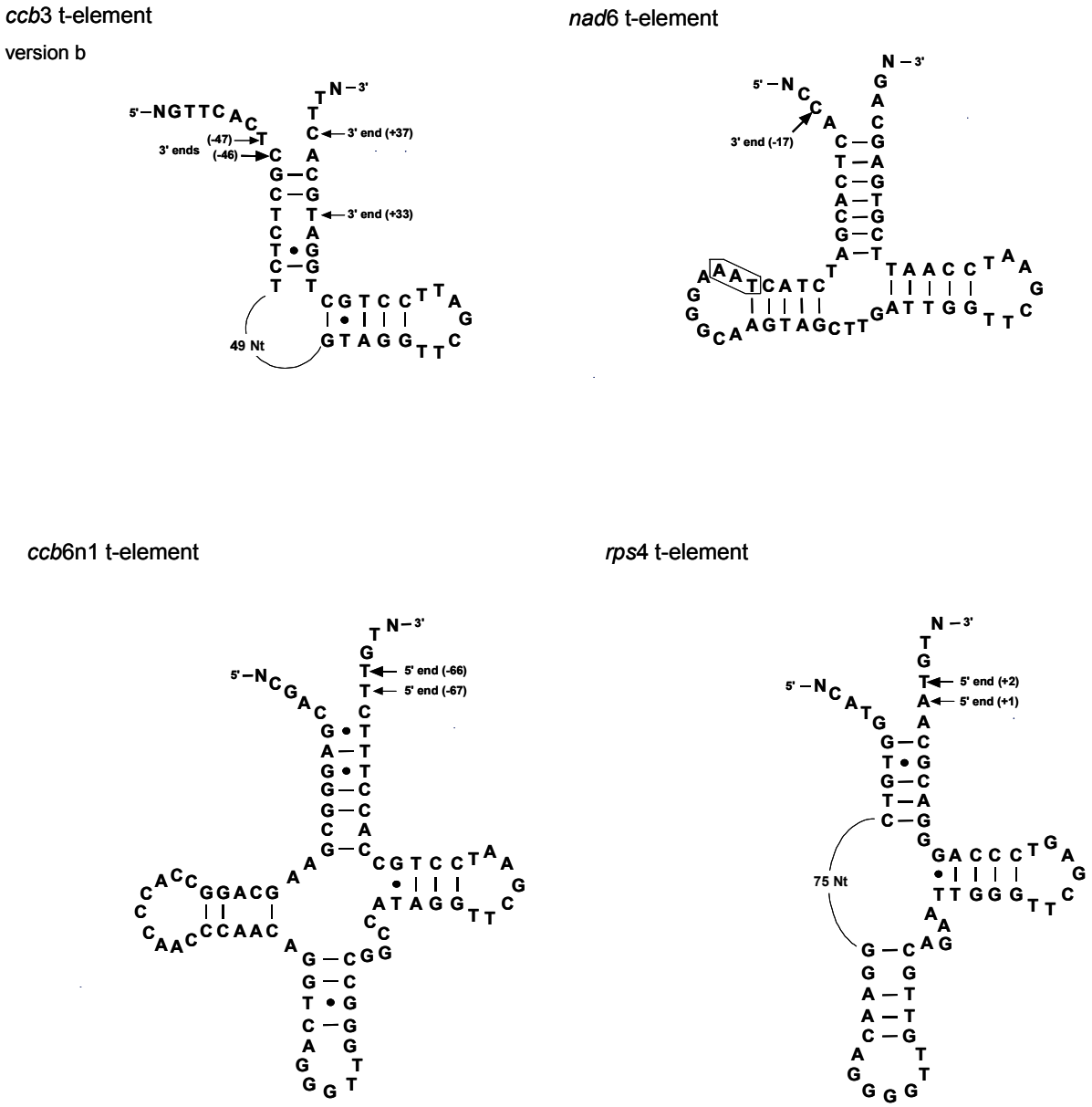


Figure 7. In the *A. thaliana* mitochondrial genome several t-elements can be identified. Either the 5', the 3' end or both ends of these t-elements coincide with 5' or/and 3' termini of mitochondrial mRNAs. This strongly suggests that these t-elements direct cleavage of the respective precursor RNAs by RNase Z and RNase P.

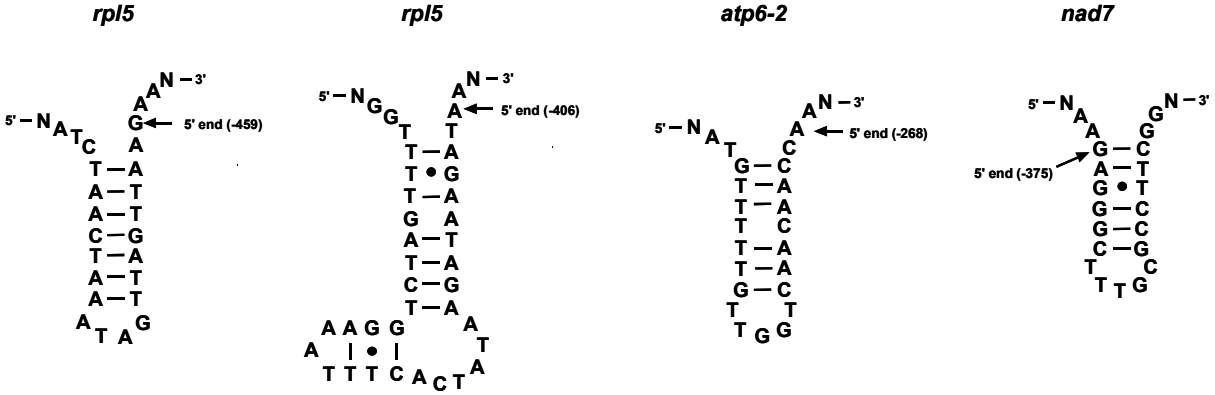


Figure 8. Stem-loop structure found at 5’ ends of mRNAs. Like the t-elements the ends of the stem-loop structures coincide with mRNA termini. These stem-loop structures might thus mimic tRNA acceptor stems.

7.2.2 Manuscript 2: "Distant sequences determine 5' end formation of *cox3* transcripts in *Arabidopsis thaliana* ecotype C24"

Polymorphisms between different ecotypes of *A. thaliana* are a powerful tool to identify the genes responsible for the differing phenotypes. In case of transcript termini in mitochondria, the observed polymorphisms can be either due to differences in the mitochondrial DNA, i.e. *cis* elements, or to proteins encoded in the nucleus and imported posttranslationally into mitochondria, i.e. *trans* factors. To distinguish between these two possibilities, the transcript ends have to be analyzed in reciprocal F_1-hybrids obtained from crosses between ecotypes with differing mRNA termini. As part of the mitochondrial DNA, *cis* elements are exclusively inherited maternally, while nuclear encoded *trans* factors are inherited bi-parentally. Genes for *trans* factors can be identified by map-based cloning while *cis* elements can be recognized by analyzing the mitochondrial DNA of the different ecotypes.

The results of such an investigation for the *cox3* gene and its transcripts in three different ecotypes of *Arabidopsis thaliana* are reported in the following publication.

Nucleic Acids Research, 2005, Vol. 33, No. 15 **4673–4682**
doi:10.1093/nar/gki774

Distant sequences determine 5′ end formation of *cox*3 transcripts in *Arabidopsis thaliana* ecotype C24

Joachim Forner, Bärbel Weber, Caterina Wiethölter, Rhonda C. Meyer[1] and Stefan Binder*

Molekulare Botanik, Universität Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany and [1]MPI für Molekulare Pflanzenphysiologie, 14424 Potsdam, Germany

Received July 1, 2005; Revised and Accepted August 1, 2005

ABSTRACT

The genomic environments and the transcripts of the mitochondrial *cox*3 gene are investigated in three *Arabidopsis thaliana* ecotypes. While the proximate 5′ sequences up to nucleotide position −584, the coding regions and the 3′ flanking regions are identical in Columbia (Col), C24 and Landsberg *erecta* (L*er*), genomic variation is detected in regions further upstream. In the mitochondrial DNA of Col, a 1790 bp fragment flanked by a nonanucleotide direct repeat is present beyond position −584 with respect to the ATG. While in L*er* only part of this insertion is conserved, this sequence is completely absent in C24, except for a single copy of the nonanucleotide direct repeat. Northern hybridization reveals identical major transcripts in the three ecotypes, but identifies an additional abundant 60 nt larger mRNA species in C24. The extremities of the most abundant mRNA species are identical in the three ecotypes. In C24, an extra major 5′ end is abundant. This terminus and the other major 5′ ends are located in identical sequence regions. Inspection of At*cox*3 transcripts in C24/Col hybrids revealed a female inheritance of the mRNA species with the extra 5′ terminus. Thus, a mitochondrially encoded factor determines the generation of an extra 5′ mRNA end.

INTRODUCTION

In seed plants, mitochondrial DNAs encode between 54 and 59 genes (1–4). The realization of this genetic information occurs in several consecutive steps, which are so far only partially understood. Transcription is initiated at various promoter sequences, some of which have been functionally characterized by *in vitro* analyses in different plant species (5,6). In dicots, an 18 bp sequence enclosing the transcription initiation site is necessary and sufficient to drive transcription *in vitro* (6,7). Other promoter motifs, summarized recently in *Arabidopsis thaliana*, diverge partially or completely from the CRTAGAGA nonanucleotide motif, but the functional requirements for these are so far unclear (8). After transcription, primary RNAs undergo a series of processing steps. For instance, many genes are transcribed into polycistronic precursor RNAs that become disassembled into smaller units (9). Apart from intron splicing, 5′ and 3′ ends are post-transcriptionally generated. The 3′ termini are at least in some instances polyadenylated, which appears to accelerate degradation of plant mitochondrial RNA (10–14). Recently, two exoribonucleases have been characterized and found to be important or essential components of the 3′ processing machinery. Both proteins are required for the generation of the correct *atp*9 3′ mRNA terminus, which is located just downstream of an inverted repeat. The knock-down of the AtmtPNPase provokes the accumulation of large 3′ extended *atp*9 transcripts. The knock-out of the AtmtRNaseII, a protein that is also involved in processing of ribosomal RNA in chloroplasts (15), causes a defect in the final 3′ trimming of *atp*9 mRNAs, which are a few nucleotides longer in the mutant plants than their wild-type counterparts (16).

In plant mitochondria, the 5′ ends of tRNAs and rRNAs are also post-transcriptionally generated. An RNaseP-like enzyme has been found to execute 5′ endonucleolytic processing on tRNA precursor molecules, a processing event which can similarly be expected to occur at the 5′ ends of rRNAs (17). Indeed, such an endonucleolytic generation of the 18S rRNA 5′ end has recently indirectly been substantiated by the identification of a 5′ leader sequence as a product of an endonucleolytic cut (18). 5′ processing has also been observed at plant mitochondrial mRNAs, but in contrast to 3′ processing the mode and mechanism of this process are unclear and might be manifold. For instance the sequence surrounding the 5′ end of the 26S RNA in *Oenothera* has been found conserved at a 5′ terminus of an *atp*9 mRNA and has thus been suggested to function as *cis*-element for the recognition and/or processing of the respective precursor molecules (19). But it can also be speculated that certain secondary structures are required for

*To whom correspondence should be addressed. Tel: +49 (0) 731 502 2625; Fax: +49 (0) 731 502 2626; Email: stefan.binder@biologie.uni-ulm.de

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

processing although no supporting experimental evidence has been found so far. The function of 5′ processing is also unclear. In analogy to such processing events in chloroplasts of *Chlamydomonas reinhardtii*, it can be speculated that the generation of certain 5′ termini might be a prerequisite of translation; however, no experimental data that would substantiate this hypothesis are yet available for plant mitochondria (20).

In the course of the characterization of the complete mitochondrial transcriptome in the model plant *A.thaliana*, we have investigated the gene arrangements and the steady state transcripts of the mitochondrially encoded *cox*3 gene in three different ecotypes. While no differences were observed in the 3′ region, within the gene and 5′ flanking sequence up to position −584 sequences further upstream differ between the three ecotypes. Our analyses suggest that these far upstream sequences influence the generation of 5′ ends located ~140 nt downstream in sequences, which are identical in all three investigated ecotypes.

MATERIALS AND METHODS

Preparation of nucleic acids from *A.thaliana*

A.thaliana seedlings were grown under a 16 h/8 h light/dark regime (100 μmol/m^2 s) at 23°C for 15 or 16 days. Total cellular DNA was isolated from above ground parts of the seedlings and cell suspension culture with PhytoPure kit according to the manufacturer's instructions (Amersham Bioscience). Total DNA from 28-day C24/Col hybrid old plants (21), which were gown under 16 h/8 h light/dark regime (60 μmol/m^2 s), was isolated with DNeasy Plant Mini kits (Qiagen).

Total RNA was isolated from the same plants with an RNeasy Plant Mini kit (Qiagen). Extraction procedures followed the manufacturer's recommendations.

Mitochondria were isolated from cell suspension culture following a protocol published previously (22). mtDNA and mtRNA were extracted according to a protocol described previously (23). For the isolation of mtDNA, 3 M sodium acetate (pH 7.0) was used instead of 2 M sodium acetate, pH 4.0.

Northern and Southern hybridization

For Southern blot analysis, ~10 μg of total DNA were digested with HindIII and size fractionated on 1% (w/v) agarose gels. Southern transfer and hybridization were performed with porablotNY$^+$ membranes following a protocol given by the manufacturer (Macherey-Nagel). The DNA probe corresponding to the *cox*3 N-terminal region (positions 3–232 with respect to the ATG) was generated by PCR with primers cox3-5PS and cox3-3PS. Labeling of the probe was performed with Rediprime™ II Random Prime Labelling System and 50 μCi [α-^{32}P]dCTP according to the manufacturer's protocol (Amersham Bioscience).

Southern blot analysis of 1–3 μg of mtDNA was performed under the same conditions as described above. 5′ end labeling of oligonucleotide probes was performed using polynucleotide kinase following the standard procedures (24). An mtDNA library established in pBluescript II was screened after colony transfer to porablotNY$^+$ membranes according to the manufacturer's protocol.

For northern blot analysis, ~10 μg of total RNA was sizes fractioned on agarose gels under denaturing conditions in the presence of glyoxale (24). Blotting and hybridization was performed using Duralon UV as recommended by the manufacturer (Stratagene).

PCR analyses

PCRs were performed with *Taq* DNA polymerase (Promega), BD Advantage™ 2 PCR Enzyme System (BD Bioscience) and Phusion™ High-Fidelity DNA polymerase (Finnzymes) according to conditions specified by the companies. PCR parameters for *Taq* and BD Advantage™ DNA polymerase were as follows: 3 min at 95°C, 35 cycles of 1 min at 94°C, 1 min at oligonucleotide-specific annealing temperature, 1 min (per kb product) at 72°C (*Taq*) or 68°C (BD Advantage DNA Pol) and a final elongation step of 3 min at the same temperature. Parameter for amplification reactions performed with Phusion™ High-Fidelity DNA polymerase were 30 s at 98°C, 35 cycles of 8 s at 98°C, 25 s at oligonucleotide-specific annealing temperature, 30 s (per kb product) at 72°C and a final elongation step of 5 min. Primers used in the individual reactions are indicated in the text. Primer sequences are available on request.

CR–RT–PCR was performed as described previously (25). Primers used are given in the text and indicated in Figure 5, respectively.

Standard methods

Primer extension experiments, restriction analyses and other basic methods in molecular biology were carried out following the standard procedures (24) or protocols of the manufacturer. Sequence analysis was carried out with Thermo Sequenase sequencing chemicals and ALF express sequencers (Amersham Bioscience). *In silico* sequence analyses were carried out with different tools at the NCBI server (26,27).

RESULTS

A 1.8 kb insertion is present in the upstream region of the nuclear *cox*3 gene in *A.thaliana* ecotype Col

In the course of a systematic analysis of transcription in mitochondria of *A.thaliana*, mitochondrial sequences from ecotype C24 were compared with those of an mtDNA copy present in chromosome 2 in the nuclear DNA of ecotype Col (1,28). This nuclear mtDNA copy comprises ~618 kb, of which ~270 kb are sequenced. It is almost identical with the genuine mitochondrial DNA sequence determined in ecotype C24 (28,29). This comparison revealed the presence of an ~1.8 kb insertion in the *cox*3 upstream region in the Col sequence (Figure 1). The insertion of 1790 bp is flanked by a 9 bp direct repeat, which is present in a single copy in the C24 mitochondrial sequence and which could potentially be involved in the insertion of this DNA fragment. The insertion is assembled from mtDNA fragments of various origins, a phenomenon frequently observed in these organelles (30). Major parts originate from *rpl*16/*rps*3, *atp*9, orf275 and orf262 sequences (1).

This insertion displaces a potential mitochondrial promoter present ~640 bp upstream of the *cox*3 gene in ecotype C24 by 1.8 kb and thus could potentially influence transcription of this

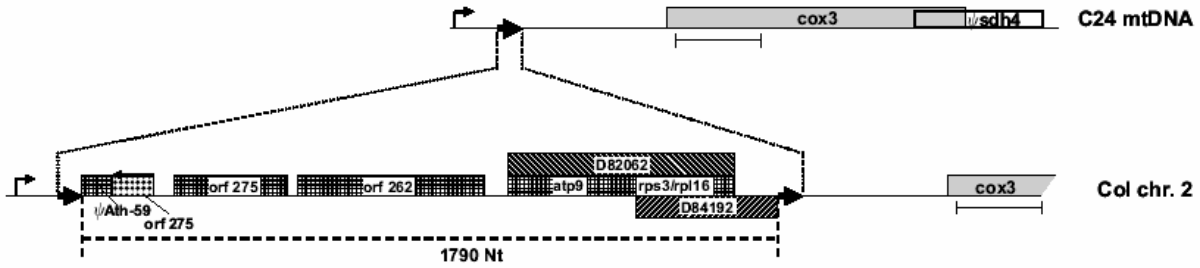


Figure 1. Scheme of *cox3* gene arrangements in the mitochondrial DNA of *A.thaliana* ecotype C24 (C24 mtDNA) and in the chromosome 2 mtDNA copy in the nuclear DNA of ecotype Col (Col chr. 2). A 1790 bp insertion flanked by nonanucleotide direct repeats (bold arrows) is present in Col. A putative *cox3* promoter of the CNM-type (bent arrow) is found ~640 (C24) and 2430 bp (Col) upstream of the *cox3* reading frame (gray box), which overlaps with a Ψsdh4 gene (white box) in C24 and Col, respectively. Parts of this insertion are found in other genomic regions of the mtDNA of C24 (indicated by checkered boxes). An orf275 sequence in reverse orientation is given as rhomboid box with an arrow. *rpl16/rps3* sequences found in the mtDNA of a maternal distorted leaf mutant are given by hatched boxes with the respective accession numbers (38). ΨAth-59 refers to a sequence with high similarity to a small non-coding RNA (44). The hybridization probe used in the Southern blot analysis of total DNA is given beneath the *cox3* gene.

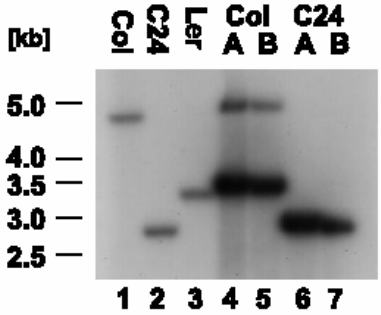


Figure 2. A *cox3* probe detects different DNA fragments in *A.thaliana ecotypes* C24, Col and L*er*. Total DNA isolated from green seedlings of ecotypes Col (lane 1), C24 (lane 2) and L*er* (lane 3) as well as total DNA from two different cell suspension culture lines each (A and B) from Col (lanes 4 and 5) and C24 (lanes 6 and 7) were digested with HindIII (H) and hybridized with an At*cox3* probe.

gene in ecotype Col. However, since this insertion was detected in the chromosomal copy the question arises whether the mtDNA of Col corresponds to the genomic arrangement observed in the nucleus or whether it corresponds to the mtDNA configuration present in C24 mitochondria.

Southern blot analysis reveals different At*cox3* environments in ecotypes Columbia (Col), C24 and Landsberg *erecta* (L*er*)

To investigate the genomic environment of the *cox3* gene in these ecotypes, total DNA isolated from green seedlings was hybridized with a probe corresponding to the N-terminal part of the *cox3* reading frame. The detection of HindIII fragments of 4.7 kb in Col, 2.9 kb in C24 and 3.4 kb in L*er* identifies individual *cox3* gene arrangements in the three ecotypes (Figure 2, lanes 1–3). Analogous results were obtained with EcoRV-digested DNAs (Col: 4.5 kb; C24 2.7 kb; L*er* 7.2 kb) (data not shown). The fragment sizes detected in C24 and Col correspond to those expected from the sequences of mtDNA (C24) and chromosome 2 (Col), respectively (1,28). These hybridizations indicate that in green plants only single *cox3* arrangements exist in the individual ecotypes, strongly suggesting that at least in Col the *cox3* genomic environment is identical in mtDNA and in the mtDNA copy in chromosome 2. The presence of an mtDNA copy homologous to the one detected in ecotype Col was investigated by PCR. The results obtained strongly suggest that such homologous mtDNA copies do not exist in ecotypes C24 and L*er*. This indicates that the insertion of the mtDNA copy in chromosome 2 in Col is an evolutionary recent event (Supplementary Figure S1).

A different hybridization pattern is observed in the Southern analysis of total DNA extracted from individually propagated cell suspension cultures. While again the expected fragments were detected in C24 DNA (Figure 2, 2.9 kb HindIII fragment in lanes 6 and 7; 2.7 kb EcoRV fragment, data not shown), each two fragments were visualized in the DNA of Col. Relatively weak signals correspond to the expected sizes observed in DNA from green plants (Figure 2, 4.7 kb HindIII fragments in lanes 4 and 5; 4.5 kb EcoRV fragments, data not shown). Additional strong signals are almost identical to those observed in green plants from L*er* (Figure 2, 3.4 kb HindIII fragments in lanes 4 and 5; 7.2 kb EcoRV fragments, data not shown). An amplification of the 3.4 kb HindIII fragment by PCR followed by restriction digestion confirmed the identity of these fragments in Col cell suspension culture and L*er* green plants (Supplementary Figure S2).

The different intensities of the signals in Col cell suspension culture could be due to the origin of the corresponding DNA fragments from the nucleus (weak signals) and the mtDNA (strong signals), which is present in much higher copy number in total DNA preparations. Thus, mtDNAs were isolated from cell suspension cultures (C24 and Col) and examined by Southern hybridization. An At*cox3*-specific oligonucleotide probe (cox3-3PS) detects the expected DNA fragments in both HindIII- and EcoRV-digested DNAs from C24, which were identical with those found in all other hybridization analyses of this ecotype (Figure 3, center part; HindIII: 2.9 kb; EcoRV: 2.7 kb). In Col mtDNA also single fragments were detected (Figure 3, middle panel; HindIII: 3.4 kb; EcoRV: 7.2 kb). These correspond to the strong signals observed in the hybridization analysis of total DNA confirming the mitochondrial origin of these fragments (Figure 2, lanes 4 and 5). This also indirectly confirms that the relatively

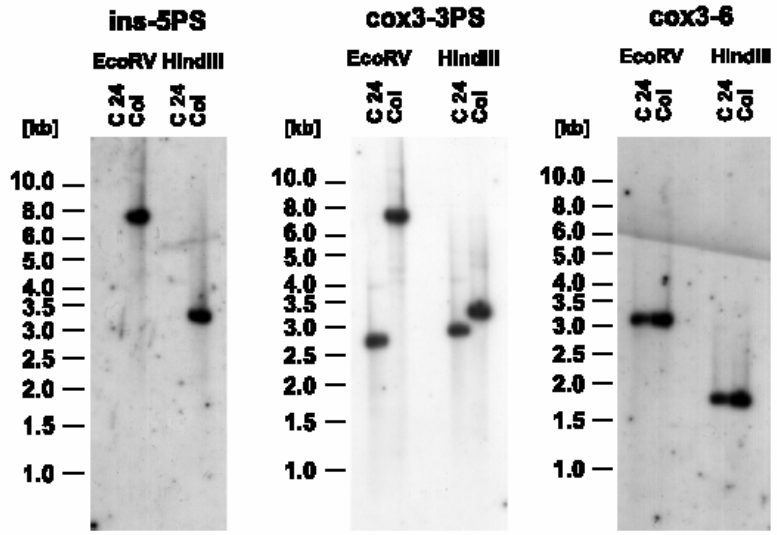


Figure 3. Southern blot analysis of mtDNA isolated from cell suspension cultures of ecotypes Col and C24. mtDNA was digested with HindIII and EcoRV, respectively, and hybridized with oligonucleotide probes specific for the 1790 bp insertion in Col (ins-5′PS, left panel), for the 5′ part of the *cox*3 reading frame (cox3-3PS, middle panel) and for the region downstream of the *cox*3 gene (cox3-6, right panel).

weak signals visualizing fragments with the expected size of 4.7 kb (HindIII) and 4.5 kb (EcoRV) are indeed of nuclear origin. An oligonucleotide probe specific for the insertion in Col (ins-5PS) only detects the predicted fragments in Col, confirming that this insertion or at least parts of it are also present in the mtDNA from this ecotype (Figure 3, left panel; HindIII: 3.4 kb; EcoRV: 7.2 kb). No hybridization of this probe was observed in the analysis of C24 mtDNA, suggesting that this insertion is completely absent from mtDNA of this ecotype. No differences were observed between the different ecotypes when hybridization probes correspond to At*cox*3 3′ regions, suggesting that the different At*cox*3 genomic environments detected in the investigated ecotypes as well as in different tissues of these (i.e. green plants and cell suspension cultures) are due to rearrangements in the 5′ regions (Figure 3, right panel; HindIII: 1.8 kb; EcoRV 3.1, data not shown).

These experiments indicate that in Col cell suspension culture a *cox*3 arrangement becomes predominant that corresponds to the single detectable arrangement in green plants of ecotype L*er*. This suggests that this arrangement pre-exists in Col most likely as a sublimon. Such substoichiometric molecules have previously been identified in maize mitochondria (31,32).

To exclude the possible contamination and/or mixtures of different ecotypes in the cell suspension culture, the ecotypes of the Col cell suspension cultures as well as of Col and L*er* plants were verified using specific single nucleotide polymorphisms CER462304 and CER428393 (33) (data not shown).

The rearrangement in the mtDNA of Col suspension culture relocates a complete *atp*9 gene into the *cox*3 5′ region

For a detailed investigation of the *cox*3 genomic environment in the mtDNA of Col suspension cells, which might very similarly or identically also exist in ecotype L*er*, a HindIII mtDNA genomic library was established and screened using the *cox*3-specific oligonucleotide probe (cox3-3PS). This identified several clones which contained 3.4 kb HindIII fragments, one of which was almost completely sequenced (data not shown).

The *cox*3 5′ region found in cell suspension culture is co-linear up to nucleotide position −1352 in respect to the ATG with the sequence of the mtDNA copy in chromosome 2 (Supplementary Figure S3), which is most likely identical to the original mtDNA in green plants of this ecotype (see above and Figures 2 and 3). This includes 65 bp of the 3′ terminal part of the *atp*9 gene and a 186 nt downstream region, which almost exactly corresponds to the 3′ non-translated region of the *atp*9 mRNA in both green plants and cell suspension culture [Forner and Binder, unpublished data, (16)]. In contrast to the sequence in chromosome 2 (Figure 1), a complete *atp*9 gene including a 5′ region containing two CNM-type promoters is present in the cell suspension specific *cox*3 arrangement in Col.

Different *cox*3 mRNAs are present in the three investigated *A.thaliana* ecotypes

The analysis of the genomic environments of the *cox*3 gene in the three different ecotypes revealed substantial variations in the 5′ regions. To investigate the potential influence of these rearrangements on transcription, At*cox*3 steady state transcripts from green seedlings were analyzed by northern hybridization. A 1.45 kb probe corresponding to the complete *cox*3 reading frame as well as 341 and 306 nt 5′ and 3′ flanking sequences detects mRNAs of ~1500 nt from all three ecotypes (Figure 4). In C24, an additional major steady state mRNA species of ~1560 nt and a minor RNA molecule of ~2200 nt are detected. Additional transcripts are also detected in Col,

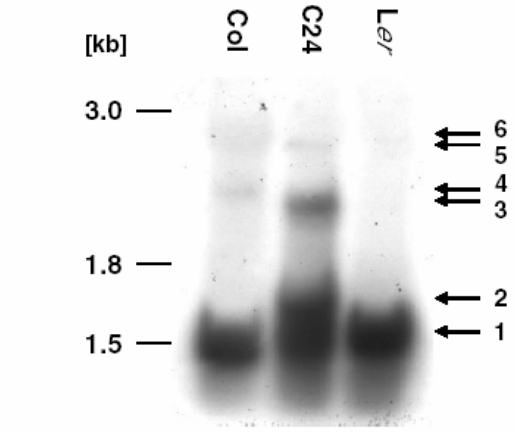


Figure 4. Northern blot analysis of At*cox*3 mRNA in total RNAs isolated from green seedlings of *A.thaliana* ecotypes C24, Col and L*er*. A DNA probe representing the complete *cox*3 reading frame as well as 341 and 306 nt 5′ and 3′ flanking sequences detects transcript of ~1500 nt (1) in all three ecotypes. A second prominent transcript of ~1560 nt (2) is solely detectable in C24. In this ecotype, an ~2.2 kb prominent (3) and an ~2.8 kb low abundant (5) potential precursor RNA, respectively, are also observed. Such potential primary transcripts (4 and 6) with slightly different sizes are also present in Col. The individual RNAs are indicted by arrows and numbered in the right margin.

but these are different from those found in C24. Thus, *cox*3 mRNAs of different sizes are present in the three ecotypes.

Different 5′ termini are found in identical *cox*3 upstream sequences in the three ecotypes

In order to determine the 5′ and 3′ extremities of the *cox*3 mRNA in the different ecotypes, CR–RT–PCR analyses were performed with total RNAs from green seedlings and mitochondrial RNA from cell suspension cultures. In all RNAs from all ecotypes and tissues identical major 3′ ends 314 (+/2) nucleotides downstream of the *cox*3 reading frame were found (Figure 5A–E). In Col cell suspension culture, additional major 3′ ends of shorter RNAs were present (Figure 5E). These ends were found 21 (±6) and 101 (±2) nucleotides, respectively, downstream of the *cox*3 stop codon and were likewise found in green seedlings of the same ecotype, however, at a much lower frequency (Figure 5B, 2 and 1 out of 29 clones).

In contrast to the 3′ termini, a complex situation is found for the 5′ ends. In green seedlings, major 5′ ends are found at position −379 (±1) in all investigated ecotypes (Figure 5A–C). These ends were detected after two PCRs with primer pairs b/d and a/e after initiating cDNA synthesis with primer c. In C24, two additional ends are detected at −384 (±2) and −437 (Figure 5A). With RNA from the cell suspension culture of C24 5′ ends were also found at −384 (±1) and −444 /−437 (Figure 5D). With RNA from cell suspension culture from Col again a major 5′ end is detected at −379 (±2) but also a minor end at −383 (±1), which is not detected in green seedling of this ecotype. A CR–RT–PCR of cell suspension culture RNA was performed with 5′ primer g annealing 295–316 nt upstream of the oligonucleotide a used for the analysis of RNA from green seedlings. However, both primers essentially detect the same ends (Figure 5). In addition, in the analyses with primer a no ends were detected downstream of primer g. Taken together the CR–RT–PCR with these primer pairs detects four different 5′ ends. Two of these ends locate in close vicinity at −379 and −384 and can be considered as a single end, although the frequencies with which the individual ends were found differ between the investigated ecotypes. This 5′ end and the 3′ terminus at +314 define the major transcript of ~1500 nt, detected in all three ecotypes (Figure 4). The other 5′ ends at −444 and −437, respectively, also represent probably a single terminus, which is unique to C24. This could be the 5′ end of the shorter of the two additional RNA species that are exclusively found in this ecotype (Figure 6, middle lane C24). To identify 5′ ends further upstream located, CR–RT–PCR with primer f located upstream of the detected major 5′ termini was performed with mtRNA from cell suspension culture from C24 and Col. These analyses detect identical 5′ ends at −537 (±2) and −538 (±5) (Figure 5D and E), respectively, in both ecotypes, and a Col-specific end at −561/−562. All the different 5′ ends found in the ecotypes investigated are located in a region in which the sequences are identical.

In steady state RNAs from cell suspension cultures from C24 and Col additional differing *cox*3 5′ ends were detected further upstream with 5′ primer f. However, these are located in a sequence region that differs between these ecotypes. In C24, these map to positions −642 and −652 directly at a predicted CNM-type promoter or 10 nt upstream, respectively. In Col 5′ termini at −627 and −773 (±4) are detected in another CR–RT–PCR and are situated in the 1790 bp insertion.

Another ecotype-specific 5′ end is detected at positions −561/−562 in Col cell suspension culture. Since this end has so far not been detected by the primer extension analysis, an experimental artifact can presently not be excluded.

*Cox*3 transcript analysis in C24/Col hybrids

Since the major At*cox*3 5′ ends are located in identical sequence environments, two factors might be responsible for the generation of these varying termini: either a mitochondrially encoded factor which could act in *trans* or *cis*, or (a) *trans* factor(s) encoded in the different nuclear backgrounds of the ecotypes. We thus analyzed the At*cox*3 5′ ends in individual plants of reciprocal crossings between C24 and Col. The identity of the nuclear background of the parent plants and these hybrids was confirmed using SNPs MASC03783 and MASC03308 (34) (data not shown). In addition, the configuration and thus origin of the mtDNA was tested by the presence or absence of the 1790 (+9 bp repeat) bp insertion by PCR (Supplementary Figure S4). In all plants tested, the configuration of the mtDNA corresponds to the female parent, which is in agreement with a strict maternal inheritance of mitochondria.

Total RNAs isolated from the hybrids and the parental ecotypes were investigated by northern hybridization. The *cox*3 probe used in these experiments has been described in the above mentioned northern analysis. In ecotypes C24 and Col the same mRNAs are detected as described above (Figure 4 and Figure 6A, lanes 1 and 2). In the hybrids, the pattern of *cox*3 mRNAs is always consistent with those found in

4678 Nucleic Acids Research, 2005, Vol. 33, No. 15

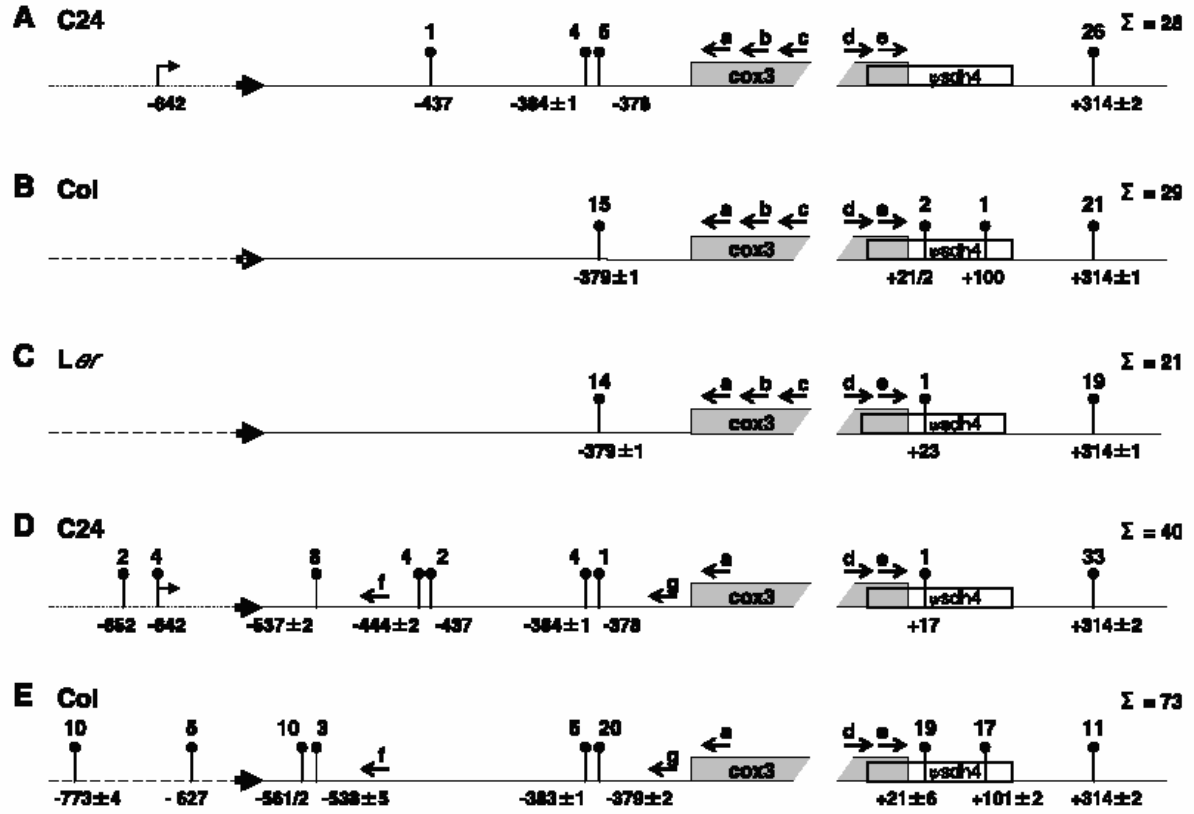


Figure 5. CR–RT–PCR mapping of the 5′ and 3′ extremities of At*cox3* transcripts from ecotypes C24, Col and L*er*. Analyses were performed with total RNA from green seedlings (**A**–**C**) and mtRNA from cell suspension cultures (**D** and **E**). The number of individual clones analyzed is given in the right margin. Prominent ends found in several clones are indicated by pins with number of clones above. The continuous line represents sequences identical in all investigated ecotypes; dashed lines are identical in Col and L*er*; the dotted line is specific for C24 in this position. The locations of the ends are given in respect to the ATG or the stop codon of the At*cox3* gene. Primers (open arrows) used in the CR–RT–PCR are a: Atcox3-1; b: Atcox3-2; c: Atcox3-3PS; d: Atcox3-4; e: Atcox3-5; f: ATB2174-2; g: Atcox3-3.

the female parent, demonstrating a maternal inheritance of the C24-specific mRNA species of ∼1560 nt (Figure 6A, lanes 4–7).

Investigation of these mRNAs by primer extension analysis with oligonucleotide Atcox3-3 revealed a corresponding result. 5′ termini detected in the parental lines are consistent with those found in the CR–RT–PCR described above, except for C24. Here, primer extension signal 1 corresponds to the 5′ ends mapped at −444/−437; however, signal 2 corresponding to −384/−378 is not visible, although it has been clearly detected in the CR–RT–PCR and a corresponding mRNA is detected in the northern analysis (Figure 6B, lane 1; Figure 5A and D). In Col and L*er* extension, signal 2 matches to the −378 end, while signal 1 corresponding to the 5′ end at −444/−437 is not observed. This is consistent with the results of the CR–RT–PCR analyses (Figure 6B, lanes 2 and 3; Figure 5B, C and E). The 5′ ends mapped in the hybrid plants are identical with those of the female parent. In hybrid plants with a C24 female parent, primer extension signal 1 is clearly detectable (Figure 6B, lanes 4 and 5) while this 5′ end is not found in hybrid plants with a Col female parent. Thus, the generation of the −444/−437 5′ end in C24 is independent from the ecotype-specific nuclear background but follows the maternal lineage, which is consistent with the northern analysis. This strongly suggests a cytoplasmic inheritance of this trait. Thus, a mitochondrial factor is responsible for the generation of the additional 5′ end located within the sequence identical in C24, Col and L*er*.

DISCUSSION

A nonanucleotide direct repeat might be responsible for the integration or excision of the 1790 bp DNA fragment

Our analysis of the At*cox3* upstream sequence revealed three different genomic arrangements in the far upstream region of this gene. In Col, a 1790 bp insert is found to be flanked by a nonanucleotide direct repeat (5′-CTTTACGAG-3′), which is present as a single copy at this locus in C24 (Figure 1). These arrangements suggest that this nonanucleotide repeat is involved in the recombinatorial integration or excision of the insert. Such small repeats have been previously found to be involved in recombination events (35). For instance, in *Oenothera* a decanucleotide repeat is responsible for circularization and formation of a small subgenomic molecule,

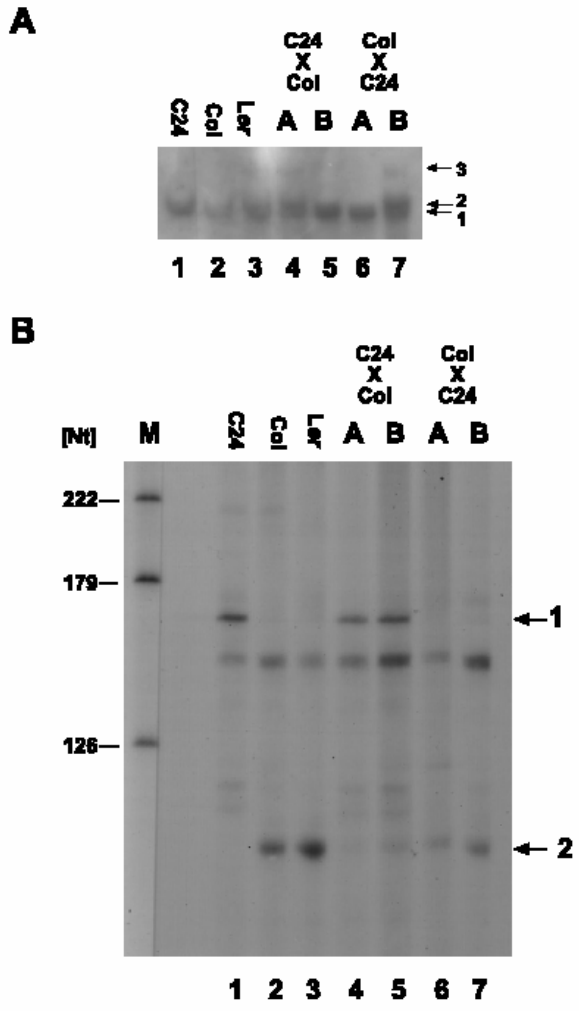


Figure 6. Transcript analysis of C24 × Col and Col × C24 hybrids. (**A**) Northern hybridization of total RNA from ecotypes C24, Col, L*er* (lanes 1–3) and each two plants (A and B) of C24 × Col and Col × C24 hybrids (lanes 4–7). The C24 specific RNAs of ~1560 and 2800 nt are detected in the C24 parental line (lane 1) and in hybrids with C24 female parents (lanes 4 and 5). (**B**) Primer extension analysis of the same RNAs as described in (A) with primer Atcox3-3. Products of ~160 nt (1) were generated on RNAs of C24 (lane 1) and of plants A and B with C24 as female parental line (lanes 4 and 5). These products correspond to 5′ ends at positions around −440, which are exclusively detected in the CR–RT–PCR of this ecotype (Figure 5A and D). A second prominent product (2) of ~100 nt is detected in almost all plants, except in C24 in lane 1. This ends corresponds to 5′ termini around −380, which was found in the CR–RT–PCR in all ecotypes (Figure 5).

which contains the 3′ part of the 26S rRNA (36). Similarly, the 1790 bp insertion might be excised from the mtDNA genome by an intramolecular recombination of the nonanucleotide repeats in green plants of ecotype Col. However, no corresponding signal has been observed in the hybridization of total DNA of Col plants. This is most likely due to the low abundance of such a molecule, which would be present in substoichiometric amounts as described for sublimons (31). The identified nonanucleotide can be found in four other places within the mitochondrial genome of C24 (five places in total) and four times within the chromosome 2 sequence in Col at locations identical with those in C24, but there are so far no experimental data for the involvement of one of these repeats in a recombination event.

A third configuration is present in L*er*

The 1790 bp insert in Col is composed of mtDNA fragments of different origins (Figure 1). Smaller sequence stretches with sizes between 32 and 71 bp are unique and are not found in the mitochondrial (C24), chloroplast (Col) or nuclear (Col) genome of *A.thaliana*. However, since the mitochondrial DNA copy within chromosome 2 of Col is not completely sequenced, it is possible that homologous sequences are present in another locations in the mtDNA (1,28,37). All other parts are duplicated sequences, whose origins are indicated in Figure 1. These are parts of other genes or ORFs. The insert is only partially conserved in the L*er* configuration, which is also the predominant arrangement in the Col suspension culture (Supplementary Figure S3). Here, most likely a recombination has occurred within the *atp*9 sequence, which established a complete reading frame of this gene. The insert sequences downstream of the *atp*9 gene are conserved in fragments of a Maternal Distorted Leaf (MDL) mutant established in L*er* (accession numbers: D82062 and D84192) (38). This supports that identical arrangements exist in L*er* and Col. Most likely, the set up of the cell suspension culture has shifted the stoichiometry of the two pre-existing *cox*3 arrangements in Col, similar to the changes observed in the chm1-1 mutant (39,40), a phenomenon that has been observed in sublimons of other plant species and that contributes to the complexity of plant mitochondrial genomes (31,32,40–42). Interestingly, the size of the BamHI fragment deduced from the 3.4 kb HindIII fragment and the L*er* sequence of data bank entry D82062 is 1.56 kb, which fits with the predominant fragment hybridizing with an *atp*9 in the chm1-1 mutant (40). Thus, the set up of the cell suspension culture might have somehow altered *chm* gene expression, which may have caused the shift in stoichiometry. However, detailed comparative studies of the *atp*9 arrangements in the different ecotypes are necessary to elucidate these complex genomic conditions. Differing mtDNA configurations have also been observed in other ecotypes and other genomic environments (43).

MtDNA copies in different ecotypes?

Three different 5′ arrangements are found in the mtDNA of the three investigated ecotypes. This raises the question whether potential nuclear mtDNA copies, if they exist, correspond to the individual mtDNA configurations or whether they all correspond to the configuration of the mtDNA copy in Col. In ecotypes C24 and L*er*, our PCR analyses do not indicate the presence of an mtDNA copy homologous to the one in Col. This does not generally exclude other ecotypes to contain an mtDNA copy homologous to the one in Col; however, we assume that such a copy might only be present in ecotypes closely related to Col. Furthermore, the data suggest the insertion of the mtDNA in chromosome 2 of Col to be an evolutionary recent event. More studies with a large variety of different ecotypes are required to finally clarify this issue.

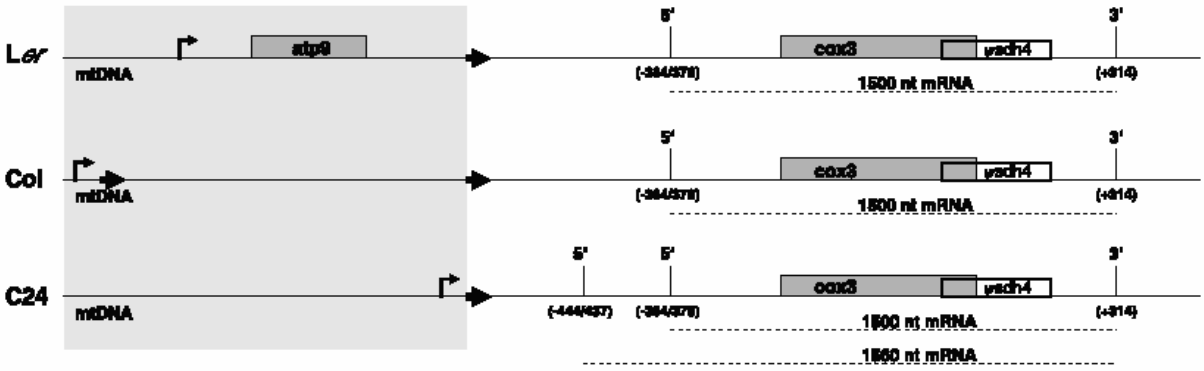


Figure 7. Genomic arrangements and transcripts of the *cox*3 genes in three different *A.thaliana* ecotypes. The localization of the *atp*9, *cox*3 (gray boxes) and the Ψ*sdh*4 genes (transparent box) are indicated in the mtDNAs of the investigated *A.thaliana* ecotypes here. The major 5′ and 3′ termini of the predominant mRNAs (dashed lines) are given including their position in respect to the ATG (5′ ends) and stop codon (3′ ends). Genomic regions that differ between the ecotypes are highlighted by a gray background. Mitochondrial promoters are given as bent arrows. Bold arrows represent the nonanucleotide direct repeat.

Different precursor RNAs are detectable in Col and C24

The northern analysis revealed different potential precursor RNAs in the different ecotypes. These might be attributed to transcription initiation at different promoters (Figure 7). In C24, a CNM-type promoter can be predicted with a potential transcription start point 642 bp upstream of the At*cox*3 reading frame. Indeed, 5′ ends were mapped at this site confirming the functional importance of this promoter in this ecotype. However, other 5′ termini are mapped 10 nt upstream, suggesting a further 5′-located transcription initiation site. This promoter as well as the CNM-type sequence are also present in the At*cox*3 arrangement in Col. Although the 1790 bp insertion (+9 bp repeat) increased the distance to the *cox*3 reading frame, these promoters are probably similarly active. However, it is also possible that transcription initiation sites are present within the insertion, which in addition could contribute to the transcription of this gene in Col. Several 5′ termini have been mapped around positions −773 and −627 within the insertion (Figure 5E); however, it is unclear whether they originate from processing or transcription initiation. Sequences at these 5′ ends do not show any striking similarity to new *A.thaliana* promoter sequences identified recently (8); however, considering their variability the presence of promoters at these ends can well be possible. The sequences at these ends are also present in the L*er* arrangement. Thus, as in Col transcription initiation at these sequences may occur, although it is more likely that the *cox*3 is co-transcribed with the 5′-located *atp*9 gene, for which several promoters have been identified (8).

A mitochondrial factor is responsible for the generation of an alternative 5′ end

In all ecotypes, a major transcript of ~1500 nt can be detected (Figure 4). The 3′ ends are consistently found 314 nt downstream of the *cox*3 stop codon (Figure 5). The generation of these ends either by transcription termination or by post-transcriptional processing thus seems to be independent from 5′ end formation, which is different in the investigated ecotypes. A major 5′ terminus is also consistently found in the three investigated ecotypes and crosses of them (Figures 4–6). Only in C24 the 5′ end at −384/−378 is not detectable in the primer extension analysis (Figure 6B, lane 1), although both the detection of the 1500 nt transcript and the CR–RT–PCR demonstrate the existence of this end (Figures 4, 5 and 6A). Thus, both *cis* and *trans* factors determining this end are identical in these ecotypes. While the latter are unknown, the *cis* elements are most likely located within the sequences identical in all three ecotypes ranging from 584 bp upstream of the *cox*3 reading frame to the 3′ ends.

In C24, an additional prominent 5′ end ~60 nt upstream of the normal major 5′ terminus is detected (Figures 5 and 6B). This end is also located in a sequence region, which is identical in the investigated ecotypes (Figure 7). A nuclear encoded factor responsible for the generation of this end can be excluded, since the appearance of this end is maternally and thus mitochondrially inherited. Thus, a mitochondrially encoded *cis* element or *trans* factor specifies the formation of this 5′ end. Considering the coding capacity of all known mitochondrial genomes, a mitochondrially encoded *trans* factor required for specific transcription initiation or processing is highly unlikely. Instead, we assume that the differing sequences located at least 140 bp upstream are responsible for the appearance of this alternative 5′ end. These sequences most likely differently influence the folding of the precursor RNA, which in the case of C24 creates a secondary structure that allows a processing in this sequence region. Thus, there is influence across a distance of >140 nt. Generally two different scenarios are feasible: either the alternative folding allows base pairing at this −444/437 site and a processing by a double-strand-specific ribonuclease or in contrary the normally base paired region is made single stranded and thus accessible to a single-strand-specific ribonuclease.

Taken together, the detailed analysis of mitochondrial transcripts in different ecotypes is an additional useful tool to elucidate and characterize the requirements for transcript end formation in plant mitochondria.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

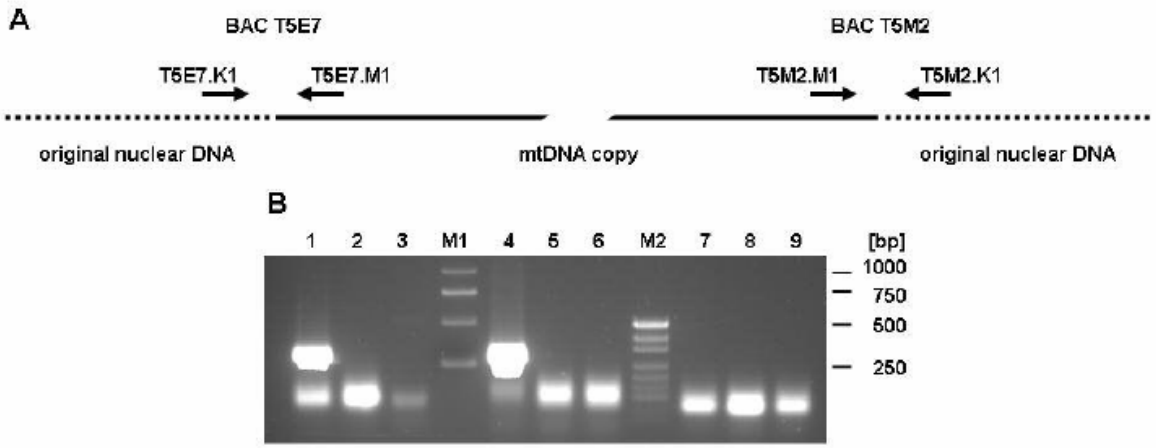
The authors thank Carmen Schilling-Kolle for taking care of the cell suspension cultures. This work is supported by DFG grant Bi 590/6-1 and a fellowship of the Studienstiftung des Deutschen Volkes to J.F. Funding to pay the Open Access publication charges for this article was provided by the DFG.

Conflict of interest statement. None declared.

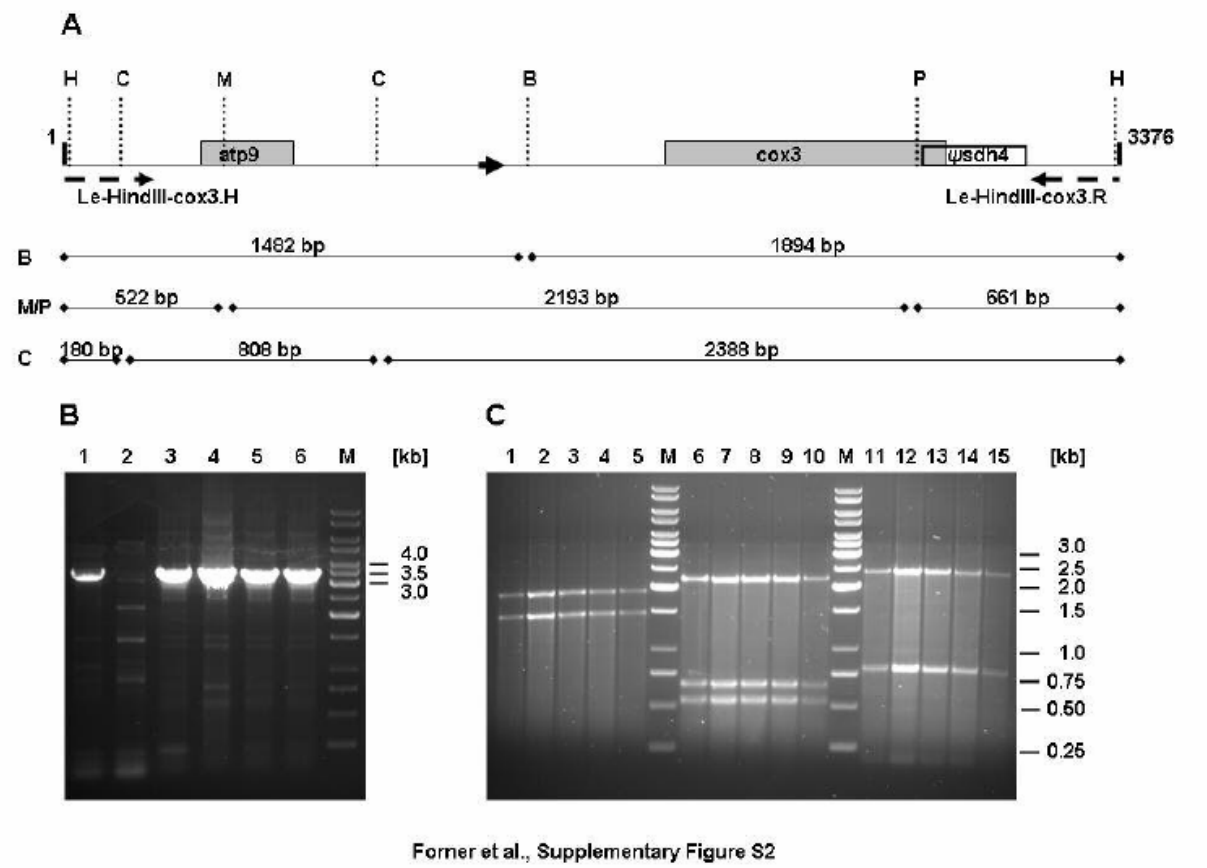
REFERENCES

1. Unseld,M., Marienfeld,J.R., Brandt,P. and Brennicke,A. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genet.*, **15**, 57–61.
2. Notsu,Y., Masood,S., Nishikawa,T., Kubo,N., Akiduki,G., Nakazono,M., Hirai,A. and Kadowaki,K. (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics*, **268**, 434–445.
3. Kubo,T., Nishizawa,S., Sugawara,A., Itchoda,N., Estiati,A. and Mikami,T. (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic Acids Res.*, **28**, 2571–2576.
4. Handa,H. (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.*, **31**, 5907–5916.
5. Rapp,W.D. and Stern,D.B. (1992) A conserved 11 nucleotide sequence contains an essential promoter element of the maize mitochondrial atp1 gene. *EMBO J.*, **11**, 1065–1073.
6. Dombrowski,S., Hoffmann,M., Guha,C. and Binder,S. (1999) Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria. *J. Biol. Chem.*, **274**, 10094–10099.
7. Hoffmann,M. and Binder,S. (2002) Functional importance of nucleotide identities within the pea *atp9* mitochondrial promoter sequence. *J. Mol. Biol.*, **320**, 943–950.
8. Kühn,K., Weihe,A. and Borner,T. (2005) Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic Acids Res.*, **33**, 337–346.
9. Hoffmann,M., Kuhn,J., Däschner,K. and Binder,S. (2001) The RNA world of plant mitochondria. *Prog. Nucleic Acid Res. Mol. Biol.*, **70**, 119–154.
10. Kuhn,J., Tengler,U. and Binder,S. (2001) Transcript lifetime is balanced between stabilizing stem–loop structures and degradation-promoting polyadenylation in plant mitochondria. *Mol. Cell. Biol.*, **21**, 731–742.
11. Dombrowski,S., Brennicke,A. and Binder,S. (1997) 3′-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *EMBO J.*, **16**, 5069–5076.
12. Gagliardi,D. and Leaver,C.J. (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO J.*, **18**, 3757–3766.
13. Gagliardi,D., Perrin,R., Marechal-Drouard,L., Grienenberger,J.M. and Leaver,C.J. (2001) Plant mitochondrial polyadenylated mRNAs are degraded by a 3′- to 5′exoribonuclease activity, which proceeds unimpeded by stable secondary structures. *J. Biol. Chem.*, **276**, 43541–43547.
14. Lupold,D.S., Caoile,A.G. and Stern,D.B. (1999) Polyadenylation occurs at multiple sites in maize mitochondrial *cox2* mRNA and is independent of editing status. *Plant Cell*, **11**, 1565–1578.
15. 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.
16. Perrin,R., Meyer,E.H., Zaepfel,M., Kim,Y.J., Mache,R., Grienenberger,J.M., Gualberto,J.M. and Gagliardi,D. (2004) Two exoribonucleases act sequentially to process mature 3′-ends of atp9 mRNAs in *Arabidopsis* mitochondria. *J. Biol. Chem.*, **279**, 25440–25446.
17. Marchfelder,A. (1995) Plant mitochondrial RNase P. *Mol. Biol. Rep.*, **22**, 151–156.
18. Perrin,R., Lange,H., Grienenberger,J.M. and Gagliardi,D. (2004) AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in *Arabidopsis thaliana* mitochondria. *Nucleic Acids Res.*, **32**, 5174–5482.
19. Schuster,W. and Brennicke,A. (1989) Conserved sequence elements at putative processing sites in plant mitochondria. *Curr. Genet.*, **15**, 187–192.
20. Marchfelder,A. and Binder,S. (2004) Plastid and plant mitochondrial RNA processing and RNA stability. In Daniell,H. and Chase,C. (eds), *Molecular Biology and Biotechnolgy of Plant Organelles*. Springer, Dordrecht, Netherlands, pp. 261–294.
21. Meyer,R.C., Torjek,O., Becher,M. and Altmann,T. (2004) Heterosis of biomass production in *Arabidopsis*. Establishment during early development. *Plant Physiol.*, **134**, 1813–1823.
22. Klein,M., Binder,S. and Brennicke,A. (1998) Purification of mitochondria from Arabidopsis. *Methods Mol. Biol.*, **82**, 49–53.
23. Binder,S. (1995) Mitochondrial nucleic acid purification and analysis. *Methods Mol. Biol.*, **49**, 383–389.
24. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Habror Laboratory Press, Cold Spring Harbor, NY.
25. 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.
26. Tatusova,T.A. and Madden,T.L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.*, **174**, 247–250.
27. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
28. Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.I., Town,C.D., Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M. *et al.* (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature*, **402**, 761–768.
29. Stupar,R.M., Lilly,J.W., Town,C.D., Cheng,Z., Kaul,S., Buell,C.R. and Jiang,J. (2001) Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc. Natl Acad. Sci. USA*, **98**, 5099–5103.
30. Knoop,V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.*, **46**, 123–139.
31. Small,I., Isaac,P.C. and Leaver,C.J. (1987) Stoichiometric differences in DNA molecules containing the atpA gene suggest mechanisms for the generation of mitochondrial diversity in maize. *EMBO J.*, **6**, 865–869.
32. Small,I., Suffolk,R. and Leaver,C.J. (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell*, **58**, 69–76.
33. 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.
34. Schmid,K.J., Sorensen,T.R., Stracke,R., Torjek,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.
35. Andre,C., Levy,A. and Walbot,V. (1992) Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.*, **8**, 128–132.
36. Manna,E. and Brennicke,A. (1986) Site-specific circularization at an intragenic sequence in *Oenothera* mitochondria. *Mol. Gen. Genet.*, **203**, 377–381.
37. Sato,S., Nakamura,Y., Kaneko,T., Asamizu,E. and Tabata,S. (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.*, **6**, 283–290.
38. Sakamoto,W., Kondo,H., Murata,M. and Motoyoshi,F. (1996) Altered mitochondrial gene expression in a maternal distorted leaf mutant of *Arabidopsis* induced by chloroplast mutator. *Plant Cell*, **8**, 1377–1390.
39. Martinez-Zapater,J.M., Gil,P., Capel,J. and Somerville,C.R. (1992) Mutations at the Arabidopsis CHM locus promote rearrangements of the mitochondrial genome. *Plant Cell*, **4**, 889–899.
40. Abdelnoor,R.V., Yule,R., Elo,A., Christensen,A.C., Meyer-Gauen,G. and Mackenzie,S.A. (2003) Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc. Natl Acad. Sci. USA*, **100**, 5968–5973.
41. Janska,H., Sarria,R., Woloszynska,M., Arrieta-Montiel,M. and Mackenzie,S.A. (1998) Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell*, **10**, 1163–1180.

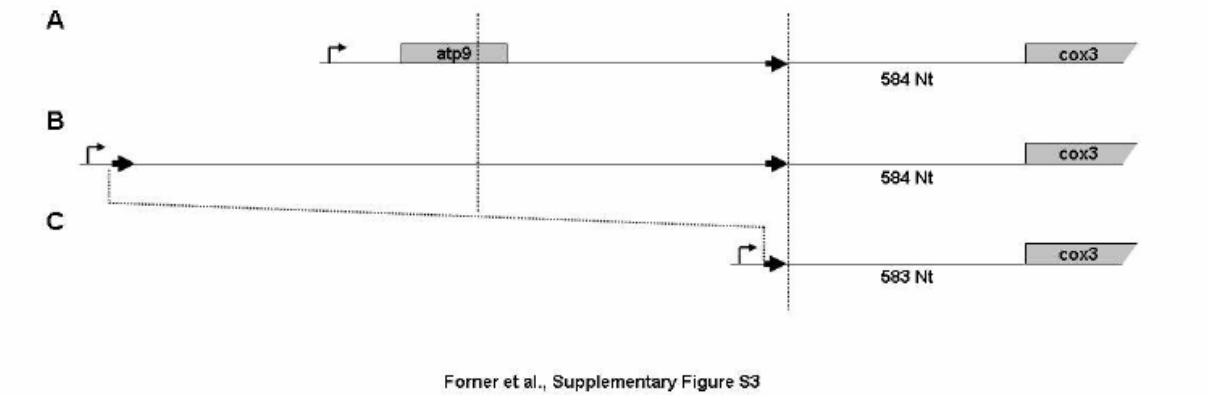
42. Mackenzie,S. and McIntosh,L. (1999) Higher plant mitochondria. *Plant Cell*, **11**, 571–586.
43. Ullrich,H., Lattig,K., Brennicke,A. and Knoop,V. (1997) Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. *Plant Mol. Biol.*, **33**, 37–45.
44. Marker,C., Zemann,A., Terhorst,T., Kiefmann,M., Kastenmayer,J.P., Green,P., Bachellerie,J.P., Brosius,J. and Huttenhofer,A. (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.



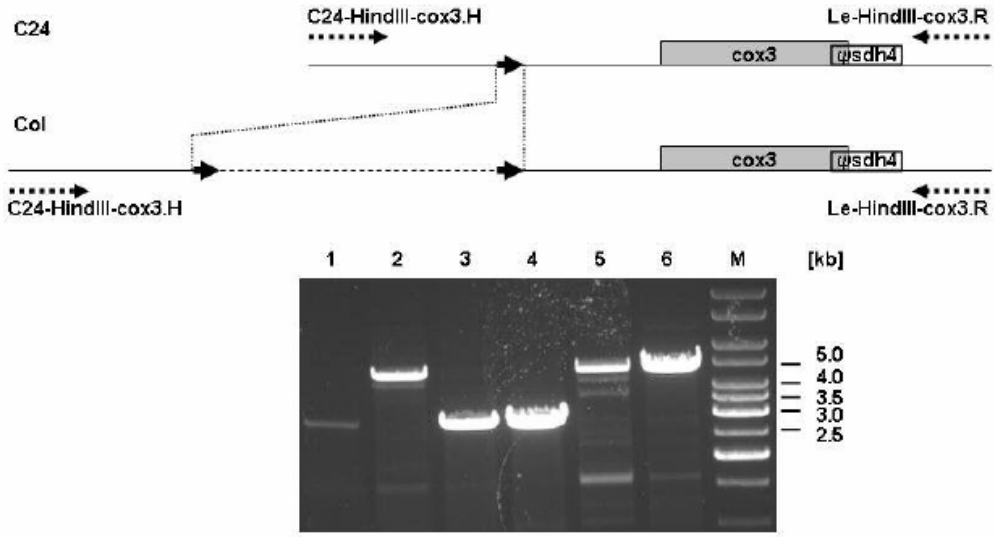
Supplementary Figure S1. PCR analysis of the border sequences of the mtDNA copy in chromosome 2. Three primer pairs T5E7.K1/T5E7.M1 (lanes 1 to 3), T5M2.M1/T5M2.K1 (lanes 4 to 5) and T5E7.K1/T5M2.K1 (lanes 7 to 9) covering both mtDNA borders and their flanking original nuclear DNA (A) were used to investigate total DNA from ecotypes Col (lanes 1, 4 and 7) C24 (lanes 2, 5 and 8) and L*er* (Lanes 3, 6 and 9). PCR products of expected sizes of 283 bp in lanes 1 to 3 and 276 bp in lanes 4 to 6 are only obtained with Col DNA. Conversely a 259 bp indicative of the absence of the mtDNA copy in chromosome 2 is not found in Col as expected. However, no such product is observed in C24 and L*er* indicating differing genomic arrangements in these ecotypes. This indicates that probably no homologous mtDNA copy exists in ecotypes C24 and L*er*, however, the sequences present surrounding the potential integration site in these ecotypes are not co-linear with the sequence in Col (without the mtDNA copy). DNA size markers are given in [kb].



Supplementary Figure S2. The 3.4 kb *Hin*dIII fragment is identical in Col and L*er*. (A) Genomic map of an At*cox*3 3.4 kb HindIII fragment isolated from a mtDNA library from Col cell suspension culture. Restriction sites are given for *Hin*dIII (H), *Cla*I (C), *Mun*I (M), *Pvu*II (P) and *Bam*HI (B). Primers Le-HindIII-cox3.H and Le-HindIII-cox3.R are given as dashed arrows, the bold arrow indicates the location of the 9 bp repeat. Fragments expected upon digestion of this fragment with *Bam*HI (B), *Mun*I/*Pvu*II (M/P) and *Cla*I (C) are indicated with sizes of expected fragments. (B) Amplification reactions with primers mentioned above were performed with total DNA isolated from green seedlings from ecotypes Col (1), C24 (2), L*er* (3), total DNA from two cell suspension culture lines from Col (5, line A; 6, line B) and the cloned fragment (lane 4). While products with the expected size of 3.4 kb are obtained with DNA from Col and L*er*, no such DNA fragment is observed in C24 (lane 2). (C) The 3.4 kb *Hin*dIII fragment was digested with *Bam*HI (lanes 1 to 5), *Mun*I/*Pvu*II (lanes 6 to 10) and ClaI (lanes 11 to 15). This generates identical restriction fragments from PCR products obtained from Col (green plant, lanes 5, 10 and 15; suspension cultures A and B lanes 3, 4, 8, 9, 14 ad 15), L*er* (green seedlings, lanes 1, 6 and 11) and from the cloned fragment (lanes 2, 7 and 11). Sizes of marker fragments are given in kilobases [kb].



Supplementary Figure S3. Structure of the 3.4 kb At*cox*3 *Hin*dIII fragment isolated from mtDNA from Col cell suspension culture (A). A complete *atp*9 gene is found 1547 nucleotides upstream of the *cox*3 reading frame (grey boxes). The same fragment is probably also present in L*er*. For comparison the *cox*3 5' genomic environments in Col (4.7 kb *Hin*dIII) and in C24 (2.9 kb *Hin*dIII) are shown (B and C). Identical insert sequences in the different ecotypes are enclosed by dotted lines. At*cox*3 upstream sequences identical between the three ecotypes comprise 584 bp. In C24 (C) a single nucleotide is missing most likely due to a sequence error in a G-stretch. The nonanucleotide direct repeat is given as bold arrows. Potential CNM-type promoters are given as bent arrows.



Forner et al., Supplementary Figure S4

Supplementary Figure S4. PCR analysis of the mtDNA configuration in C24/Col hybrid plants. The *cox*3 5' regions, which differ in C24 and Col (upper part) are used as a marker to identify the origin of the mtDNA in individual plants obtained by crossing C24 and Col and vice versa. Amplification reactions with primer C24-HindII-cox3.H, annealing 973 to 946 nucleotides upstream of the 1,790 bp insertion and primer Le-HindII-cox3.R complementary to sequence 556 to 584 nucleotides downstream of the *cox*3 stop codon generate fragments of 2948 bp on C24 mtDNA and 4748 bp on Col mtDNA. PCRs were carried out with total DNA from C24 (1), Col (2), two plants A and B of C24xCol (lanes 3 and 4) and two plants A and B of ColxC24 hybrids (♀ x ♂). Size markers (M) are given in [kb].

7.2.3 Manuscript 3: "The red fluorescent protein eqFP611: application in subcellular localization studies in higher plants"

Except for a few components of the ribosomes, all proteins involved in the mitochondrial RNA metabolism in higher plants are encoded in the nucleus and imported posttranslationally into mitochondria. Therefore, a basic step in the characterization of a protein supposed to fulfill such a function is to investigate whether it is really imported into mitochondria. This is often done by expressing green fluorescent protein (GFP) or red fluorescent protein (RFP) fused to the protein of interest. In such chimeric proteins, the GFP or RFP moiety still shows intrinsic fluorescence, while the subcellular localization of the fusion protein is determined by the non-fluorescent partner polypeptide.

However, the respective subcellular compartment has to be unequivocally identified. A convenient way to do so is to express a spectrally different second fluorescent protein with a known targeting signal in parallel. If the spatial distribution of both fluorescences is identical, the subcellular localization of the protein of interest has been determined succesfully.

The following manuscript describes the introduction of the red fluorescent protein eqFP611 as a GFP-compatible reporter gene in higher plants. Especially the establishment of a tobacco line stably expressing mitochondrially targeted eqFP611 is a preparative for the future detailed characterization of yet unidentified components of the RNA processing apparatus in plant mitochondria

The red fluorescent protein eqFP611: application in subcellular localization studies in higher plants

Joachim Forner[1] and Stefan Binder[1§]

[1]Molekulare Botanik, Universität Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany

[§]Corresponding author

Email addresses:
JF: joachim.forner@uni-ulm.de
SB: stefan.binder@uni-ulm.de

Abstract

Background

Intrinsically fluorescent proteins have revolutionized studies in molecular cell biology. The parallel application of these proteins in dual- or multilabeling experiments such as subcellular localization studies requires non-overlapping emission spectra for unambiguous detection of each label. In the red spectral range, almost exclusively DsRed and derivatives thereof are used today. To test the suitability of the red fluorescent protein eqFP611 as an alternative in higher plants, the behavior of this protein was analyzed in terms of expression, subcellular targeting and compatibility with GFP in tobacco.

Results

When expressed transiently in tobacco protoplasts, eqFP611 accumulated over night to levels easily detectable by fluorescence microscopy. The native protein was found in the nucleus and in the cytosol and no detrimental effects on cell viability were observed. When fused to N-terminal mitochondrial and peroxisomal targeting sequences, the red fluorescence was located exclusively in the corresponding organelles in transformed protoplasts. Upon co-expression with GFP in the same cells, fluorescence of both eqFP611 and GFP could be easily distinguished, demonstrating the potential of eqFP611 in dual-labeling experiments with GFP. A series of plasmids was constructed for expression of eqFP611 in plants and for simultaneous expression of this fluorescent protein together with GFP. Transgenic tobacco plants constitutively expressing mitochondrially targeted eqFP611 were generated. The red fluorescence was stably transmitted to the following generations, making these plants a convenient source for protoplasts containing an internal marker for mitochondria. To test this application as a mitochondrial marker, eqFP611 was used exemplarily to study the subcellular localization of MTH1 and MTH4, two *Arabidopsis thaliana* homologs of the human mitochondrial transcription termination factor mTERF. The identical signal patterns of mitochondrially targeted eqFP611 and MTH-GFP fusions indicate the mitochondrial localization of both MTH1 and MTH4.

Conclusions

In plants, eqFP611 is a suitable fluorescent reporter protein. The unmodified protein can be expressed to levels easily detectable by epifluorescence microscopy without adverse affect on the viability of plant cells. Its subcellular localization can be manipulated by N-terminal signal sequences. eqFP611 and GFP are fully compatible in dual-labeling experiments.

Background

Since the cloning of the green fluorescent protein (GFP) cDNA and its first heterologous expression in the early 1990s [1,2], the use of intrinsically fluorescent proteins (IFPs) has become one of the most powerful tools in molecular and cell biology. These proteins are applied as reporters in gene expression studies, as indicators of intra-cellular physiological changes, for monitoring dynamics of organelles and proteins, for investigation of protein-protein interactions *in vivo* and as fusion partners in studies of the subcellular localization of proteins [3,4].

From the very beginning, many efforts have been made to optimize various features of the native GFP with the aim to improve its application in biological research. These modifications include for instance improved folding efficiency, higher expression level or increased solubility [3]. Cyan and yellow fluorescent derivatives of GFP have been created for investigations requiring the simultaneous distinguishable tagging of more than one protein at a time [5,4]. These are used to compare the spatial distribution or the expression pattern of two or more proteins and for the analysis of protein-protein interactions by FRET. So far no red fluorescent variant of GFP has been reported. Recently, investigation of several non-bioluminescent anthozoan species has led to the isolation of various true red fluorescent proteins (RFPs) [6]. Among these, DsRed and its derivatives are the most commonly used in molecular and cell biological research [7].

Since plants contain a large number of multi-gene families, comparisons of the subcelluar localizations of the individual members are necessary as part of the comprehensive analysis of these proteins. The possibility to label several proteins with different fluorescent proteins is a great advantage when analyzing their respective subcellular localization. As a crucial prerequisite for such studies, the compartments to which the fusion proteins are targeted have to be unequivocally identified. This is often done by staining with compartment-specific dyes. Mitochondria for instance can be visualized by staining with the red fluorescent dye MitoTracker® Red CM-H2Xros (Molecular Probes, Eugene, OR) which specifically interacts with the respiratory chain. The staining procedure, however, is time-consuming, invasive and short-lived and can be replaced simply by co-expression of a spectrally different second fusion protein with a defined subcellular localization. Additionally, the fused target sequence of the fluorescent marker protein can be readily exchanged, which allows selective labeling of nearly every subcellular structure under investigation without the need to have a specific dye for the different compartments.

Despite the discovery of a mulitplicity of fluorescent proteins in the red spectral range in recent years [6], so far almost exclusively different forms of DsRed have been used for studies in molecular cell biology in plants [8-12]. These proteins are applied in dual-labeling experiments together with GFP or alone to report on promoter activity or as a marker in transgenic plants. To introduce an alternative RFP for the application in plant cells and to expand the palette of red fluorescent reporters for plant research, we tested the suitability of the red fluorescent protein eqFP611 from the sea anemone *Entacmaea quadricolor* as a marker in subcellular localization experiments in plants.

We demonstrate that native eqFP611 can be expressed in plant cells. Fusions of this protein with respective N-terminal signal sequences can be efficiently targeted to mitochondria and peroxisomes. We performed co-expression experiments with eqFP611 and GFP, created vectors for the straightforward introduction of the eqFP611 gene into plants and showed the suitability of eqFP611 as a mitochondrial marker in studies of the subcellular localization of two proteins with potential functions in mitochondria.

Results and Discussion

eqFP611 can be functionally expressed in plant cells

Recently, eqFP611, the gene for a red fluorescent protein from the sea anemone *Entacmaea quadricolor*, has been cloned and characterized [13,14]. The protein shows red fluorescence with excitation and emission maxima at 559 nm and 611 nm, respectively, an extinction coefficient of 78,000 [M^{-1} cm^{-1}] and a quantum yield of 0.45. It forms tetramers at physiological concentrations and has a maturation half-time of 4.5 h. eqFP611 has been succesfully expressed in bacteria and animal cells [13], but not yet in plants.

To test its use as a marker in plants, the native eqFP611 cDNA was cloned into a pUC19-based vector containing plant promoter and terminator sequences. This allows the transient expression of eqFP611 from the resulting plasmid peqFP611 in tobacco protoplasts. Upon inspections of cells transformed with this plasmid in the epifluorescence microscope, the red fluorescence was clearly detectable with a filter set (HQ545/30/HQ 610/75) usually used for visualization of MitoTrackerRed and here later referred to as MitoTracker filter set (Fig. 1). The protein accumulates in the nucleus and in the cytosol, where it is evenly distributed and does not form any visible aggregates, but is clearly excluded from the chloroplasts. No such fluorescence was detectable in untransformed control cells, confirming that the red fluorescence indeed originates from the expression of the introduced eqFP611. Protoplasts were analysed 16 hours after transformation. Incubation for an additional 24 hours did not markedly increase the intensity of the red fluorescence, suggesting the maturation process to be completed within 16 hours after transformation. Protoplasts expressing eqFP611 looked perfectly normal and did not show any sign of detrimental effects of this fluorescent protein.

These results show that eqFP611 can be readily used in plants, since the functional protein accumulates to detectable levels without any obvious adverse effects. In contrast to GFP, whose original jellyfish-derived cDNA was misspliced specifically in plants at a cryptic splice site [15], no modification of the eqFP611 coding sequence is necessary for efficient expression in plants.

As expected from its spectral characteristics, its fluorescence is easily detectable with a filter set (see above) that excludes the red autofluorescence of chlorophyll, a crucial advantage for an RFP applied in mesophyll cells. Similar to GFP, the native eqFP611 accumulates in the nucleus and in the cytosol in plant cells [16]. Thus, it should be suited to investigate protein targeting into e.g. mitochondria, peroxisomes and plastids within plants. In HeLa cells, native, unmodified eqFP611 was also found in the nucleus and the cytosol [13].

Targeting eqFP611 to mitochondria

To investigate whether eqFP611 can indeed be used as reporter protein for the analysis of subcellular protein sorting, import into plant mitochondria was exemplarily tested. To this end, the presequence of the mitochondrial isovaleryl-CoA-dehydrogenase (IVD) was added to the N-terminus of eqFP611 (plasmid pIVD145-eqFP611). The IVD presequence was chosen because it has previously been found to efficiently target a GFP fusion protein exclusively to mitochondria [17]. In addition, the protein has been repeatedly detected in proteomic analyses of this organelle, demonstrating its unambiguous localization in mitochondria [18-20]. Inspection of the protoplasts transformed with pIVD145-eqFP611 using the MitoTracker filter set revealed the red fluorescence to be restricted exclusively to rod-shaped structures of 1 – 2 µm in length distributed throughout the cell (Fig. 2A), a pattern characteristic of a mitochondrial localization of the fusion protein. No red fluorescence was detectable in other parts of the protoplasts. Thus, eqFP611 can be efficiently targeted to plant mitochondria, its subcellular localization being exclusively determined by the targeting information of the signal peptide fused to its N-terminus. Furthermore, this result confirms that eqFP611 is efficiently transported through two membranes while retaining its ability to fold properly for effective fluorescence. Similar to the native eqFP611, prolonged incubation of the protoplasts did not increase the intensity of the fluorescence.

The picture of the transformed protoplast displayed in Fig. 2A demonstrates nicely that the use of the MitoTracker filter set is appropriate to easily detect the red fluorescence of eqFP611 while effectively blocking chlorophyll autofluorescence. The latter is clearly visible through the FITC (fluorescein isothiocyanate) filter set (HQ 470/40/HQ 500 LP), which in turn blocks the fluorescence of eqFP611 (Fig. 2B). This autofluorescence in the chloroplasts exactly fits to the areas without fluorescence in Fig. 2A. Furthermore, the untransformed cells surrounding the eqFP611-expressing protoplast in Fig. 2A clearly show that no other autogenous fluorescence is visible through the MitoTracker filter set.

To assess the relative stability of the eqFP611 fluorescence in plants, we qualitatively compared the time elapsed until bleaching of the red fluorescence in protoplasts transiently expressing IVD145-eqFP611 and of MitoTracker® Red CM-H2Xros (Molecular Probes, Eugene, OR) used for staining of untransformed protoplasts. This latter mitochondria-specific fluorescent dye has excitation/emission maxima of 579 nm and 599 nm, respectively. When individual cells of both approaches were inspected under identical light conditions in the fluorescence microscope, the fluorescence of IVD145-eqFP611 was at least as stable as the

fluorescence of MitoTracker, which further demonstrates the usability of eqFP611 as marker at least in plant mitochondria.

Co-expression of eqFP611 and smGFP4 in tobacco protoplasts

Experiments like subcellular localization studies in which one of the fluorescent proteins is used to mark a distinct cellular compartment, require the simultaneous expression of two different fluorescent proteins. If eqFP611 is to be used routinely in such applications, its expression must be fully compatible with other IFPs, e.g. GFP. To test whether co-expression of both fluorescent proteins is indeed useful, tobacco protoplasts were simultaneously transformed with the constructs pIVD145-eqFP611 and pIVD145-smGFP4. Both plasmids contain identical mitochondrial targeting sequences fused to the N-termini of eqFP611 or smGFP4, respectively. Most of the succesfully transformed protoplasts incorporated both plasmids and expressed both eqFP611 and smGFP4. Identical patterns of the red and the green fluorescence in these protoplasts confirmed the co-expression of both proteins in the same cell (Fig. 3). In addition to the GFP-derived green fluorescence in the mitochondria, the red chlorophyll autofluorescence in the chloroplasts is seen with the FITC filter set (Fig. 3B).

To examine whether the transport into mitochondria of both fusion proteins occurs independent of each other and to exclude a possible chance "piggy back" effect during subcelluar transport of the two chimeric proteins, tobacco protoplasts were transformed with a different combination of plasmids. This time, pIVD145-smGFP4 was co-transformed with plasmid pKAT2-eqFP611, which latter encodes a recombinant protein of the peroxisomal targeting signal 2 (PTS2) [21] of 3-keto-acyl-CoA thiolase 2 (KAT2) [22] N-terminally fused to the eqFP611 reading frame. Red and green fluorescences were again found exclusively in the expected organelles (Fig. 4). The green fluorescence is observed in mitochondria, while the red fluorescence is visible in approximately 2 μm large roundish structures, a shape expected for leaf peroxisomes. No green fluorescence is seen in these organelles and conversely no red fluorescence is detected in mitochondria. This strongly suggests that if there is any a interference, it does not disturb the correct the targeting of the individual fusion proteins. Thus, eqFP611 and smGFP4 can be used in parallel to study protein sorting to different organelles within the same plant cell.

To verify that the KAT2-eqFP611 fusion protein was indeed targeted to peroxisomes, pKAT2-eqFP611 was transiently co-transformed together with p35S-N-TAP2(G)pex. The latter plasmid encodes a GFP fusion protein targeted to peroxisomal membranes by the C-terminal 36 amino acids of cotton ascorbate peroxidase (APX). As shown in Fig. 5, the

patterns of the green and the red fluorescence overlap, indicating the correct peroxisomal localization of KAT2-eqFP611. Green fluorescence seems to be more intensive at the boundaries of the peroxisomes, while the red fluorescence is equally distributed within the organelles. This is consistent with the predicted intra-peroxisomal localization of the APX and KAT2 proteins, respectively. No green or red fluorescence is visible outside the peroxisomes. These experiments demonstrate that the N-terminal peroxisomal targeting signal 2 efficiently directs eqFP611 to the corresponding organelle and that this RFP can thus be exployed to study protein sorting into peroxisomes in plants.

Thus, as observed in both mitochondria and peroxisomes, eqFP611 is a suitable partner for GFP in double-labeling experiments. When the two IFPs are co-expressed in the same cell, no mutual interference regarding development of fluorescence or intracellular sorting is observed. Additionally, their different emission spectra allow easy distinguishion between eqFP611 and GFP fluorescence. The previously reported minor green fluorescence of eqFP611 was undetectable under the conditions used (Fig. 2B and 4B) [13].

Furthermore, despite the tendency of eqFP611 to form tetramers [13], its fusion proteins can be efficiently and reliably targeted to organelles. The transport across single (peroxisomes) or double (mitochondria) membranes does not interfere with the formation of the higher order structure necessary for emitting fluorescence. In addition, the fusion of a signal sequence to its N-terminus does not interfere with the red fluorescence of eqFP611.

Expression of both eqFP611 and smGFP4 from a single plasmid

Transformation of *Nicotiana benthamiana* leaves by injection of *Agrobacterium tumefaciens* [23] containing IFP fusion genes is another fast and simple method for the analysis of the subcellular localization of a protein. This procedure is presumably closer to the *in vivo* conditions than protoplast transformation, since the transformed cells remain in the original tissue context. In addition, this approach does not require the relatively laborious preparation of protoplasts. In this case, expression of the two fusion proteins from the same plasmid is advantageous, since a single transfection event is sufficient to ensure that every transformed cell contains both IFP genes. Apart from that, expressing both fluorescent proteins from the same plasmid should generate equal amounts of RFP and GFP within a cell. The entire procedure should be easier since only a single construct has to be handled. To investigate the feasibilty of this procedure, plasmid pIVD144-eqFP611-IVD145-smGFP4 containing both the eqFP611 and the smGFP4 genes with mitochondrial presequences each under control of a CaMV 35S promoter was constructed and first tested by transformation into tobacco

protoplasts. Again, both red and green fluorescence could easily be detected in the same cell (Fig. 6). The fluorescence is found exclusively in mitochondria, the patterns of both red and green fluorescence being identical. This result is indistinguishable from the experiment with the same eqFP611 and smGFP4 expression cassettes encoded on two different plasmids (Fig. 3), but this time every transformed protoplast expressed both eqFP611 and smGFP4.

For co-expression of eqFP611 and smGFP4 in *Nicotiana benthamiana*, a binary vector suitable for plant transformation by agrobacteria was generated. The RFP-GFP-expression cassette from pIVD144-eqFP611-IVD145-smGFP4 was transferred into pBI121, creating pIVD144-eqFP611-IVD145-smGFP4-pBI121. *Agrobacterium tumefaciens* containing the latter plasmid was then injected into *N. benthamiana* leaves. After transformation, both red and green fluorescence were visible in mitochondria of epidermal cell layers (data not shown), demonstrating the convenient use of the corresponding vector in this system.

Tobacco plants stably expressing mitochondrially targeted eqFP611

A third way to use eqFP611 as a mitochondrial marker in plant cells is the generation of transgenic plants constitutively expressing mitochondrially targeted eqFP611. To create such plants, the RFP-expression cassette of pIVD145-eqFP611 was cloned into pBI121. The resulting plasmid pIVD145-eqFP611-pBI121 was stably transformed into tobacco by leaf disc transformation. Several independent plant lines were regenerated from transgenic calli and screened for bright red fluorescence in mitochondria. Red fluorescent mitochondria were observed in all T_0 transformants, but expression levels varied between individual plants. In addition, segregation was observed in the next generation. Thus, only the offspring of the most strongly fluorescent T_1 plant was used in subsequent experiments. The transgenic plants completed their life cycle like wild-type plants and the red fluorescence in mitochondria was stably transmitted up to the T_3 generation, the last generation analysed. No phenotypic differences were observed between the transgenic and wild-type plants. Thus, eqFP611 obviously causes no cytotoxic or other detrimental effects
even upon constitutive expression over several generations.

Investigation of the subcellular localization of putative mitochondrial transcription termination factors

In the experiments described so far, the application of eqFP611 as a reporter in plants had been tested with well characterized targeting sequences. To test the practical use of eqFP611 as mitochondrial marker in the analysis of the subcellular localization, two hitherto uncharacterized proteins were investigated.

These proteins belong to the familiy of homologs of the the human mitochondrial transcription termination factor mTERF [24] in *Arabidopsis thaliana*. For this species, the presence of at least 33 mTERF-like proteins has been reported [25], but no details have been given. Our own *in silico* survey using the BLAST algorithm (http://www.ncbi.nlm.nih.gov/BLAST/) [26] with mTERF as probe revealed 25 such proteins in *Arabidopsis thaliana*, which were accordingly named mTERF homologs 1 to 25 (MTH1 to MTH25). For nomenclature see Additional file 1 "Arabidopsis thaliana mTERF homologs". For the analysis of the subcellular localization, MTH1 and MTH4, encoded by the genes At1g61960 and At1g61990, respectively, were chosen. Both are predicted to be targeted to mitochondria via the amino acids at their N-termini by the TargetP (http://www.cbs.dtu.dk/services/TargetP/) [27], MITOPROT (http://ihg.gsf.de/ihg/mitoprot.html) [28] and Predotar (http://urgi.versailles.inra.fr/predotar/predotar.html) [29] algorithms.

In the first approach, wild-type tobacco protoplasts were transformed with both a plasmid encoding the N-terminal 120 amino acids of MTH4 fused N-terminally to smGFP4 (pMTH4-smGFP4) and a plasmid encoding mitochondrially targeted eqFP611 (pIVD144-eqFP611-pUC19). In this experiment, the intensity of the fluorescence of the MTH4-smGFP4 fusion protein was very weak and it was impossible to document the GFP-derived signal against the background of the autofluorescence of the chloroplasts (data not shown). To avoid the autofluorescence, the plasmids were injected into protoplasts from white *rpoB*$^-$ tobacco leaves [30]. In these plants the *rpoB* gene for an essential subunit of the plastid-encoded RNA polymerase is disrupted by insertional mutagenesis and the biosynthesis of chlorophyll is abolished. The lack of chlorophyll autofluorescence now allowed the detection of the relatively weak fluorescence of MTH4-smGFP4 in these protoplasts. Inspection of the transformed cells revealed identical patterns of the red and of the green signals (Fig. 7), suggesting that MTH4 is indeed a mitochondrial protein.

Likewise, a fusion protein of the first 120 amino acids of MTH1 and smGFP4 exhibited a weak fluorescence in tobacco protoplasts. In an alternative approach, the smGFP4 gene was replaced by smRSGFP [31] in which the substitution of a serine at position 65 by a threonine results in a higher fluorescence yield and a slight red shift. The resulting plasmid pMTH1-smRSGFP was transformed into protoplasts derived from a tobacco plant stably expressing an IVD145-eqFP611 fusion protein as mitochondrial marker (see above). The patterns of the red and green fluorescence observed in the transformed protoplasts (Fig. 8) revealed an identical distribution of the two fusion proteins which is consistent with their transport into

mitochondria. This again strongly suggests a subcellular targeting of MTH1 into mitochondria.

Thus, the 120 N-terminal amino acids of both MTH1 and MTH4 direct import of the respective fusion proteins into mitochondria, suggesting MTH1 and MTH4 to be mitochondrial proteins. In human mitochondria, mTERF mediates both initiation and termination of transcription of the mitochondrial rDNA [32]. So possibly MTH1 and MTH4 have similar functions in plant mitochondria.

The presequences of MTH1 and MTH4 show high similarity on the nucleotide level (80 %). Interestingly, in both cases the fluorescence observed during transient expression of the MTH-GFP fusion proteins was very weak. This is most likely caused by the MTH moiety of the chimeric proteins, since a bright fluorescence was seen upon expression of the IVD145-GFP fusion protein. Thus, the N-terminal part of the MTHs either reduces translation efficiency or interferes with proper folding or stability of the fusion protein

Furthermore, the succesful analysis of the subcellular localization of the previously uncharacterized proteins MTH1 and MTH4 nicely exemplifies that eqFP611 is a suitable mitochondrial marker in dual-labeling experiments in plant cells**.**

Tools for expressing eqFP611 in plants

In the course of this work, we have created several plasmid tools to express eqFP611 alone or together with smGFP4 in plants. Some of these plasmids contain fusions with mitochondrial and peroxisomal targeting sequences which can be used as background markers for the visualization of the corresponding organelles. In all vectors, the targeting sequences can be easily and (in the case of the "double vector" pIVD144-eqFP611-IVD145-smGFP4) independently exchanged to create N-terminal fusions with any protein. Furthermore, all IFP expressing cassettes in these vectors can be transferred by HindIII/EcoRI digestion into the plant transformation vector pBI121 and derivatives thereof. These can be used for *A. tumefaciens*-mediated transformation in agroinfiltration experiments with *N. benthamiana* leaves and for stable transformation. Furthermore, we created a tobacco line stably expressing easily detectable levels of mitochondrially targeted eqFP611. This line should be quite useful for laboratories working on plant mitochondria. It is ideal for experiments in which a mitochondrial marker is required like in analyses of protein subcellular localization by GFP fusions. Additionally, due to the spectral properties of eqFP611, its fluorescence can be visualized using a filter set adequate for detection of the commonly used MitoTrackerRed dye. All of these materials are freely available to the scientific community upon request

Conclusions

Our results consistently demonstrate that eqFP611 meets all requirements for a potential fluorescent reporter protein for application in plants. It can be expressed in plant cells from the unmodified cDNA sequence to levels easily detectable by epifluorescence microscopy without any adverse affect on viability. eqFP611 fluorescence can readily be separated from the red chlorophyll autofluorescence by using appropriate filter sets. Its subcellular localization can be efficiently controlled by N-terminal signal sequences. eqFP611 and GFP are fully compatible in dual-labeling experiments since there is no cross-interference with regard to expression and intra-cellular sorting and their fluorescence spectra can be clearly distinguished. Therefore, eqFP611 represents a true alternative to DsRed and can be added into the tool box of red fluorescent proteins for use in plants.

Methods

Plasmid construction/cloning strategy

The eqFP611 wild type coding sequence (696 bp) was PCR amplified from a respective cDNA clone [13] with primers eqFP611-H 5‘-cacccgggATGaactcactgatcaagg-3‘ (in which the EcoRI site at nucleotide position 4 relative to the start codon was eliminated) and eqFP611-R 5‘-tcgagctcTCAaagacgtcccagtttg-3‘. The PCR product was digested with XmaI and SacI and cloned into the respective site in the vector pIVD145-smGFP4, in which eqFP611 replaced the smGFP4 gene [17]. The resulting plasmid pIVD145-eqFP611 was used for studying mitochondrial targeting.

The plasmid peqFP611 for the expression of eqFP611 without presequence was obtained by excision of the IVD presequence from pIVD145-eqFP611 by BamHI digestion followed by religation.

To follow targeting into peroxisomes pKAT2-eqFP611 was constructed as follows: Primers KAT2-5‘-2 5‘-tctagaccATGgagaaagcgatcgag-3‘ and KAT2-3‘-2 5‘-cccgggagggtcacctacttcacttgg-3‘ were used to amplify the N-terminal part (297 bp) of the 3-keto-acyl-CoA thiolase 2 (KAT2, At2g33150) coding sequence using total oligo(dT) primed cDNA from *A. thaliana* seedlings. The PCR product was cloned using the pGEM®-T Vector System I kit (Promega), sequenced, excised with XbaI and SmaI and ligated into plasmid peqFP611. The 99 amino-acid long N-terminal part from KAT2 including the peroxisomal targeting signal 2 (from amino acids 1 to 34) is now fused in frame upstream the eqFP611 coding sequence [21,22].

To study subcellular targeting of two fusion proteins simultaneously, a plasmid carrying two genes for different fluorescent proteins fused to identical mitochondrial targeting sequences (pIVD144-eqFP611-IVD145-smGFP4) was constructed. Briefly, IVD-eqFP611 and IVD-smGFP4 fusions both under control of a CaMV 35S promoter were introduced into the same plasmid in head-to-head orientation separated by a spacer sequence. Both presequences can be exchanged separately by XhoI (eqFP611) and BamHI (smGFP4) restriction digestion, respectively. A cloning intermediate, pIVD144-eqFP611, containing a PCR-amplified eqFP611 expression cassette was used as mitochondrial reference for the analysis of the subcellular localization of the MTH4-smGFP4 fusion protein. Cloning details are available on request.

For constitutive expression of eqFP611 and GFP fusion proteins in plants, plasmids suitable for agrobacteria-mediated transformation were constructed. To generate pIVD145-eqFP611-

pBI121, the HindIII-EcoRI fragment containing the eqFP611 expression cassette was removed from plasmid pIVD145-eqFP611 by cutting with EcoRI and partial digestion with HindIII. This DNA fragment was ligated into pBI121 digested with the same enzymes, which replaces the GUS cassette in this vector.

An analogous approach was used to generate pIVD144-eqFP611-IVD145-smGFP4-pBI121 from pIVD144-eqFP611-IVD145-smGFP4 and pBI121, except that the HindIII digestion was complete.

To examine the subcellular localization of the putative mitochondrial proteins MTH4 (mTERF homolog 4, At1g61990) and MTH1 (mTERF homolog 1, At1g61960), cDNA fragments corresponding to the N-terminal 120 amino acids of each protein were amplified with primer pairs MTH4+08GFP-H 5'-gatgatcaccATGtattctctgattctcc-3'/MTH4GFP-R 5'-cgtgatcacaagtgctttcttagcatc-3' and MTH1GFP-H 5'-gatgatcaccATGtatgctctgatacac-3'/MTH1GFP-R 5'-cgtgatcacaatggatttctcagcatc-3' using cDNA clones [EMBL:AF462848] (MTH4) and [EMBL:AK117461] (MTH1) [33,34], respectively, as DNA template. The products were cloned into pGEMT from where they were excised by BclI digestion. The inserts were then ligated into pIVD145-psmGFP4 digested with BamHI where they replaced the IVD145 presequence, generating pMTH4-smGFP4 and pMTH1-smGFP4, respectively.

To get pMTH1-smRSGFP used for the investigation of the subcellular localization of MTH1, firstly psmRSGFP(BclI)-pGEMT was constructed by amplification of the smRSGFP reading frame (717 bp) from psmRSGFP [31] with smRSGFP.H 5'-cccgggATGagtaaaggagaagaac-3' and smRSGFP.R 5'-gagctc*tgaTCA*tttgtatagttcatccatgc-3' followed by ligation into pGEMT. The additional BclI site was added for easy discrimination between smGFP4 and smRSGFP. Finally, the smGFP4 reading frame in pMTH1-smGFP4 was replaced by the smRSGFP coding sequence from psmRSGFP(BclI)-pGEMT via the SmaI/SacI sites.

The vector backbone of psmGFP4 (sometimes also designated psmGFP) has been reported to be based on pUC118 and to contain the sequence *ggatcc*aaggagatataaca ATG agt (smGFP4) around the smGFP4 start codon [GenBank:U70495] [31]. Our plasmid psmGFP4 and all its derivatives deviate from the published configuration in some aspects. Sequencing of pIVD145-smGFP4 shows the sequence downstream of the CaMV 35S promoter to be *tctaga*ggatcct ATG cag…(IVD).. gat a*gg atc c*cg ccc ggg ATG agt (smGFP4). PCRs with one primer binding in the vector backbone and the other one in the CaMV 35S promoter or smGFP4 coding sequence in our psmGFP4 clearly show that the multiple cloning site is not orientated like in pUC118 and pUC18 but like in pUC119 and pUC19 (data not shown).

The absence of a 473 bp fragment in a digestion of the plasmid pIVD144-eqFP611 with RsaI (data not shown) rather indicates a pUC19-like instead of a pUC119-like configuration of the psmGFP4-derived vector-backbone.

Polymerase chain reactions

All PCRs were performed with BD Advantage™ 2 Polymerase Mix (Becton Dickinson GmbH, Heidelberg, Germany), Phusion™ High-Fidelity DNA Polymerase (BioCat GmbH, Heidelberg, Germany) or self-produced Taq polymerase, respectively. Amplifications were done in 22 to 35 cycles under conditions recommended by the manufacturer (BD Advantage2, Phusion). Reactions with self-produced Taq were done following standard protocols [35].

All PCR-derived DNA fragments were sequenced after cloning, except the expression cassette in pIVD144-eqFP611. In this case, only the IVD144 mitochondrial presequence used as reference for analysis of the subcellular localization of MTH4 was analyzed.

Transformation procedures

PEG-mediated transient transformation of protoplasts was essentially carried out as described previously [36]. For transformation of single constructs, 60 µg DNA were used. In case of simultaneous transformation of two separate plasmids, 30 µg to 60 µg of each plasmid DNA were used.

Transgenic *Nicotiana tabacum* L., cv Petit Havana plants were generated essentially as described elsewhere [37]. Expression of IVD145-eqFP611 in the T_0, T_1, T_2 and T_3 plants was followed by fluorescence microscopic analysis of parts of the lower epidermis.

Agrobacteria-mediated transformation of *N. benthamiana* by leaf infiltration was performed as described elsewhere [23].

Strain GV2260 of *Agrobacterium tumefaciens* was used for experiments requiring T-DNA transfer.

Fluorescence microscopy

A Carl Zeiss Axioplan I microscope and the axiovision software (Carl Zeiss, Oberkochen, Germany) were used for visualization and documentation of eqFP611 and GFP fluorescence. The microscope was equipped with FITC (fluorescein isothiocyanate) (HQ 470/40/HQ 500 LP) and MitoTracker (HQ545/30/HQ 610/75) filter sets obtained from AHF (Tübingen, Germany) for GFP and eqFP611 analysis, respectively.

Authors' contributions

JF designed and constructed the plasmids, carried out the microscopic analyses and drafted the manuscript.

SB conceived and supervised the project and worked over the draft version of the manuscript.

Acknowledgements

We thank Jörg Wiedenmann for providing the pQE32-based eqFP611 cDNA clone and critical reading of the manuscript, Jeff Harper for generous gift of plasmid p35S-N-TAP2(G)pex and Edyta Bocian for subcloning of the eqFP611 expression cassette from pIVD145-eqFP611 into pBI121. The authors are also grateful to Bärbel Weber for excellent technical assistance with protoplast preparation, protoplast transformation and fluorescence microscopy and Carmen Schilling-Kolle for transformation of pIVD145-eqFP611-pBI121 into tobacco as well as cultivation of the wild-type tobacco plants. This work was supported by the Deutsche Forschungsgemeinschaft, Rudolf und Clothilde Eberhardt-Stiftung and a fellowship of the Studienstiftung des deutschen Volkes to JF.

References

1. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ: **Primary structure of the Aequorea victoria green-fluorescent protein.** *Gene* 1992, 111: 229-233
2. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC: **Green fluorescent protein as a marker for gene expression.** *Science* 1994, 263: 802-805
3. Yang TT, Cheng L, Kain SR: **Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein.** *Nucleic Acids Res* 1996, 24: 4592-4593
4. Lippincott-Schwartz J, Patterson GH: **Development and use of fluorescent protein markers in living cells.** *Science* 2003, 300: 87-91
5. Tsien RY: **The green fluorescent protein.** *Annu Rev Biochem* 1998, 67: 509-544
6. Verkhusha VV, Lukyanov KA: **The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins.** *Nat Biotechnol* 2004, 22: 289-296
7. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY: **Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.** *Nat Biotechnol* 2004, 22: 1567-1572
8. Jach G, Binot E, Frings S, Luxa K, Schell J: **Use of red fluorescent protein from Discosoma sp. (dsRED) as a reporter for plant gene expression.** *Plant J* 2001, 28: 483-491
9. Dietrich C, Maiss E: **Red fluorescent protein DsRed from Discosoma sp. as a reporter protein in higher plants.** *Biotechniques* 2002, 32: 286-293
10. Goodin MM, Dietzgen RG, Schichnes D, Ruzin S, Jackson AO: **pGD vectors: versatile tools for the expression of green and red fluorescent protein fusions in agroinfiltrated plant leaves.** *Plant J* 2002, 31: 375-383
11. Mirabella R, Franken C, van der Krogt GN, Bisseling T, Geurts R: **Use of the fluorescent timer DsRED-E5 as reporter to monitor dynamics of gene activity in plants.** *Plant Physiol* 2004, 135: 1879-1887
12. Nishizawa K, Kita Y, Kitayama M, Ishimoto M: **A red fluorescent protein, DsRed2, as a visual reporter for transient expression and stable transformation in soybean.** *Plant Cell Rep* 2006 Jul 14; [Epub ahead of print]
13. Wiedenmann J, Schenk A, Röcker C, Girod A, Spindler KD, Nienhaus GU: **A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria).** *Proc Natl Acad Sci U S A* 2002, 99: 11646-11651
14. Wiedenmann J, Vallone B, Renzi F, Nienhaus K, Ivanchenko S, Rocker C, Nienhaus GU: **Red fluorescent protein eqFP611 and its genetically engineered dimeric variants.** *J Biomed Opt* 2005, 10: 14003
15. Haseloff J, Siemering KR, Prasher DC, Hodge S: **Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly.** *Proc Natl Acad Sci U S A* 1997, 94: 2122-2127
16. Köhler RH, Zipfel WR, Webb WW, Hanson MR: **The green fluorescent protein as a marker to visualize plant mitochondria in vivo.** *Plant J* 1997, 11: 613-621
17. Däschner K, Couée I, Binder S: **The mitochondrial isovaleryl-coenzyme a dehydrogenase of arabidopsis oxidizes intermediates of leucine and valine catabolism.** *Plant Physiol* 2001, 126: 601-612
18. Kruft V, Eubel H, Jansch L, Werhahn W, Braun HP: **Proteomic approach to identify novel mitochondrial proteins in Arabidopsis.** *Plant Physiol* 2001, 127: 1694-1710

19. Millar AH, Sweetlove LJ, Giege P, Leaver CJ: **Analysis of the Arabidopsis mitochondrial proteome.** *Plant Physiol* 2001, 127: 1711-1727
20. Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH: **Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins.** *Plant Cell* 2004, 16: 241-256
21. Kato A, Hayashi M, Kondo M, Nishimura M: **Transport of peroxisomal proteins synthesized as large precursors in plants.** *Cell Biochem Biophys* 2000, 32: 269-275
22. Germain V, Rylott EL, Larson TR, Sherson SM, Bechtold N, Carde JP, Bryce JH, Graham IA, Smith SM: **Requirement for 3-ketoacyl-CoA thiolase-2 in peroxisome development, fatty acid beta-oxidation and breakdown of triacylglycerol in lipid bodies of Arabidopsis seedlings.** *Plant J* 2001, 28: 1-12
23. Voinnet O, Rivas S, Mestre P, Baulcombe D: **An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus.** *Plant J* 2003, 33: 949-956
24. Fernandez-Silva P, Martinez-Azorin F, Micol V, Attardi G: **The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions.** *EMBO J* 1997, 16: 1066-1079
25. Linder T, Park CB, Asin-Cayuela J, Pellegrini M, Larsson NG, Falkenberg M, Samuelsson T, Gustafsson CM: **A family of putative transcription termination factors shared amongst metazoans and plants.** *Curr Genet* 2005, 48: 265-269
26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, 25: 3389-3402
27. Emanuelsson O, Nielsen H, Brunak S, von Heijne, G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, 300: 1005-1016
28. Claros MG, Vincens P: **Computational method to predict mitochondrially imported proteins and their targeting sequences.** *Eur. J. Biochem* 1996, 241: 779-786
29. Small I, Peeters N, Legeai F, Lurin C: **Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences.** *Proteomics* 2004, 4: 1581-1590
30. Krause K, Maier RM, Kofer W, Krupinska K, Herrmann RG: **Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome.** *Mol Gen Genet* 2000, 263: 1022-1030
31. Davis SJ, Vierstra RD: **Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants.** *Plant Mol Biol* 1998, 36: 521-528
32. Martin M, Cho J, Cesare AJ, Griffith JD, Attardi G: **Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis.** *Cell* 2005, 123: 1227-1240
33. Seki M, Carninci P, Nishiyama Y, Hayashizaki Y, Shinozaki K: **High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper.** *Plant J* 1998, 15: 707-720
34. Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K: **Functional annotation of a full-length Arabidopsis cDNA collection.** *Science* 2002, 296: 141-145

35. Sambrook J, Russel DW: *Molecular Cloning: A Laboratory Manual, Ed 3*. Cold Spring Harbor, Cold Spring Harbor Laboratory Press 2001
36. Koop HU, Steinmüller K, Wagner H, Rossler C, Eibl C, Sacher L: **Integration of foreign sequences into the tobacco plastome via polyethylene glycol-mediated protoplast transformation.** *Planta* 1999, 199: 193–201
37. Horsch R, Fry J, Hoffman N, Wallroth M, Eichholtz D, Rogers S, Fraley R: **A simple and general method for transferring genes into plants.** *Science* 1985, 227: 1229-1232

Figures

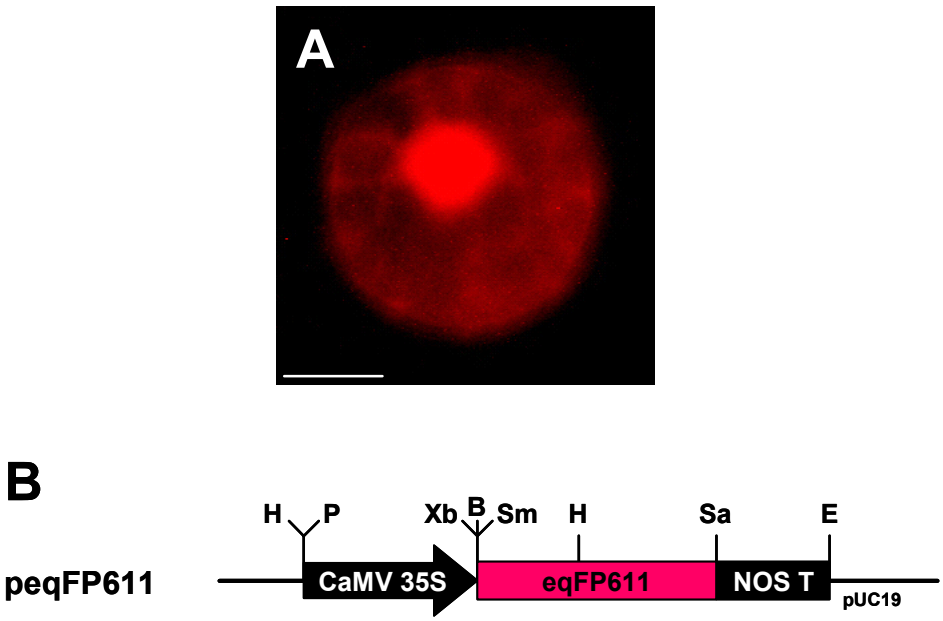


Figure 1 - eqFP611 without presequence.

Transient expression of original eqFP611 without presequence in *N. tabacum* wild type protoplasts.

(A) Image taken through MitoTracker filter set. Scale bar: 10 µm.

(B) Plasmid peqFP611 used for transformation. Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; red box, eqFP611: eqFP611 coding sequence; black box, NOS T: nopaline synthase terminator. H: HindIII, P: PstI, Xb: XbaI, B: BamHI, Sm: SmaI, Sa: SacI, E: EcoRI restriction sites.

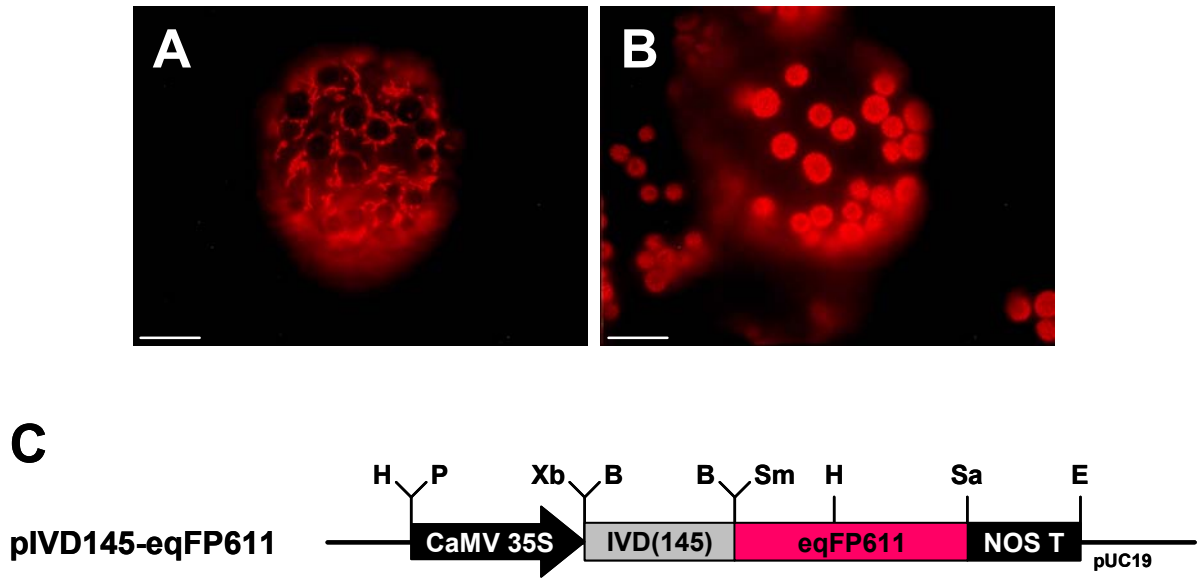


Figure 2 - Mitochondrially targeted eqFP611.

N. tabacum wild type protoplasts expressing a fusion protein of eqFP611 and the N-terminal 48 amino acids of IVD.

Pictures showing the same cell were taken through (A) MitoTracker and (B) FITC filter sets, respectively. Scale bars: 10 µm.

(C) Map of plasmid pIVD145-eqFP611 used for transformation. Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey box, IVD(145): N-terminal 145 nucleotides of the IVD coding sequence; red box, eqFP611: eqFP611 coding sequence; black box, NOS T: nopaline synthase terminator. H: HindIII, P: PstI, Xb: XbaI, B: BamHI, Sm: SmaI, Sa: SacI, E: EcoRI restriction sites.

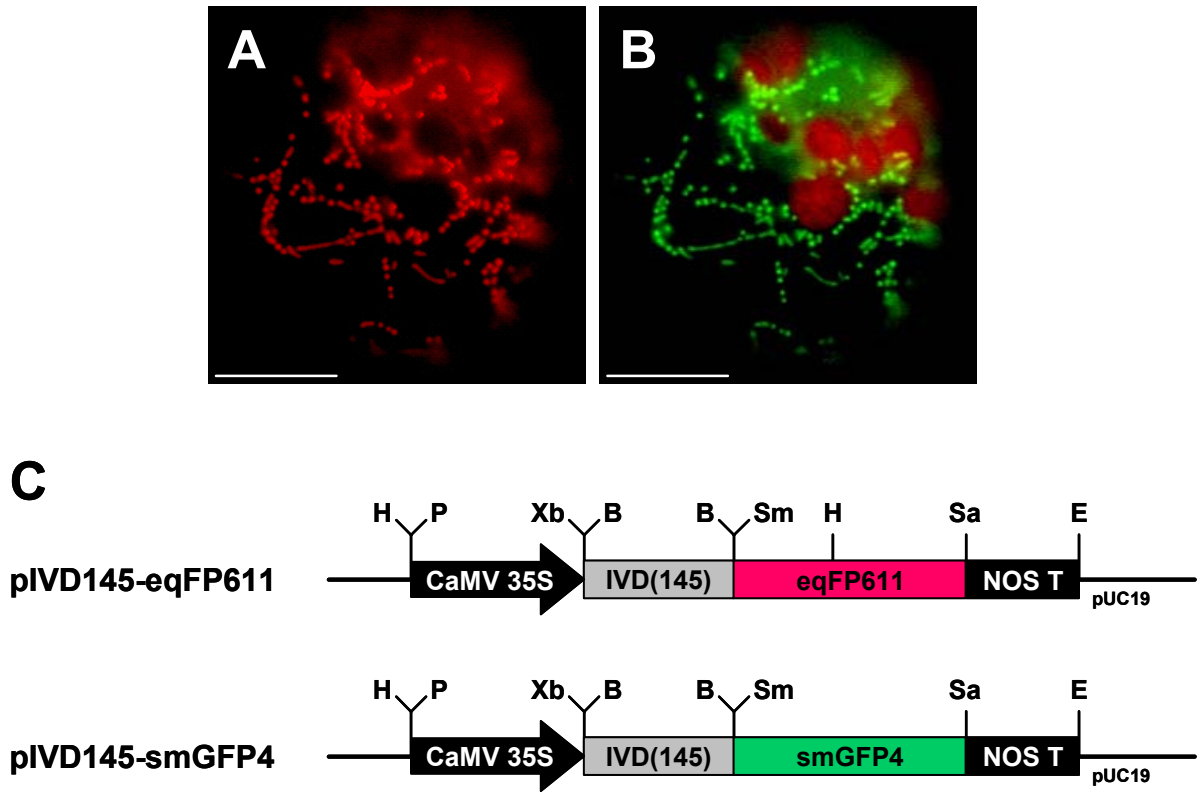


Figure 3 - Co-expression of eqFP611 and smGFP4 fusion proteins targeted to mitochondria.

Tobacco protoplasts transiently transformed with plasmids pIVD145-eqFP611 and pIVD145-smGFP4. The eqFP611 and smGFP4 fusion proteins contain the mitochondrial presequence corresponding to the N-terminal 48 amino acids of IVD.

Transformed protoplast seen through (A) MitoTracker and (B) FITC filter sets, respectively. Scale bars: 10 µm.

(C) Map of plasmids pIVD145-eqFP611 and pIVD145-smGFP4. Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey box, IVD(145): N-terminal 145 nucleotides of the IVD coding sequence; red box, eqFP611: eqFP611 coding sequence; green box, smGFP4: smGFP4 coding sequence; black box, NOS T: nopaline synthase terminator. H: HindIII, P: PstI, Xb: XbaI, B: BamHI, Sm: SmaI, Sa: SacI, E: EcoRI restriction sites.

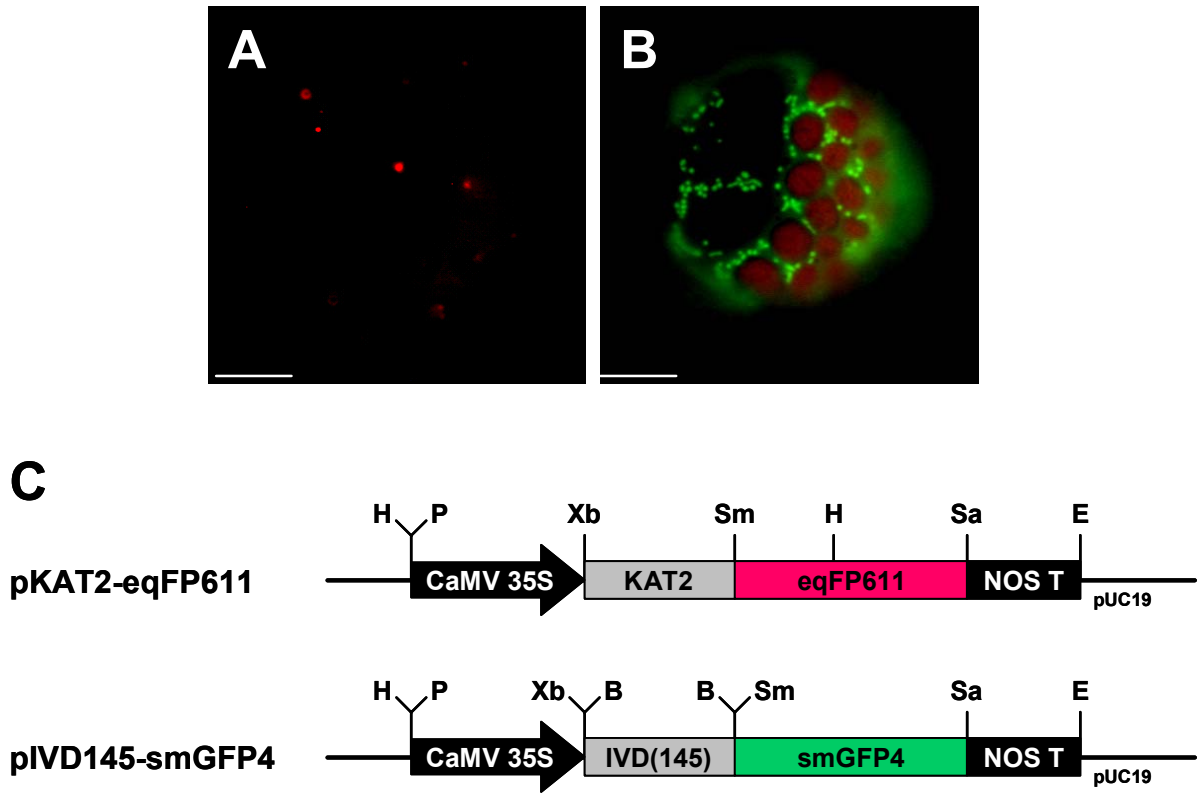


Figure 4 - Co-Expression of peroxisomally targeted eqFP611 and mitochondrially targeted smGFP4.

Transient co-transformation of 2 separate plasmids encoding eqFP611 with a peroxisomal targeting signal 2 (pKAT2-eqFP611) and smGFP4 with a mitochondrial presequence (pIVD145-smGFP4) into *N. tabacum* wild type protoplasts.

View of a doubly transformed cell through (A) MitoTracker and (B) FITC filter sets, respectively. Scale bars: 10 µm.

(C) Plasmids pKAT2-eqFP611 and pIVD145-eqFP611 used for transformation. Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey box, KAT2: N-terminal 297 nucleotides of the KAT2 coding sequence; grey box, IVD(145): N-terminal 145 nucleotides of the IVD coding sequence; red box, eqFP611: eqFP611 coding sequence; green box, smGFP4: smGFP4 coding sequence; black box, NOS T: nopaline synthase terminator. H: HindIII, P: PstI, Xb: XbaI, Sm: SmaI, Sa: SacI, E: EcoRI, B: BamHI restriction sites.

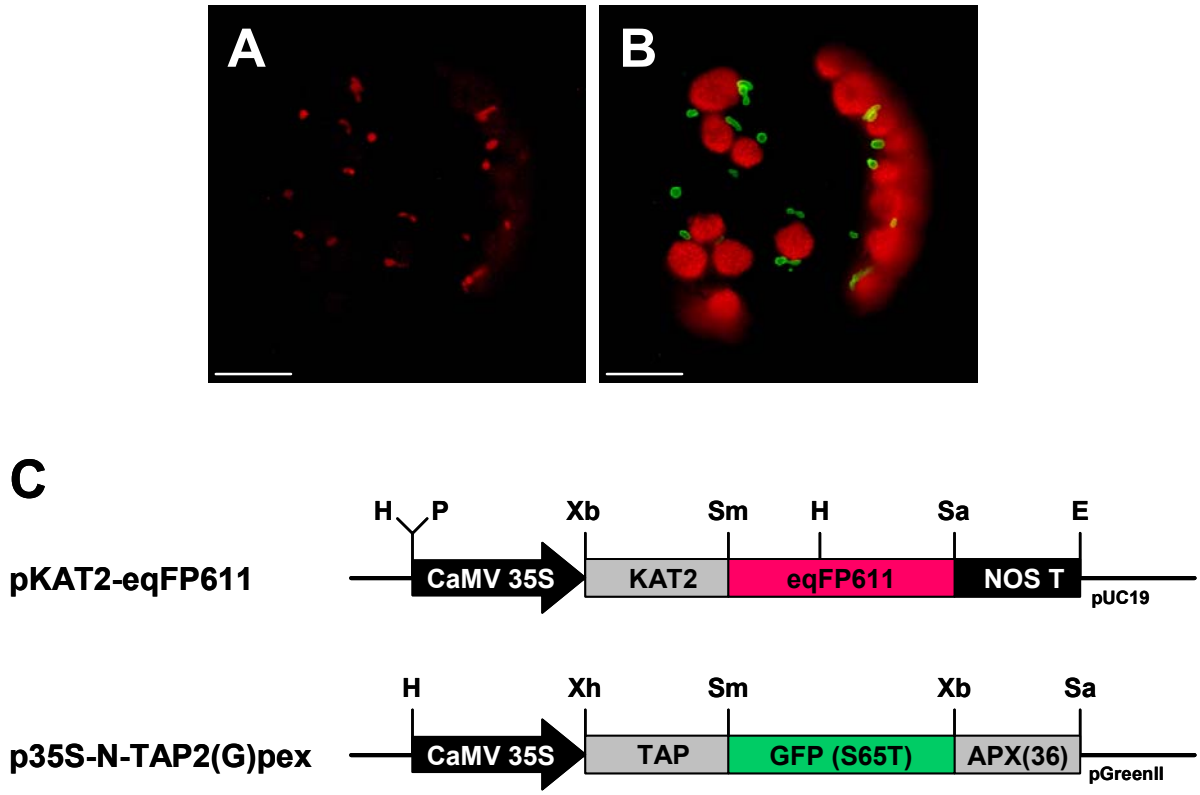


Figure 5 - Co-transformation of tobacco protoplasts with plasmids encoding eqFP611 and GFP targeted to peroxisomes.

Transient transformation of 2 separate plasmids encoding eqFP611 with a peroxisomal targeting signal 2 (pKAT2-eqFP611) and GFP targeted to the peroxisomal membrane (p35S-N-TAP2(G)pex) into the same *N. tabacum* wild type protoplasts.

Pictures of the same protoplast taken through (A) MitoTracker and (B) FITC filter sets, respectively. Scale bars: 10 µM

(C) Plasmid maps.

Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey box, KAT2: N-terminal 297 nucleotides of the KAT2 coding sequence; grey box, TAP: chimeric sequence for tandem affinity purification; red box, eqFP611: eqFP611 coding sequence; green box, GFP(S65T): GFP coding sequence including the S65Tmodification; grey box, APX: sequence encoding the C-terminal 36 amino acids of cotton ascorbate peroxidase; black box, NOS T: nopaline synthase terminator.

H: HindIII, P: PstI, Xb: XbaI, Sm: SmaI, Sa: SacI, E: EcoRI, Xh: XhoI restrictions sites.

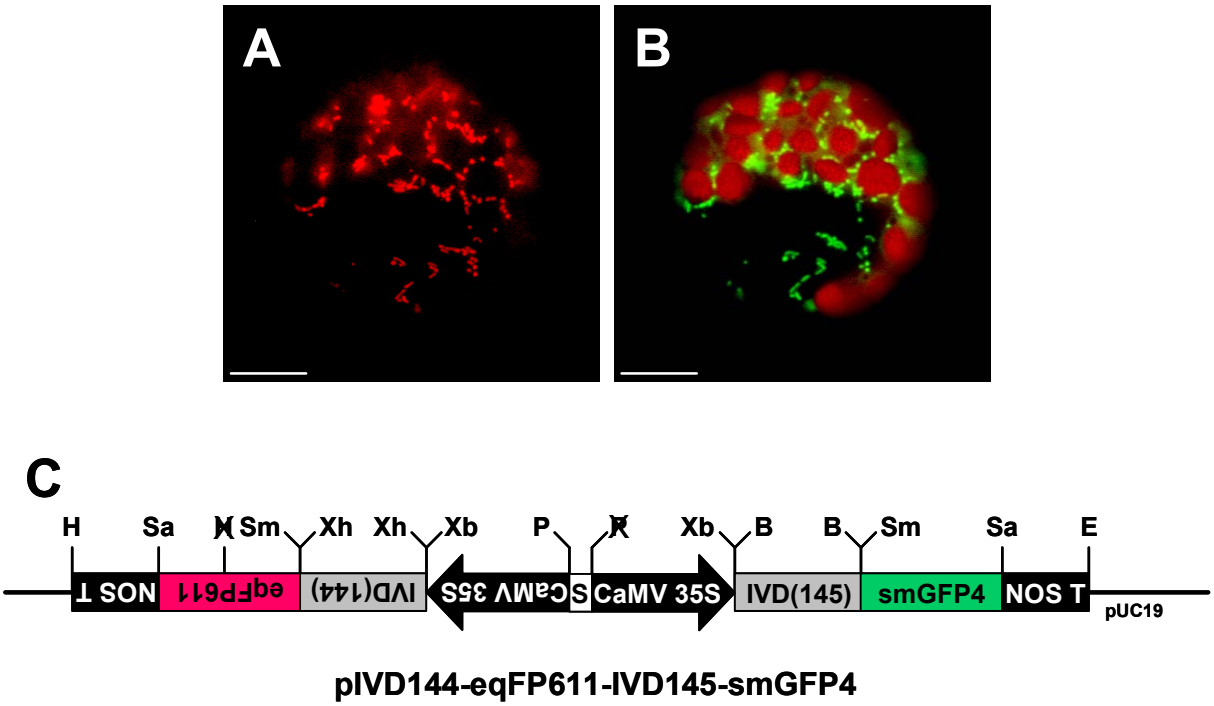


Figure 6 - Mitochondrially targeted eqFP611 and smGFP4 expressed from the same plasmid

Transient transformation of 1 single plasmid encoding both eqFP611 and smGFP4 with mitochondrial presequences (pIVD144-eqFP611-IVD145-smGFP4) into *N. tabacum* wild type protoplasts.

Pictures of the same cell, taken through either (A) MitoTracker or (B) FITC filter sets, respectively. Scale bars: 10 µm.

(C) Plasmid used for transformation.

Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey boxes, IVD(144)/(145): N-terminal 145 and 144 nucleotides of the IVD coding sequence, respectively; red box, eqFP611: eqFP611 coding sequence; green box, smGFP4: smGFP4 coding sequence; black box, NOS T: nopaline synthase terminator; white box, S: spacer sequence.

H: HindIII, Sa: SacI, Sm: SmaI, Xh: XhoI, Xb: XbaI, P: PstI, B: BamHI, E: EcoRI restriction sites; X: restriction site eliminated by PCR with modified primers (HindIII) or ligation of DNA ends derived from PstI and Mph1103I digestion (PstI).

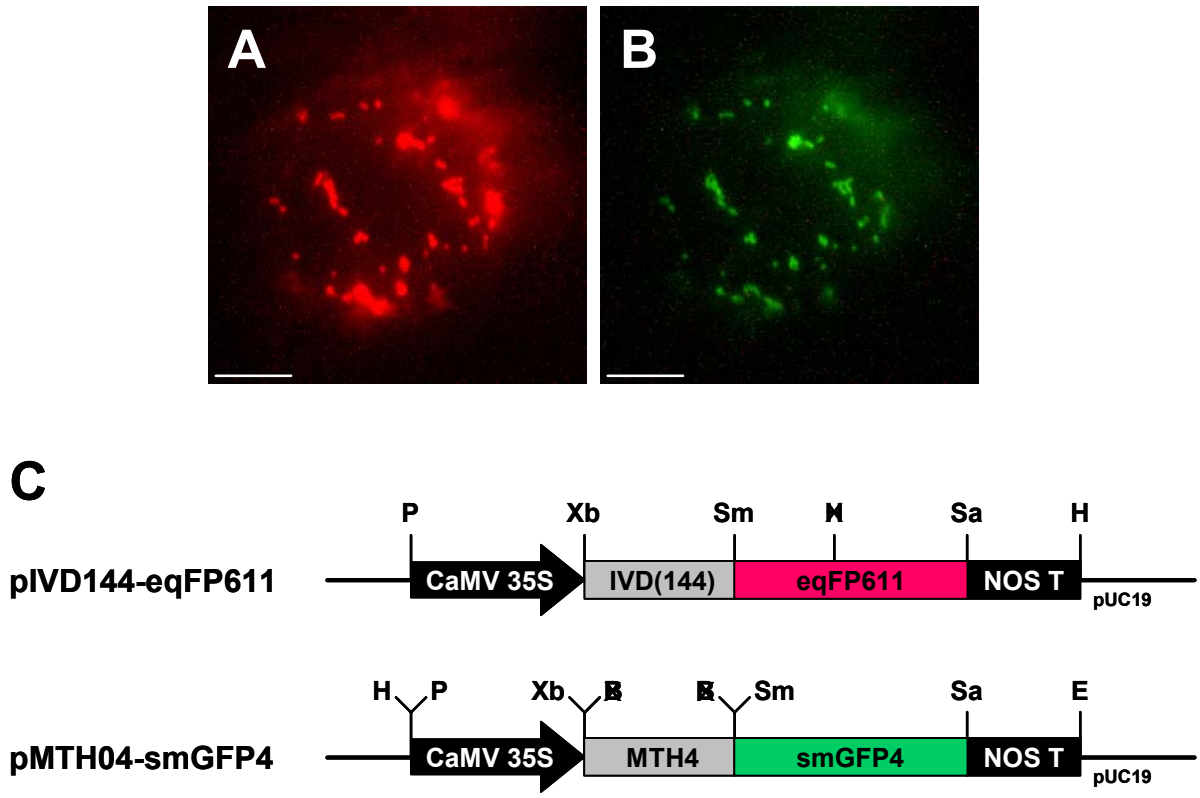


Figure 7 - Co-transformation of mitochondrially targeted eqFP611 and smGFP4 with MTH4 presequence.

Transient transformation of 2 separate plasmids encoding eqFP611 with a mitochondrial presequence (pIVD144-eqFP611) and smGFP4 with the MTH4 presequence (pMTH4-smGFP4) into the same *rpoB*$^-$ albino *N. tabacum* protoplasts [30].

Pictures of the same cell, taken through (A) MitoTracker and (B) FITC filter sets, respectively. Scale bar: 10 µm.

(C) Maps of the respective plasmids.

Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey box, MTH4: N-terminal 361 nucleotiedes of mTERF homolog 4 (MTH4); grey box, IVD(144): N-terminal 144 nucleotides of the IVD coding sequence; red box, eqFP611: eqFP611 coding sequence; green box, smGFP4: smGFP4 coding sequence; black box, NOS T: nopaline synthase terminator.

P: PstI, Xb: XbaI, Sm: SmaI, Sa: SacI, H: HindIII: BamHI, E: EcoRI restriction sites. X: restriction site eliminated by ligation of DNA ends derived from BamHI and BclI digestion or by PCR with modified primers (HindIII).

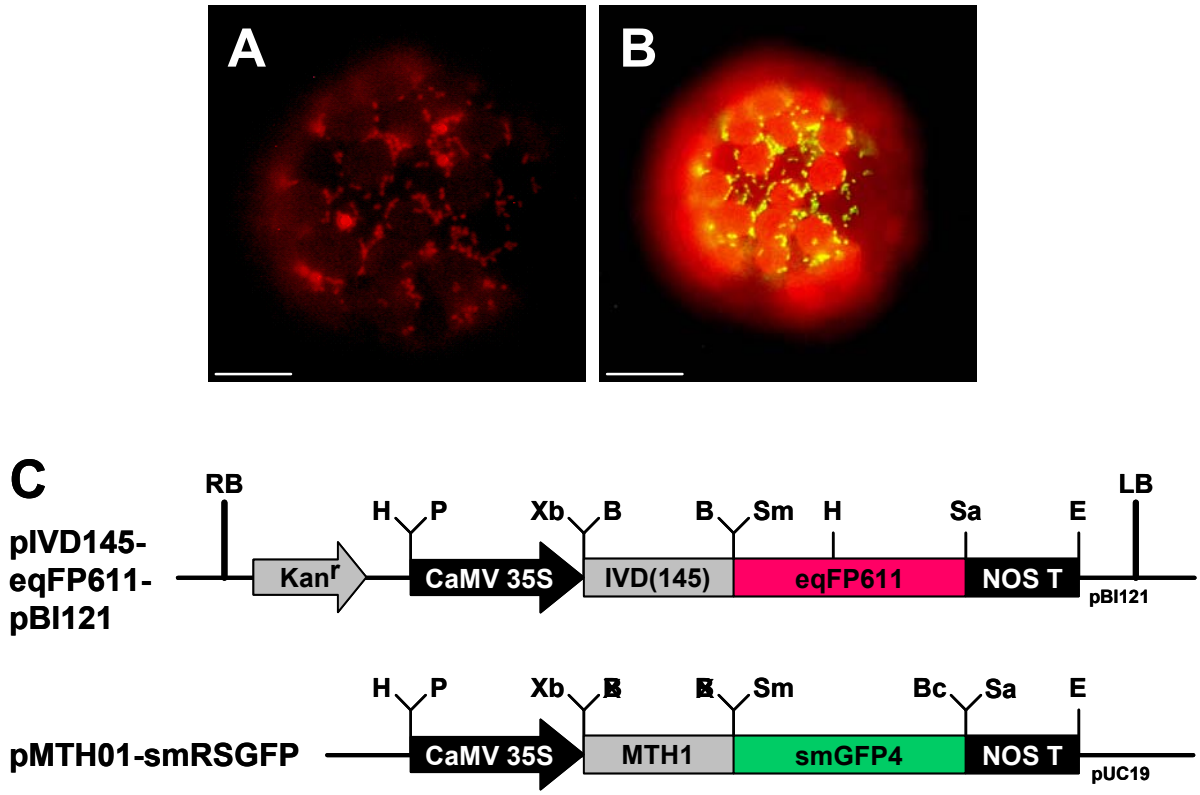


Figure 8 - Transient expression of MTH1-smRSGFP in protoplasts constitutively expressing mitochondrially targeted eqFP611.

Transient expression of a plasmid encoding smRSGFP with the MTH1 presequence (pMTH1-smRSGFP) in protoplasts derived from stably transformed *N. tabacum* plants constitutively expressing mitochondrially targeted eqFP611 (pIVD145-eqFP611-pBI121).

A transgenic protoplast additionally transformed with the MTH1-smGFP4 construct. Images taken through (A) MitoTracker and (B) FITC filter sets, respectively. Scale bar: 10 µm.

(C) Delineation of the plasmids used for transformation.

Black arrow, CaMV 35S: cauliflower mosaic virus 35S promoter; grey box, MTH1: N-terminal 361 nucleotiedes of mTERF homolog 1 (MTH1); grey box, IVD(145): N-terminal 145 nucleotides of the IVD coding sequence; red box, eqFP611: eqFP611 coding sequence; green box, smRSGFP: smRSGFP coding sequence; black box, NOS T: nopaline synthase terminator; grey arrow, Kan^r: kanamycin resistance cassette (NOS promoter, neomycin phosphotransferase II, NOS terminator). RB: right border, LB: left border. H: HindIII, P: PstI, Xb: XbaI, B: BamHI, Sm: SmaI, Sa: SacI, E: EcoRI, Xh: XhoI, Bc: BclI restriction sites. X: restriction site eliminated by ligation of DNA ends derived from BamHI and BclI digestion.s.

Additional files

Additional file 1 –Arabidopsis thaliana mTERF homologs

This file contains a list of 25 *Arabidopsis thaliana* mTERF homologs found *in silico* and an assignment of names.

Arabidopsis thaliana MTERF HOMOLOGS (AtMTHs)

Gene name	**Locus**	
AtMTH1	At1g61960	*
AtMTH2	At1g61970	*
AtMTH3	At1g61980	*
AtMTH4	At1g61990	*
AtMTH5	At1g62010	*
AtMTH6	At1g62085	*
AtMTH7	At1g62110	*
AtMTH8	At1g62120	*
AtMTH9	At1g62150	*
AtMTH10	At1g79220	
AtMTH11	At2g03050	
AtMTH12	At2g34620	
AtMTH13	At5g07900	
AtMTH14	At1g74120	
AtMTH15	At2g44020	
AtMTH16	At3g46950	*
AtMTH17	At3g60400	
AtMTH18	At4g02990	
AtMTH19	At4g14605	
AtMTH20	At5g23930	*
AtMTH21	At5g64950	
AtMTH22	At1g56380	*
AtMTH23	At1g62490	*
AtMTH24	At1g21150	
AtMTH25	At3g58480	

* Genes marked with an asterisk show high similarity to each other on the nucleotide level (≥ 75 %).

7.3 Acknowledgements

Firstly I'd like to thank Prof. Dr. Stefan Binder for the opportunity to work in his lab and continuous well-grounded and constructive advice during the experimental and writing phase of this thesis.

Special thanks to Bärbel Weber for excellent technical assistance and many years of smooth and succesful teamwork.

I'm also much obliged to all the others who have worked in the transcriptome project group, namely Steffen Wildum, Caterina Wiethölter, Markus Schwarzländer, Sabine Thuß, Christian Jonietz, Angela Hölzle and all the countless practical students, as without their dedicated collaboration this project could not have been succesfully completed in the allocated time.

I am grateful to all present and former members of the Binder lab, namely Joachim Schuster, Stephanie Schmidt-Gattung, Edyta Bocian, Michaela Hoffmann, Oliver Schmid, Philipp Gyaja, Joachim Frank, Annemarie Matthes, Daniela Thornton, Cornelia Guha, Tanja Knill, Daniela Köhler, Carolin Müller and Ulrike Tengler for the pleaseant working atmosphere, numerous courtesies with practical problems in the daily laboratory work and enlightening discussions.

I also thank the whole institute for molecular botany for good working conditions and innumerable little supportings.

Many thanks to Prof. Dr. Axel Brennicke and Prof. Dr. Hans-Peter Braun who agreed to be the second and third referees for this thesis and to all members of the Promotionskommission for their attendance to take over this task.

I'd also like to express my thankfulness to Rhonda Meyer and Otto Törjek in Golm for the collaboration with ecotypes, hybrids and mapping.

I'm as well appreciative of the scholarship granted to me by the Studienstiftung des deutschen Volkes to secure my subsistence during the work on this thesis.

And last but not least I'd like to say „Thank you" to all those who are not mentioned here by name but who nevertheless helped in any way to ensure and enable the succesful preparation of this thesis.

7.4 Curriculum vitae

ENTFÄLLT

7.5 List of publications

Original publications:

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

Kempken F, Bolle N, **Forner J** and Binder S (2007) *Transcript End Mapping and Analysis of RNA Editing in Plant Mitochondria.* In: Leister D and Herrmann J (eds) Methods in Molecular Biology, vol. 372: Mitochondrial Genomics and Proteomics Protocols. Humana Press Inc., Totowa, NJ, USA. In press

Forner J and Binder S (2007) *The red fluorescent protein eqFP611: application in subcellular localization studies in higher plants.* Under review

Forner J, Weber B, Thuß S, Wildum S and Binder S (2007) *Mapping of mitochondrial mRNA termini in Arabidopsis mitochondria: t-elements contribute to 5' and 3' end formation.* Accepted for publication in Nucleic Acids Research

Oral presentations:

Forner, J (2003) *Transcriptomics in Arabidopsis thaliana mitochondria.* MitEuro D57 course "Manipulation of mitochondrial gene expression in fungi", Orsay, France

Forner, J (2004) *Towards a Complete Transcript Map in Mitochondria of Arabidopsis thaliana.* MitEuro WP4&5 Workshop, Rom, Italy

Posters:

Forner J, Weber B, Wiethölter C and Binder S (2004) *Towards a complete transcript map in Arabidopsis thaliana mitochondria.* Poster, T09-016. 15th International Conference on Arabidopsis research, Berlin, Germany

Forner J, Weber B, Wiethölter C and Binder S (2005) *Characterization of cox3 and other mitochondrial transcripts in Arabidopsis thaliana.* Poster, P28. International Congress on Plant Mitochondrial Biology, Obernai, France

Forner J and Binder S (2006) *Analysis of mitochondrial mRNA ends in different ecotypes of Arabidopsis thaliana.* Poster, #2. 19. Tagung Molekularbiologie der Pflanzen, Dabringhausen, Germany

7.6 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 oder inhaltlich übernommene Stellen als solche erkenntlich gemacht habe.

Ulm, den