Thèse présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur Strasbourg I

En biologie moléculaire et cellulaire
à la Faculté des sciences de la vie
par Sarah HOLEC

Polyadenylation and RNA degradation: from mitochondria to the nucleus of *Arabidopsis thaliana*

Polyadénylation et dégradation de l'ARN : des mitochondries au noyau d'Arabidopsis thaliana

soutenue publiquement le 7 mars 2008 devant la commission d'examen :

Directeur de thèse : Dr. Dominique Gagliardi, IBMP (Strasbourg, France)
Rapporteur interne : Pr. Mario Keller, IBMP (Strasbourg, France)
Rapporteur externe : Pr. Gadi Schuster, Technion (Haifa, Israel)
Rapporteur externe : Pr. Stefan Binder, Universität Ulm (Ulm, Allemagne)
Examinateur : Dr. Cécile Bousquet-Antonelli, LGDP (Perpignan, France)

Institut de Biologie Moléculaire des Plantes
(CNRS UPR2357)

First of all, I would like to express all my gratitude to the members of the jury who accepted to read and judge my work : Pr. Gadi Schuster, Pr. Stefan Binder, Dr. Cécile Bousquet-Antonelli and Pr. Mario Keller.

Un merci tout particulier à Mario Keller qui m'a donné goût à la biologie moléculaire et qui, le premier, m'a introduit dans le bâtiment de l'IBMP, lorsque je n'imaginais pas encore que cet institut deviendrait ma deuxième maison pendant 5 ans!

Dans l'ordre chronologique, j'aimerais ensuite adresser mes plus chaleureux remerciements à Marie-Edith Chabouté pour m'avoir accueillie dans son équipe lors de mon DEA, ainsi qu'à Guy Houlné, mon premier Papa de sciences, si sage. J'en profite pour dire toute mon affection et ma reconnaissance pour les bons moments passés aux membres du laboratoire 512 que j'ai côtoyé : Fred bien sûr, Valérie, Hélène, Roxanna, Julien, Ondrej, Lénine, Chamseddine, ainsi qu'aux voisins du 509 : Anne-Catherine, Etienne, Jean-Luc, Natacha, Virginie : ne changez rien! Jean-Luc, je t'adresse tout particulièrement ma gratitude pour toutes les fois où tu m'as sauvée (de la crise informatique), avec une patience et une gentillesse sans failles.

Je ne pourrais jamais assez remercier Gag de tout ce qu'il a fait pour moi. Mais tout de même : j'apprécie la chance énorme que j'ai eu d'avoir été accueillie en thèse par toi dans le laboratoire 323 (faire confiance à une blonde comme moi, comment a-tu fais?). Je te remercie pour la patience infinie que tu as eu avec moi, pour la rigueur que tu m'as enseignée (ch'u pas maniaque, mais...), pour les conversations interminables quand je doutais de moi, pour ta bonne humeur et ton humour. Merci aussi aux autres membres de l'équipe : Jean pour tes conversations qui me font voyager dans l'espace et le temps, et Heike, ma colocataire de box. Heike, I want to deeply thank you for everything you did for me. I am particularly grateful for the reading and correction of the decorated-with-flowers first draft of my thesis, but also for all the conversations we had (especially the non-scientific ones), for your big heart and your delicious meals. Je voudrais aussi dire un gros merci aux autres membres du laboratoire qui ont été des voisins de paillasse adorables : Patrice, Monique et José, ainsi que notre petite Brésilienne Carolina. Sans oublier tous les membres de la mito team of Jean-Michel Grienenberger qui savent créer une ambiance de travail agréable. Je remercie particulièrement les voisines que je suis souvent allé embêtées et qui ne m'ont jamais renvoyé : Laurence, Anne-Marie, Thalia (grande soeur de thèse), Claire (je demanderais à Brad s'il peut apporter des fraises Tagada pour ton pot de thèse), Samira ainsi que Romaric et François. Un salut spécial aussi aux anciens de Winless et compagnie qui ont fait des pauses de midi des moments délirants (Pierre, Clarisse, Alban, Ben, petit Roro...).

Je voudrais aussi remercier tous les membres de l'IBMP qui font de ce lieu un endroit auquel je suis attachée, ainsi que tous mes amis qui m'ont soutenue et supportée pendant ces années, Manu, Gabrielle, Claire, Matthieu, Isabelle, Jeanne, Marie, Stéphane (un énorme merci pour ta logistique infaillible qui a contribué sans aucun doute à la réussite de ce travail, et pour ton soutien).

Laurent, je te remercie infiniment pour toute l'aide que tu m'a apportée dans la rédaction de ce manuscrit mais aussi pour ton aide technique et tes idées scientifiques. Merci aussi pour tous ces derniers mois, pour ta patience et ton soutien, tout aurait été plus difficile sans toi.

Enfin, je voudrais exprimer toute ma reconnaissance et mon affection à mes parents, qui m'ont supportée et soutenue tout au long de mon parcours, et qui m'ont permis d'écrire ces lignes aujourd'hui.

Résumé de la thèse en français

Introduction

Les mitochondries sont des organelles vitales pour la plupart des eucaryotes, impliquées dans la production d'énergie et dans de nombreuses voies métaboliques. Elles proviennent d'un unique événement d'endosymbiose d'un ancêtre des α-protéobactéries par une cellule hôte hétérotrophe. Malgré cette origine monophylogénétique, les mitochondries ont évolué de manière très divergente d'un organisme à l'autre. Cette divergence est particulièrement remarquable dans l'organisation du génome et dans la régulation de son expression.

La différence de taille est frappante : alors que celles des génomes mitochondriaux des vertébrés vont de 15 à 20kb et que celle du génome de la levure *Saccharomyces cerevisiae* est de 76kb, la taille des génomes mitochondriaux des plantes varie de 180kb à 2400kb. Cette augmentation de taille n'est pas corrélée à un accroissement de la capacité codante, mais s'explique par l'accumulation d'introns et de larges régions intergéniques, absentes chez les animaux. En effet, la densité des gènes est très faible chez les génomes mitochondriaux de plante, avec en moyenne un gène pour 8kpb. La grande taille de ces génomes est aussi la conséquence de duplications fréquentes.

Le génome mitochondrial humain est une molécule circulaire double brin, très compacte, avec une seule petite région non codante et sans introns. Au contraire, chez les plantes, le "cercle maître" est plutôt un modèle théorique. De longues séquences répétées permettent des recombinaisons donnant lieu à de nombreuses molécules sub-génomiques circulaires, branchées, ou linéaires. Les gènes sont éparpillés sur ces molécules, avec de grandes régions intergéniques. De plus, les génomes mitochondriaux de plante contiennent des insertions d'ADN étranger viral, nucléaire, chloroplastique ou d'origine inconnue.

En conséquence de cette diversité dans l'organisation des génomes, les mécanismes contrôlant l'expression du génome diffèrent remarquablement. Chez l'homme, la transcription est très simple : chacun des deux brins d'ADN est transcrit à partir d'une seule région promotrice, donnant lieu à deux transcrits polycistroniques, qui sont ensuite clivés de part et d'autre des ARN de transfert pour libérer les ARN ribosomiques et messagers. Chez la plante, l'analyse des unités transcriptionnelles est complexe du fait de l'existence de promoteurs multiples ou alternatifs, de nombreux sites de maturation des transcrits précurseurs, de cas de trans-épissage et de l'absence de sites connus de terminaison de la transcription. Les ARN polymérases mitochondriales, de type phagique, sont capables de reconnaître les promoteurs, d'initier la transcription et de procéder à l'élongation des transcrits par elle-mêmes, mais elles requièrent des facteurs spécifiques pour reconnaître tous les sites d'initiation de la transcription *in vivo*. De fait, la transcription est initiée à de nombreux sites pour un même gène et à partir de séquences variées qui peuvent être situées dans

des régions intergéniques ou sur le brin opposé à un gène. De plus, l'absence de terminaison de transcription donne lieu à des transcrits extrêmement grands, contribuant à la transcription de régions intergéniques. On ne peut pas exclure que certains transcrits soient des ARN non-codant fonctionnels. Cependant, les transcrits intergéniques ne sont pas conservés, même entre deux espèces proches et proviennent souvent de séquences recombinées d'origine variée. C'est pourquoi l'on s'attend à ce que la plupart des transcrits intergéniques corresponde à des transcrits "illégitimes", ou "cryptiques".

Le contrôle de l'abondance de chaque ARN est essentiel pour le fonctionnement correct de la cellule. Il est le résultat de l'équilibre entre la synthèse et la dégradation des ARN. La polyadénylation (addition d'une série d'adénosines à l'extrémité 3' d'un ARN) joue ici un rôle crucial. Dans le cytoplasme des cellules eucaryotes, les ARN messagers sont protégés et donc stabilisés par une longue queue polyA. Cependant, cela est loin d'être une généralité. Dans la mitochondrie de plante, comme dans la bactérie et dans le chloroplaste, la polyadénylation est un signal qui dirige l'ARN vers la dégradation.

La polynucléotide phosphorylase (PNPase) est une 3'-5' exoribonucléase qui dégrade les ARN polyadénylés. La PNPase mitochondriale d'*Arabidopsis thaliana,* caractérisée dans notre laboratoire, est une protéine essentielle pour la plante, du moins à certaines étapes de son développement puisque les mutants sous-exprimant la PNPase ne sont pas viables. Cependant, lors de l'introduction d'un transgène pour la sur-expression de la PNPase, des événements aléatoires de co-suppression (extinction de l'expression du gène endogène ainsi que du gène exogène) ont permis d'obtenir des plantules pour lesquelles l'expression de la PNPase était abolie, appelées PNP-. Avant mon travail de thèse, il avait été démontré au laboratoire et dans d'autres équipes que des intermédiaires de dégradation et de maturation de certains transcrits s'accumulaient sous forme polyadénylée dans les mitochondries des plantes PNP-. Cependant, seuls des transcrits choisis par une approche gène-candidat ont été étudiés. Ainsi, le gène chimérique *atpA-orf522* du tournesol stérile PET1, les gènes *cox2* du maïs, *atp9* du pois et de la pomme de terre, *atp1* de *Oenothera* et le gène ribosomique *18Srrn* d'*Arabidopsis* ont été étudiés.

1er axe de recherche : la détermination du rôle biologique de la PNPase dans les mitochondries de plante.

Afin de mieux définir les rôles biologiques de la PNPase, nous avons entrepris une recherche à grande échelle des substrats de la PNPase. Pour cela, nous avons élaboré une banque de petits ARN polyadénylés à partir de plantes PNP-. Nous avons sélectionné les ARN de 100-150 nucléotides pour minimiser la contamination par les ARN cytoplasmiques, portant des queues polyA de plus de 200 nucléotides. 433 clones ont été séquencés, et l'exploitation de ces séquences a montré que 70% des clones étaient d'origine mitochondriale et que la banque

présentait une bonne complexité puisque seulement 0,02% des clones sont représentés plusieurs fois. L'information générée par ces petits marqueurs de dégradation a été utilisée pour identifier et déterminer, par Northern blot, l'accumulation des transcrits en absence de la PNPase par comparaison avec le contexte sauvage, ainsi que la taille de ces transcrits. Ainsi, plusieurs sortes de substrats se dégagent de cette étude.

Tout d'abord nous avons montré que la PNPase dégrade les sous-produits de la maturation des ARN ribosomiques et de transfert, ce qui permet de généraliser le rôle de la PNPase dans l'élimination d'intermédiaires de maturation. Nous avons pu observer un certain nombre de clones arrangés de manière consécutives, suggérant l'action d'une endonucléase. Cependant, l'analyse par Northern blot montre que la plupart des ARN qui s'accumulent correspondent à la séquence complète entre deux transcrits matures, indiquant que l'activité endonucléolytique a un effet mineur par rapport à la PNPase. La dégradation des ARN mitochondriaux est donc principalement exoribonucléasique.

Ensuite, de façon surprenante, nous avons mis en évidence des ARN provenant de régions qui, bien qu'étant exemptes de gènes connus, sont parfois très fortement transcrites. C'est le cas d'un transcrit de 500 nucléotides, appelé *Non-Coding Overexpressed* (*NCO*), généré par une partie de l'une des grandes régions répétées, ainsi que de transcrits de différentes tailles issus d'une région intergénique contenant 230 nucléotides provenant de la duplication d'une séquence contenant le promoteur de l'ARN 26S. Ainsi, la PNPase dégrade des transcrits intergéniques produits par un contrôle relâché de la transcription. *NCO* est un exemple de transcrit pour lequel le niveau d'accumulation est extrêmement élevé en absence de PNPase et la stabilisation de ce transcrit pourrait s'avérer délétère pour la mitochondrie car son accumulation pourrait mobiliser des facteurs indispensables aux étapes post-transcriptionnelles telles que la maturation de transcrits fonctionnels. D'autres transcrits présentent un danger potentiel pour les fonctions de la mitochondrie. C'est le cas de ceux qui contiennent des cadres de lectures (ORF) chimériques créés par recombinaison (dont certains peuvent mener à un phénotype connu de stérilité mâle cytoplasmique), ou des ARN antisens transcrits à partir du brin opposé à celui d'un gène connu. Nos résultats confirment que des ORF chimériques tels que *orf240a*, *orf294* et *orf315* sont exprimés dans les mitochondries d'Arabidopsis. De même, nous avons pu mettre en évidence des promoteurs initiant la transcription en antisens de chacun des cinq gènes que nous avons examinés : *NCO*, *atp9*, *nad4*, *nad5*, *nad7*. La combinaison de grandes séquences intergéniques et de la duplication des promoteurs par recombinaison ainsi que de la spécificité relâchée des promoteurs et l'absence de terminaison de transcription conduisent à une production massive de transcrits "cryptiques" dans les mitochondries de plante. Ces transcrits illégitimes et potentiellement délétères sont donc produits en permanence dans la mitochondrie d'Arabidopsis mais rapidement dégradés par la PNPase.

Ainsi, le transcriptome des mitochondries d'Arabidopsis thaliana ne dépend pas directement de la transcription, mais du contrôle de la stabilité des ARN qui s'appuie sur la polyadénylation et la dégradation par la PNPase. La PNPase est

essentielle pour la viabilité de la plante, ce qui peut être expliqué, du moins en partie, par son rôle crucial dans une voie de surveillance des ARN mitochondriaux impliquant la polyadénylation, qui contrebalance le manque de contrôle de la transcription.

De plus, l'analyse des séquences des clones de la banque nous a permis de révéler l'addition d'extensions non codées par le génome en amont des queues polyA de certains clones. Ces extensions pourraient être dues à une activité RNA polymérase RNA-dépendante, portée par exemple par les ARN polymérases de type phagique de la mitochondrie, mais leur éventuelle fonction reste mystérieuse.

2ème axe de recherche : l'étude de la maturation des extrémités 3' des ARN mitochondriaux de plante.

Au cours de l'exploitation de la banque de petits ARN mitochondriaux polyadénylés, nous avons mis en évidence l'accumulation du transcrit *NCO* en absence de la PNPase, issus de la grande répétition I du génome. La séquence codant pour ce transcrit est composée de parties d'ADN recombinées et n'est pas conservée dans d'autres espèces proches d'Arabidopsis. C'est pourquoi il paraît peu probable que *NCO* porte une fonction, même si cela est difficile à prouver expérimentalement.

Certaines caractéristiques spécifiques de ce transcrit nous ont cependant poussé à l'utiliser comme modèle d'étude de la maturation des ARN mitochondriaux. En effet, nous avons montré par Northern blot que *NCO* est transcrit à un taux extraordinairement élevé et est dégradé très rapidement par la PNPase. Mais, de manière intéressante, ce même Northern blot, ainsi que des expériences de RT-PCR circulaire ont montré que le transcrit *NCO* a des extrémités matures bien définies, même en absence de PNPase. Ceci indique non seulement que *NCO* est maturé à ses extrémités 5' et 3' mais aussi que la PNPase n'est pas responsable de sa maturation. En effet, son taux de renouvellement très élevé garanti que la maturation de ce transcrit est bien indépendante de la PNPase et non pas un événement précédant l'extinction de l'expression de la PNPase dans la plante au moment de la cosuppression, ce qui en fait un bon modèle pour étudier les processus de maturation indépendants de la PNPase.

La maturation en 5' de *NCO* se fait très probablement par l'intermédiaire d'un clivage endonucléolytique, impliquant une activité RNase Z opérant en 3' des ARNt. En effet, un élément ayant une forte homologie avec l'ARNt^{Phe}, *Ath-59*, se situe exactement en 5' du transcrit *NCO* et la machinerie de maturation des ARNt pourrait être responsable de la maturation en 5' de *NCO.* Ceci reste à confirmer par un test *in vitro* ou *in organello.*

Pour vérifier si l'extrémité 3' mature de *NCO* était générée par terminaison de la transcription ou par une maturation post-transcriptionnelle, nous avons procédé à des expériences de "run-on" pour évaluer l'activité de transcription de la séquence du *NCO* et de la séquence en aval. Ces expériences ont permis de montrer que la transcription de *NCO* ne s'arrêtait pas à son extrémité 3' mature mais que *NCO* était bien maturé à partir d'un précurseur plus long. Un long transcrit de 1,8kb, indétectable chez Arabidopsis, est trouvé dans des cellules cybrides *A. thaliana-Brassica napus* avec une séquence génomique mitochondriale identique à celle d'Arabidopsis pour cette région et avec le génome nucléaire de *B. napus*. Cette observation montre que la maturation de l'extrémité 3' de *NCO* requiert facteur nucléaire.

Pour tenter d'élucider le mécanisme de maturation en 3' de *NCO*, nous avons voulu déterminer la taille et la position des transcrits en aval de *NCO*, dans l'objectif de détecter d'éventuels produits d'un clivage endonucléolytique. Les expériences de 5' et 3' RACE (Rapid Amplification of cDNA Ends) des transcrits en aval de *NCO* ainsi que de RPA (Ribonuclease Protection Assay) ont révélé la présence d'une population hétérogène de transcrits en aval de *NCO.* Ces transcrits sont polyadénylés et s'accumulent en absence de PNPase, indiquant que la PNPase est responsable de leur élimination. Ces résultats ne permettent pas de conclure quant à une activité endonucléolytique en 3' de *NCO,* qui contribuerait à la maturation en 3' de *NCO.*

Ainsi, différentes options sont possibles pour expliquer la maturation en 3' de *NCO* :

- la maturation se fait par une dégradation exoribonucléasique effectuée par une activité distincte de la PNPase et de la RNaseII, l'autre 3'-5' exoribonucléase mitochondriale identifiée, car des expériences de RT-PCR circulaires ont montrés que ces deux enzymes n'étaient pas responsables de la maturation de *NCO*. Cependant, cette hypothèse est peu probable car elle impliquerait une activité spécifique qui ne compenserait pas l'absence de PNPase ou de RNaseII.
- la maturation de l'extrémité 3' de *NCO* est due à un clivage endonucléolytique en 3' de *NCO,* suivi d'une dégradation très rapide du produit de clivage en aval. Le mécanisme de dégradation de ce produit reste inconnu. Il pourrait être effectué soit par une 3'-5' exoribonucléase spécifique, soit par une 5'-3' exoribonucléase, soit par une vague d'endoclivages, initiée par ce premier clivage.

Nos résultats montrent que la PNPase est responsable de la dégradation, et de NCO, et des transcrits en aval. La maturation de l'extrémité 3' du transcrit NCO dépend d'un facteur spécifique en trans. Ce transcrit est maturé en absence de PNPase. En contexte sauvage, la maturation de NCO est très efficace mais celui-ci est toutefois instable. Cet exemple montre pour la première fois que la maturation en 3' et la stabilisation d'un transcrit mitochondrial de plante peuvent être découplées. Il est ainsi possible qu'il y ait compétition à l'extrémité 3' mature des transcrits entre la polyadénylation et la stabilisation par un facteur spécifique,

ce qui expliquerait pourquoi la majorité des sites de polyadénylation sont trouvés à l'extrémité 3' des transcrits dans les mitochondries de plante.

3ème axe de recherche : l'investigation d'une éventuelle voie de dégradation des ARN nucléaires impliquant la polyadénylation, chez les plantes.

De façon intéressante, l'exploitation des données de la banque de petits ARN polyadénylés nous a permis de révéler l'existence d'ARNs nucléaires non-codants polyadénylés. Ceci nous a poussé à étudier la possibilité d'une voie de dégradation des ARN impliquant la polyadénylation dans les noyaux de plante.

La polyadénylation des ARN messagers cytoplasmiques est nécessaire à leur export du noyau, leur stabilité et leur traduction. Cependant, il a été montré que la dégradation d'ARN intergéniques et non-codants dans le noyau de levure était activée par leur polyadénylation par un complexe appelé TRAMP (pour Tr4f/Air2p/ Mtr4p Polyadenylation complex). La machine de dégradation de 3' en 5' des ARN nucléaires est l'exosome, un complexe composé d'un anneau de 6 sous-unités ayant un domaine de la RNase PH d' *E. coli* ainsi que 3 sous-unités ayant des domaines de liaison à l'ARN. Deux 3'-5' exoribonucléases y sont associées : l'une de type RNase II, l'autre, spécifique au noyau, de type RNase D, appelée Rrp6p. L'implication de Rrp6p dans la maturation d'ARNr ainsi que de sno/snRNA (pour small nucleolar/small nuclear RNA) a été montré chez la levure. Chez l'homme, des transcrits de gènes ribosomiques, probablement intermédiaires de dégradation, et des précurseurs de la β-globine ont été trouvés sous forme polyadénylée. De plus, l'ARN 5S, intermédiaire de maturation peu abondant des ARN ribosomiques a été récemment détecté sous forme polyadénylée chez des plantes de l'espèce *Nicotiana*.

Nous avons confirmé par RT-PCR que des ARN non-codants nucléaires, tels que des snoRNA et des transcrits intergéniques ribosomiques sont polyadénylés chez Arabidopsis. Nous avons cartographié les sites de polyadénylation sur plusieurs de ces transcrits : certains des ces sites se situaient dans l'ARN mature, d'autres à son extrémité 3' ou en aval du transcrit. Ces résultats indiquent que des précurseurs comme des transcrits mature peuvent être polyadénylés. Cette polyadénylation pourrait cibler des ARN précurseurs pour leur maturation ou des transcrits aberrants détectés par un système de contrôle de qualité. Elle pourrait aussi s'appliquer à des ARN en excès. La prochaine étape dans l'étude de la polyadénylation nucléaire sera l'identification d'autres espèces polyadénylées pour déterminer l'implication de la polyadénylation dans la maturation et la dégradation des ARN nucléaires. D'autre part, l'identification du facteur de polyadénylation chez la plante est indispensable pour l'étude de la fonction de la polyadénylation : l'existence d'un homologue de TRAMP chez la plante est à l'étude dans le laboratoire.

Pour identifier des espèces polyadénylées dans les noyaux de plantes, nous nous sommes attaché à caractériser un facteur spécifique de la dégradation nucléaire des ARN : l'homologue végétal de Rrp6p.

Nous avons caractérisés trois homologues de Rrp6p, que nous avons appelés respectivement RRP6L1, RRP6L2 et RRP6L3. On retrouve une telle famille multigénique chez les autres plantes étudiées, le riz et le peuplier alors qu'un seul gène existe chez l'homme et chez la drosophile, comme chez la levure. Ainsi, la multiplication des gènes Rrp6 pourrait être une spécificité végétale. Une analyse phylogénique nous a permis de regrouper RRP6L1 et RRP6L2 tandis que RRP6L3 se détache. Alors que RRP6L3 apparaît spécifique aux plantes, RRP6L1 et RRP6L2 sont plus proches des protéines RRP6 animales et de levure.

Pour tenter d'élucider les fonctions de RRP6L1 et RRP6L2, nous avons caractérisé des plantes mutées pour l'une ou l'autre de ces protéines, par insertion de T-DNA. Nous avons détecté l'accumulation d'un substrat de Rrp6p chez la levure, l'intermédiaire de maturation d'ARN ribosomique 5'ETS (pour 5' External Transcribed Sequence) dans les lignées mutantes *rrp6l2*. Cet ARN est trouvée sous sa forme polyadénylée. Cependant, ce transcrit ne s'accumule pas en absence de RRP6L1 et on ne peut pas détecter d'augmentation de son accumulation dans le double-mutant *rrp6l1-rrp6l2.* Ces résultats prouvent que la fonction de RRP6L2 dans la dégradation du 5' ETS n'est pas partagée par RRP6L1.

Ces résultats nous ont poussé à analyser la localisation intra-cellulaire des trois protéines RRP6L1, RRP6L2 et RRP6L3. Pour ce faire, nous avons fusionné chacune de ces protéines à la GFP (Green Fluorescent Protein) et nous avons introduit ces constructions, tout d'abord de manière transitoire dans des cellules isolées de tabac BY2 (pour Bright Yellow 2) et dans des cellules de feuilles de *Nicotiana benthamiana*, puis de manière stable, dans des plantes d'Arabidopsis thaliana.

Le signal de localisation obtenu avec les fusion-GFP exprimées de façon transitoire a été confirmé par celui obtenu par expression stable : RRP6L1 et RRP6L2 sont toutes deux très fortement représentées dans le noyau, une faible proportion de RRP6L2 étant cytoplasmique. Au contraire, RRP6L3 est exclue du noyau et sa localisation est strictement cytoplasmique.

Une observation plus attentive des cellules de racines d'Arabidopsis nous a permis de préciser la localisation de RRP6L1 et RRP6L2 dans le noyau. De manière très intéressante, les protéines RRP6L1 et RRP6L2 paraissent occuper des territoires distincts du noyau : alors que la fusion RRP6L1-GFP est détectée dans la vacuole nucléolaire et le nucléoplasme, la fusion RRP6L2-GFP fluoresce majoritairement dans le nucléole, avec un faible signal dans le nucléoplasme.

Cette exclusion mutuelle de l'espace occupé par RRP6L1 et RRP6L2 suggère que ces deux protéines n'ont pas des fonctions qui se recouvrent mais plutôt des substrats différents, localisés dans des structures distinctes du noyau. Comme la biogenèse des ARNr prend place dans le nucléole, la localisation nucléolaire de RRP6L2 est cohérente avec sa fonction dans la maturation des ARN ribosomiques.

La localisation cytoplasmique de RRP6L3 suggère que cette protéine a évolué pour remplir d'autres fonctions que les fonctions nucléaires attribuées à Rrp6p chez la levure. Cependant, l'expression cytoplasmique d'une protéine de la famille Rrp6 est très intrigante et soulève la possibilité de l'existence d'une voie

de dégradation des ARN dans le cytoplasme des plantes, impliquant une protéine de la famille Rrp6. Il serait très intéressant de suivre l'existence d'un telle voie de dégradation dans le cytoplasme, et notamment de déterminer si la polyadénylation est impliquée ou non. Pour ce faire, la détection d'ARN polyadénylés générés par un virus dont le cycle est uniquement cytoplasmique est à l'étude au laboratoire.

Ainsi, notre travail a permis d'apporter plusieurs arguments pour confirmer le caractère ubiquitaire de la polyadénylation et de son rôle dans le ciblage des ARN pour leur dégradation, conservé dans les systèmes génétiques allant des bactéries aux noyaux des plantes. De plus, nous avons caractérisé une famille de trois protéines Rrp6 spécialisées, ce qui suggère une complexité des fonctions de dégradation des ARN unique chez les plantes.

Conclusion

L'objet de cette thèse est l'étude des voies de dégradation de l'ARN impliquant la polyadénylation chez les plantes. Le premier système génétique analysé est la mitochondrie d'*Arabidopsis thaliana*. En effet, des études antérieures ont montré que certains ARN mitochondriaux peuvent être polyanédylés de façon non constitutive pour activer leur dégradation par la PNPase mitochondriale.

Le premier objectif de ce travail de thèse a été de déterminer le rôle biologique de la PNPase d'une manière globale. Nos résultats ont montrés non seulement que la PNPase est indispensable pour l'élimination des sous-produits de la maturation des ARN ribosomiques et de transfert, mais aussi que la polyadénylation et la PNPase sont impliqués dans une voie de surveillance des ARN qui contrebalance le manque de contrôle de la transcription dans les mitochondries de plante et défini un transcriptome fonctionnel.

Le transcrit *NCO* identifié au cours de cette analyse nous a servi de modèle pour étudier la maturation de l'extrémité 3' des ARN mitochondriaux. Nous avons pu démontrer pour la première fois que la maturation et la stabilisation d'un ARN mitochondrial de plante peuvent être des étapes découplées.

De plus, les résultats obtenus lors de l'étude du rôle de la PNPase dans la mitochondrie nous ont poussé à mener des investigations sur l'existence d'une voie de polyadénylation impliquée dans la dégradation d'ARN nucléaires. Nos résultats donnent de bonnes indications d'une telle voie dans le noyau des plantes. Nous avons aussi caractérisé une famille de protéines homologues à la protéine Rrp6p de levure, associée à l'exosome nucléaire, une machinerie de dégradation exoribonucléolytique des ARN. La spécialisation de ces trois protéines RRP6L1, RRP6L2 et RRP6L3 indique une complexité spécifique au règne végétal.

Ainsi, les résultats obtenus lors de cette thèse ont contribué à montrer le caractère ubiquitaire d'une voie de dégradation dépendant de la polyadénylation.

INTRODUCTION

The object of this thesis is the study of RNA degradation pathways involving polyadenylation. The first material of investigation I used is plant mitochondria. My work contributed to better understand the roles of polyadenylation for RNA degradation in plant mitochondria. Therefore, in the first part of the introduction, I give an overview of the plant mitochondrial genetic system and focus on processes that are important for the comprehension of my results. Detailed reviews on the plant mitochondrial genetic system have been published previously : Gagliardi & Binder, 2007; Mackenzie, 2007; Hanson and Bentolila, 2004; Gagliardi *et al*, 2004; Binder and Brennicke, 2003; Burger *et al*, 2003.

The results I obtained in plant mitochondria led me to investigate an eventual RNA degradation pathway involving polyadenylation in plant nucleus. The link between RNA degradation and polyadenylation seems widespread in all kingdom of life. Therefore, the second part of the introduction gathers major points of knowledge on RNA degradation, from bacteria to the nucleus of eukaryotes.

I. Overview of the plant mitochondrial genetic system.

1.1. What are mitochondria?

The name “mitochondrion”, given in 1898 by Benda, derives from the Greek “mitos” (thread) and “chondros” (grain). It refers to the distinctive double-membrane enclosed organelle found in almost every eukaryotes, commonly called the “power plant of the cell” because it generates the major source of energy used by the cell : ATP. The number of mitochondria in a cell varies widely depending on organisms, tissue types and state of development. It can reach millions in energy demanding organs like the heart.

Mitochondria are ubiquitous throughout the eukaryotic domain and it is now commonly admitted that they originate from a single endosymbiotic event, when a prokaryotic cell, related to todays α-proteobacteria, was taken up by a heterotrophic host cell, about 2 billion years ago (Gray *et al,* 1999) (Figure i1).

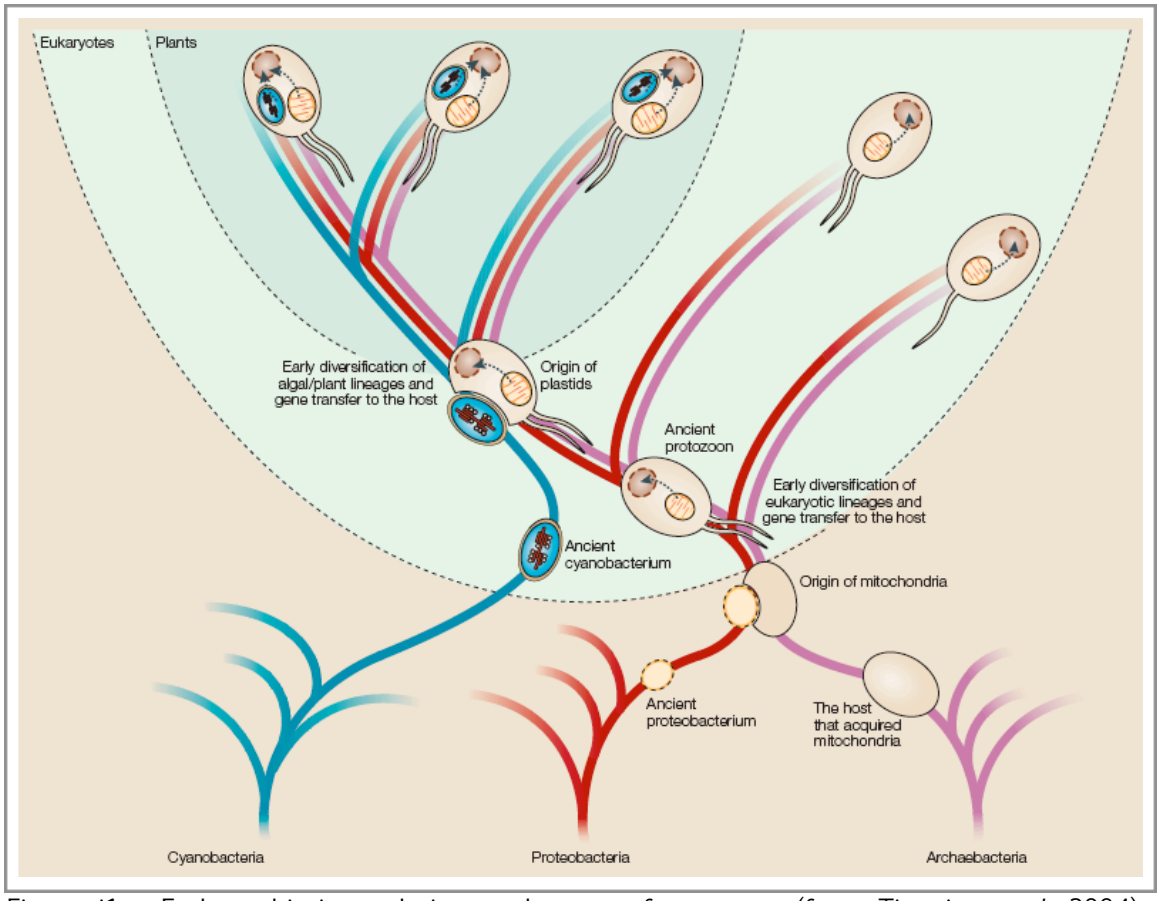


Figure i1 : Endosymbiotic evolution and trees of genomes (from Timmis *et al*, 2004). Intracellular endosymbionts that originally descended from free-living prokaryotes have been important in the evolution of eukaryotes by giving rise to two cytoplasmic organelles. Mitochondria arose from α-proteobacteria and chloroplasts from cyanobacteria. The figure shows a diagram of the evolution of eukaryotes, highlighting the incorporation of mitochondria and chloroplasts into the eukaryotic lineage through endosymbiosis and the subsequent co-evolution of the nuclear and organelle genomes. The nature of the host cell that acquired the mitochondrion (lower right) is fiercely debated among cell evolutionists.

As a consequence, mitochondria possess their own genome. They cannot be created *de novo* but divide by fission. Mitochondria are transmitted to daughter cell, in almost every case by the female gamete.

Some unicellular eukaryotes lack mitochondria : the microsporidians, metamonads, and archamoebae (Cavalier-Smith, 1991). These so-called amitochondriates appear as the most primitive eukaryotes on phylogenic trees based on rRNA sequence comparison. However, they possess genes of mitochondrial origin or organelles derived from mitochondria, like mitosomes in *Giardia* (Henze and Martin, 2003) and hydrogenosomes (Embley *et al*, 2003). The host that acquired the endosymbiont is thus a common ancestor to all eukaryotes and is generally assumed to be an archeabacteria (Timmis *et al*, 2004).

One of the major functions of mitochondria is the production of ATP. This is achieved by the ATP synthase, that uses the proton gradient force driven by the electron transport, coupled to the oxidation of NADH. A set of 4 protein complexes forms the electron transport chain, seated in the internal membrane of mitochondria. These complexes are assembled from hundreds of subunits, of which only a few are mitochondrial encoded. Plant mitochondria harbor additional enzymes such as NADH dehydrogenases and alternative oxidases (Figure i2).

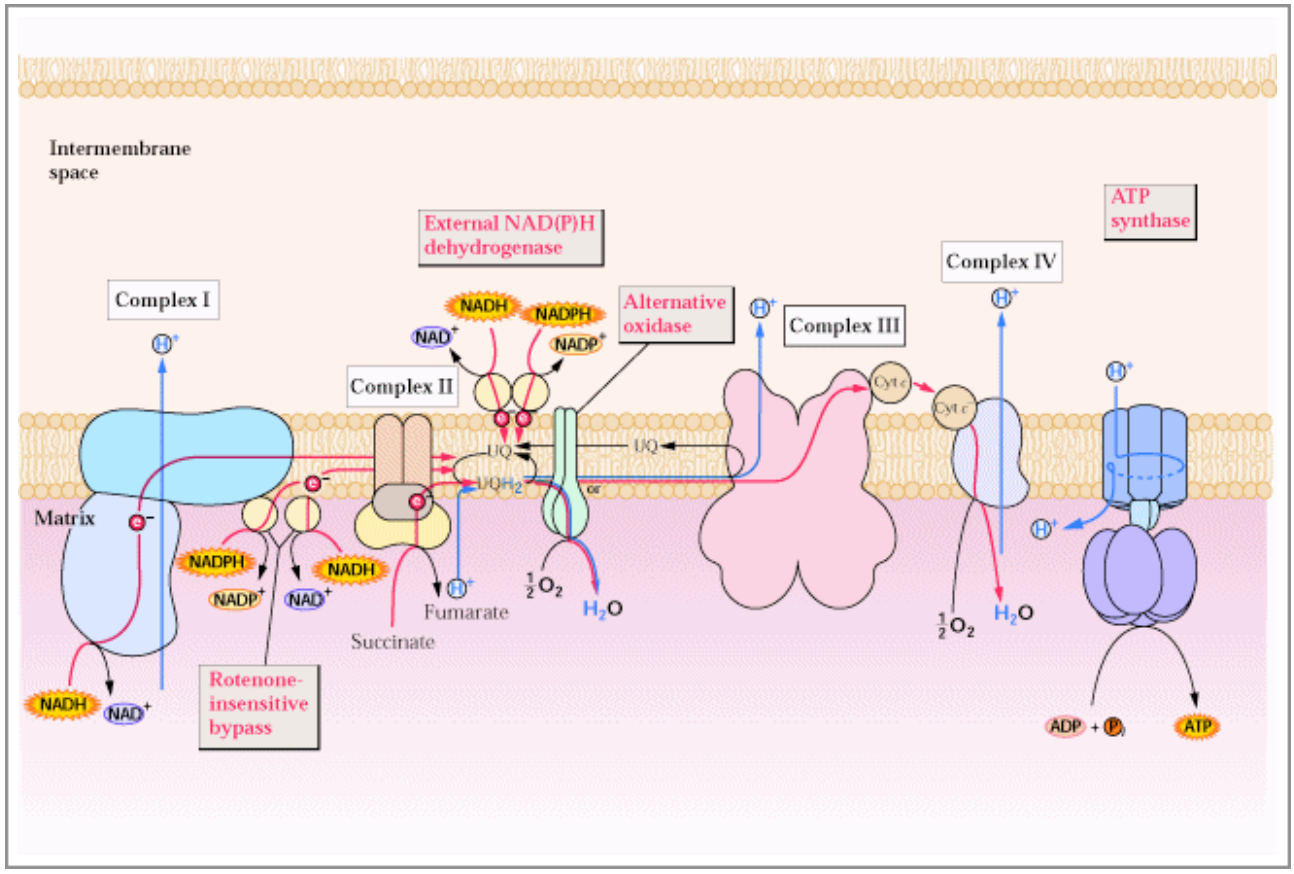


Figure i2 : Organization of the electron transport chain and ATP synthesis in the inner membrane of plant mitochondria. In addition to the four standard protein complexes found in nearly all other mitochondria, the electron transport chain of plant mitochondria contains five additional enzymes (depicted in yellow, and green for the alternative oxidase). Red arrows indicate electron transfer, blue arrows indicate proton translocation (from Buchanan, Gruissem and Jones, 2000).

None of these additional enzymes pumps protons. This multiplicity of by-passes in plants, where animals have only the uncoupling protein, gives a greater flexibility to plant energy coupling. Because plants have the alternative pathways, they can survive exposure to inhibitors of the respiratory complexes (Buchanan *et al*, 2000).

Mitochondria also host the TriCarboxylic Acid (TCA) cycle, or Krebs cycle, that produces electron donators for the electron transport chain. In addition, mitochondria are essential for the synthesis of lipids and various cofactors, such as biotine, folate and lipoic acid. Mitochondria also have an important role in ion homeostasis, in particular in calcium storage, in apoptosis and in amino acids metabolism (Schuster and Binder, 2005).

1.2. Plant mitochondrial genome.

As a inheritance of their prokaryote ancestor, mitochondria have their own genetic system, encoding a small number of genes (40-50 on average). This number does not correlate with the at least thousand mitochondrial proteins that are required to perform the large variety of biochemical reactions (Millar *et al*, 2005). During evolution, most of the endosymbiont original genome was transfered to the nucleus -and still is (Adam and Palmer, 2003). Thus, most of the mitochondrial proteome is imported from the cytoplasm. Interestingly, gene transfer is not unidirectional and sequences of nuclear, chloroplastic and viral origin, are found in plant mitochondrial genomes (Marienfeld *et al*, 1999, Burger *et al*, 2003).

It is noteworthy that despite a unique origin and very conserved gene content, mitochondrial genomes are extremely diverse in size, structure and expression mechanisms. In animals, mitochondrial genomes have an average size of 15-60 kbp. The smallest genome, with 6 kbp, was found in the protist *Plasmodium* sp. Higher plant mitochondrial genomes are usually large, varying mostly between 218 kbp (*B. hirta*) and 570 kbp *(Z. mays)* and contain 54 to 60 genes (see Table i1). However, single species as muskmelon apparently possess mitochondrial genome of more than 2400 kb (Ward et al, 1981).

Plant	Latin name	Size (bp)	Number of encoded genes (proteins, rRNAs, tRNAs)	Reference
Wheat	*Triticum aestivum*	452 528	55 (35, 3, 17)	Ogihara *et al*, 2005
Tobacco	*Nicotiana tabacum*	430 597	60 (36, 3, 21)	Sugiyama *et al*, 2004
Maize	*Zea mays*	569 630	58 (33, 3, 21)	Clifton *et al*, 2004
Rice	*Oryza sativa*	490 520	55 (35, 3,17)	Notsu *et al*, 2002
Rapeseed	*Brassica napus*	221 853	54 (34, 3, 17)	Handa , 2003
Sugar beet	*Beta vulgaris*	368 799	59 (29, 5, 25)	Kubo *et al*, 2000
Thale cress	*Arabidopsis thaliana*	366 924	57 (32, 3, 22)	Unseld *et al*, 1997

Table i1 : To date sequenced mitochondrial genomes of higher plants.

1.2.1. Genome structure.

The plant mitochondrial genome can be assembled on a circular chromosome, called "master circle". Such a structure has been observed in *Brassica hirta,* with 218 kbp, that has the smallest mitochondrial genome among angiosperms (Palmer and Herbon, 1987). However, such a master circle was not detectable in most species. Plant mitochondrial DNA is more commonly multipartite, organized as a complex mix of subgenomic circular and linear molecules (Backert *et al,* 1997). These molecules result from inter- and intramolecular recombination events that are particularly frequent in plant mitochondria (Mackenzie, 2007). Recombination events involve large repeated sequences in inverse orientation and result in sequence izomerisation or "flip-flop" (see Figure i3).

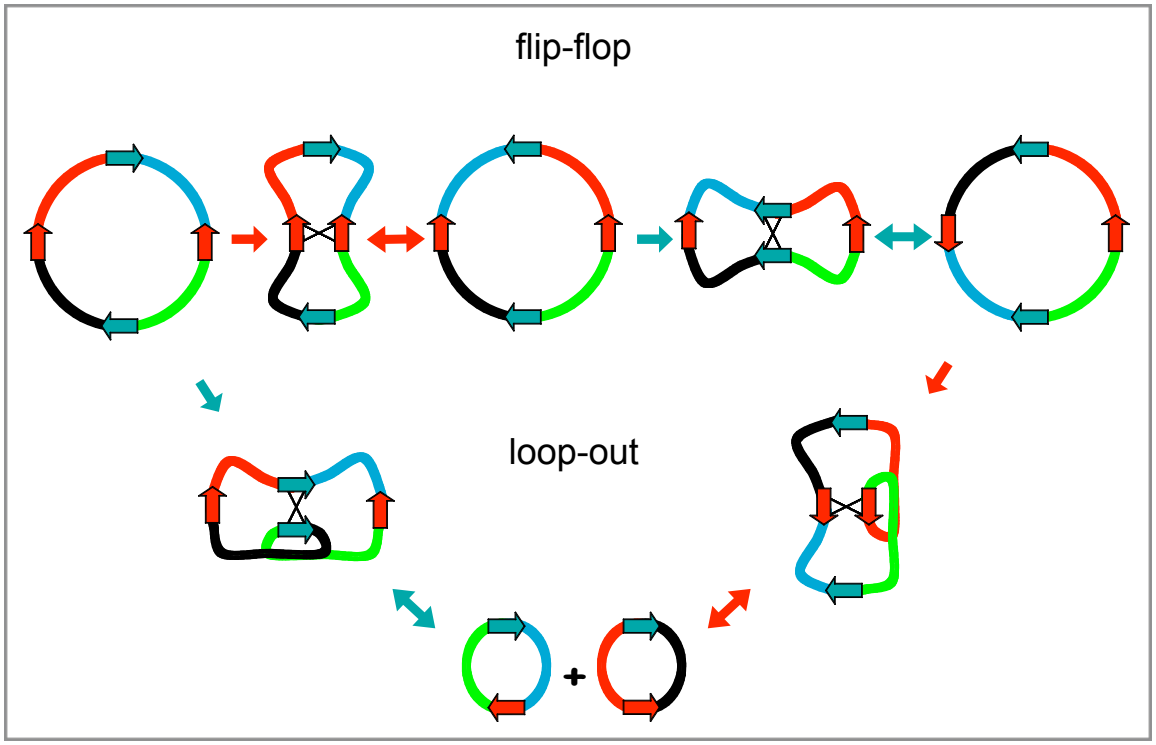


Figure i3 : Model of mitochondrial recombination events. Top : recombination between inverted repeat sequences give rise to "flip-flop" or sequence izomerisation. Bottom : recombination between repeat sequences in direct orientation gives rise to "loop-out", resulting in sub-genomic molecules (modified from Klein *et al*, 1994).

Other repeats, that can be as small as 7 nucleotides, give rise to ectopic intragenic recombinations (André *et al*, 1992). Such recombinations can create chimeric genes, found associated with every case of Cytoplasmic Male Steritily (CMS, due to the inability to produce functional pollen) studied so far. Some molecules, called sub-genomic, appear to be very low abundant. Their number can be increased by "loop-outs" of direct repeats (Figure i3). That phenomenon is called substoichiometric shifting (SSS) and appears to be unique to higher plants. This process is controlled by nuclear genes and can directly influence plant

growth development and fertility. The mitochondrial protein MSH1, a homolog of *E. coli* MutS, has been shown to play a role in the suppression of such illegitimate recombinations (Mackenzie, 2007). Another mitochondrial protein, OSB1 has been described to participate in controlling the stoechiometry of alternative mtDNA forms generated by recombination (Zaegel *et al*, 2006).

Knowing the complexity of the mitochondrial genome, the mechanism of replication is puzzling : as the mtDNA is multipartite, each of molecules has to be replicated to maintain a complete information. In vertebrates and trypanosomes, the replication is of thêta type but in plants, little information is available to this day. Two mechanisms have been postulated. The mtDNA likely replicates via rolling circle mechanism, according to electron microscopic analyses (Backert and Börner, 2000). On the other hand, in 2006, Manchekar and colleagues showed that replication initiation occurs by strand invasion in a recombination-mediated way (Manchekar *et al*, 2006).

Thus, the plant mitochondrial genome has a complex structure, due to a high degree of recombination activity that subdivides the genome and increases genetic variation.

1.2.2. Genome content.

The exceptionally large size of plant mitochondrial genomes does not correlate with an increased coding capacity. As shown in Figure i4, about half of the Arabidopsis mitochondrial genome is composed of non-coding sequences, and repetitive sequences account for about 7% of the genome. The ORF of genuine genes are very well conserved throughout the plant kingdom.

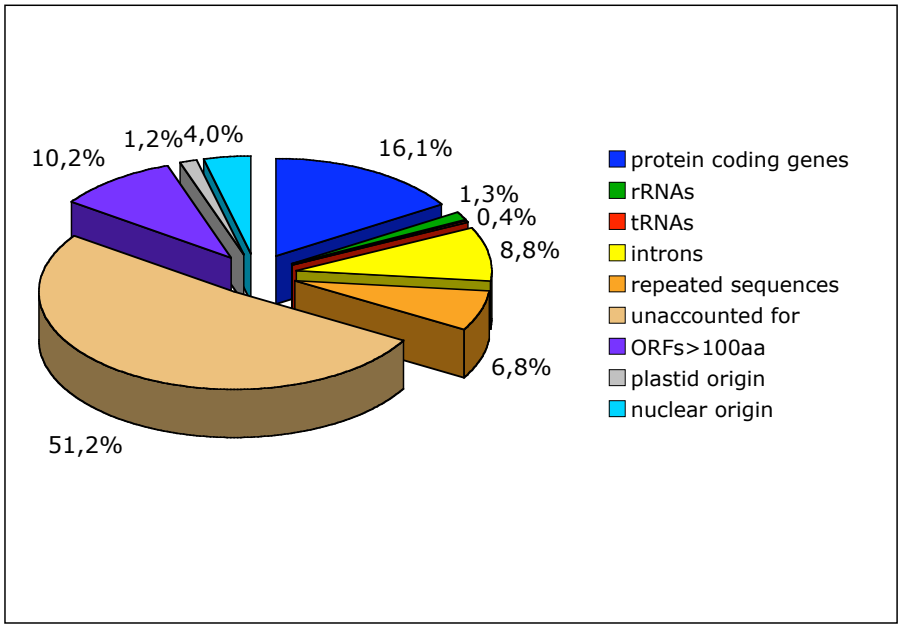


Figure i4: Genome sequences in *Arabidopsis thaliana*. abbreviations : aa : amino acids, ORFs : open reading frames, (modified from Marienfeld *et al*, 1999).

Most plant mitochondrial genomes resemble *Arabidopsis thaliana* genome, that contains 57 genes encoding 32 proteins, 22 tRNAs and 3 rRNAs (Marienfeld *et al*, 1999). The protein coding genes can be grouped in functional categories : those involved in the respiratory metabolism, genes required for *c* cytochromes biogenesis and those required for expression of the mitochondrial genome (ribosomal proteins and a putative maturase) (Farré and Araya, 1999).

In addition to functional genes, plant mitochondrial genomes contain a variable number of unknown ORFs. 85 of such ORFs of more than 100 codons are present in the mitochondrial genome of *Arabidopsis thaliana* (Unseld *et al*, 1997). Whether these ORFs encode functional proteins or not is unlikely as they are not conserved between species.

Hence, plant mitochondrial genomes are not only structurally complex, but also contain a huge amount of DNA that was sometimes called "junk DNA" because no encoded function could be established. Are these non-coding sequences devoid of any function for the mitochondria? Are these regions transcribed? If yes, do they accumulate or are they actively degraded?

1.3. Transcription in plant mitochondria.

1.3.1. Transcriptional units.

Human mitochondrial genes are transcribed as two large polycistronic RNAs that are then cleaved by endonucleases to release mature transcripts. In yeast, transcription is initiated from about 20 promoters, each driving transcription of polycistrons (Christianson and Rabinowitz, 1983). In plant mitochondria, genes are scattered along the large DNA molecule. For example, in *Arabidopsis thaliana*, the gene density is 8 kbp per coding region on average but genes are unevenly distributed and intergenic spacers can reach up to 57kbp (Unseld *et al*, 1997). In contrast to yeast and human mitochondria, monocistronic genes are the norm in plant mitochondria. However, some polycistronic genes exist. Most of them are specific for one or a few species, like *nad5-nad4L-orf25* in *Arabidopsis thaliana* (Brandt *et al*, 1992), *apt9-rps13* in tobacco, *coxIII-orf25* and *rps3-rpl16-nad3-rps12* in rice (Lelandais *et al*, 1996). Some of them, like the *5S-18S rrn, nad3-rps12* and *rps3-rpl16,* are conserved between species.

1.3.2. Promoter features.

The multiple yeast mitochondrial promoters comprise a conserved nonanucleotide motif : ATATAAGTA (Shadel and Clayton, 1993). In humans, mitochondrial genome contains only three promoters, LSP, HSP1 and HSP2. LSP and HSP2 drive

transcription of the light strand and the heavy strand, respectively, and rRNA genes are also transcribed from HSP1 (Bonawitz *et al*, 2006).

In plant mitochondria, transcription is driven by numerous promoters, scattered all along the genome. Analysis of the surrounding sequences of the initiation sites by *in vitro* capping experiments allowed to determine a consensus promoter sequence for tRNA, rRNA and mRNA, first in monocotyledons (Mulligan *et al*, 1988; Brown *et al*, 1991), then in dicotyledons (Binder and Brennicke, 1993b). The core element of this consensus is the tetranucleotide CRTA (R= A or G), that is found in monocotyledons as well as in dicotyledons. However, more plasticity was observed in monocotyledons where this motif can be TRTA, YYTA (Y= T or C) or RRTA (Fey and Maréchal-Drouard, 1999). *In vitro* transcription analyses identified further nucleotide identities important for the promoter function, clustering near the transcription start site and at position -12 (Rapp et al, 1993). In dicotyledones, a 18nt sequence encompassing the transcription initiation site is necessary and sufficient to drive transcription *in vitro* (Dombrowski *et al* 1999, Hoffmann *et al*, 2002).

Kühn and colleagues showed more recently by 5' RACE and *in vitro* capping experiments, that the CRTA motif can be embedded in two types of conserved nonanucleotide motif (CNM) in dicotyledones : type-1 CNM, CRTAAGAGA and type-2 CNM, CGTATATAA (Kühn *et al*, 2005). However, promoters that do not fit to the classical consensus were previously described (Binder *et al*, 1994; Remacle and Maréchal-Drouard, 1996; Fey and Maréchal-Drouard, 1999) and this study confirmed that many other sites than the conserved promoters can drive transcription.

In addition, *in vitro* capping experiments have identified several transcription starts for a single gene, as is exemplified by the *atp9* gene that has six different initiation sites in maize (Mulligan *et al*, 1988) and two in *Oenothera berteriana* (Binder and Brennicke, 1993a). Kühn and colleagues have shown that multiple promoters for a single gene is a common feature in *Arabidopsis thaliana* mitochondria (Kühn *et al*, 2005).

Thus, considerable flexibility is observed in plant mitochondrial transcription system, with variable promoter structures and multiple promoters for one gene, that could lead to different rates of transcription.

1.3.3. The transcriptional machinery : RNA polymerases and transcription factors.

The *Arabidopsis* nuclear genome encodes three T7/T3 phage-like RNA polymerases, that have been identified by homology with yeast mitochondrial Rpo41. RpoTm (encoded by AtRPOT1), targeted to mitochondria, RpoTmp, presumably chloroplastic and mitochondrial (encoded by AtRPOT2) and RpoTp, also named NEP for Nuclear Encoded Polymerase, (encoded by AtRPOT3) that is strictly chloroplastic. Two different in frame start codons in RpoTmp are the reason for alternative translation initiation responsible for the targeting of RpoTmp protein to ei-

ther chloroplast or mitochondria (Hedke, 2000). In monocotyledones, this RpoTmp does not exist (Emanuel *et al*, 2004). A recent *in vitro* study showed that RpoTmp displays no significant promoter specificity whereas RpoTm initiated alone transcription from a subset of promoters (Kühn, *et al.* 2007). This implies that auxiliary factors are required for efficient initiation of transcription at all promoters *in vivo*.

T7 RNA polymerase is a single polypeptide that can perform all steps of the transcription on its own : recognition of the initiation site, initiation and elongation of transcription. However, human and yeast phage-type polymerases are not selective and need specific cofactors. These cofactors are of 2 types : the mtTFA proteins (or ABF2, or TFAM) that have 2 High Mobility Groups (HMG) and the mtTFB proteins (mtTF1 in yeast or TFB1M and TFB2M in human) that are related to rRNA methyltransferases (McCulloch *et al*, 2002). As plant mitochondrial polymerases are homolog to yeast and human ones, one could expect similar cofactors to be present in plants. No mtTFA homologs but 2 putative mtTFB were found in *Arabidopsis* by genetic approaches, one being targeted to mitochondria. These TFB-like proteins display unspecific DNA-binding activity but their functions remain to be solved (Gagliardi & Binder, 2007).

Biochemical studies have led to the identification of a wheat 63kDa protein that has affinity for the *cox2* promoter (Ikeda and Gray, 1999). This P63 belongs to the large family of Pentatrico Peptide Repeat (PPR) proteins that are mostly invovled in RNA metabolism. Two other proteins of 32 and 44 kDa show affinity for the *atp9* promoter in pea (Hatzack *et al*, 1998). More recently, proteomic studies have identified potential transcription factors and a supposed transcription initiation factor (Heazlewood *et al*, 2004).

Thus, little is known about plant mitochondrial transcription factors and the extent of their influence on promoter specificity, except that nuclear-encoded factors seem to play an important role.

1.3.4. The termination of the transcription.

To date, little is known about transcription termination in plant mitochondria. Bacterial and human mitochondria express transcripts with a 3' stem-loop that serve as Rho-independent terminator. Some plant mitochondrial transcripts have similar 3' stem-loops but they rather play a role in processing than in termination (Dombrowski *et al,* 1997). Signals for the transcription termination are thus unknown. However, a number of plant genes code for proteins that share similarities with the human mTERF termination factor. mTERF proteins form a family that is only found in plants and metazoans (Linder *et al*, 2005). MOC1, a mTERF homologue has been identified in *Chlamydomonas reinhardtii.* MOC1 is targeted to the mitochondrion, and its expression is up-regulated in response to light. Loss of MOC1 causes a high light-sensitive phenotype and disrupts the transcription and expression profiles of the mitochondrial respiratory complexes. (Schonfeld *et al,* 2004).

However, no efficient transcription mechanism seem to exist in plant mitochondria, and transcription, once initiated, can give rise to primary transcripts that are far longer than mature transcripts. For instance, *atp9* and *atp8/orfB* primary transcripts can have ten times the size of the mature transcript in *Arabidopsis thaliana* (Perrin et al, 2004a).

1.3.5. Regulation at the transcriptional level.

Run-on transcription assays were used to study regulation of gene expression at the transcriptional level. A first study showed that promoter strength is the main determinant of transcripts abundance, rRNA genes being more transcribed than mRNAs. However, the absence of correlation between steady-state abundance of transcripts and transcriptional activity suggested involvement of post-transcriptional processes (Mulligan *et al*, 1991). Studies using run-on and steady-state analysis showed that transcript steady-state levels are more homogeneous than transcriptional activities, indicating that RNA turnover is a compensating mechanism buffering asymmetries of transcription (Giegé *et al*, 2000; Leino *el al*, 2005). In addition, tissue-specific regulation of transcription appears to be absent in plant mitochondria (Kühn *et al*, 2005).

Furthermore, Finnegan and Brown described unexpected transcription of non-coding sequences, including 21kbp flanking rRNA genes, supporting the idea that either transcription can initiate from promoter with relaxed specificity in plant mitochondria or there is a lack of transcription termination (Finnegan & Brown, 1990).

Hence, plant mitochondrial genomes are prone to recombination and insertion events, resulting in large and complex structures containing high amounts of non-coding sequences. Transcription initiation and termination are not specified accurately, and rather present relaxed features. Regulation of transcripts abundance is not determined only by promoter strength, demonstrating the importance of post-transcriptional processes. As plant mitochondrial lack control at the transcriptional level, post-transcriptional processes appear crucial for correct gene expression.

1.4. Post-transcriptional processes.

Mitochondrial primary transcripts undergo many post-transcriptional processes, some of which are specific to plant mitochondria. Transcripts have to be spliced, edited at specific sites by a C to U conversion (in higher plants), and matured at their 5' and at their 3' ends. Some transcripts undergo also other nucleotide modifications. Finally, control of RNA stability is an important parameter in genome expression.

1.4.1. Editing.

For a long time, it was thought that the plant mitochondrial genetic code was different from the universal code. In 1989, editing was discovered simultaneously by three different groups (Covello and Gray 1989 ; Gualberto *et al* 1989 ; Hiesel *et al* 1989). RNA editing is a molecular processes in which the information content is altered in a RNA molecule through a chemical change in the base. Generally, the first or the second base of the codon is edited, leading to the change of amino acid, resulting in a better amino acid conservation of the protein. In plant mitochondria and chloroplasts of higher plants, it consists in a C to U conversion. In *Tryponosoma bruceii*, it consists in addition/deletion of uridine. Edition can give rise to the formation of an initiation codon or a stop codon.

In *Arabidopsis thaliana* mitochondria, 456 editing sites were identified (Giegé and Brennicke, 1999). Most editing sites have been found in ORF sequences (441 sites), and only a few sites were located in 5' and 3' UTR, in introns and tRNAs. In chloroplasts, editing is far more limited and only about 30 sites are edited (Tillich *et al*, 2006). Plant mitochondrial electroporation assays showed that, in wheat, nucleotides between -16 and +6 of the editing site are necessary for the recognition by the editing machinery but that they are not conserved (Farré *et al*, 2001). *In vitro* RNA editing system from pea has narrowed the sequence requirements for the site specificity to the region 5–20 nucleotides upstream of this site in the atp9 mRNA (Takenaka *et al*, 2004).

As there are numerous editing sites, it is unlikely that there is a specific factor for each of them. But if one factor could recognize several related sites, a family of factors could exist. For instance, it was recently demonstrated in tobacco chloroplast that an editing factor recognizes several sites (Kobayashi *et al*, 2007).

Proteins of the PentatricoPeptide Repeat (PPR) family are good candidates for such factors. PPR proteins, containing repeats of a degenerate motif of 35 amino acids, are molecular adapters recognizing a specific RNA sequence through these motif repetitions (Small and Peeters, 2000). *PPR* genes constitute one of the largest gene family in plants, with about 450 members in *Arabidopsis* and similar number in rice. By contrast, only a few PPR proteins are present in non-plant eukaryotes and none in prokaryotes, apart one exception, probably due to horizontal transfer (Lurin *et al*, 2004). The majority of the members are targeted to chloroplast or mitochondria and there is more and more evidence that specific PPR proteins are linked to various post-transcriptional processes such as editing, processing or translation as well as stabilizing transcripts (Andrés *et al*, 2007). However, the roles of PPR proteins were so far deduced from genetic links and the mechanisms of action remain to be demonstrated.

In the chloroplast, two PPR proteins were shown to specifically direct editing of *ndhD* transcripts : CRR4 in the translation initiation codon (Kotera *et al*, 2005)

and CRR21 in the site 2 of *ndhD* (Okuda *et al*, 2007), demonstrating that PPR proteins are indeed involved in organellar editing.

Thus, PPR proteins are likely to be important factors specifying editing. Moreover, the identity of the enzyme catalyzing the deamination reaction is still unknown.

1.4.2. Splicing.

Plant mitochondria have a large number of group II introns, that are all in protein-coding genes. There is only one group I intron in the *cox1* gene of several unrelated land plants, apparently acquired by horizontal transfer (Palmer *et al*, 2000).

As a consequence of the massive rearrangement of mitochondrial genome, splicing occurs not only in *cis* but also in *trans* : *nad1*, *nad2* and *nad5* gene are fragmented and exons in *trans* are joined by splicing. Specific splicing requires nuclear encoded factors (Brangeon *et al*, 2000). One example of these factors is a group II intron-encoded maturase, At-nMat1a, that is involved in NAD4 transcript splicing (Nagakawa and Sakurai, 2006). Recently, a PentatricoPeptide Repeat gene, *OTP43*, was shown to be required for *trans*-splicing of the mitochondrial *nad1* first intron in *Arabidopsis thaliana* (Falcon de Longevialle *et al*, 2007).

1.4.3. 5' and 3' extremity maturation of mitochondrial transcripts.

Plant mitochondria primary transcripts are always longer than mature transcripts. The 5' extremity is variable and does not automatically correspond to the transcription initiation site. The lack of transcription termination results in 3' extensions that can span kilobases beyond the mature 3' end of a transcript. In some cases, cotranscribed genes are matured to give rise to monocistronic units. Therefore, maturation is a crucial step for a functional transcript and it is also a mean of regulation of the gene expression.

Theoretically, processing can be performed by 3 types of enzymes : endonucleases, 5' to 3' exoribonucleases and 3' to 5' exoribonucleases. Endonuclease activities are involved in processing of tRNAs, rRNAs and cleavage of polycistronic mRNAs precursors. But no mitochondrial gene encoding endonuclease has been identified yet. To date, there is no data supporting existence of a 5' to 3' exoribonuclease in plant mitochondria but two 3' to 5' exoribonucleases have been characterized : RNase II and PolyNucleotide Phosphorylase (PNPase) (Perrin *et al*, 2004a).

1.4.3.1. tRNA maturation.

tRNAs precursors have to undergo many post-transcriptional processes to be functional, consisting in 5' and 3' extremities maturation, nucleotide modification (including editing) and addition of CCA nucleotides at 3' ends, that are not genomically encoded in plant mitochondria.

trn genes are arranged in single units, in gene cluster, or are cotranscribed with protein coding genes. Mature 5' and 3' extremities are obtained by endonucleolytic cleavages. Processing at the 5' side of bacterial tRNA is catalyzed by the quasi-universal ribozyme RNase P. Similarly, the first step of tRNA maturation in mitochondria is 5' end processing by an RNase P-like activity (Marchfelder *et al*, 1990). In *E. coli*, where all tRNAs have an encoded CCA motif, the 3' extremity is processed in an exoribonucleolytic way, by a number of redundant 3' to 5' exonucleases, including RNase PH, RNase T, RNase D and RNase II (Nicholson, 1999). In *B. subtilis,* tRNA lacking an encoded CCA motif are matured by an RNase Z endocleavage. RNase Z is an ubiquitous enzyme also known in *E. coli* as RNase BN, where is can mature tRNAs in absence of exonucleases (Condon, 2007). Mitochondrial 3' end maturation consists firstly in a RNase-Z type enzyme cleavage exactly at the so-called "discriminator base", or one nucleotide upstream. Then, CCA nucleotides are added by a nucleotidyl transferase (see Figure i5). *A priori*, 5' processing maturation precedes 3' processing (Kunzmann *et al*, 1998).

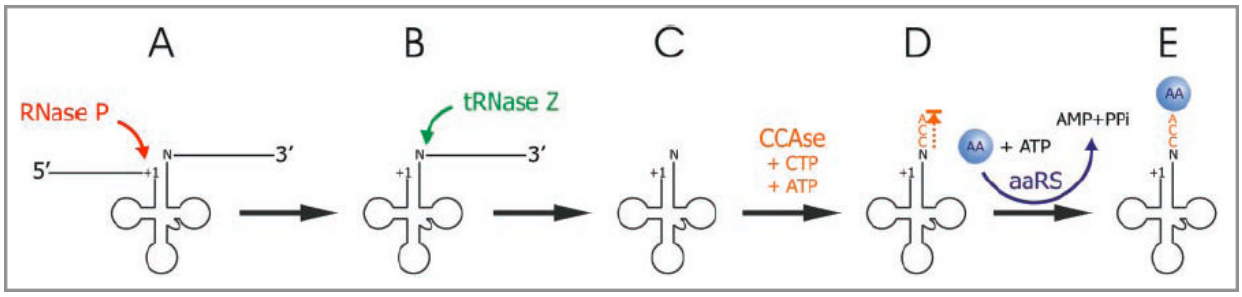


Figure i5 : tRNA end processing followed by aminoacylation. (A) tRNA is transcribed as a precursor, with a 5' end leader and a 3' end trailer. (B) RNase P endonucleolytically cleaves the tRNA at +1 position (C) tRNase Z endonucleolytically cleaves the precursor on the 3' side of the discriminator base (N). (D) CCA-adding enzyme (CCAse) adds CCA to the 3' end of the tRNA (N) produced by tRNase Z cleavage. (E) tRNA is charged with the cognate amino acid by a specific aminoacyl-tRNA synthetase (aaRS). From Levinger et al, 2004.

To date, genes coding for plant mitochondrial tRNA processing machinery have not been cloned but a candidate gene encoding a putative mitochondrial nucleotidyl-transferase has been identified (Martin and Keller, 2004).

In comparison to mRNAs, very few tRNA are edited. *In vitro* studies showed that editing is a prerequisite for processing of Oenothera $tRNA^{Phe}$ and larch $tRNA^{His}$ (Marchfelder *et al*, 1996; Maréchal-Drouard *et al*, 1996). A recent *in organello* confirmation came from a study showing that unedited larch $tRNA^{His}$ is expressed but not matured in potato mitochondria, whereas the edited version of the same tRNA could undergo maturation (Placido *et al*, 2005).

1.4.3.2. rRNA maturation.

Three *rrn* genes are present in the plant mitochondrial genome : 26S rRNA, 18S rRNA and 5S rRNA coding genes. 18S rRNA and 5S rRNA are always cotranscribed whereas the 26S rRNA is at a distinct locus. rRNAs are generated as precursors and are matured at their extremities. Generally, the 5' end has to be matured by excision of the 5' external transcribed sequence (ETS) : transcription is initiated 900 nucleotides upstream of the 26S rRNA and 150 nucleotides upstream of the 18S rRNA in *Arabidopsis* (Perrin *et al,* 2004b, Kühn *et al,* 2005). The 5' end maturation of the 18S rRNA is performed by an endonucleolytic cleavage at the mature 5' extremity or one to two nucleotide upstream (Perrin *et al*, 2004b). An exception is the potato 26S rRNA which mature 5' extremity corresponds to the transcription initiation (Binder *et al*, 1994).

Little is known on the 3' end maturation of rRNAs. It could be achieved by sole exonucleolytic trimming or by combination of endo- and exonuclease activities. To date, there is no evidence of any endonucleolytic cut generating the 3' end of an rRNA. 3' end maturation of the 18S rRNA is better understood. It is transcribed as a long precursor consisting sequentially in a ETS, the 18S rRNA, an Intergenic Transcribed Spacer (ITS), the 5S rRNA and a trailer sequence. Endocleavages at the 5' end of 18S rRNA and 5S rRNA generate secondary precursors. The 18S rRNA-ITS secondary precursor accumulates upon down-regulation of mitochondrial PNPase, suggesting that PNPase has an essential role in the 3' processing of the 18S rRNA (see Figure i6). Similarly, PNPase was shown to be essential for chloroplastic 23S rRNA 3' end maturation (Walter *et al,* 2002).

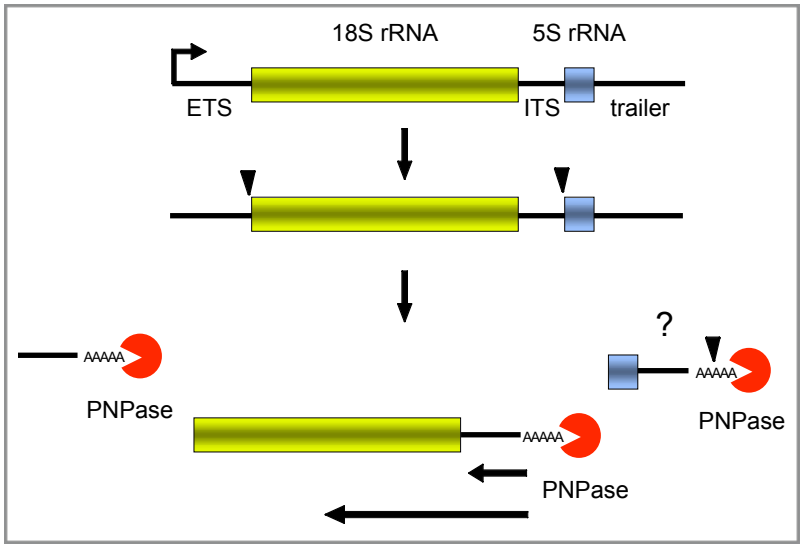


Figure i6 : Maturation of 18S and 5SrRNA in Arabidopsis. The endonucleolytic cleavages are indicated by triangles. The RNA fragments generated by endonucleolytic cuts are polyadenylated and subsequently degraded or matured by PNPase (arrows at the bottom). The 5SrRNA maturation process is unknown for the moment (from Gagliardi & Binder, 2007).

1.4.3.3. mRNA maturation.

5' and 3' ends of messenger RNAs are diverse and do not contain identified conserved motif (Forner *et al*, 2007), except for a short conserved sequence in 5' of some mRNAs and 26S rRNA (Schuster and Brennicke, 1989), indicating that other factors must determine the mature 5' and 3' generation. A *cis*-element shared by few transcripts is a 3' end stem-loop structure. In bacteria and human mitochondria, such stem-loops serve as signal for transcription termination. It is not the case in plant mitochondria, where stem-loops rather play a role in transcripts 3' end processing. *In vitro* assays on pea transcripts showed that a 3' to 5' exoribonuclease trims the transcript until the stem-loop (Dombrowski *et al*, 1997). The ability of stem-loops to impede the progression of 3' to 5' exoribonucleases such as PNPase and RNase II also plays a role for RNA stability (Bellaoui *et al*, 1997, Kuhn *et al*, 2001). However, 3' stem-loop containing RNAs are exceptions and most transcripts do not have such a structure but are nevertheless processed (Forner *et al*, 2007). This implies that other *cis*-elements determine RNA 3' end processing.

Cis-element involved in mRNA processing are likely to be recognized by *trans* factors. The CMS/Restoration factor (Rf) system provides a good tool to identify such factors. Indeed, the CMS-linked transcript processing is often affected upon restoration of the fertility by the Rf factor. To date, in all the cases where the Rf was identified, it was found to be a PPR protein (Hanson and Bentolila, 2004). For example, it has been demonstrated that a PPR protein is responsible for the processing of the CMS-associated transcript in BT (Boro II) rice (Wang *et al*, 2006).

Investigation of transcripts processing is complicated by the heterogeneity of the transcript population because mitochondrial genes exhibit multiple transcript initiation and processing sites. Primary and processed transcripts are distinguished by their 5' tri- or monophosphate, respectively, detected by *in vitro* capping experiments using guanylyltransferase (Binder *et al*, 1993a) or modified 5' RACE PCR technique (Kühn *et al*, 2005). As there is no evidence for a 5' to 3' exoribonuclease, 5' processing is likely performed by endonucleases. That implies that a *cis*-element has to be recognized, whether it is structural or in the RNA sequence. This applies particularly to the polycistronic units that are matured into monocistronic transcripts. In human mitochondria, processing of the majority of polycistronic units is believed to be generated by the tRNA processing machinery (Ojala *et al*, 1981). This was also described in plant mitochondria, for *atp6* mRNA that is cotranscribed with tRNA or tRNA-like structure in radish and *Brassica napus* : the cleavage in 3' of the tRNA is responsible for the 5' maturation of *atp6* (Makaroff *et al*, 1989; Singh and Brown, 1991). Recently, it was confirmed that tRNA processing machinery also contributes to mRNA 5' or 3' end formation, when tRNA or tRNA-like structures are cotranscribed with mRNAs and matured by endocleavages (Forner *et al*, 2007).

A recent study showed the importance of distant upstream sequences for the 5' maturation of the *cox3* transcript : the mature *cox3* main 5' extremity is con-

served in *Ler, Col* and C24 ecotypes. Only in C24, where the upstream sequence differs from other ecotypes, an additional 5' end is observed. Surprisingly, this additional 5' extremity is maternally inherited, and therefore, likely due to mitochondrial genome. This indicates that distant upstream genomic sequences can influence 5' extremities by a unknown mechanism (Forner *et al*, 2005).

Thus, maturation of plant mitochondrial transcripts relies on exonucleolytic activities and probably also on yet unidentified endonucleolytic ones. At least endonucleases of the tRNA machinery contribute to 3' and 5' ends formation when a signal for this is present. Transacting factors, such as PPR proteins, are required to mediate processing of mitochondrial transcripts and other yet unknown mechanisms might also play a role.

1.4.4. 3' addition of nucleotides.

Some processed transcripts carry non genomically encoded nucleotides in 3'. That is the case of tRNA with addition of CCA. Some transcripts also contain non-encoded nucleotides, apart from non-constitutively added polyadenosine tails, that will be discussed in details later. For example, analysis of the maize *rps12* mRNA showed that 95% of the clones representing truncated species contain between 1 and 4 added residues in 3', and that 83% are CCA. 60% of *cox2* mRNAs were also presenting additional residues, mostly A and C, but also G and U (Williams *et al*, 2000). Kuhn and colleagues also detected maximally 3-nucleotide-long nonencoded extensions of pea *atp9* transcripts, most frequently of adenosines combined with cytidines (Kuhn *et al*, 2001). It is not clear if these additions have any function but the prevalence of CCA added argues for a non-specific activity of the nucleotidyl transferase maturing tRNAs. Primer extension assays showed that these transcripts with a high degree of 3' modification are low abundant, suggesting that these RNAs are prone to degradation. Similarly, addition non-encoded nucleotides, including CCA, have been described in 3' of a chloroplastic transcript, *ndhD* mRNA (Zandueta-Criado and Bock, 2004).

1.5. Translation and post-translational processes.

Expression of mitochondrial mRNAs requires a fully functional machinery, including ribosomes, tRNAs and translating factors, but also post-translational modifications and specific protein degradation for their regulation.

Mitochondrial genome encodes only 3 rRNAs, a few ribosomal proteins and an incomplete set of tRNAs. Therefore, the greater part of the translating machinery has to be imported from the cytoplasm, including translation factors, aminoacyl-tRNA synthetases, most of the ribosomal proteins and missing tRNAs (Salinas *et al,* 2006).

Typically, the translation initiation codon is an AUG in plant mitochondria. Alternative start codons have been described, as for instance, GTG for the sunflower *cob* gene (Siculella *et al*, 1998) or ACG for the maize *cox2* gene (Dong et al, 1998). However, the mechanism of translation initiation is not known to date. Neither Shine-Dalgarno sequence, nor 5' cap structures are present in mitochondrial mRNAs. Ribosomes are unlikely to scan linearly the mRNA as long 5' UTRs contain many AUGs. Few data are known on translation initiation : some PPR proteins are translational activators (reviewed in Andrés *et al*, 2007) with similar roles to those described in chloroplast (Schmitz-Linneweber e*t al*, 2005). Editing can create new initiation codons and thus regulate translation (Chapdelaine and Bonen, 1991).

Trans factors are indispensable for translation. Some of them have been identified by sequence homology to yeast factors, like mEF-Tu, TS and G elongation factors and also an EF-1-alpha (At1g07920), that is found in *Arabidopsis* proteome (Heazlewood *et al*, 2004).

In addition to translational regulation, post-translational processes contribute to protein maturation. This is the case for legumes COX2 which undergoes presequence maturation steps (Daley *et al,* 2002) and for the RPS2 maize protein of which the C-terminal part is cleaved off to give rise to the mature protein (Perrotta *et al*, 2002). General analysis of post-translational modification has just started (Millar *et al*, 2005).

Unspliced or partially spliced mRNAs are often detected and can be particularly abundant in early stages of germination as shown in wheat (Li-Pook-Than *et al,* 2004). Partially edited or unspliced transcripts can be detected in polysomes, indicating a lack of mRNA quality-control during translation (Phreaner *et al*, 1996 ; Lu and Hanson, 1996).

Degradation of proteins is the last crucial point of control of genetic expression, as unassembled or damaged (by reactive oxygen species, for example) proteins have to be eliminated, in addition to the usual turnover. For this purpose, a whole assortment of proteases is found in mitochondria : the proteases responsible for the presequence degradation of imported proteins, but also Clp, Lon and FtsH, or AAA proteases. Gene regulation at post-translational level exists for example in the bean CMS system where the ORF239 protein is selectively degraded in vegetative organs by a Lon-type protease (Sarria *et al,* 1998).

In conclusion, many progresses have been made in the recent years in the comprehension of plant mitochondria post-transcription and translation. PPR proteins obviously play an important role for several aspects that need specificity including splicing, editing, stability and translation initiation. Transcription control appears to be relaxed in plant mitochondria and post-transcriptional steps are crucial. Steady-state level of transcripts is defined by the equilibrium between transcription and degradation rates. RNA degradation is thus at the center of plant mitochondria gene expression.

II. Polyadenylation and RNA degradation.

RNA degradation is essential for cell viability. It has a housekeeping role in the normal turnover of mRNA that adjusts the message population to the needs in specific proteins and can play an anti-viral defense role. Moreover, defective transcripts can be removed by quality-control mechanisms. Degradation enzymes also contribute to processing primary transcripts into functional molecules.

The major effectors of these processes are ribonucleases (RNases). RNases can be separated in two categories : the enzymes cleaving inside the RNA molecule, named endoribonucleases, and the enzymes trimming the RNA molecule from one of its extremities, named exoribonucleases. Exoribonucleases can be differentiated by their enzymatic mechanism. Hydrolytic RNases use water to cleave the nucleotide-nucleotide bond, releasing nucleoside monophosphates. Phosphorylases use inorganic phosphate to cleave the same bond and release nucleoside diphosphates (see Figure i7).

These enzymes are modulated and activated by transacting factors that direct access to the RNA, such as helicases that unwind RNA molecules and are found associated with RNases throughout all the living kingdom. RNA stability is also influenced by *cis*-elements that regulate accessibility to degrading machines. These elements can be intrinsic features as secondary structures, or specific modification, as polyadenylation, that plays an essential role in RNA degradation regulation by targeting RNA for degradation.

The following part of the introduction will first give an overview of prokaryotic RNA degradation as the basic knowledge in this field comes from bacteria. I will then detail similitudes and differences in prokaryotic-derived organelles, describing differences between mitochondria from various organisms. Finally, I will review eukaryotic RNA degradation, focusing on the exosome.

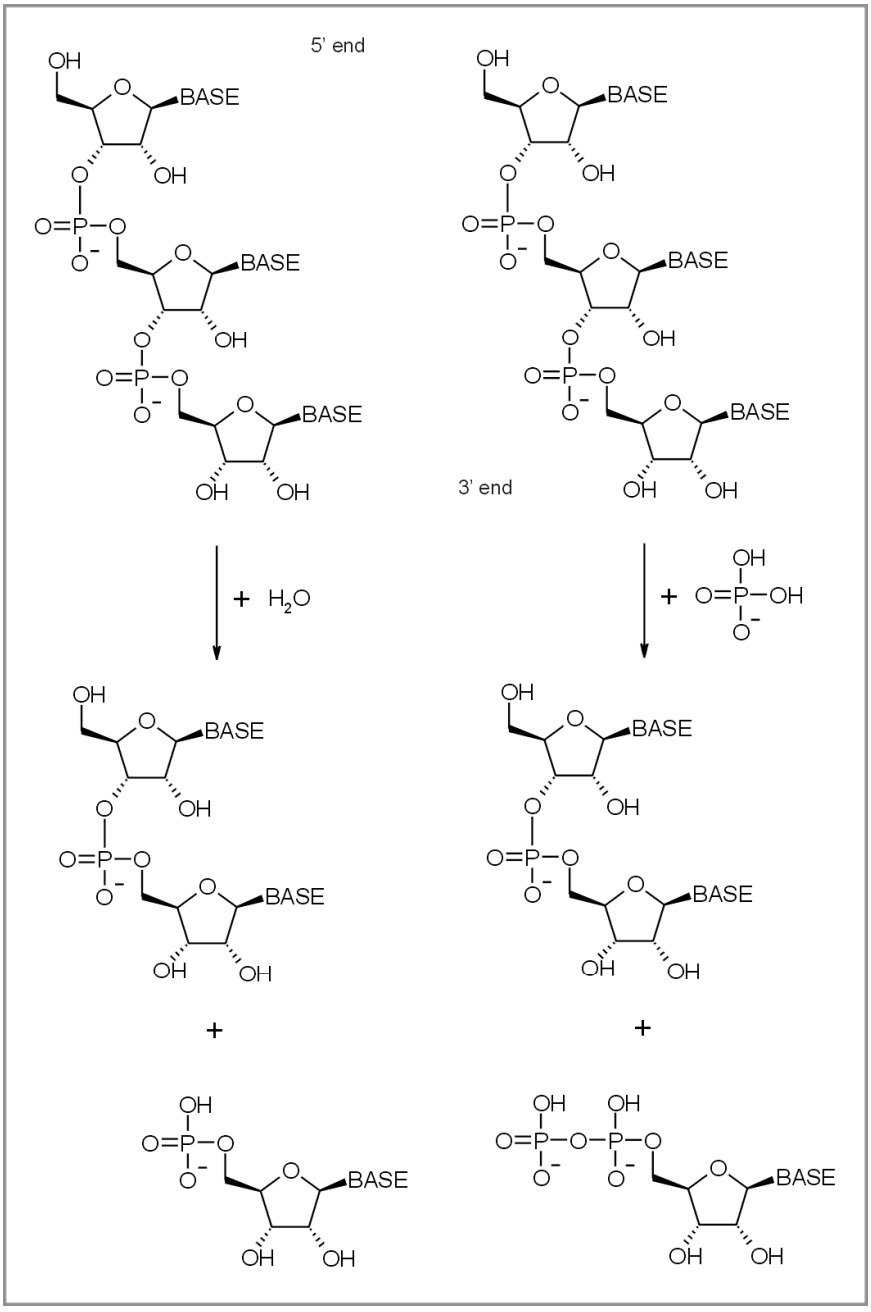


Figure i7 : Reaction diagrams for both hydrolytic (left) and phosphorolytic (right) 3'-5' exoribonuclease degradation of RNA.

2.1. Overview of RNA degradation in bacteria.

2.1.1. mRNA decay.

Endonucleolytic cleavages are essential for mRNA decay in bacteria. Released fragments are then eliminated by 3' to 5' exoribonucleases. No 5' to 3' exoribonuclease has been identified in *Escherichia coli* to date (Figure i8).

RNase E, an essential enzyme, initiates endonucleolytic cleavage in *E. coli*, inactivating the messenger for translation in a rate-limiting step. Following this

event, additional downstream cleavages between ribosome-protected areas are directed by the same enzyme. A number of other endoribonucleases also participate in mRNA decay but with a minor contribution, as they are more specialized in other processes. These include RNase G, homologue of RNase E, RNase III, specifically acting on double-stranded structures, RNase P and RNase Z, predominantly acting on tRNA precursors.

RNase E prefers accessible RNA 5' extremity : at least 4 single-stranded nucleotides at the 5' end are required for its activity and 5' monophosphate residues are also favored upon triphosphate residues. It has been recently shown that, in *E. coli*, a pyrophosphatase acts on mRNAs to enable RNase E to launch its attack, in a way that strikingly resembles eukaryotic decapping mRNAs (Celesnik *et al*, 2007). This work explains the 5' termini influence on RNase E cleavage and shows for the first time that *E. coli* mRNA decay initiation is not the endonucleolytic cleavage itself but the rate-limiting conversion of triphosphorylated to monophosphorylated 5' terminus.

Bacterial mRNA 3' ends are protected against 3' to 5' exoribonucleases by secondary structures, often Rho-independent transcription terminator. The fragments generated by endocleavages have new unprotected 3' extremities that are sensitive to 3' to 5' exoribonucleases degradation.

In *E. coli*, three processive exoribonucleases predominantly degrade mRNAs : PolyNucleotide Phosphorylase (PNPase), RNase II and RNase R. PNPase is a phosphorolytic enzyme whereas RNase II and RNase R possess hydrolytic activity. RNase II degrades only single-stranded RNA and is completely inhibited by secondary structures. RNase R and PNPase are not impeded by secondary structures, as long as a 3' toehold is there. This toehold is often provided by poly(A) tails. Polyadenylation is catalyzed by a poly(A) polymerase, PAP1.

Interestingly, even if they have distinct catalytic properties, none of the 3 exoribonucleases is essential, suggesting overlapping roles. However, bacteria can survive in the absence of RNAs R and RNase II, but not when both PNPase and RNase R or PNPase and RNase II are lacking.

Stable RNA structures, as repeated extragenic palindromic (REP) containing mRNAs, may require several rounds of polyadenylation and exoribonucleolytic trimming by PNPase/RNase R. On the other hand, RNaseII can antagonize PNPase/RNase R degradation by eliminating poly(A) tails in 3' of the stable secondary structure, thus eliminating the toehold for PNPase/RNase R.

mRNA can also be targeted for degradation by non-coding RNAs (ncRNAs) that have a role of translational regulators of gene expression. The translation initiation of the target RNA is inhibited and target and regulatory RNA are both degraded. Degradation can in this case be initiated by RNase III or RNase E.
The short 2-5nt that are produced by exoribonucleases are further degraded by oligoribonuclease (Orn), an essential exoribonuclease in *E. coli* (see model in Figure i8 (a)).

In *Bacillus subtilis*, the situation is different as key enzymes of *E. coli* RNA degradation are lacking : *B. subtilis* has no homologs of RNase II, Orn and RNase

E. In contrast to *E. coli* where 3' to 5' degradation is primarily hydrolytic, it is principally a phosphorolytic process in *B. subtilis.* Although PAP1 homologs are lacking, polyadenylation has been observed in *B. subtilis*. Until very recently, 5' to 3' exoribonucleolytic activity was thought to be absent from Bacteria but the essential enzyme J1 (and its paralog J2) of *B. subtilis* is now known to bear both an endonucleolytic and a 5' to 3' exonucleolytic activity (Mathy *et al*, 2007). Thus, mRNA decay can occur from either the 5' or the 3' extremity, as in eukaryotes. As the 5' to 3' exonucleolytic activity of RNase J1 prefers single-stranded monophosphorylated 5' ends, it may degrade the products of its own endonucleolytic activity (Figure i8 (b)).

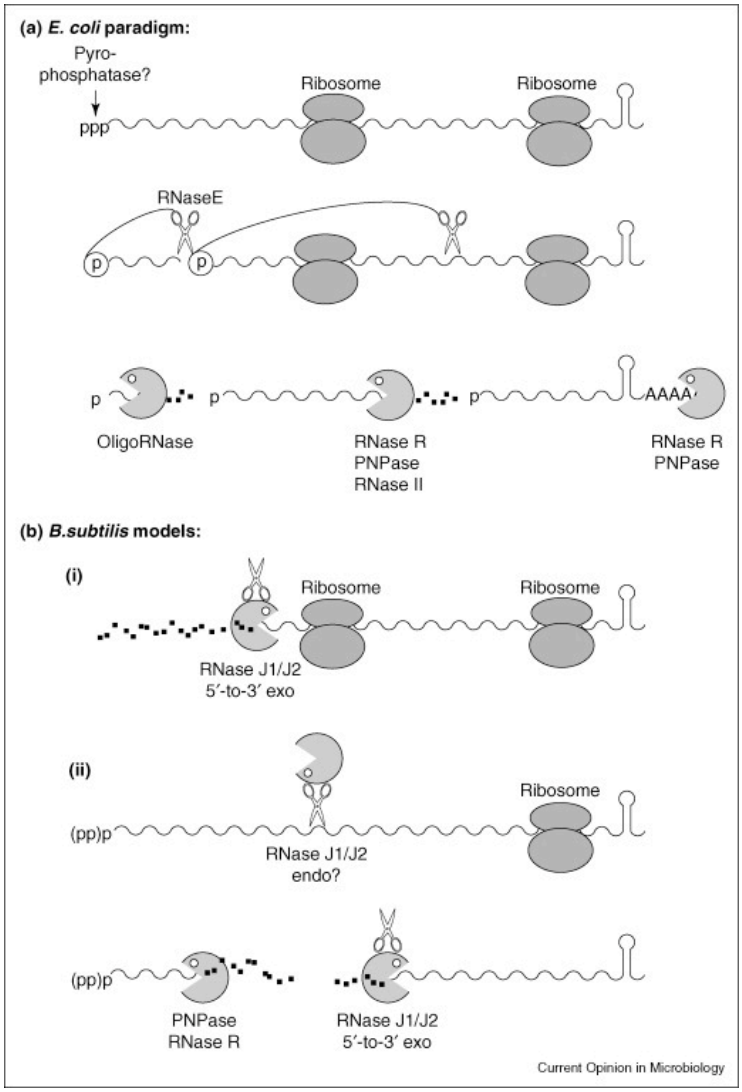


Figure i8 : Principal mRNA decay pathways in *E. coli* (a) and *B. subtilis* (b). Endonucleolytic activity is represented by scissors and exoribonucleolytic activities by "pacmans". Two different models (i) and (ii) are possible for *B. subtilis* (from Condon, 2007).

2.1.2. The RNA degradosome.

A small portion of both RNase E and PNPase is found in a multimeric protein complex called RNA degradosome. Other components are RhlB, an RNA helicase, and enolase, a glycolytic enzyme whose role in mRNA degradation is still mysterious. Structural analyses have shown that RNase E is a large multi-domain protein with an N-terminal catalytic region and a C-terminal non-catalytic region that is mostly natively unstructured. (Callaghan *et al*, 2005). This unstructured region provides a platform for the assembly of the other components of the RNA degradosome and contains RNA binding sites. The composition of the degradosome is variable and can be modulated, e.g. in response to cold shock. Alternative forms containing the CsdA cold-shock protein, polyphosphate kinase (PKK), DnaK or GroEL heat shock proteins have been described. Multiprotein complexes containing RNase E were also identified in Proteobacteria although the composition of the associated proteins appeared to be different. For example, the degradosome-like complex of *Pseudomonas syringae*, harbors RNase R instead of PNPase (Carpousis, 2007).

Though the association of an endoribonuclease, an exoribonuclease and an RNA helicase in a complex represents a somehow ideal machinery to degrade RNAs, the assembly of the exosome is not essential in *E. coli* and does not affect processing of ribosomal and transfer RNA (Ow *et al*, 2000). However, microarray analyses have shown that the degradosome is required for degradation of some mRNAs (Bernstein *et al*, 2004). This indicates that other pathways than the degradosome-involving pathway exist to achieve the same processes or partially overlap. Clearly, degradation by the exoribonuclease RNase R is a major pathway in *E. coli* RNA degradation (Cheng and Deutscher, 2005).

2.1.3. Degradation of stable RNAs.

Beside mRNAs, stable RNAs, namely rRNA and tRNA, are also submitted to degradation, predominantly upon stress conditions. Stable RNA are generally not turned over during exponential growth. However, with 98% of cellular RNA, they represent a storehouse of nutrients that can be used in case of starvation.

The mechanisms that protect ribosomal RNA during normal conditions but allow degradation under slow-growth conditions are not known. RNase I is a non specific endoribonuclease, most active in the absence of divalent cations that is primarily in the periplasmic space of *E. coli.* Cell permeabilization triggers extensive RNA degradation by allowing the entry of the RNase I and the loss of ribosome-stabilizing ions such as Mg^{2+}. PNPase can be induced by cold shock in *E. coli* (Beran & Simons, 2001). RNase R levels are also dramatically increased in response to a variety of stress conditions (Chen & Deutscher, 2005).

Severe stress conditions can lead to activation of the toxin-antitoxin systems in *E. coli.* Toxins and antitoxins are encoded by the same operon. The toxins are

RNases or facilitors of RNase activity. The antitoxin binds the toxin and blocks its activity but it is very labile and requires constant *de novo* synthesis. An arrest of translation will lead to activation of RNase, leading to destruction of most cellular RNAs by a phenomenon called shut-down decay (SDD) (Condon, 2006).

2.1.4. Role of polyadenylation in Bacteria.

The role of polyadenosine tails in stabilizing eukaryotic mRNA is well known. On the contrary, polyadenylation in prokaryotes has not been understood until 1995 (Hajnsdorf *et al*, 1995 ; O'Hara *et al*, 1995). This is due to the difficulties to detect polyadenylated RNA, leading to the idea that polyadenylation is a eukaryotic feature.

Nevertheless, the importance of polyadenylation for the regulation of bacterial mRNA stability is now well established. It provides a toehold for PNPase and RNase R exoribonucleases and this process is essential for the elimination of structured RNAs. Thus, polyadenylation is an indispensable tool for the turnover of transcripts stabilized by a 3' hairpin or other secondary structure.

Aberrant stable RNA are degraded by quality-control processes. Quality-control occurs mostly at the level of RNA precursors. In the absence of poly(A) polymerase, large amounts of defective precursors accumulate, indicating that aberrant precursors are polyadenylated prior to their degradation. Degradation then is carried out primarily by PNPase (Li *et al*, 2002). In absence of PNPase and RNase R, 16S and 23S rRNA accumulate, implying that they would normally be degraded by these RNases. These data indicate that the pathway involving polyadenylation and PNPase/RNase R is essential for bacterial RNA quality-control.

Polyadenylation is also involved in mediating stable RNA quality-control. As polyadenylation marks aberrant precursors for degradation, the question is : how does the poly(A) polymerase discriminate between aberrant and normal RNAs? As aberrant tRNAs are misfolded, structure recognition certainly plays a role at one point but this question needs further investigation.

2.2. RNA stability and degradation in the chloroplast.

Chloroplast are photosynthetic organelles that arose from prokaryotic endosymbionts, and therefore contain their own genome and gene expression machinery, as mitochondria. Gene expression in chloroplasts is preferentially modulated by mRNA stability rather than by transcriptional control. RNA degradation in chloroplasts is believed to follow the bacterial model : an endonucleolytic cleavage, initiates degradation (Bollenbach *et al*, 2004). Addition of a poly(A) or poly(A)-rich tail follows and triggers rapid degradation by PNPase and possibly other enzymes. A model of degradation pathway in chloroplasts is shown in Figure i9.

Cis-acting elements can act as positive or negative regulators of chloroplast RNA stability. Inverted repeat (IR) sequences at 3' ends of mRNAs form stem-loop

structures and impede the progression of exoribonucleases, thereby prolonging mRNA lifetime (Stern and Gruissem, 1987). However, CSP41a and CSP41b (Chloroplast Stem-loop binding Proteins), mediating endonucleolytic cleavage, recognize specifically these stem-loops (Bollenbach and Stern, 2003).

5' UTR elements serve as binding sites for nucleus-encoded protein complexes that protect chloroplast RNAs from 5' to 3' degradation activity (Drager *et al*, 1999; Higgs et al, 1999 ; Vaistij et al, 2000, Raynaud *et al*, 2007). These complexes resemble the cap structure protecting cytoplasmic mRNAs, as the removal of these stabilizers triggers degradation of the RNA. In addition to stabilizing complexes, Nishimura and colleagues showed that antisense RNA in chloroplasts can protect otherwise unstable transcripts from 3' to 5' exonucleolytic activity (Nishimura *et al*, 2004). They suggested that protection by antisense transcripts may occur naturally in chloroplast genome. An example of naturally occuring antisense transcript in chloroplast was given by the work of Zandueta-Criado and Bock, showing that the sequence upstream of the stem-loop at the 3' end of the *ndhD* transcript is used as a template to polymerize an extra sequence found at the transcript terminus (Zandueta-Criado and Bock, 2004).

The role of polyadenylation was first shown *in vitro*, using lysed chloroplasts. Upon treatment with poly(A) inhibitor cordycepin (3'-dATP), endonucleolytic cleavage products accumulated (Litsisky *et al,* 1997). *In vivo* confirmation was then brought by the work of Komine and colleagues using green fluorescent protein (GFP) fusion experiments. A polyadenylated green fluorescent protein gene was introduced in *Chlamydomonas reinhardtii* chloroplast such that the 3' end poly(A) tail would be exposed after RNase P cleavage. Indeed, no GFP protein nor polyadenylated GFP transcript could be detected in this strain, while the stability of the tagged mRNA and expression of GFP were relatively high in strains expressing a control construct either lacking a poly(A) tail or containing a heteropolymeric (A + U) tail (Komine *et al*, 2002).

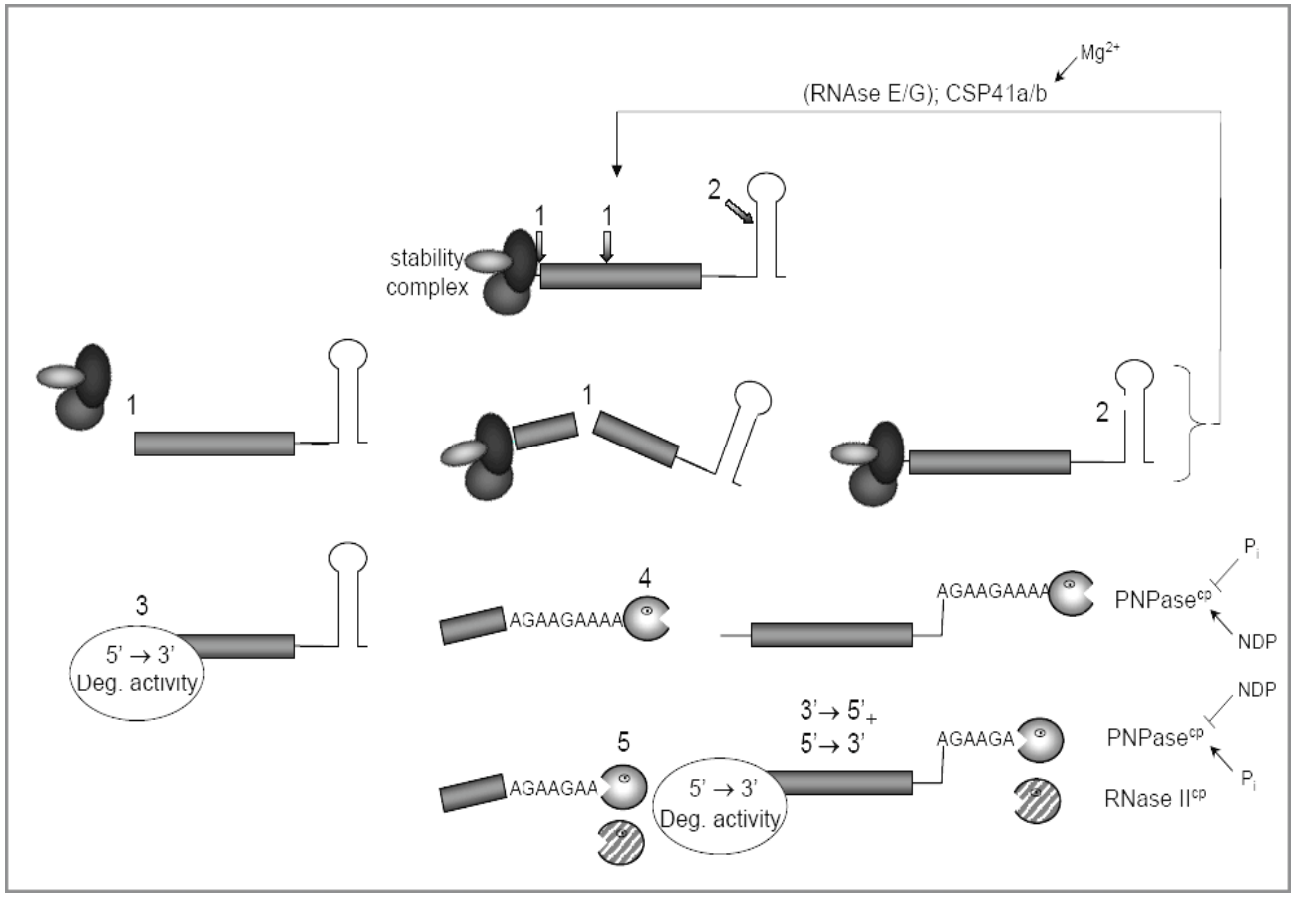


Figure i9 : Model for chloroplast mRNA degradation. Degradation is intiated by endonucleolytic cleavages within the 5'UTR, the coding region (1) or the 3' UTR (2) by either a RNase E/G-like activity, CSP41a or CSP41b. Cleavage at these positions yield distal products that may undergo further rounds of endonucleolytic cleavage, and proximal products that may be degraded by a 5' to 3' activity (3) or may be polyadenylated by PNPase (gray pacman) (4). The polyadenylated RNA is then degraded in the 3' to 5' direction by PNPase or RNaseII/R (striped pacman) (5), or simultaneously in both directions. Regulation of PNPase activity by NDP and Pi and of CSP41 by Mg^{2+} are indicated (from Bollenbach *et al*, 2004).

In contrast to short and homopolymeric poly(A) tails of *E. coli*, poly(A) tails in chloroplast can reach several hundreds nucleotides. Moreover, spinach chloroplast poly(A) tails contain about 70% adenosine, 25% guanosine and 5% of both cytidine and uridine. Most polyadenylation sites are found within the transcript, at positions corresponding to endonucleolytic cleavage sites (Lisitsky *et al*, 1996). Mature transcripts are only occasionally polyadenylated at the 3' end. It may be possible that the 3' stem-loop, by reducing polyadenylation-triggered degradation in 3' of mature transcripts, forces endonucleolytic degradation (Slomovic *et al*, 2005).

Long and heteropolymeric poly(A) tails were detected in *E. coli* upon deletion of poly(A) polymerase and were synthesized by poly(A) polymerizing activity of the PNPase (Mohanty and Kushner, 2000). It was suggested that PNPase could similarly be responsible for the poly(A) tails in the chloroplast. Indeed, purification of PAP activity from spinach chloroplasts yielded only PNPase, suggesting that

PNPase is responsible for both polyadenylation and degradation of RNA in spinach chloroplast. The reaction catalyzed by PNPase is reversible and is directly influenced by the Pi/NDP ratio as presented in the Figure i10 (Yehudai-Resheff *et al*, 2001). In *Arabidopsis* chloroplast, the poly(A) polymerase is unliklely to be PNPase as non-encoded tails are homopolymeric, but rather resembles the *E. coli* Ntr-PAP, derived from an Nucleotidyl transferase (Ntr). A bioinformatic study found several candidates for Ntr-PAP in Arabidopsis which was predicted to be imported into chloroplast and mitochondria (Martin and Keller, 2004). It is noticeable that both PNPase and *E. coli* PAP, being of the Ntr family, are able to add specifically C or A to RNA without DNA template (Yue *et al*, 1996).

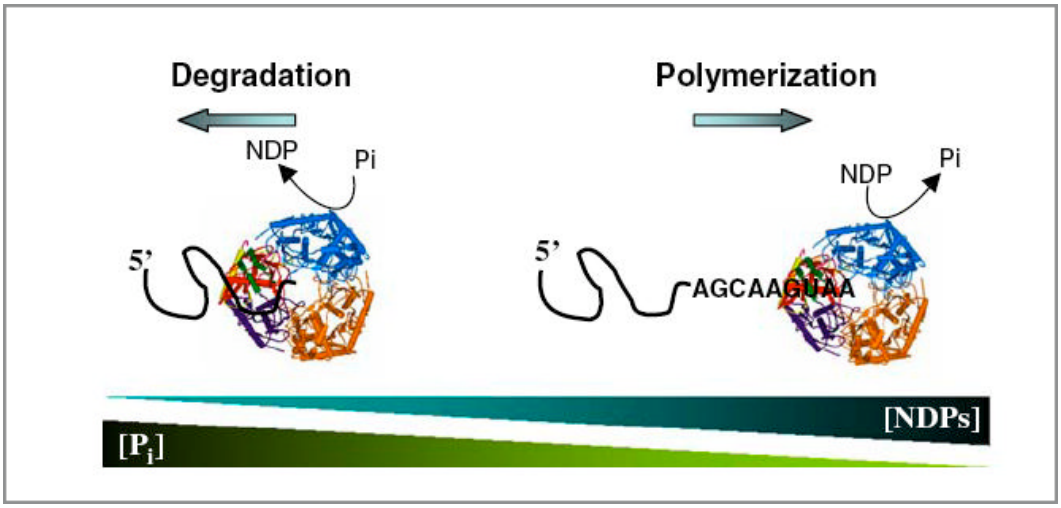


Figure i10 : Model of the possible modulation of enzymatic activities of PNPase as an exoribonuclease or a polymerase. When degrading RNA, it consumes Pi and produces nucleotides, therefore increasing the local concentration of nucleotides (left side of Figure). When polymerizing RNA, it consumes nucleotides and produces Pi (right side of Figure). PNPase is modeled as a homotrimer, according to crystal structure of the enzyme from bacteria *Streptomyces antibioticus* (from Lin-Chao *et al*, 2007).

Another argument supporting the idea that the poly(A) polymerase is distinct from PNPase in Arabidopsis chloroplasts is that polyadenylated RNAs accumulate upon down-regulation of PNPase (Walter *et al*, 2002). In the same work, new insights into PNPase functions in Arabidopsis chloroplast were brought. No phenotype was observed at the plant level upon PNPase transcript cosuppression but PNPase was shown to be indispensable for 3' end maturation of 23S rRNA transcripts. Although this enzyme was found to be essential for efficient 3' end processing of mRNAs, it was insufficient to mediate transcript degradation. In addition, this exoribonuclease was shown to function in tRNA decay (Walter *et al*, 2002).

Chloroplastic RNA degradation mechanism follows the bacterial model but 5' to 3' directed degradation plays an important role in RNA destabilization. Polyadenylation, crucial for 3' to 5' degradation, has been well studied in chloroplasts.

2.3. RNA degradation in mitochondria of various organisms : the role of polyadenylation.

As mitochondrial genomic expression systems have retained prokaryotic features, the basic principle of polyadenylation triggering RNA degradation could theorically be conserved in this organelle from all organisms. However, regulations of mitochondrial mRNA turnover are surprisingly diverse from one species to another.

2.3.1 Human mitochondria.

In human, the mitochondrial genome is transcribed as two large polycistronic RNAs that are processed by endonucleolytic cuts at tRNA extremities, which simultaneously releases mRNAs (Ojola *et al*, 1981). Some of the mRNAs do not encode a functional stop-codon but terminate with U or UA. The missing A(s) is (are) provided by 3' polyadenylation, that further extends the RNA by 50 to 60 residues. Thus, polyadenylation generates functional mRNAs. mt-tRNA are oligoadenylated and it has been suggested that polyadenylation is involved in editing of the 3'-acceptor region of some mt-tRNAs (Yakobori & Pääbo, 1997). However, the functional role of poly(A) tails in translation activation and turnover of human mitochondrial RNAs is not well understood.

The work of Temperley and colleagues (Temperley *et al*, 2003) showed that translation-dependent deadenylation enhanced degradation, suggesting that polyadenylation is required for stability of human mt-mRNA, as for cytoplasmic RNAs in eukaryotes. Conversely, Tomecki and colleagues reported that several mt-mRNA were stabilized upon deadenylation by down regulation of mt poly(A) polymerase (Tomecki *et al*, 2004). This suggests that the long poly(A) tails do not directly determine mRNA stability. Slomovic and colleagues reported that human mitochondria retained truncated and transiently polyadenylated transcripts in addition to RNAs harboring stable 3'-end poly(A) tails. They suggested that polyadenylation-dependent RNA degradation can occur in human mitochondria, and coexists with the stable 3'-end polyadenylation (Slomovic *et al*, 2005). Nagaike and colleagues suggested that the poly(A) length of human mitochondrial RNAs is controlled by both polyadenylation mitochondrial Poly(A) polymerase and deadenylation by hPNPase, and that polyadenylation is required for mt-mRNA stability (Nagaike et al, 2005). However, the recent demonstration that human PNPase is located in the mitochondrial inter-membrane space (IMS) tempers conclusions about the function of PNPase in mtRNA degradation in the matrix (Chen *et al*, 2006).

2.3.2 *Saccharomyces cerevisiae* mitochondria.

Yeast mitochondrial degradation system is particular. Indeed, mature transcripts are not polyadenylated in budding yeast but contain an encoded dodecamer sequence 5'-AAUAA(U/C)AUUCUU-3' in their 3'UTR. Primary transcripts are always matured two nucleotides downstream of the dodecamer and this sequence seems vital for mRNA stability and translatability (Butow *et a*l, 1989). A protein showing affinity to the dodecamer has been reported but its function remains unknown (Li & Zassenhaus, 1999). In fission yeast, a function analogous of the dodecamer function is attributed to the C-core motif, a conserved processing signal in 3' of the coding region. The C-core motif is conserved in all species of the genus *Schizosaccharomyces* (Schäfer, 2005). A complex comparable to the RNA degradosome, called mtEXO, has a NTP-independent 3' to 5' exoribonucleolytic activity. The mtEXO is composed of an helicase, Suv3p, that unwinds double-stranded RNA, and a single-strand specific RNase, Dss1p, similar to RNase II. As there is a strong accumulation of mitochondrial mRNA and rRNA precursors in strains lacking mtEXO, the mtEXO is believed to play a central role in degradation of aberrant and unprocessed RNAs, as part of a mitochondrial RNA surveillance system (Dziembowski *et al*, 2003). Rogowska and colleagues recently demonstrated that reduced transcription in yeast mitochondria rescues the phenotype of deficient RNA degradation, revealing the importance of maintaining the balance between RNA synthesis and degradation (Rogowska *et al*, 2006).

In fission yeast as in budding yeast mitochondria, no evidence for polyadenylation was found, indicating that yeast mitochondria have evolved a degradation pathway independent of polyadenylation.

2.3.3. Trypanosome mitochondria.

The protist *Trypanosoma brucei* is among the earliest branching eukaryotes and trypanosomal mitochondria are characterized by a unique arrangement of the mitochondrial DNA and posttranscriptional events that govern gene expression. mRNAs are massively modified by uridine insertion and deletion, a process called kinetoplast RNA editing (Simpson *et al*, 2003). This process require guide RNAs (gRNAs) that guide the insertion of uridines into mRNAs in trypanosomes. The 5' end of a gRNA hybridizes to a short region of an unedited pre-mRNA, called an anchor sequence, while its 3' end functions as a template for the editing process.

Trypanosome mitochondrial mRNAs can be polyadenylated. mRNA poly(A) tails length are developmentally regulated in a transcript-specific manner and coordinated with mRNA editing status : unedited RNA molecules can be extended with a short poly(A) tail (20nt) whereas partially and totally edited RNAs have either short or long (120-200nt) poly(A) tails (Militello and Read, 1999). The poly(A) tail length also fluctuates during life cycle progression. Moreover, polyadenylation plays a dual role in modulating RNA stability. In *in vitro* RNA turnover assays, the pres-

ence of a poly(A)$_{20}$ tail destabilizes unedited RNAs but the same modification stabilizes their partially and fully edited counterparts (Kao and Read, 2005). Another feature of trypanosomes is that degradation of polyadenylated RNAs in *T. brucei* mitochondria can occur through a unique mechanism that requires the polymerization of UTP into RNAs (Ryan and Read, 2005). Therefore, trypanosomal mitochondria appear to have evolved unique characteristics in editing, polyadenylation and RNA degradation. In addition, this is the only organism/organelle to harbor polyadenylated RNAs in absence of PNPase or similar phosphorolytic proteic complex. It is to notice that a sequence with significant homology to the yeast exoribonuclease Dss1p was found in the *T. brucei* genome (Ryan *et al*, 2003). Addition of stable poly(A) tails to full-length RNAs in organelles is unique to trypanosome and human mitochondria.

2.3.4. Higher plant mitochondria.

In higher plants mitochondria, the situation differs strikingly. Similar to the situation in *E. coli* and chloroplasts, mature transcripts are not constitutively polyadenylated. However, in 1999, two independent research groups discovered that transcripts can be polyadenylated in plant mitochondria (Gagliardi and Leaver, 1999 ; Lupold *et al*, 1999). In sunflower, the PET1-Cytoplasmic Male Sterility is correlated with the presence of a chimeric *atp1-orf522* transcript. Polyadenylation specifically triggered destabilization of this transcript in the male florets where fertility was restored (Gagliardi and Leaver, 1999). In contrast, polyadenylation did not alter the stability of maize *cox2* transcript *in vitro* (Lupold *et al*, 1999). Kuhn and colleagues later showed that the life span of pea *atp9* and *Oenopthera apt1* transcripts is balanced between the degradation-promoting polyadenylation and the stabilizing effect of the 3' stem-loop (Kuhn, 2001).

Two distinct ribonuclease activities were purified from mitochondria, one of them preferentially degraded polyadenylated substrates (Gagliardi and Leaver, 1999 ; Gagliardi *et al*, 2001). A reverse genetic approach subsequently identified the genes responsible for two 3' to 5' exoribonuclease activities in *Arabidopsis* mitochondria, with homology to *E. coli* PNPase and RNase II, respectively, and were thus named AtmtPNPase and AtmtRNaseII (Perrin *et al*, 2004a). In absence of AtmtPNPase, large transcripts accumulate and therefore AtmtPNPase could be involved in the removal of long 3' extensions, that may reach several kilobases. AtmtRNaseII degrades only unstructured RNA, and could act after AtmtPNPase, eliminating short nucleotidic extensions to generate the mature 3' ends.

RNase II can eliminate the poly(A) tail that triggers PNPase activity but is stopped by RNA structure. So, it could also have a role antagonist to PNPase, by eliminating PNPase toehold on the RNA, as it was already described in *E. coli* (Marujo et al, 2000 ; Mohanty and Kushner, 2003).

AtmtPNPase is essential, at least at some stages of development, in *Arabidopsis* as homozygous T-DNA insertion mutants are not viable. However, down-regulation of AtmtPNPase was observed in some plants where ectopic ex-

pression of the gene triggered cosupression. Plant mitochondria lacking AtmtPNPase accumulate polyadenylated *atp9* precursors but also polyadenylated *18SrRNA* fragments and precursors (Perrin *et al*, 2004b). Thus, it was concluded that AtmtPNPase substrates are polyadenylated and that AtmtPNPase is involved in processing and/or degradation of these substrates. Similarly to the prokaryotic system, the stem-loop structures at the 3' end of some transcript necessitate several rounds of polyadenylation and degradation to be eliminated (Kuhn *et al*, 2001). However, degradation is not impeded by internal secondary structures (Gagliardi *et al*, 2001). Endonucleases are possibly involved even if none has yet been identified. However, in contrast to bacteria, RNA 3' ends are prevalently polyadenylated in mitochondria, raising the idea that the 3' to 5' exoribonucleolytic activity of the PNPase is primordial for degradation.

Mitochondrial polyadenylated RNAs are low abundant, making their characterization difficult. That is why composition and size of poly(A) tails are not well documented. But circular RT-PCR experiments showed for example that 80% of unmatured *atp9* and *orfB* transcripts accumulating upon down regulation of the PNPase have tails of 5 to 17 residues (Perrin *et al*, 2004a). However, short ploy(A) tails could be preferentially amplified by this technique.

The enzyme responsible for the polyadenylation of transcripts in mitochondria is not characterized yet but bioinformatic studies have revealed some candidate genes (Martin and Keller, 2004).

In summary, polyadenylation roles and occurrence differ in mitochondria from humans, plants, trypanosomes and yeast. This indicates that various solutions were elaborated by distinct genetic systems during evolution to control RNA degradation.

2.4 Overview of eukaryotic mRNA decay.

The process of mRNA turnover is important for different aspects of eukaryotic mRNA functions : modulation of gene expression, quality control of RNA biogenesis, and anti-viral defense. Messenger-specific degradation by the silencing machinery for anti-viral defense and development functions is a field of research that has exploded these years. I will not detail here this important area of RNA degradation which is developed in the following reviews : Chapman and Carrington, 2007, Ding and Voinnet, 2007, Brodersen and Voinnet, 2006.

Most of the knowledge on housekeeping eukaryotic RNA degradation comes from *Saccharomyces cerevisiae*, where most of the components were identified. Mostly, mRNA degradation begins with deadenylation at the 3' end of the mRNA. Major components of the deadenylation pathway are Ccr4p-Pop2p-Notp complex and Pan2p-Pan3p complex in yeast. In mammals and vertebrates, an additional complex, the PARN complex, has a major role in deadenylation (Parker & Song, 2004). Following deadenylation, the 5' m^7G cap structure is removed by decapping

complexes, involving predominently Dcp1p and Dcp2p proteins. This exposes the unprotected 5' end of the transcript to digestion by 5' to 3' exonucleases as Xrn1p in yeast cytoplasm. Alternatively, deadenylated mRNA can be degraded in a 3' to 5' direction, by the exosome, a large protein complex that exists both in the nucleus and in the cytoplasm of eukaryotic cells. In this case, the cap structure is removed by the DcpS scavenger decapping enzyme (Figure i11).

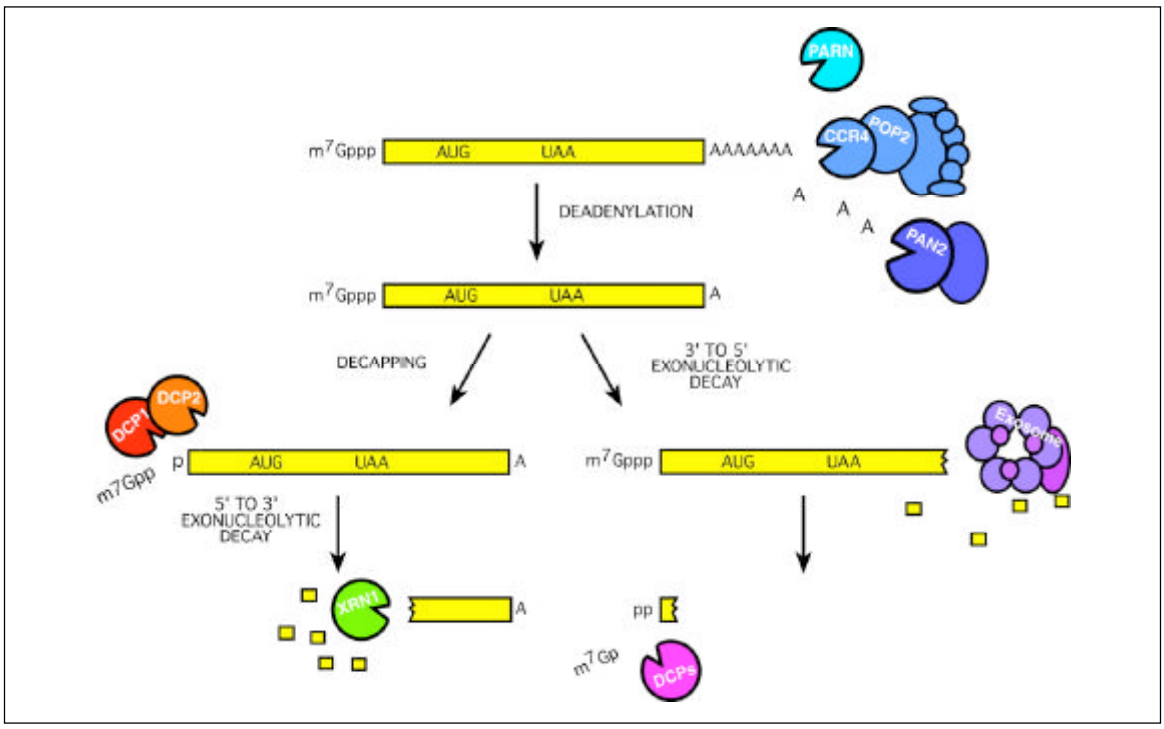


Figure i11 : Two general mRNA decay pathways. Both pathways are initiated by deadenylation by the Ccr4/Pop2/Not complex or possibly by the alternative deadenylases, Pan2/Pan3 and PARN. Poly(A) tail shortening can lead to either 3'-5' exonucleolytic digestion by the exosome or decapping by the Dcp1/Dcp2 complex. Decapping is followed by 5'-3' exonuclease digestion by Xrn1. The residual cap structure resulting from exosome digestion is cleaved by the scavenger decapping enzyme DcpS. From Decker & Parker, 2002.

5' to 3' and 3' to 5' exoribonucleolytic degradation pathways are both important for RNA decay. However, my interest focuses on polyadenylation and its involvement in 3' to 5' degradation. That is why the following paragraphs will detail an essential 3' to 5' degrading machine : the exosome.

2.5. The exosome.

2.5.1. Exosome structure.

The exosome is a ring-shaped complex conserved in all eukaryotes and in some archaea. The archaeal exosome is formed by a stable six-subunits ring composed of three heterodimers of ribosomal RNA processing (Rrp) factors,

Rrp41/Rrp42. A nine-subunit exosome also exists, where the ring is capped by three copies of cep1 synthetic lethality (Cls)4 and/or Rrp4 subunits (Büttner *et al*, 2005). Each of the ring subunits harbors a RNase PH domain but only Rrp41 subunits are catalytically active (Lorentzen *et al*, 2005). Cls4 and Rrp4 subunits harbor the KH/S1 RNA binding domains. Similarly to PNPase, the active sites of Rrp41/Rrp42 heterodimers face the interior of the ring, leading to a proposed model for RNA degradation where the substrate is recruited by the RNA binding-domain-containing subunits and then channelled into the central cavity to reach the active sites. The channelling through a constricted pore near RNA-binding sites limits access to single-stranded substrates (Lorenzten et al, 2007).

The eukaryotic exosome shows an overall architecture similar to its archaeal counterpart. Rrp41-like (Rrp41, Rrp46 and Mtr3) and Rrp42-like (Rrp42, Rrp43 -human OIP2- and Rrp45 -human PM-Scl-75) subunits form a RNase PH-like ring, capped by KH and/or S1 domain-containing subunits Rrp4, Rrp40 and Cls4. In contrast to archaeal exosome, the eukaryotic exosome require all nine subunits to form a stable structure. While the human core exosome contains only these 9 subunits, the yeast exosome includes a tenth subunit, Rrp44, that is a RNase II-like protein (Liu *et al*, 2006). Although phosphorolytic activity was reported for yeast Rrp41 and human Rrp43 (Mitchell *et al*, 1997; Jiang and Altman, 2002), it was recently shown that only Rrp44 (also named Dis3), was responsible for the activity of the yeast core exosome (Dziembowski *et al,* 2006). Similarly, the human nine-subunit exosome, is thought to be devoid of phosphorolytic activity (Liu *et al*, 2006 and 2007). The architecture of the yeast Rrp44-exosome complex suggests 2 possible routes for RNA 3' processing : RNA can go through the exosome channel and/or enter Rrp44 directly, without interaction in the central channel (Wang *et al*, 2007). Interestingly, in *Arabidopsis thaliana*, in addition to the RNase II-related AtRrp4 hydrolytic activity (Chekanova *et al*, 2002), a phosphorolytic activity is attributed to the exosome ring-subunit AtRrp41 (Chekanova *et al,* 2000).

Phosphorolytic RNases of the RNase PH superfamily can be found in a multimeric structure throughout evolution. In bacteria, RNase PH, involved in tRNA precursors 3' end trimming, forms an homohexameric complex (Choi *et al*, 2004) that has a "doughnut" shape surrounding a central channel that can accommodate a single-stranded RNA molecule. This structure is kept by prokaryotic/organellar PNPase (Symmons *et al,* 2000; Yehudai-Reseheff *et al,* 2003) and eukaryotic/archaeal exosome (Aloy *et al,* 2002; Büttner *et al*, 2005; Lorentzen *et al*, 2005), both composed of RNase PH domain subunits (see Figure i12).

PNPase and exosome ancestor certainly arose from gene duplication giving rise to duplicated core domains. In PNPase the two domains remained linked on the same polypeptide, followed by two RNA-binding domains (KH and S1 domains). 3 of such proteins form an homotrimeric structure. The catalytic activity was predicted to be harbored only by the second core domain in bacteria but the first domain exhibits degrading activity in spinach chloroplast (Yehudai-Resheff *et al*, 2003). Structure and domains of the exosome and its counterparts are presented in Figure i12.

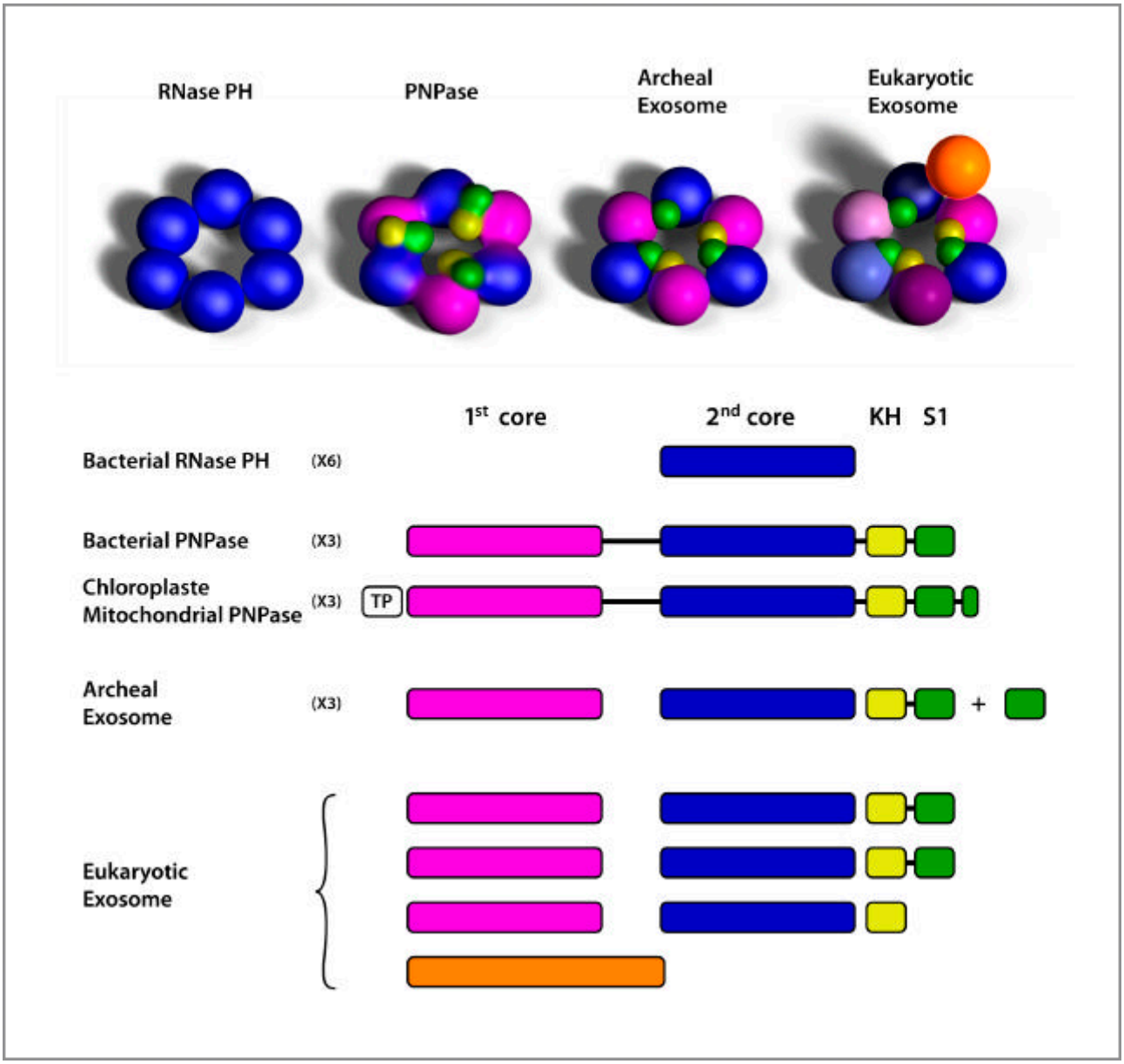


Figure i12 : Structure and domains are conserved from RNase PH to PNPase and exosome. The first and second RNase PH core domains are shown in pink and blue, respectively, KH domains in yellow and S1 domains in green. PNPase is a homotrimer, in contrast to other complexes where core domains are harbored by dinstinct subunits. The RNase II-related protein of the eukaryotic exosome is shown in orange. The domain diagram is adapted from Lin-Chao et al, 2007.

Nuclear and cytoplasmic exosomes differ by specific factors associated : the GTPase Ski7 in the cytoplasm, and Rrp6 and Rrp47 in the nucleus, a RNase D-like enzyme and a putative nucleic acid binding protein, respectively.

2.5.2. Functions of the exosome.

In the cytoplasm, the exosome is required for mRNA turnover, with the help of the superkiller (Ski) complex. Yeast Ski2p, Ski3p and Ski8p are a putative RNA helicase, a tetratricopeptide-repeat protein and a protein containing WD motif, respectively, whereas Ski7p directly mediates interaction between the complex formed by these 3 proteins and the exosome (Araki *et al*, 2001). The cytoplasmic exosome is also required for cellular defense against pathogens. For instance, it controls the levels of the LA virus double-stranded RNA in yeast (Butler, 2002).

In addition to its role in mRNA turnover, the exosome is a key component of the RNA-surveillance machinery and degrades misprocessed, misadenylated and defective transcripts in the nucleus and aberrant transcripts in the cytoplasm (see next paragraph). The nuclear Rrp6-exosome has an action coordinated with processing factors to regulate retention at the transcription site of defective mRNA, or release from this site to export or decay (Rougemaille et al, 2007). Another role of the exosome, in association with Rrp6, is precise 3' extremity maturation of small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs) and 5,8S RNA ribosomal RNA (Briggs *et al*, 1998; Allmang *et al*, 1999a; van Hoof et al, 2000). The Rrp6-exosome also degrades the non-coding 5' external transcribed spacer (ETS) fragment that is released during pre-rRNA processing (de la Cruz *et al*, 1998). Rrp6p was recently shown to interacts with Rrp47 through its N-terminal domain, PMC2NT, Rrp47 promoting its catalytic activity (Stead *et al*, 2007). In the nucleus, cryptic unstable transcripts (CUTs) are also degraded by nuclear exosome (Wyers *et al*, 2005 ; LaCava *et al*, 2005).

2.5.3. Exosome involvement in RNA quality-control.

The same components of general RNA degradation are also involved in different quality-control pathways. Pre-mRNA that fail to undergo correct processing can be retained in the nucleus and are degraded by the nuclear exosome (Bousquet-Antonelli *et al*, 2000). Mismodified pre-tRNAs as well as pre-rRNAs and pre-snoRNAs/snRNAs with processing errors are degraded in the nucleus by the exosome (Vanacova *et al*, 2005, Dez *et al*, 2006, LaCava *et al*, 2005, Doma and Parker 2007). Pre-ribosomes that undergo nuclear surveillance appear to concentrate in a sub-nucleolar focus called the No-body (Dez *et al*, 2006). mRNAs containing a premature termination codon (PTC) are degraded either by dedadenylation-independent decapping or by accelerated deadenylation and 3' to 5' exonucleolytic cleavage in a pathway called Nonsense Mediated Decay (NMD) (Mitchell and Tollervey, 2003). This was demonstrated in yeast and mammals, whereas in Drosophila, it was shown that the PTC containing mRNA degradation is initiated by endonucleolytic cleavage(s) in the vicinity of the nonsense codon. The resulting 5' fragment is rapidly degraded by exonucleolytic digestion by the exosome, whereas the 3' fragment is degraded by XRN1 (Gatfield and Izaurralde,

2004; Orban and Izaurralde, 2005). In the so-called Non Stop Decay (NSD) process, mRNAs lacking translation termination codon are recognized and rapidly degraded in the 3' to 5' direction by the cytoplasmic exosome (van Hoof *et al*, 2002). The cytoplasmic exosome is also thought to degrade endonucleolytic cleavage products of the no-go decay, a recently described process in which mRNAs with stalls in translation elongation are recognized and targeted to cleavage (Doma and Parker, 2006). The exosome is also known to mediate rapid degradation of mRNAs containing AU-rich instability elements (AREs) (Mukherjee *et al*, 2002; Chen *et al*, 2001)

2.6. Polyadenylation activates the nuclear exosome.

RRP6 was first characterized as a suppressor mutant of a temperature-sensitive allele of the PAP1 gene encoding the nuclear canonical poly(A) polymerase (Briggs *et al*, 1998). In contrast to null mutants of other exosome subunits, haploid yeast cells with an rrp6- null allele were viable but showed slow growth and a lethal phenotype at sensitive temperature. Polyadenylated RNAs accumulated in these cells (Kuai *et al*, 2004). These results suggested that Rrp6p is responsible for the degradation of polyadenylated RNAs. Indeed, recent studies proved that polyadenylated RNAs are degraded by the exosome in yeast nucleus. A new complex responsible for nuclear polyadenylation was identified. This complex was named TRAMP, for Trf4p/Air2p/Mtr4p polyadenylation complex. TRAMP complex contains Trf4p, or its homolog Trf5p, a poly(A) polymerase, Air2p and Air1p, RNA-binding zinc knuckle proteins, and Mtr4p, a RNA helicase, known as a exosome cofactor (Houseley *et al*, 2006, LaCava *et al*, 2005, Wyers *et al*, 2005; Vanacova *et al*, 2005). Recent discoveries on polyadenylation-mediated RNA degradation that were not known at the time I did my thesis work are developed in the Discussion & Perspectives chapter.

In the archaea *Sulfolobus solfataricus,* the exosome was shown to carry out polynucleotidylation (Portnoy *et al,* 2005). It appears that polyadenylation is a signal for RNA degradation that was conserved through evolution in bacteria, archaea and in the organelles and nucleus of eukaryotes. However, halophilic archaea and yeast mitochondria are exceptions as they lack exoribonucleolytic complex related to PNPase/exosome and RNA polyadenylation. In the absence of PNPase or the exosome, other degradation pathways take place, as RNase R exoribonucleolytic degradation in the halophile *Haloferax volcanii* (Portnoy *et al*, 2005).

In conclusion, few organisms lack polyadenylation. In most organisms, it seems that the ancestral role of polyadenylation in triggering RNA degradation has been conserved. At the time I am writing this manuscript, a review (Doma and Parker, 2007) relating unpublished results, evokes polyadenylation in plant nucleus.

Thesis objectives

It has been demonstrated that plant mitochondria lacking PNPase accumulate polyadenylated maturation and degradation intermediates. So far, only candidate transcripts, i.e. chosen by the researcher, were investigated for polyadenylation and degradation in plant mitochondria. These transcripts were the CMS related chimeric gene *atpA-orf522* in sunflower (Gagliardi & Leaver, 1999), protein-coding genes *cox2* in maize (Lupold *et al*, 1999), *atp9* in pea and potato (Kuhn *et al*, 2001 ; Gagliardi *et al*, 2001), and *atp1* in *Oenothera* (Kuhn *et al*, 2001) and ribosomal gene *18Srrn* (Perrin *et al*, 2004b). However, polyadenylated RNAs in plant mitochondria were never investigated as a whole.

Therefore, the first objective of my thesis was to identify PNPase substrates in a broad-spectrum approach, without preconceived ideas. To gain insight into biological roles of PNPase for plant mitochondria, identification of substrate RNAs is indispensable. I consequently elaborated a library of polyadenylated RNAs from plants lacking mitochondrial PNPase. Analysis of this library is reported in the first chapter. Most of the data obtained were published in Molecular and Cellular Biology (Holec *et al*, 2006) and commented for a non-specialized readership in Médecine/Sciences (Holec and Gagliardi, 2006). Conclusions and methods of the polyadenylated RNAs library are also detailed in reviews to be published in a special issue on RNA quality control and RNA surveillance of Biochimica et Biophysica Acta - Structure and Gene expression and Methods in Enzymology.

The data obtained from the library of polyadenylated RNA led me to conduct research projects in two radically different directions :

- on one hand, I used a model mitochondrial transcript to study transcript processing, independent of mitochondrial PNPase. This project is developed in the second chapter.
- on the other hand, the discovery of non-mitochondrial polyadenylated transcripts prompted me to investigate eventual RNA degradation pathway involving polyadenylation in the plant nucleus. This study is the object of the third chapter. Data of this chapter are currently submitted for publication.

RESULTS

Chapter I
A RNA surveillance system in plant mitochondria

1. Cloning of small polyadenylated transcripts from plants down-regulated for the mitochondrial PNPase.

To better understand the role of PNPase in plant mitochondria, we cloned a library of polyadenylated substrates from plants down-regulated for PNPase.

Homozygous T-DNA insertion mutant of PNPase could not be obtained (Perrin *et al*, 2004a), as *AtmtPNPase* gene is essential in *Arabidopsis*. However, overexpression of a tagged version of PNPase in wild-type plants triggered co-suppression of endogenous PNPase. This phenomenon, quite common in plants, consists of the extinction of the expression of one gene under ectopic expression of this gene in the same orientation. After 2-3 weeks of normal growth, development stopped, plants exhibited round, downward curled leaves and never resumed growth (Perrin *et al*, 2004 a, see Figure 1). Northern and Western analysis revealed that PNPase gene was down-regulated. Plants were named PNP-.

As we could not control co-suppression, we had to screen for PNPase down-regulation plant by plant, preventing us to isolate mitochondria from PNP- plants. I therefore tried to develop a strategy to circumvent this problem. I used a RNAi strategy to generate mutant plants that would allow down-regulation of PNPase upon oestradiol-induced expression of an artificial microRNA (amiRNA) targeting PNPase mRNA. Plants overexpressing a tagged version of PNPase were stably transformed with a plasmid containing the amiRNA precursor under the control of an oestradiol-induced promoter. Unfortunately, I could detect accumulation of the amiRNA but not under-expression of PNPase mRNA by Northern blot. This strategy was thus unsuccessful and I further used co-suppressed plants for my studies.

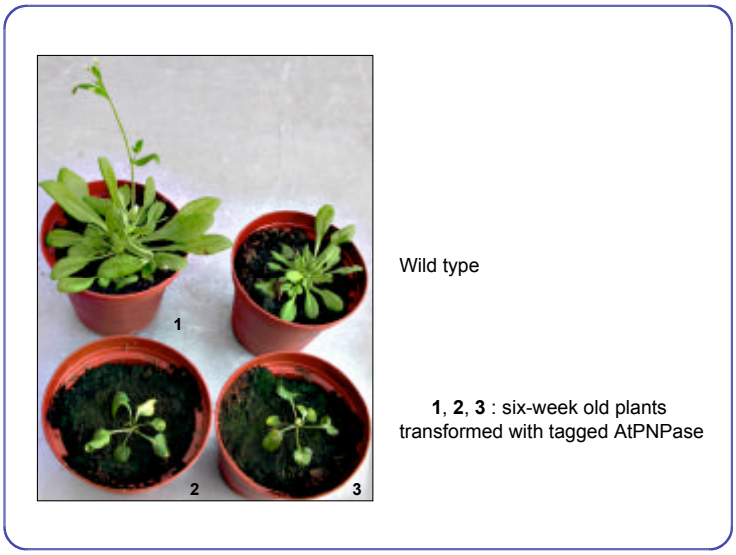


Figure 1 : PNP- plants compared to wild-type plant. **2** and **3** are co-suppressed plants . They exhibit round downward curled leaves and failed to set up new leaves (from Perrin *et al*, 2004a.)

As PNP- plants have to be identified individually, it is not possible to isolate mitochondria from them. To bypass this problem and gain access to mitochondrial RNA substrates of PNPase, we used a strategy based on polyadenylation status of PNPase substrates. We expected that mitochondrial polyadenylated substrates will be over-represented in PNP- plants (Figure 2A). Total RNA from PNP- plants was extracted and 5 µg were separated by electrophoresis on a 6% polyacrylamide gel. To limit contamination by cytoplasmic mRNAs, only the small RNAs of 100-150nt were eluted from the gel. Polyadenylated RNA were selectively amplified by reverse transcription using the SMART PCR cDNA synthesis kit (Clontech).

This kit allows incorporation of the same sequence tag at both 3' and 5' ends of polyadenylated RNA, without a ligation step (Figure 2B). When the RT reaches the 5' end of the RNA, its terminal transferase activity adds a short deoxycytidine (dC) sequence. This dC sequence then anneals to the 3' end of the SMART oligonucleotide (containing a stretch of dG), at which time the RT switches to transcribe the 5' portion of the SMART oligonucleotide. The result is single stranded cDNA flanked by a sequence-tagged oligo dT primer on the 3' end and the SMART oligonucleotide on the 5' end. This cDNA is amplified by PCR, using the SMART-adapter primer and a reverse primer. PCR products were then cloned into TOPO vectors (Invitrogen). Inserts were amplified and sequenced without further purification.

433 sequences were analyzed. Among these, only 9 were duplicated and 1 was present 4 fold, indicating a sufficient complexity of the library. 301 clones (70% of the library) were mitochondrial, proving the success of the strategy, 77 were nuclear and 37 were plastidial. Clones mapping the same genomic locus but differing by a few nucleotides in 5' or in 3' were considered as independent. Some clones mapped to sequences that are present in both nuclear and mitochondrial genomes, due to a large insertion of mitochondrial DNA in the nuclear genome of *Arabidopsis*. However, their nuclear origin could be excluded because of mitochondrial expression features such as C to U edition, expression from known mitochondrial promoter or localization near a known mitochondrial gene. For some regions highly represented by different clones, we confirmed mitochondrial expression by Northern analysis on mitochondrial RNA. Additionally, accumulation in PNP- plant versus wild-type confirmed the mitochondrial origin as PNPase is strictly mitochondrial.

Doubts subsist for 2 clones matching to the polyubiquitin gene (At1g65350) that harbors an insertion of the mitochondrial genome (clones m+nVH9F and m+nVF9F in appendix 2). Similarly, one clone could be assigned to either mitochondria or chloroplast (clone IIIH7R in appendix 2). 18 clones did not match any sequence in current databases. The origin of these sequences is unclear : it could reflect unidentified contamination or sequence variation between *Arabidopsis thaliana* ecotype C24 that was used for sequencing the mitochondrial genome and ecotype Col-0 used in this study. This idea is supported by the fact that one clone matched perfectly to the mitochondrial genome of ecotype *Landsberg* but not to the C24 sequence (clone uVC7R in appendix 2).

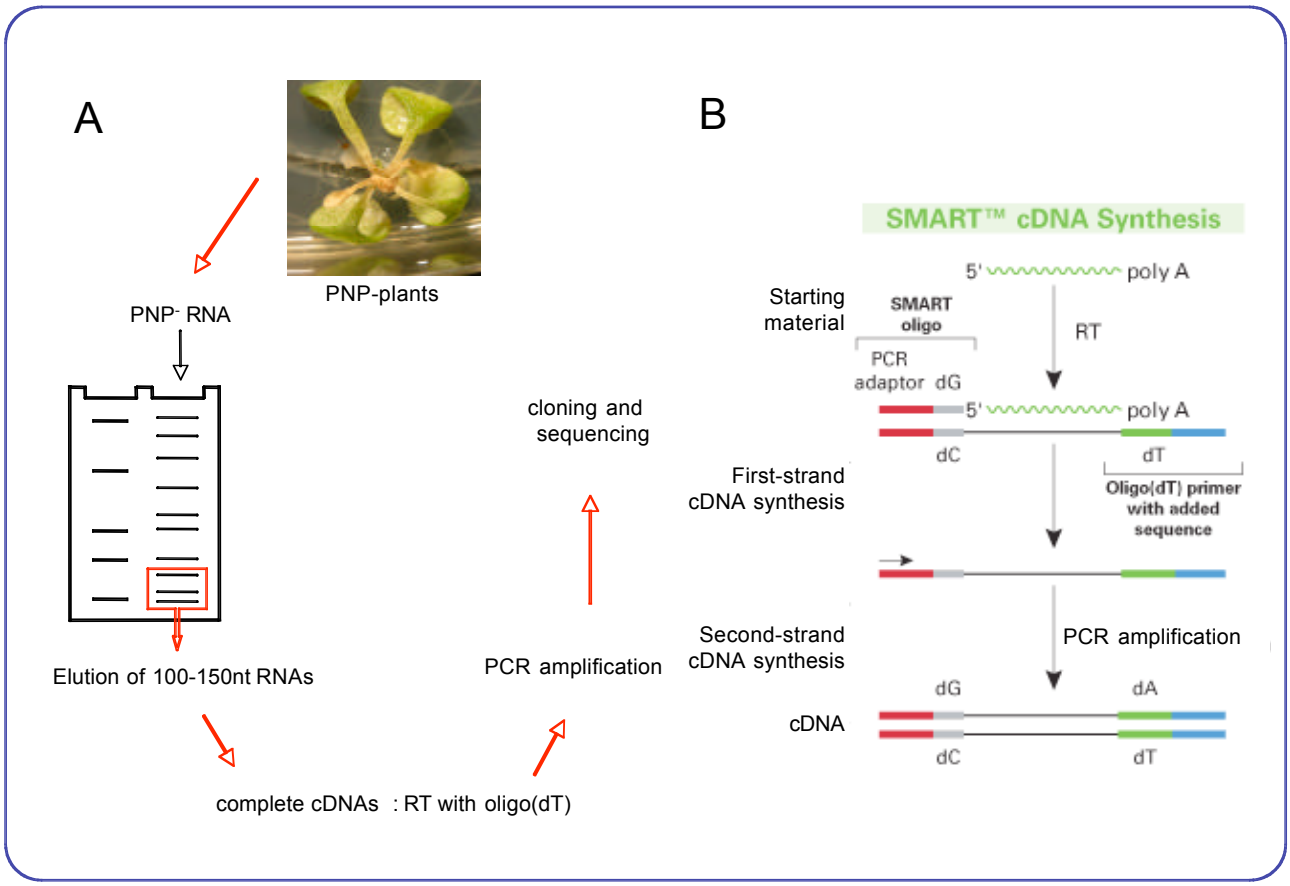


Figure 2 : **A**. Construction of a polyadenylated RNA library from PNP- plants. **B**. Model of the SMART cDNA synthesis kit used to construct the library.

The clones of the library are listed in appendix 2 (coordinates refer to the annotated mitochondrial genome sequence of ecotype C24, accession NC_001284).

As polyadenylation targets RNAs for degradation in plant mitochondria, these polyadenylated clones were considered to be degradation tags accumulating in PNP- plants. This hypothesis was confirmed by Northern blot for some transcripts, as detailed further.

2. Maturation by-products are degraded by the PNPase.

Only 17 clones mapped to rRNAs sequences (8 in the 26S rRNA, 9 in the 18S rRNA) and 20 clones were found to match mRNA sequences. So less than 13% of the library corresponds to rRNA or mRNA sequences. By contrast, 143 clones (47% of the library) were found just upstream or downstream of tRNA or rRNA sequences. These clones potentially represent maturation by-products of tRNA and rRNA synthesis. For the 5S rRNA, for example, degradation tags located downstream of the mature transcript are represented in the Figure 3.

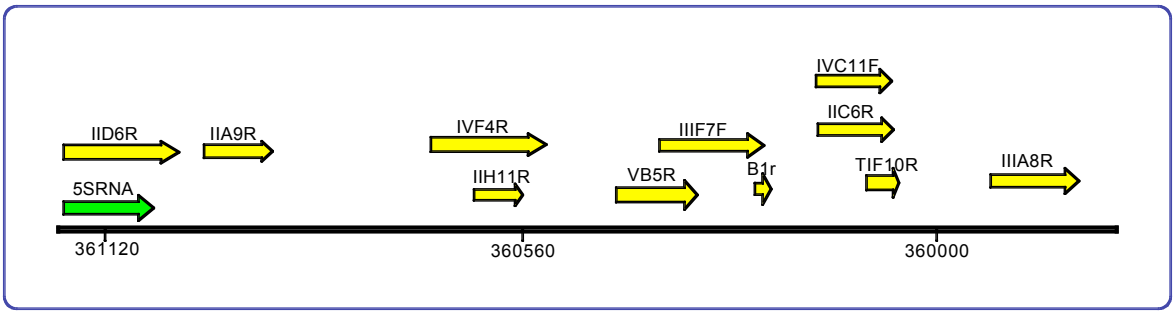


Figure 3 : Position of the degradation tags at the 5SrRNA locus. The mature 5SRNA is represented by the green arrow, degradation tags are represented by yellow arrows. Names of the clones are mentioned. Coordinates on the genomic sequence are indicated.

Polyadenylated degradation tags are located downstream 5S rRNA 3' end, on a distance of about 2kbp, indicating that, transcription proceeds into downstream sequences, but that resulting transcripts are polyadenylated and degraded. Similarly, many degradation tags match sequences of the 26S rRNA leader. This suggests that, as it was the case for the 18S rRNA (Perrin *et al*, 2004b), PNPase is involved in the degradation of by-products of 5S and 26S ribosomal RNA synthesis or processing.

Many small polyadenylated RNAs are found 5' and/or 3' to tRNA genes, such as *trnK, trnG, trnf-M, trnI, trnQ, trnM*. This is particularly striking for tRNA gene clusters *trnS-trnY-trnP-trnC-trnN-trnY* and *trnS-trnY-trnE*, where almost all intergenic sequences are represented in the library (Figure 4).

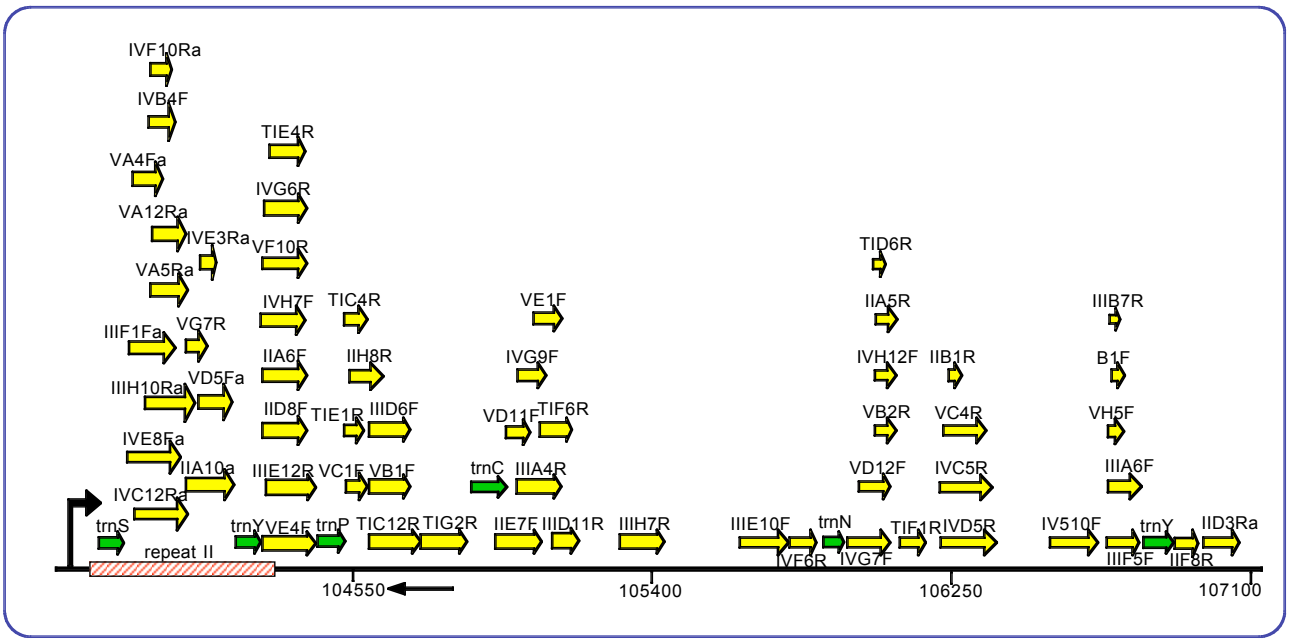


Figure 4 : Position of the degradation tags at *trn* genes cluster locus. *trn* genes are represented with a green arrow, degradation tags are represented with yellow arrows. Names of the clones are mentioned. Promoter is represented by a bent arrow. Black arrow represents the probe used for the Northern analysis Coordinates on the genomic sequence are indicated.

. Such an accumulation strongly supports the idea that, once tRNAs are excised from polycistronic precursor transcripts by RNase P and RNase Z, maturation by-products are polyadenylated and degraded by the PNPase.

Some degradation tags are consecutive. This suggests the involvement of an endonucleolytic activity in the degradation of these maturation by-products. However, this arrangement is not systematically observed.

To confirm the role of the PNPase in the degradation of tRNA maturation by-products, Northern analysis was performed (Figure 5). We compared the accumulation of the *trnP-trnC* intergenic sequence in PNP- plants versus wild-type plants. As expected, we could detect this intergenic transcript in PNP- plants but not in wild-type plants, proving that PNPase is indeed responsible for the degradation of the *trnP-trnC* intergenic sequence. Most of the signal is detected as a band of 350nt, although smaller, discrete bands are also visible. This size corresponds to the length of the full-length intergenic sequence, supporting the idea that the major part of the degradation is achieved by the PNPase and that contribution of any endonucleolytic activity would be minor.

This analysis shows that even if the cloning was restricted to small RNAs of 100-150nt, our strategy can result in identification of much larger substrates of PNPase.

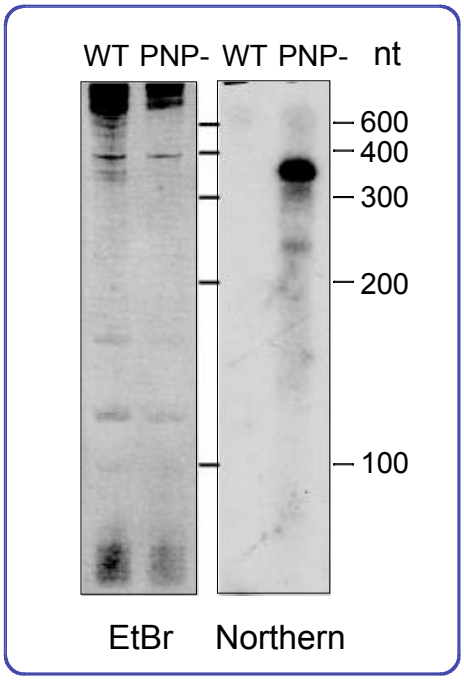


Figure 5 : Northern analysis of the trnP-trnC intergenic sequence. The probe used is a radiolabeled RNA represented with a black arrow on Figure 4. EtBr staining is showed for loading control.

These results strongly suggest that PNPase degrades RNA maturation by-products in plant mitochondria.

3. Non coding genomic regions are highly transcribed.

In the *Arabidopsis* mitochondrial genome, up to 10% of the sequences are predicted to be ORFs of at least 100nt. As these ORFs are not conserved between plant species, they unlikely represent functional genes. Several clones of the library corresponded to *orf240a, orf315, orf275, orf135b, orf106f, orf294, and orf 145c.* These results show that putative non functional ORFs are indeed expressed from the *Arabidopsis* mitochondrial genome, but are degraded by the polyadenylation-dependent pathway.

We found also tags in regions that do not contain any known genes. Some of these regions are expressed at low levels, as indicated by the low number and scattered distribution of the tags. This could reflect a basal expression of the mitochondrial genome, probably due to a relaxed control of transcription. In contrast, other regions are highly expressed. For instance, a region of 230nt containing the *rrn26S* promoter has been duplicated and inserted into a genomic region devoid of any gene. Specific sequences downstream of the duplicated region are found in the screen, indicating that the inserted 26S promoter drives transcription of these sequences (Figure 6A).

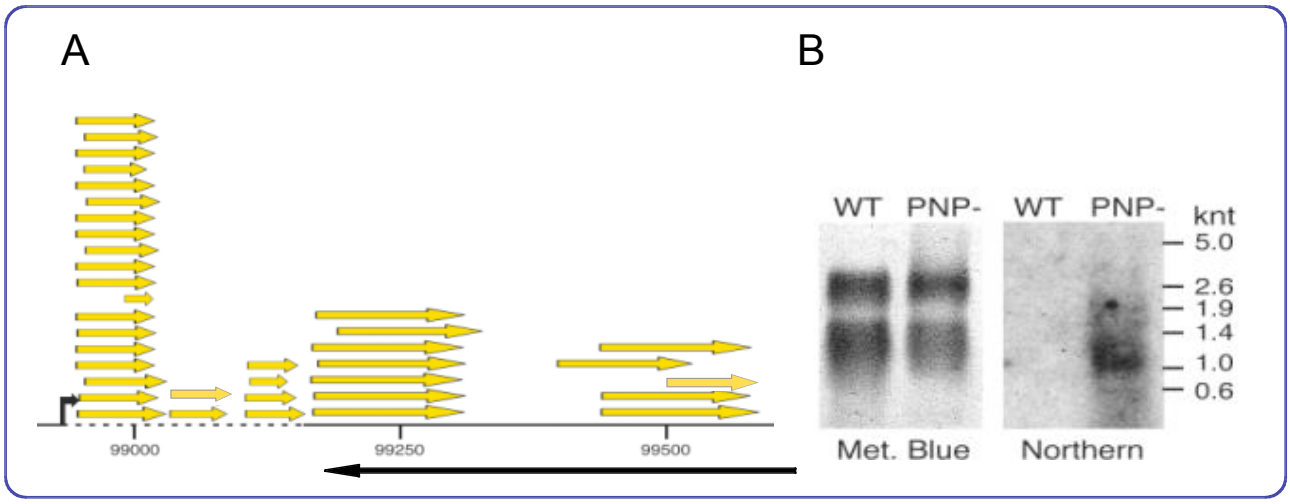


Figure 6 : PNPase degrades transcripts expressed from a region without known gene. **A.** Location of the degradation tags on the genomic sequence. They map a region spanning position 98900 to 99600. The promoter is represented by a bent arrow, the dotted line corresponds to the duplicated region containing the *rrn26S* promoter. Tags on this sequence are transcribed from either one or the other region. Tags in 3' are specific to this region. **B.** Northern blot analysis of the transcripts expressed at this locus. Each lane contains 5µg of RNA from PNP- or wild-type (WT) plants. The probe consisted of an RNA complementary to nt 99179 to 99750 of the mitochondrial genome (black arrow). The methylene blue staining (Met.Blue) is shown for loading control.

We evaluated accumulation of transcripts of this region by Northern blots, using RNA from PNP-plants and wild-type plants. An RNA probe complementary to the region downstream of the duplicated *rrn26S* promoter region was used. Results are shown in Figure 6B. We could not detect transcripts in wild-type plants, indicating that transcripts corresponding to this region are not stable. In absence

of PNPase, RNA species of different lengths, detected as a smear, accumulate, indicating that transcripts of various length are produced. Taken together, these results show that this region is highly transcribed but that resulting RNAs are rapidly degraded by the PNPase.

Another region characterized by a high number of degradation tags, and thereby apparently highly expressed is a part of repeat I, a repeated sequence of 4196nt that spans positions 44698 to 48894 and 178863 to 183059 of the mitochondrial genome, ecotype C24 (accession number NC_001284). This region does not contain any known gene, but contains a putative promoter that fulfills all previously established criteria for plant mitochondrial promoters. Many degradation tags are found downstream of this promoter but some also are in 5', indicating that transcription may also start upstream (Figure 7A).

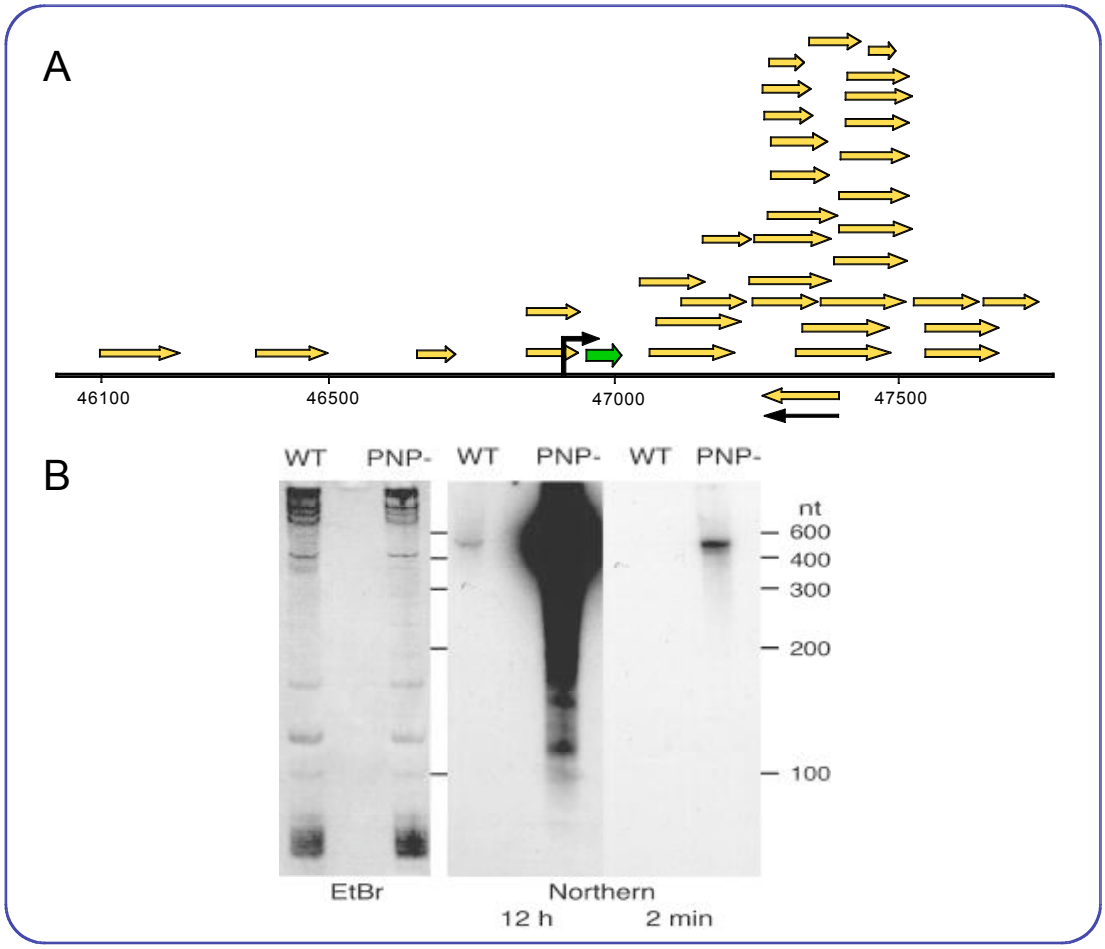


Figure 7 : PNPase degrades transcripts from the repeated region I. **A.** Location of the degradation tags on the region spanning nt 46100 to 47900. The putative promoter is represented by a bent arrow, Ath-59, a previously described expressed transcript is represented by a green arrow (Marker *et al*, 2002). Degradation tags, represented with yellow arrows, could be transcribed from either nt 44698 to 48894 or nt 178863 to 183059. **B.** Northern analysis of accumulation of the *NCO* transcript in WT and PNP- plants. The probe used is a radiolabeled *in vitro* transcribed RNA complementary to the 47264 to 47395 region of the mitochondrial genome (black arrow). Two exposure times of the same experiment are presented, indicated at the bottom of each panel. A picture of the EtBr stained gel is shown for loading control.

The high number of degradation tags prompted us to analyze the accumulation of this region's transcripts by Northern blot (Figure 7B). A discrete band of 500nt was detected both in WT and PNP- plants, but the signal was considerably stronger in PNP- plants. This transcript was named Non-Coding Overexpressed (NCO) transcript. A detailed description of it will be given in the coming chapter that will focus on this species.

Taken together, these data show that a novel NCO transcript is highly expressed but efficiently degraded by PNPase, that prevents its accumulation.

Thus, these results show that PNPase efficiently degrades transcripts resulting from highly transcribed regions of plant mitochondrial genome devoid of any coding function.

4. A putative RNA dependent RNA polymerase activity extends transcripts 3' end.

Our analysis of the library of small polyadenylated clones gave some surprising results. For instance, some clones carried non-encoded 3' extensions. This is the case for 41 clones out of 301 (13,6%) of the mitochondrial clones. These clones are listed in table 1. Non-encoded nucleotides others than homopolyadenosines are highlighted in bold. Clones carried between 1 and 36 extra nucleotides and the average length was 9-10nt per clone. It cannot be excluded that these additional nucleotides are due to cDNA synthesis, PCR artifacts or sequencing errors, in particular when only one or two nucleotides were added, as for clone mIIF8R, mIIIA11R, mVB1F, mVA3R, mVC3R, mVG6F, mVH7F, mVG8F or mVD10F. However an artifactual origin is more doubtful for longer heteropolymeric extensions upstream of poly(A) tails.

Table 1 : Clones exhibiting non-encoded nucleotides that are not homopolymeric adenosines tails. The complete sequence of the clones are presented, non-encoded nucleotides that are not homopolymeric adenosines are in bold.

clone name	sequence
mIIE2F	CACCTCAGCACAAGTTACTCGAGTAGCACAGAAGCCATCATCAATAGCTTGCTTGCCTGGGAGCTCAAC CACGAATTCTCTGTCTTCTAACCTTTCTATCACTTGAGGATTATCTACGCAATTCGTAAA**CAATTAAC**AAAA AAAAAAAAAAAAAAAAAAAAAAAAAA
mIIF2F	CCAGTAGTAGAAGGGGATCTTGAATCAAAGAGTTCCTGTCCAACAAGCAAGGAATGTCAGTCATAAATAA TGAAATCCATATAACTCTACTCCGGGTTGTGAGAACATCGTGCCTGTTGCGATTGTGTGG**AACACGC**AAA AAAAAAAAAAAAAAAAAAAAAA
mIIH5R	TTTTATTTCTTTTGGGAAAATAATACATAAACCTTTTGAGTCACAGCCACCTCAGCACAAGTTACTCGAGTA GCACAGAAGCCATCATCAATAGCT**TTAAACACTATAAAAGAAATAGCG**AAAAAAAAAAAAAAAAAAAAAA AAAAAAA
mIID7F	GCCAGTAGTAGAAGGGGATCTTGAATCAAAGAGTTCCTGTCCAACAAGCAAGGAATGTCAGTCATAAATA ATGAAATCCATATAACTCTACTCCGGGTTGTGAGAACATCGTGCCTGTTGCGATTGTGTGG**AACACGC**AA AAAAAAAAAAAAAAAAAAAAAAAA
mIIE7F	CACCTTCATGACCAGAAATAACTGGAACAAAGGGAAAGCCCACTCAACTAGTAACATACGGACACAAACA ACAAAACTACTTGTTACAAATAACATCTTGTTGTTGCATCGAAACATATAAATGTTTACTAATCTGGCTT**AA GCCGG**AAAAAAAAAAAAAAAAAAAAAAAA
mIIF8R	TGTTGTAGTTTTAGCTTGTATATGTGAAAAAAACGAGATTTTCTTTCATAGAATAACTCTTTCACATAAGAAA GTG**GG**AAAAAAAAAAAAAAAAAAAAAAAAA
mIIC10F	GGAGATGTTTATAGCTTCTTCATGGCAGGGGTGTTCTTGTTTTATGGTA**CTGACCG**AAAAAAAAAAAAAAA AAAAAAAAAAAAA
mIIB11F	CGGATCCCTTTTTTCTTTGCCTTTACGAG**NCCAAT**AAAAAAAAAAAAAAAAAAAAAAAAA
mIIIB4F	GGGAGGGGAAGCACTTCCTTCTACGGGCCCTAACGGGGCATCTTCCTATTCAGAGTGGTTTACCTACAC GTCTGATTTGGAGGATTCGGCCAGTTCCGGGCGTACCTCCTCGTCGGTGAATCAACCGATTCAGAGGG AACAGGCTGGGCCATCCAATGCCT**CAG**AAAAAAAAAAAAAAAAAAAAAAAAA
mIIIG4F	GATGTGTTTTGTCGAGTTTGCTTCGTCCTCTTTTTTTTTCTTATTTTCCTTCCTTTAGGCGCTTGCTCTCAA AAAAAAGGGAAAGGGGTCAAATAAAAAGCTACCTTTTATCAAGGCTAA**AAACAAAAGGAACCC**AAAAAA AAAAAAAAAAAAAAAAAAA
mIIIG9F	CGGATAGAGCAGATGGTCCAACTACATAACTTTTTCTTTTTTATTATTTTTATGGTCGTGCTTTGTGGCACG GCAGCACCCGTACTATTGAAATGGTTCGTCAGTAGAGATGTTCCCACGGGTGCCCCTTTTT**TTAATGGC**A AAAAAAAAAAAAAAAAAAAAAA
mIIIH9R	GATCGTTTTTAGAAAAGAAAGAACGTTTTGATTCGAGGCGGATCCCTTTTTTCTTTGCCTTTACGAGTTGA AAAGAAA**AACACC**AANAAAAAAAAAAAAAAAAAAAAAA
mIIIA11R	TAGCTTAAGTGTTCACGTAGGTAAAATAGCTTCTATAGCTCCATCCAATAGTAATCAACGGAGATAGAGTC CAGCGGTTCAACCAACGCTTCTAAGGAGAGCGGGGCAAGCAAGAAAGCAGGCAAAGTCATTGAGCCTA TTCTATTCCNAA**C**AAANAAAAAAAAAAAAAAAAAAA
mIIIB11R	GGTTGACTTCACCTCCGGTGACCGGCTTCGCGGGATCGCCTTACGTATAAGGACCTGATCCACTCTAGG CGAAGCAGGTCTACGGTGTCCACGATTGTATCCGACTATTGATTGAGTGAGGCCAACTTC**ACTCAATAAT ATAAAATCCCAAC**AAAAAAAAAAAAAAAAAAAAAA
mIVB4F	GAGCCCGTTGCGCGTTTTTTTGTTTGACCGGCCTATCTTCATAAGTAAGCTCCCTATGGCCGTCCAGTCC CTGG**AACT**AAAAAAAAAAAAAAAAAAAAAAAAAAAA
mIVE4R	TATTTGACCCGAAACCGATCGATCTAGCCATGAGCAGGTTGAAGAGAGCTCTAACAGGCCTTGGAGGAC CGAACCCACGTATGTGGCAAAATACGGGGATGACTTGTGGCT**AGATCGATCGGTTATCG**NNNNNNNNN NNNNNAANNAAAAAAAAAAAAAAAAA
mIVG6R	CTCGGAGCTGAGGTATATGAAGAATGGCCTTTTGGTCCTTTGGACACGTAGCCGAGAGCGAGCCGGATT TTTAAAATTCAAGTGAAAGAAGGAAGAAGGCAAAAAAGCCGTATAGTGCAGGAGGCA**ACTTTCT**AAAAAA AAAAAAAAAAAAACCCCAAAAAAAAAAAAAAAAAAAA
mIVC9F	GAGGACCGGGCAATCGCCCTCTTTTTTAATGAAGAAGCGGGCTAGTCCCCGAAAATGCCCGTTAATCAA GCAAGTTGGGGA**AAAAAATCTTACCTATTAGACGGAAANG**AAAAAAAAAAAAAAAAAAAAA
mIVG10F	ACACTCAGTGCTAGCAGGAGTACTTTCTCGGATACGATCCCTTTTTATAGGTTTGCTGCCCTGCGACACG TGGCAGAAGAGGACTCAAGGACACATGCCAGATCTTTGGGTCGGTATTTCTTCCTT**TTCTAAACCACGA AATACGCAACCCGTGCACATC**ANAAAAAAAAAAAAAAAAAAAAAAAA
mIVA11R	GTTGAGCCAGTTGAAAGACTAGTCGAGAAGCCCTATTCAGAAAGGTAGATGAGAGCTTGCGGAGAAGG AAAGTTCTCATTTGTCCAATCCACCGAATACTATCCTTTCCTTCTTTTACCAATTTTATCGTCGAGTTGCTG ATACGTTTGTTGAACAGCAAC**NGAAC**ANANAAAAAAAAAAAAAAAAAAAAA

mIVD12F	TGCCTGCCAAGCCAATAGGCGAAAGGGCGAAGCAACTCAAAGCTCTCCGGGGT**AACAACCCCGG**AAA AAATAAAAAAAAAAAAAAAAAAAAAAAANGNAAAAAAAAAAAAAAAAAAAAAAA
mIVE12R	CATTCTGTGGCTCGAAATCCATCATTGGCGACCACCACTGTTTTAGTGGTTACACTTACACTCCTCGGCG GGGTCGCCGCCTTTTATTTGCATTCCTTCCGG**AATGC**AAAAAANAAAAAAAAAAAAAAAAAAAA
mVB1F	AAGAAAATGGGGCTTATGTTTTTCAGCCGCATCGCACTGCTGCATAAACAATAGATCCGCTGGGCTGGAT GAGCAACCTTCTATCTGGCCTCTGTACTAGTAGTAGAGTTGCTTGCTACTTTCAA**CC**AAAAAAAAAAAAAAA AAAAAAAAAAAAAAA
mVE1F	ATCGAAACATATAAATGTTTACTAATCTGGCTAACATTGAACTTGGTAAAATTAAATGGTTGAATAATTTAAA AATTAGATTATTCGG**ACAAAAACCC**AAAAAAAAAAAAAAAAAAAAAAAAAAAA
mVG1F	GACTGTATTGGAAAAACAAAAAGTTCAATATGGCGGGAGATCCATAATTTTATTGAGAGCTTTCAATCAAAT TAGATTGTAAAGTAGTAGTCCTG**AAAACTACC**AAAAAAAAAAAAAAAAAAAAAAA
mVA3R	TCGGGTTTTTATTCCTAACTTCTTTCATCTTCGTTTATAATCGATAATTCCTTATATAAGAAGTTTATAATCGA TAATTCCTTATATAAGAAGATTCGAATTC**TT**AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAANNNAAAAA
mVC3R	GGCTTGCGGTTCATTACACCAAAGGCTTTCAGTGAACTATTGAAGCTATCGAGAGAGTACCAAATGCTGA CGGTGACGAAGGTTACCGACTTTCGGTGGACGCGGAGCCA**CC**AAAAAAAAAAAAAAAAAAAAAAAAAAAA AA
mVC4R	TAGGTAGGATAGTGCGATGTAGTCCAATGGCTAAAGCTCTGCCAGCTTCTTGTAGACTGAACTCTCTTTA GGCTCCTAGTGGTTTTCGGGTGGGATGGTTGAAATTTTATTCGCCACTAGTGGCAACGA**AAACCAC**AAA AAAAAAAAAAAAAAAAAAAAAAAA
mVD5F	GAAGTTTTCCCTATTAATTAGATTAGTAAAGGGCTTTTCCCTTACTAGTCAAGTGGTAAGGTAGGACGCTC TTCGATGAAGAAAAGAAGAGACTTT**ACT**AAAAAAAAAAAAAAAAAAAAAAAAAAAA
mVE5R	GAGTCTAGAATAAACCTTAGAAGTGGCAAAAACGACTTACTTTGTTGCAACGGGAACTACTCGCCCCGG GCACTGGTGAAATAGAATCTTATGTGTTGCAGCTACTTATGCTTTCCCAACTAGAAATCAAACGGAACAAA ACCG**TTTAAC**AAAAAAAAAAAAAAAAAAAAAAA
mVC6F	TTGAATGGAATCAGATCATCCGTTTCTGGTATGTTGGGTGTCATCCTATCCAGTACTGGATGCAGTAGAA GGTGCTCTCGGGCGAATTCCCGGCTAAAAAGTACACTGAGTTAGGCGAGAATTGTCTCG**AAAAAAAAAA AATTAAAAAAAAAGACAATCAACATACC**AAAANNAAAAAAAAAAAAAAAAAA
mVG6F	TATCTATCTATAAACAATATTAAAATCATTATTTTATCTAATTGGGCATTCCTGGGAAGAGGAAGAAGCAGAT AGAGCAAAGGCCTCCCCTTTGCTTGCGTCCGCTCTTCCCGAAGTGAGCAAATTGCATGTAGAGATCC**CA** AAAAAAAAAAAAAAAAAAAAAAAAAAAA
mVH6F	CCAGGGACTTGAAATCCTTTTTGTGTCAGCTCTTTGGAAGTGCCTTTTCCTTTACTTCAGTAAGGAGTCT CAAAATAGGAAATACCTA**TTTTGAGAAACACAAAT**AAAAAAAAAAAAAAAAAAAAAAAA
mH7F	TACTCATTCTCGGCCGCTTTCAGTTAGTGTTCATTGCTGAGGGATCCCGAGATCTCGAAGCAGA**AT**AAAA AAAAAAAAAAAAAAAAAAAAAAAA
mG8F	TGAGAAGGAAGGACGCTTTCAGAGGCGAAAGGCCATGGGGAGATACCGTCTGTGATCCATGGATCTCC GATCGGGAAACCGTATCCAAGCTCCGTGGCTAGTCTGCGCTCTTT**AG**AAAAAAAAAAAAAAAAAAAAAAA AAAAAAA
mVD10F	TAGGTTCCCGCATCACGTAAATTTCATTTCTTTTGGGAAAATAATACATAAACCTTTTGAGTCACAGCCAC CTCAGCACAAGTTACTCGAGTAGCACAGAAGCCATCATCAATAGCTTGCTTGCCTGGGAGCTCAACCAC GAATTCTCTGTCTTCTAACCTTTCTAT**AC**AAAAAAAAAAAAAAAAAAAAAAAAAA
mVA11R	TTAGTTGTTTGATGAGCCTGCGTAGTATTAGGTAGTTGGTTAGGTAAAGGCTGACCAAGCCGATGATGCT TAGCCGGTTAGAGCAAAGGACTTGAAATCCTTAGAGCAAAGCCAG**AGCCCC**AAAAAAAAAAAAAAAAAAA AAAAAAA
mVD11F	AAGGGAAAGCCCACTCAACTAGTAACATACGGACACAAACAACAAAACTACTTGTTACAAATAACATCTTG TTGTT**AAAATCCGTAAAAAACTG**AAAAAAAAAAAAAAAAAAAAAAAAAA
mVG11R	TAGTTAGTTTCATCGATATTTTTGTGGTGTTCAGTGTACCCTGAGTACAAGATC**ACT**AAAAAAAACAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAA
mVA12R	CGCGTTTTTTTGTTTGACCGGCCTATCTTCATAAGTAAGCTCCCTATGGCCGTCCAGTCCCTGGGCGGCT CTCGGTTCTTGAGCATGTTGGGAGA**AACAAAAATACTCCCAACATG**AAAAAAAAAAAAAAAAAAAAAAA AAA
mVE12F	AGCACAGAAGCCATCATCAATAGCTTGCTTGCCTGGGAGCTCAACCACGAATTCTCTGTCTTCTAACCTT TCTATCACTTGAGGATTATCTACGCAATTCGTAAATGATCA**CTACGC**AAAAAAAAAAAAAAAAAAAAAAA

Each type of residue is represented in these extensions, but adenosines account for the majority of non-encoded nucleotides. However, homopolymeric parts constituted of 30 adenosines or less are due to the primer used for cDNA synthesis that contains an oligo(dT)$_{30}$, preventing any statistics on residues composing non-encoded extensions. In the following clones, exclusively A and C residues are found : mIIIH9R, mIIIA11R, mVB1F, mIVE1R, mVC3R, mVG6F, mVD10F.

Previous studies also described non-genomically encoded extensions with a prevalence of adenosines and cytidines. In maize mitochondria, 94% of *rps12* cDNAs contained one to four non-genomically encoded C or A residues (Williams *et al*, 2000). In the same line, analysis of pea *atp9* transcripts revealed extensions of a maximum of 3 nucleotides, most frequently adenosines and cytidines. (Kuhn *et al*, 2001). These results suggested that mitochondrial transcripts could be extended by a nucleotidyltransferase activity, such as CCA:tRNA nucleotidyltransferase.

However, extensions found here are not only composed of A and C, and exceed a few nucleotides, excluding a CCA:tRNA nucleotidyltransferase activity. A possibility is that the polymerase responsible for the polyadenylation of mitochondrial clones is not strictly specific for adenosine polymerization but occasionally incorporates other nucleotides. For instance, bacterial and chloroplastic PNPase possess both polymerase and exoribonuclease activities and polymerize RNA in a non template-specific manner (Mohanty and Kushner, 2000 ; Yehudai-Resheff *et al*, 2003). However, I observed these heteropolymeric tails upon depletion of PNPase, thus PNPase is unlikely a candidate.

Interestingly, in some cases, the heteropolymeric extension was complementary to the 3' part of the encoded sequence. This is illustrated by RNA folding prediction softwares, based on structure probabilities. The secondary structures presented in Figure 8A were obtained on the Bibiserv site (http://bibiserv.techfak.uni-bielefeld.de/rnashapes) and confirmed by mfold (http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html). Non-encoded heteropolymeric sequences are highlighted in pink.

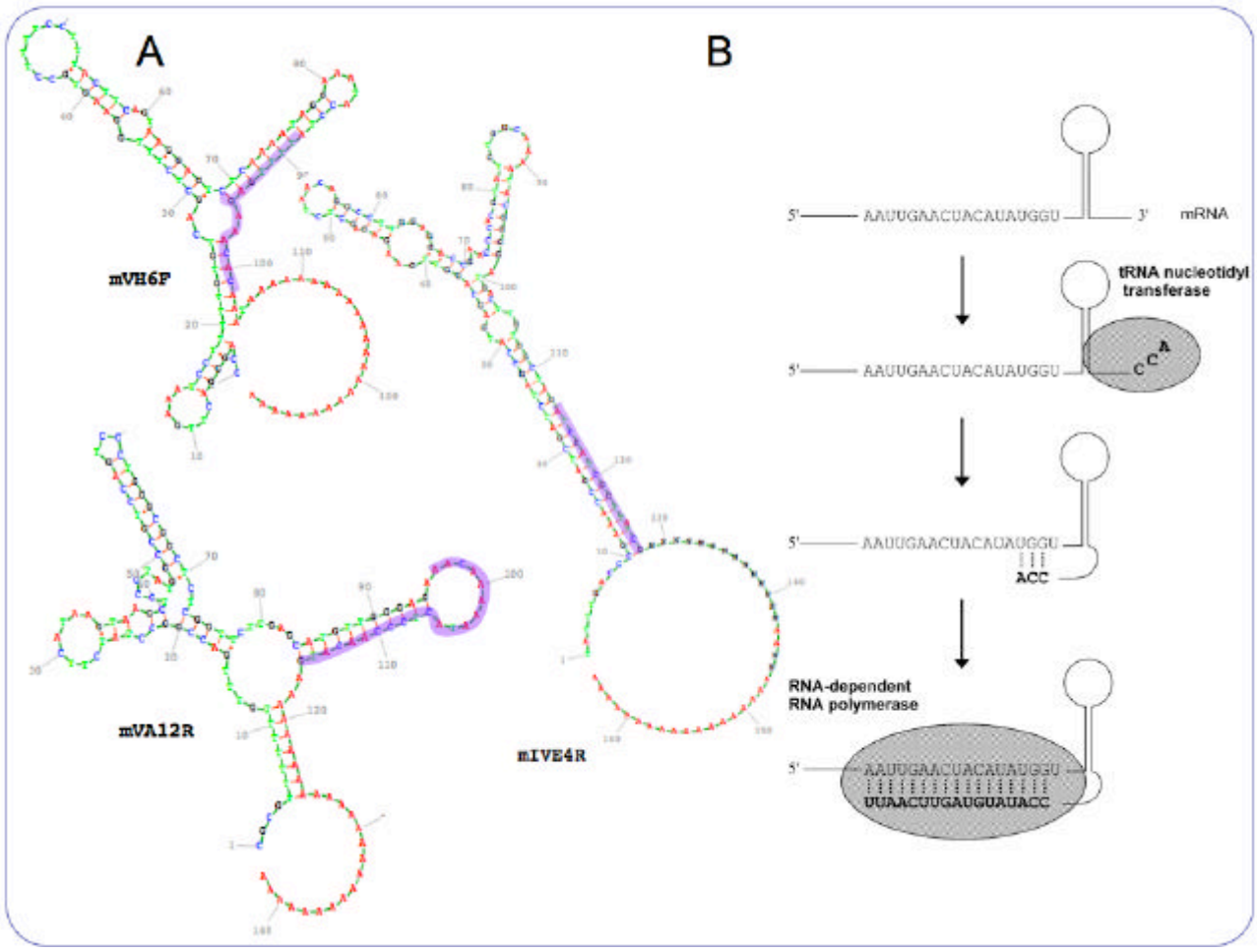


Figure 8 : A. Predicted secondary structure of 3 clones with non-encoded extensions. Predictions are made on the Bibiserv internet site (http://bibiserv.techfak.uni-bielefeld.de/rnashapes). The non-encoded nucleotides that are not polyA are highlighted in pink. B. Model for the addition of 3' non-encoded sequences to plastid *ndhD* transcripts, from Zandueta-Criado and Bock, 2004.

The three clones presented here show that the 3' extension can anneal to the genomically encoded sequence of the transcript, forming a stem or stem-loop structure. This is interesting, as stem-loop structures are known to play a stabilizing role for a restricted number of plant mitochondrial transcripts (Kuhn *et al*, 2001). Thus, such extensions could counteract the destabilizing effect of the polyA tail, as observed for pea *atp9* (Kühn *et al*, 2001). A previous study in chloroplast described the existence of 3' extensions to the *ndhD* transcript complementary to the sequence immediately downstream of the 3' hairpin, suggesting that this sequence can serve as a template for a RNA-dependent RNA polymerizing activity (Figure 8B). The presence of a CCA sequence starting the non-encoded extension of some clones suggested initial involvement of the CCA-adding enzyme (tRNA nucleotidyltransferase). Complementarity of the free CCA end with the sequence immediately upstream of the stem-loop structure may prime RNA-dependent RNA polymerization using the mRNA as a template (Zandueta-Criado and Bock, 2004).

Some of the extensions, as illustrated in Figure 8, can be explained by an RNA-dependent RNA polymerase (RDRP) activity. Such an activity could be en-

coded by the mitochondrial genome : a putative RDRP was found to be encoded in Arabidopsis mitochondria (Hong *et al,* 1998). Another possibility is that the RDRP activity is encoded by the nuclear genome of Arabidopsis and targeted to mitochondria. Two nuclear phage-type RNA-dependent RNA polymerases are targeted to the mitochondria : RpoTm and RpoTmp (Hedtke *et al,* 1999). These phage-type RDRP are good candidate because of their ability, as T7 polymerase, to polymerize RNA without template. In order to identify factors involved in the generation of heteropolymeric extensions, I tested candidate RNA polymerases for RNA-dependent RNA polymerase activity *in vitro.* These experiments were done in collaboration with Kristina Kühn in Thomas Börner's laboratory at the Humboldt Universität in Berlin.

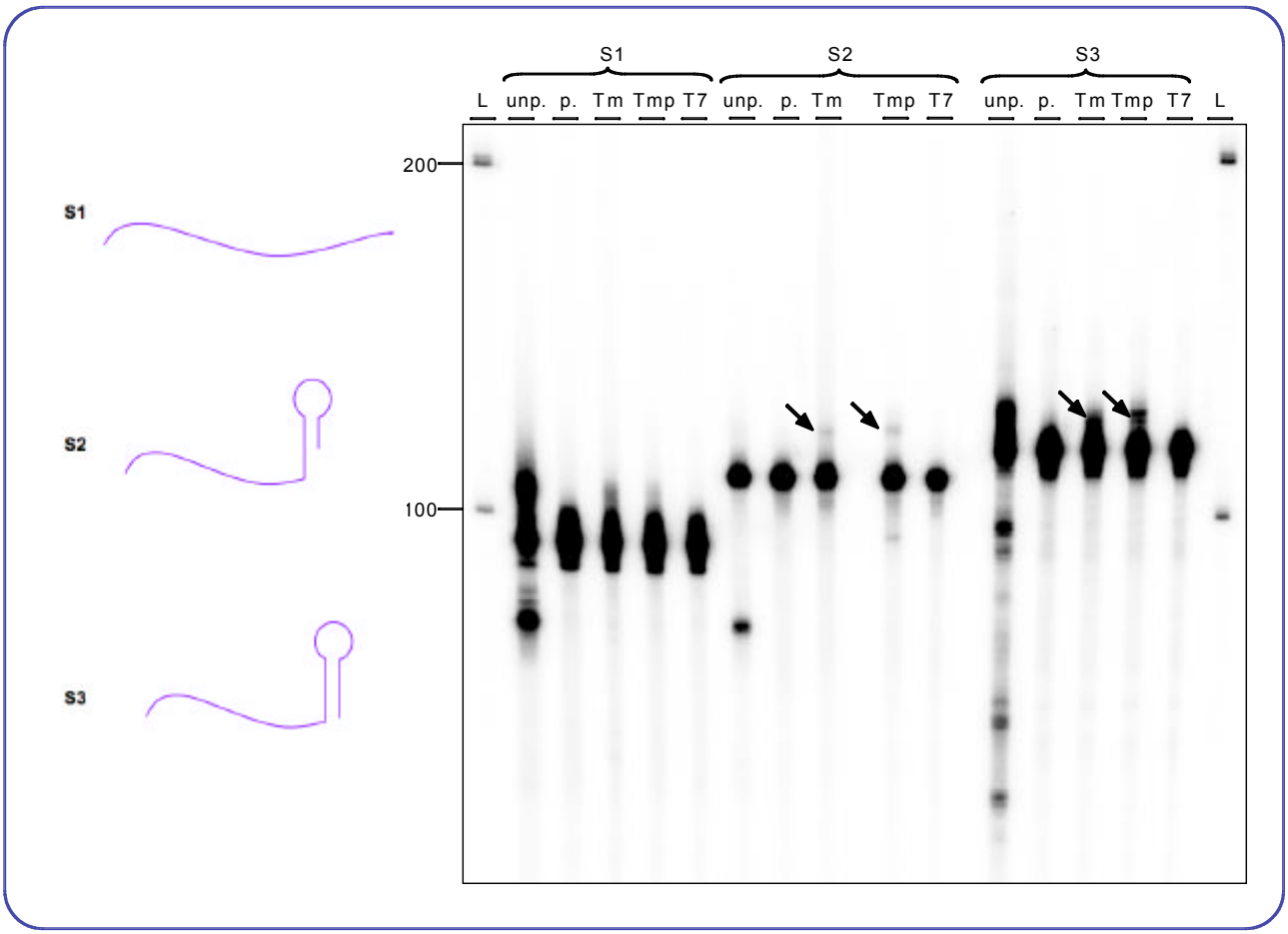


Figure 9 : In vitro RNA-dependent RNA polymerization assay. Each of the three [^{32}P]UTP radiolabeled RNA substrates S1, S2 and S3 were purified by electrophoresis fractionation followed by gel excision, 50ng were incubated 50 minutes at 30°C with overexpressed RNA polymerases RpoTm (0,08pmol) or RpoTmp (0,02pmol) or with the T7 polymerase from Ambion (0,3U), as a negative control. RNAs were extracted and loaded on acrylamide gel. unp. : unpurified subtrate not incubated with any polymerase, p. : purified substrate not incubated with any polymerase, Tm : purified substrate incubated with RpoTm polymerase, Tmp : incubation with RpoTm polymerase, T7 : incubation with T7 polymerase. 3' parts of secondary structures of the S1, S2 and S3 substrates are shown at the left of the autoradiography.

Three different radiolabeled RNA substrates were tested for *in vitro* elongation : S1, carried an unstructured 3' terminus, S2 had an incomplete stem-loop at the 3' end and S3 is terminated by a hairpin (see Figure 9). These substrates were incubated with overexpressed RpoTm and RpoTmp RNA polymerases, and T7 polymerase (Ambion) as a negative control. Resulting RNA products were extracted and separated by electrophoresis, along with unpurified (unp.) and purified (p.) substrates alone (see Figure 9). The autoradiography of the gel shows that species of higher molecular weight than the substrates were detected upon incubation with RpoTm and RpoTmp, but not T7. These species migrated as a smear for S1 and S3 but a discrete band of a higher molecular weight than the substrate was detected for S2. A band of lower molecular weight was also detected, probably corresponding to a degradation intermediate. Unfortunately, we were not able to clone these products for further analysis.
These first results suggests that RpoTm and RpoTmp are indeed able to elongate RNA substrates. S2 substrate, presenting an incomplete hairpin, was apparently a preferred substrate. To determine the precise elongation activity of RpoTm and RpoTmp, further *in vitro* analysis would be required, as for instance RNA-dependent RNA polymerization assay using cold probes, in presence of radiolabelled nucleotides.

Taken together, these results suggest that some plant mitochondrial transcripts are not only polyadenylated but also extended by heteropolymeric sequences. Phage-type polymerases could be responsible for these extensions. The function of these extensions remains mysterious. It would be interesting to investigate if the RNA polymerizing activity responsible for them is antagonist to the degrading activity of the PNPase. However, we did not observe heteropolymeric extensions to transcript of known function for the moment.

5. Non-coding transcripts are edited.

A second unexpected result of the analysis of the small polyadenylated clones library was the presence of editing sites in non-coding transcripts. So far, 441 C to U editing sites have been characterized in ORFs, 8 in introns and 7 in leader or trailer sequences (Giegé *et al*, 1999), but editing site in intergenic sequence were not described. Here, we observed editing sites in expressed intergenic regions. Mitochondrial clones for which editing has been observed are listed in the table 2. 39 clones out of 301 (13%) are edited, spanning 33 editing sites. 25 of these editing sites were not identified previously. However, as the library was done in Col0 background and the sequenced mitochondrial genome is the C24 ecotype sequence, a microgenetic variation cannot be excluded for some of these sites.

A number of editing sites were detected in intergenic sequences and a few novel editing sites were found in unknown ORFs.

clone name	location	fraction of edited clones
mIID5R	tRNA-fMet/ 3' tRNA-fMet	1/1
mVA3R	3' 5S	1/1
mTIF10R	3' 5S (nt 360064)	3/3
mIVC11F	3' 5S (nt 360064)	3/3
mIIC6R	3' 5S (nt 360064)	3/3
mVB5R	3' 5S (nt 360335)	2/2
mIIIF7F	3' 5S (nt 360335)	2/2
mTIF2R	3' orf 315	1/1
mVC1F	3' tRNA Pro	1/4
mIVE8F	3' tRNA Ser	1/1
mVG7R	3' tRNA Ser	1/5
mVD12F	3'tRNA Asn (nt 106058)	3/5
mTID6R	3'tRNA Asn (nt 106058)	3/5
mIIA5R	3'tRNA Asn (nt 106058)	3/5
mIIA6F	3'tRNA Tyr	1/8
mVF5F	5' nad 3 (3 sites)	1/1; 1/1; 1/1
mVF7R	5' NCO	1/1
mB14f	5' orf 315	1/1
mVB12F	5' tRNA Ile	1/5
mVA2R	5'26SrRNA gene	1/2
mIIIB7F	5'26SrRNA gene	1/2
mIIIG9F	ccmFc (2sites)	1/1 and 1/2
mIVC6F	cob	1/1
mVE9R	intergenic region between orf 106e and orf 107f	1/1
mVF1F	intergenic region between orf 215a and 215b	1/1
mIIIH2R	intergenic region between orf154 and repeat II	1/4
mIVH6R	NCO (nt 47302; nt 47307; nt 47324)	1/9; 1/9; 9/9
mIVA3R	NCO (nt 47324; nt 47341)	9/9 2/9
mIVA8F	NCO (nt 47324; nt 47341)	9/9 2/9
mIIH1F	NCO(nt 47324)	9/9
mIIIH11F	NCO(nt 47324)	9/9
mTIH3R	NCO(nt 47324)	9/9
mB5f	NCO(nt 47324)	9/9
mTID12F	NCO(nt 47324)	9/9
mTIE2R	NCO(nt 47324)	9/9
mVD8F	orf 240a	1/1
mIIIE1F	orf 275	1/4
mTIA7R	orf135b	1/2
mIVB3R	rpl16 (2sites)	1/1

Table 2 : List of *Arabidopsis* mitochondrial edited clones of the library. The genomic location of editing sites are indicated and positions are mentioned in brackets for some clones. Genomic regions in green correspond to previously described sites by P.Giegé *et al.* (Giegé *et al*, 1999). Fraction of edited clones for a single position is indicated.

Figure 10 illustrates editing sites found in intergenic sequences. At the *trn* gene cluster, C to U conversions are not present in all clones at a given position, excluding a possible microgenetic variation between C24 and Col0 ecotypes. Moreover, the C to U editing site at position 106058, downstream of $tRNA^N$ was observed in three independent clones, indicating that the nucleotide modification is not due to microgenetic variation neither to artefact. The NCO locus contains a hot spot of editing sites. The editing site at position 47324 (or 181489 in the other repeat) is edited in all nine corresponding clones. Additional editing sites were detected in some but not in all clones spanning the region.

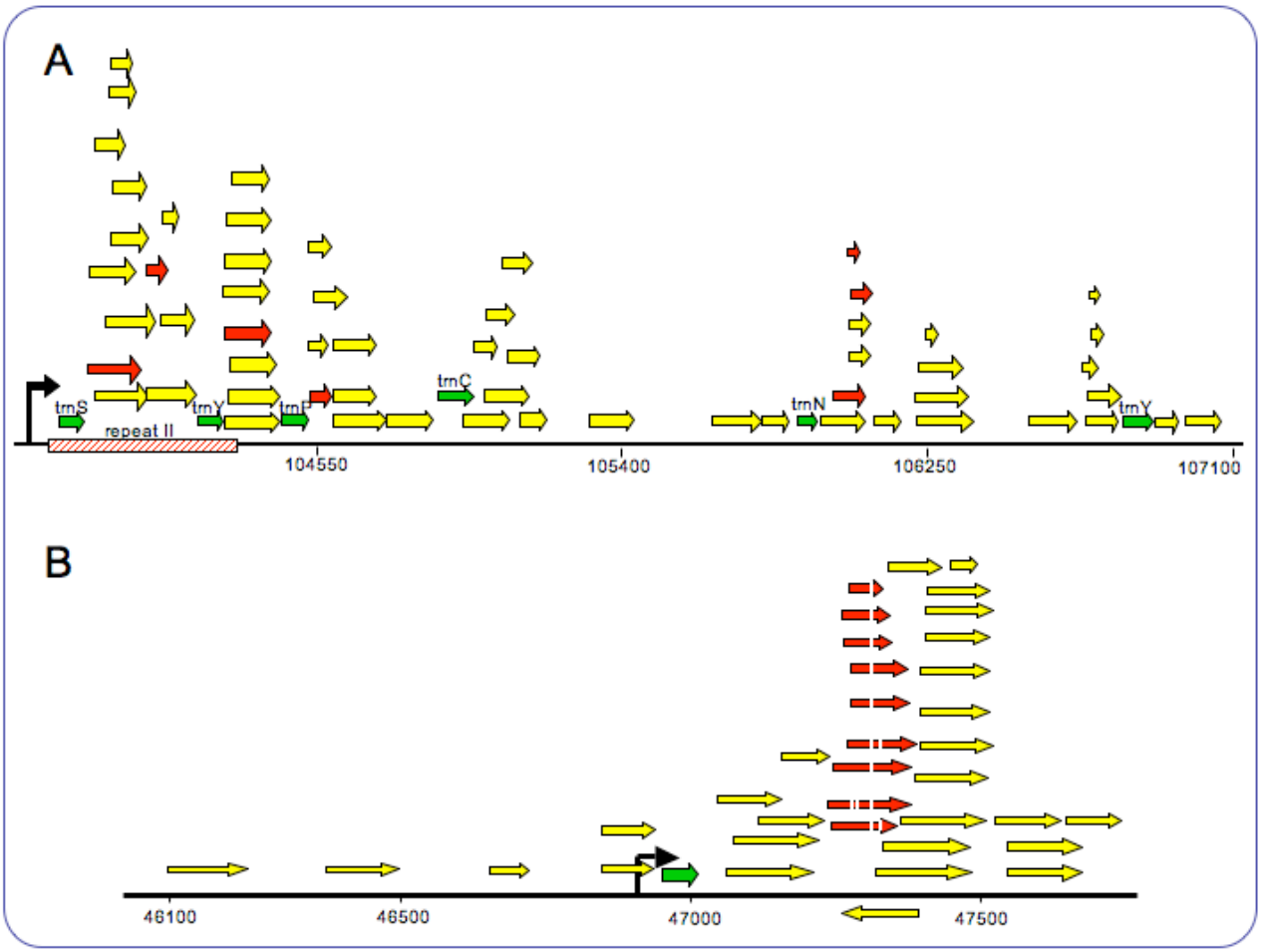


Figure 10 : Intergenic transcripts can be edited. **A**. Location of the edited and non edited clones at the *trnS-trnY-trnP-trnC-trnN-trnY* locus. *trn* genes are represented by green arrows. **B**. Location of the edited and non edited clones at the NCO locus. Exact position of the editing sites are indicated by white bars. The green arrow represents the Ath-59 transcript. Edited clones are represented by red arrows, non edited clones by yellow arrows. Putative promoters are represented by bent arrows.

As one editing site of the NCO locus was found in all clones, I verified the genomic sequence at this locus in the Col0 ecotype to exclude a possible micro-genetic variation between Col0 and C24 ecotypes. The relevant portion of the mitochondrial genome of PNP- plants was amplified by PCR, cloned and sequenced. A picture of agarose gel separating the PCR products is shown on the left side of Figure 11. As presented in the table of the same figure, this genotyping confirmed that the genomic C nucleotide of this site is converted to a U in PNP- transcripts.

100% of the transcripts investigated in PNP- plants are edited at this site. But is it also the case in wild type plants? To answer this question, I performed RT-PCR on wild-type mitochondrial RNA, followed by cloning and sequencing (see Figure 11, on the right). In 20% of the clones, this position was edited, indicating partial editing of this RNA in wild type plants. This result indicates that these RNAs are degraded before editing could operate in wild-type background. Compe-

tition between editing and degradation probably explains the reduction of edited RNA pool.

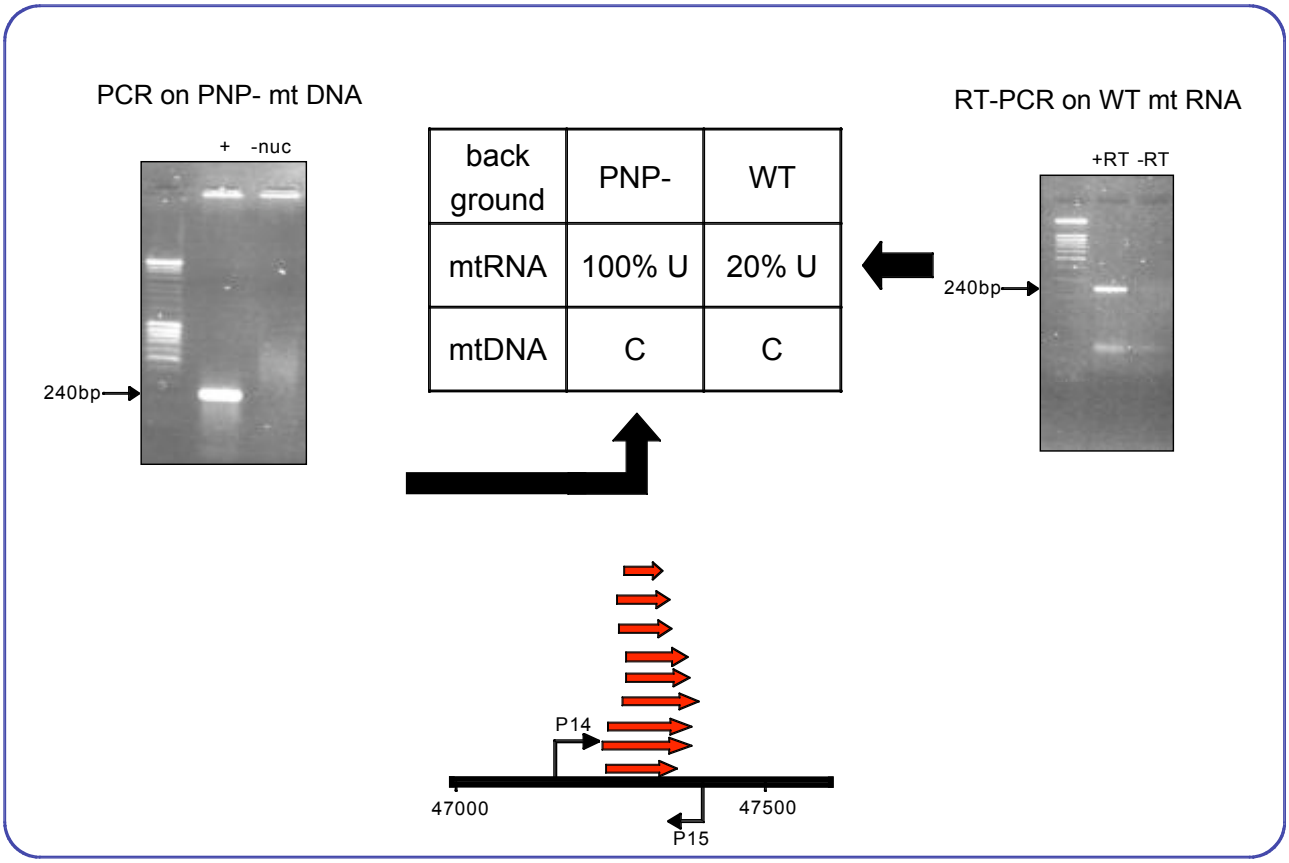


Figure 11 : The position 47324 (and corresponding 181489) is edited in PNP- plants and is not a microgenic variation between Col0 and C24 ecotypes. PCR Products on mitochondrial DNA from PNP- plants is shown on the left. Lane + : PCR products of the reaction using P14 and P15 primers, which position is indicated on the scheme. Lane -nuc : negative control for nuclear contamination, using primers on the nuclear genome. On the right, products of the RT-PCR on wild-type mitochondrial RNA are presented. Lane +RT shows PCR products of the reaction using P14 and P15 primers on cDNA synthesized with P15 primer. Negative control of the reverse transcription is shown in lane -RT.
The table sums up results of cloning and sequencing of the PCR products obtained for the two experiments.

The biological significance, if any, of these novel editing sites remains unknown for the moment. However, the results presented here suggest that more sites than previously thought are edited or, at least, “editable” as long as the target RNA is stabilized by the lack of PNPase.

Thus, these results raise the possibility that PNPase, by eliminating intergenic transcripts, would prevent titration of the editing machinery by abundant non coding transcripts. Competition between transcript for editing machinery has been already described in chloroplast (Chateigner-Boutin & Hanson, 2002). As mitochondrial editing features are similar to chloroplastic ones, competition for editing machinery could exist in mitochondria as well.

6. Antisense transcripts are expressed in mitochondria.

Among the degradation tags detected by our screen of small polyadenylated mitochondrial RNAs, we found one clone in antisense (AS) orientation to the *NCO* transcript (see location on the Figure 12A).

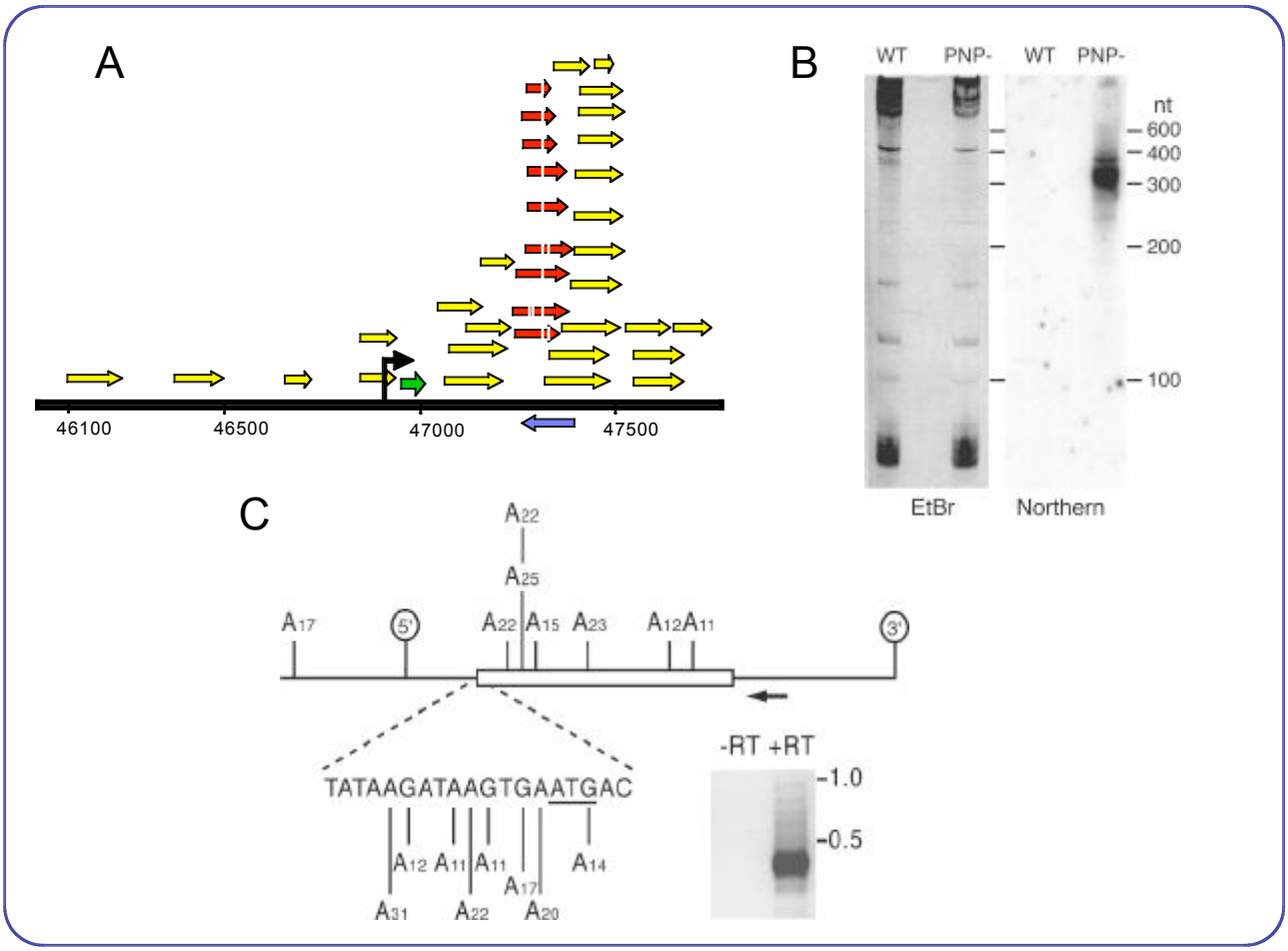


Figure 12 : RNA is synthesized in antisense orientation to *NCO* and *atp9* transcripts. **A.** Location of the antisense degradation tag (blue arrow) and the sense degradation tags (yellow and red arrows) at the NCO locus. Ath-59 is represented by a green arrow. The promoter is represented by a bent arrow. Edited sites are represented by white bars in the red arrows. **B.** Northern analysis of the accumulation of the AS RNA to the *NCO* transcript in PNP- plants. Radiolabeled RNA corresponding to nt 47264 to 47395 of the *Arabidopsis* mitochondrial genome was used as probe. The ethidium bromide (EtBr)-stained gel is shown for loading control. **C.** Mapping of polyadenylation sites of *atp9* AS transcripts by oligo(dT) adapter-primed RT-PCR. The white box represents the *atp9* ORF, 5' and 3' end are indicated. The position and length of the poly(A) tails are indicated on the sequence for each clone analyzed. The black arrow represents the gene-specific primer used in combination with oligo(dT) adapter-primer for the RT-PCR. The EtBr stained agarose gel of RT-PCR products is shown. The presence or absence of the reverse transcriptase is indicated by +RT or -RT.

This clone is complementary to the part harboring the 100% edited site but exhibits a G at the relevant position. Thus it is unlikely to result from an artifac-

tual amplification using the forward transcript as template. Rather, it could originate from transcription on the opposite strand.

To test antisense transcription at the NCO locus, we collaborated with Kristina Kühn in Thomas Börner's group. Using a modified 5' RACE technique allowing identification of primary transcripts, she was able to detect synthesis of RNA in antisense orientation to the *NCO* transcript. The initiation site of these RNAs was located downstream of a CGTA sequence, a typical core motif of promoters in *Arabidopsis* mitochondria. This result confirmed that transcription in antisense to the *NCO* transcript occurs.

Accumulation of this antisense RNA was then addressed by Northern analysis (Figure 12B). A strong signal was detected in PNP- plants whereas no signal was observed in wild-type plants. This indicates that PNPase degrades the *NCO* antisense RNA. Interestingly, the transcripts antisense to *NCO* are detected as a single band, indicating that they have defined extremities.

The presence of this antisense transcript in *Arabidopsis* mitochondria prompted us to look for other RNAs in antisense orientation. In a first approach, we used oligo-dT adapter-primed RT-PCR to detect polyadenylated RNA in antisense to *atp9* mRNA (Figure 12C). Location and length of poly(A) tails confirmed the antisense orientation of these transcripts. Positions corresponding to editing sites harbored a guanosine, excluding artifactual synthesis from sense transcript. These results suggest the activity of a promoter in antisense orientation to *atp9.* Indeed, the detection of antisense primary transcripts by modified 5' RACE technique confirmed the presence of such a promoter. However, nested-PCR was required to detect these transcripts, indicating that the promoter activity is apparently very weak. Consistently, transcripts antisense to *atp9* could not be detected by Northern blot. Using again modified 5' RACE technique, Kristina Kühn could also confirm that promoters drive antisense RNA synthesis for 3 other genes tested : *nad4*, *nad5* and *nad7* (see publication in appendix for further details).

These results indicate that antisense RNA to known genes are common in *Arabidopsis* mitochondria. Polyadenylated status of *atp9* and *NCO* antisense transcripts and accumulation of antisense transcripts in PNP- plants suggest that these RNA are degraded by PNPase.

Thus, these results suggest that PNPase prevents the accumulation of antisense RNA. Whether these antisense RNAs are deleterious for mitochondrial functions or not remains to be determined.

7. Conclusion

Large transcripts and rRNAs precursors were previously shown to accumulate upon depletion of mitochondrial PNPase (Perrin *et al*, 2004 a and b). The present analysis identified novel polyadenylated substrates of PNPase. The characterization

of small degradation tags generated sequence information that was then used to determine the sizes of RNA species accumulating in PNP- plants by Northern analysis. The results of these experiments support previous findings and allows a generalization of the role of PNPase. First, we found that PNPase removes maturation by-products of rRNAs and tRNAs. Moreover, PNPase degrades spurious transcripts generated by the relaxed control of transcription in plant mitochondria.

We observed degradation tags arranged in a consecutive manner (see Figure 4) and small RNA species, detected by Northern analysis upon long exposure (see Figure 7), that are shorter than full-length maturation by-products. This suggests that an alternative pathway involving an unidentified endonuclease could also be acting in plant mitochondria. However, mostly full-length RNA accumulate in PNP- plants, indicating that endonucleolytic activity plays a minor role. Hence, degradation in plant mitochondria is mostly exonucleolytic and performed by PNPase.

AtmtPNPase is essential for viability, most likely because of its role in turnover and maturation of stable RNAs. However, its role in the removal of highly transcribed non functional RNAs may also be crucial for plants.

We identified transcripts that were expressed to extraordinary high levels, illustrating the relaxed control of transcription initiation in *Arabidopsis* mitochondria. These transcripts could be detrimental for mitochondrial functions. For example, high accumulation of the *NCO* transcript, could mobilize editing machinery and processing machinery, and thus compete with functional RNAs.

Other RNAs with a deleterious potential are chimeric ORFs and antisense RNAs. Expression of chimeric ORFs can have important biological consequences, as most CMS systems are associated with such ORFs. Our results show that some chimeric ORFs, as *orf240a*, *orf294,* and *orf315* are expressed in *Arabidopsis* mitochondria. This is in agreement with previous studies describing transcription of mitochondrial ORFs without known function (Giegé *et al*, 2000 ; Leino *et al*, 2005).

Moreover antisense RNA are commonly produced in *Arabidopsis* mitochondria, as all five regions investigated, *nad4*, *nad5*, *nad7, atp9* and *NCO*, exhibited transcription on the opposite strand. Effects of these antisense RNAs are not documented to date but it has been demonstrated recently that expression of an antisense RNA to the *Rpo* transcript in chloroplast results in abnormal growth and development of the plant (Hegeman *et al*, 2005). The molecular reason of this impairment is unknown as neither editing status nor steady-state levels of RpoB protein and its transcript was affected. However, this example illustrates the potentially deleterious effect of antisense RNA in a plant organelle.

An intriguing result of this analysis is addition of non-genomically encoded heteropolymeric extensions upstream of the poly(A) tail displayed by some degradation tags. The potential RNA-dependent RNA polymerization activity unveiled here could similarly have negative effects on *Arabidopsis* mitochondria by ectopically stabilizing spurious transcripts through structural modifications.

Polyadenylation status of the degradation tags, supported by accumulation of RNA species in PNP- plants, clearly demonstrate that PNPase plays an important role in degrading transcripts in antisense to known genes, chimeric ORFs, and non

functional RNAs that are highly transcribed, but have a deleterious potential for mitochondria.

Thus we can assume that, in addition to its role in maturation and turnover of stable RNAs, PNPase is involved in a general surveillance pathway operating in Arabidopsis mitochondria to counterbalance relaxed transcription (Figure 13).

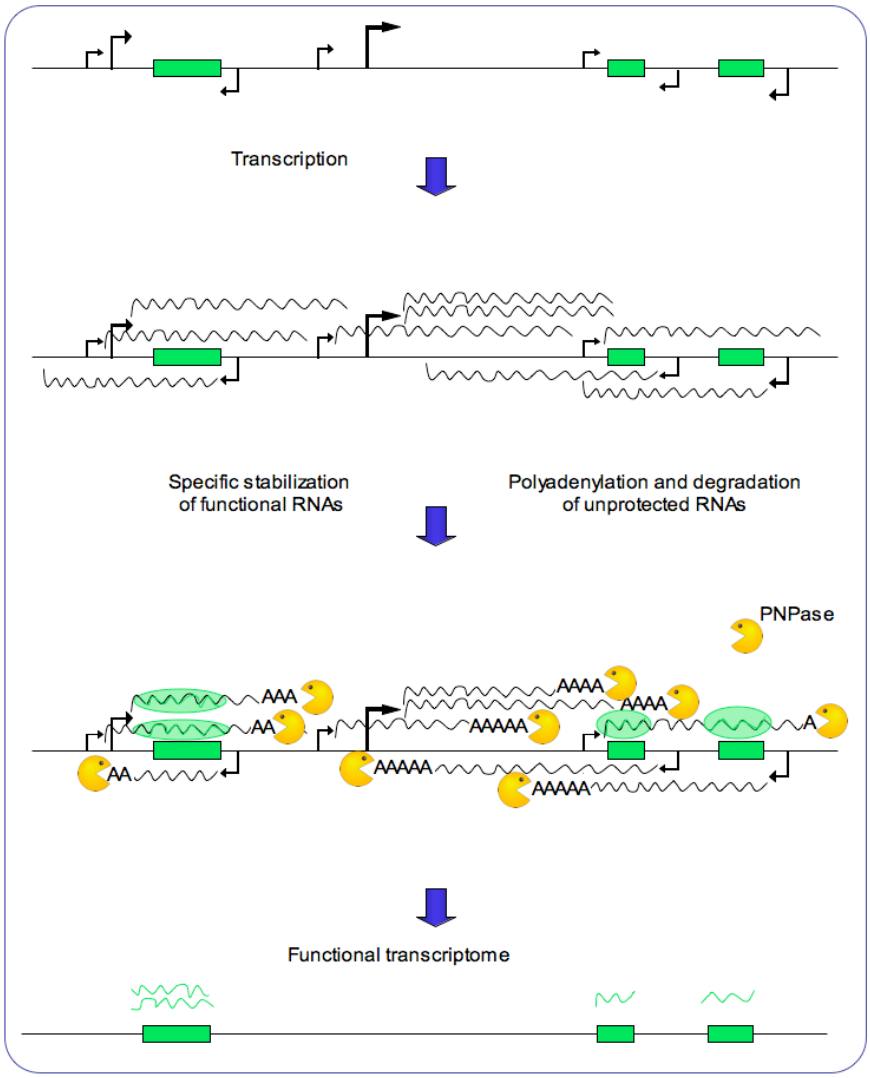


Figure 13 : Cryptic transcripts are degraded by PNPase to ensure a functional plant mitochondrial transcriptome. Transcription initiated in intergenic regions, or in antisense to genes, as well as lack of transcription termination lead to accumulation of cryptic transcripts. Whereas legitimate transcripts are protected by specific factors, other RNAs are degraded by PNPase, leading to a transcriptome that is functional for plant mitochondria. Genes are depicted by green rectangles, promoters by bent arrows. Functional transcripts, specifically protected, are depicted in green (from the review to be published in Biochimica et Biophysica Acta - Structure and Gene Expression).

Chapter II :
A model RNA to study extremity maturation of a plant mitochondrial RNA

While determining PNPase substrates in *Arabidopsis* mitochondria, we characterized a transcript that accumulate to high levels upon PNPase down-regulation. Interestingly, this transcript has defined mature extremities. Its features distinguish this transcript from previously studied mitochondrial transcripts : because of its high turnover, the transcript processing unambiguously occurs independently of PNPase.

Indeed, as PNPase is essential for the plant, down-regulated mutants were uniquely obtained by cosuppression. As cosuppression is an uncontrolled event, the moment of PNPase down-regulation is unknown. Thus, for an RNA having a low turnover, it is not possible to know if this RNA has been processed before PNPase down-regulation or if it can be processed independently of PNPase.

Because of its high turnover, this novel transcript is without doubts a good model to analyze events taking place in absence of PNPase. As it is accumulating as a mature specie, extremities processing is independent from PNPase. We thus attempted to understand processing mechanisms leading to extremities maturation of this transcript.

1. A Non-Coding Overexpressed (NCO) transcript of unknown function is present in Arabidopsis mitochondria.

As we were characterizing RNAs accumulating in absence of PNPase, we found transcripts generated from a region in the *Arabidopsis* mitochondrial genome, that does not code for any known gene. This region is situated in one of the two large repetitive sequences that are thought to be involved in recombination events of the mitochondrial genome (Backert *et al,* 1997). This repetitive sequence is called repeat I and spans positions 44698 to 48894, and 178863 to 183059 of the mitochondrial genome, ecotype C24 (accession number NC_001284).

A consensus motif for promoters is present in position 46909 (and 181074) of the genome, and could be responsible for expression of the downstream sequence. In position 48113 to 48532 (and 182278 to 182697), an unknown ORF, ORF139, is annotated. This ORF, not conserved between species, is one of the 85 ORF larger than 100 codons annotated in the *Arabidopsis* mitochondrial genome (Unseld *et al*, 1997). The region between the putative promoter and ORF139 contains none of such ORFs. This region is not conserved between plant species and is chimeric i.e. composed of sequences likely arising from minor recombination events. Stretches similar to *18Srrn*, *tRNY* and to chloroplastic sequences can be found (Figure 14A). Taken together, the non conserved and chimeric nature of this region suggests that it does not encode any functional transcript.

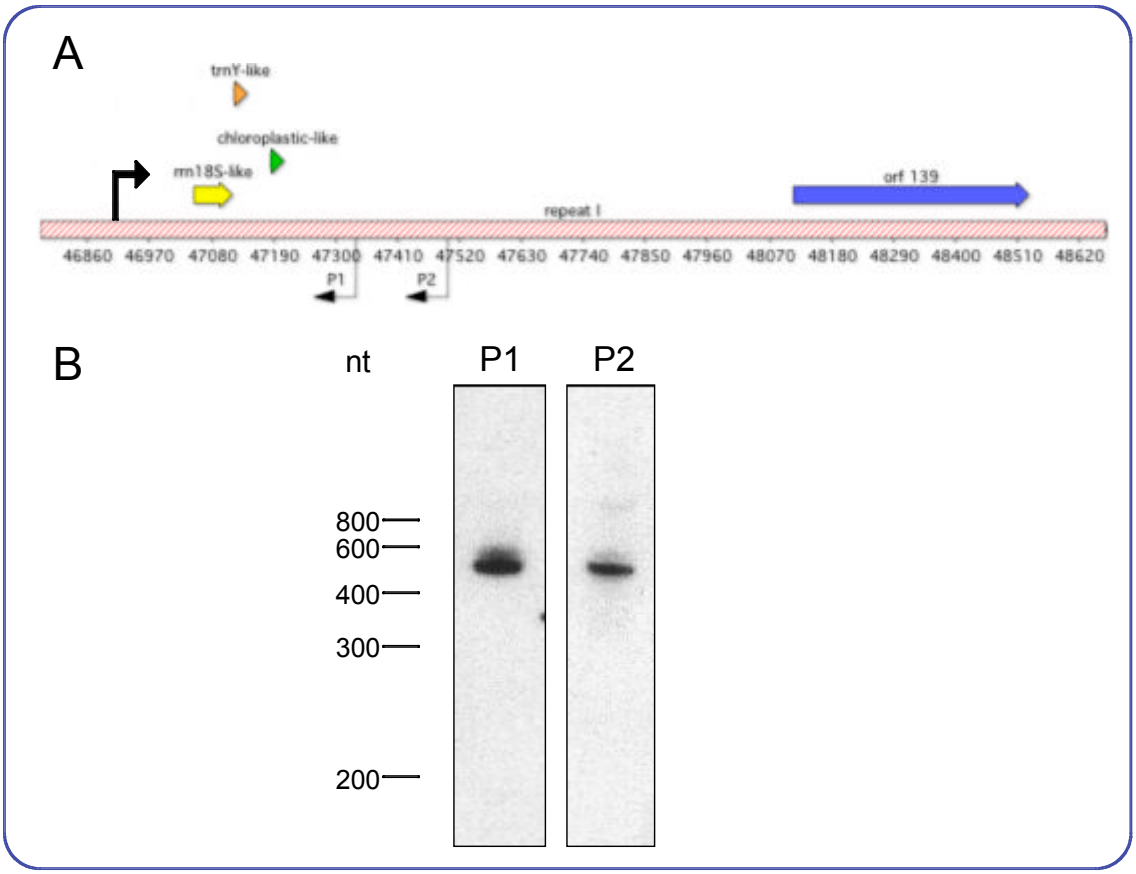


Figure 14 : **A.** Scheme of the Non Coding Overexpressed locus of the *Arabidopsis* mitochondrial genome. Position of the P1 and P2 probes used for the Northern analysis are marked with black arrows. Sequences of interest are represented by colored arrows, the putative promoter is depicted as a bent arrow. **B.** Northern analysis of the *NCO* transcript expression in WT mitochondria.

I have already mentioned that the transcript corresponding to this region can be detected in wild-type plants. This result was confirmed using two different primers upstream of orf139, namely P1 and P2 in Northern blot experiments. Indeed, I detected a transcript of about 500nt in wild-type mitochondria (Figure 14B).

Since this transcript accumulates in absence of PNPase and does not contain any ORF, I named it Non Coding Overexpressed (NCO) transcript.

2. The novel NCO transcript has defined 3' and 5' extremities.

Since this novel transcript was detected as a unique band of 500nt, in wild-type plants and in absence of PNPase, I wanted to determine its 3' and 5' extremities. In this purpose, I performed circular RT-PCR (cRT-PCR) experiments (Figure 15).

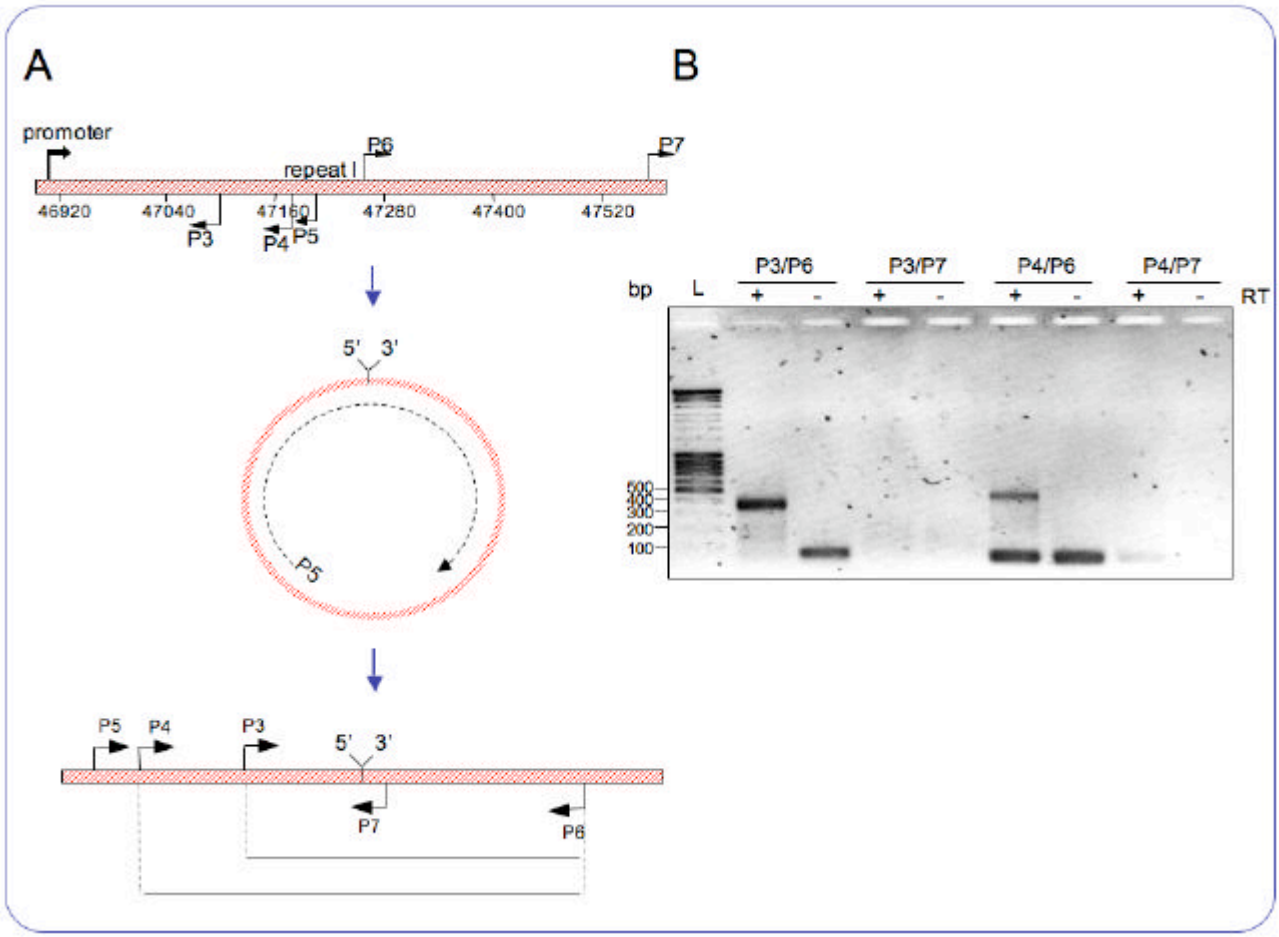


Figure 15 : **A**. Scheme of the primers used for the cRT-PCR. At the top, primers are positioned on the mitochondrial genome. Below, ligated transcript is represented : P5 is the primer used for the reverse transcription of WT mitochondrial RNA. After ligation and reverse transcription, primers used for the PCR are represented on the linear cDNA, with the two PCR products revealed on the gel in B. **B**. Products of the cRT-PCR, separated on a 1,5% agarose gel, revealed with EtBr. The pairs of primers used for the different PCRs are indicated at the top of the wells. Sample and negative control without reverse transcriptase are indicated by + and -RT, respectively.

Following circularization by T4 RNA ligase and reverse transcription of wild-type mitochondrial RNA with P5 primer, PCR with P3/P6 and P4/P6 primers gave rise to amplicons of about 410bp and 490bp respectively. The difference of size corresponds to the distance between P3 and P4 primers. In contrast, no amplification was detected with primers P3/P7 and P4/P7, indicating that P7 is downstream from the expressed transcript. The band of small size observed in -RT when using primer P6 is most probably due to primer artifact.

Cloning and sequencing of the PCR products confirmed that 3' extremity of the *NCO* transcript is upstream of P7 (see Figure 16).

Out of 14 clones, 5' extremities of 12 clones mapped to the position 47020 of the *Arabidopsis* mitochondrial genome (in the first repeat). 1 clone ended 1nt upstream and 1 clone ended 2 nucleotides downstream of this position. 5'

extremities mapped just downstream of the small non messenger RNA Ath-59, which 3' extremity is annotated at the position 47018.

Ath-59 is one of the "small non messenger RNA" detected previously by Hüttenhofer's group (Marker *et al*, 2002). It was described as an abundant, potentially functional RNA. However, we were unable to detect it by Northern blot on mitochondrial RNA. As Ath-59 has a strong homology with $tRNA^{Phe}$, a cross-reaction with $tRNA^{Phe}$ was possibly detected by Marker and co-workers. However, they characterized Ath-59, indicating a steady-state level sufficient for cloning.

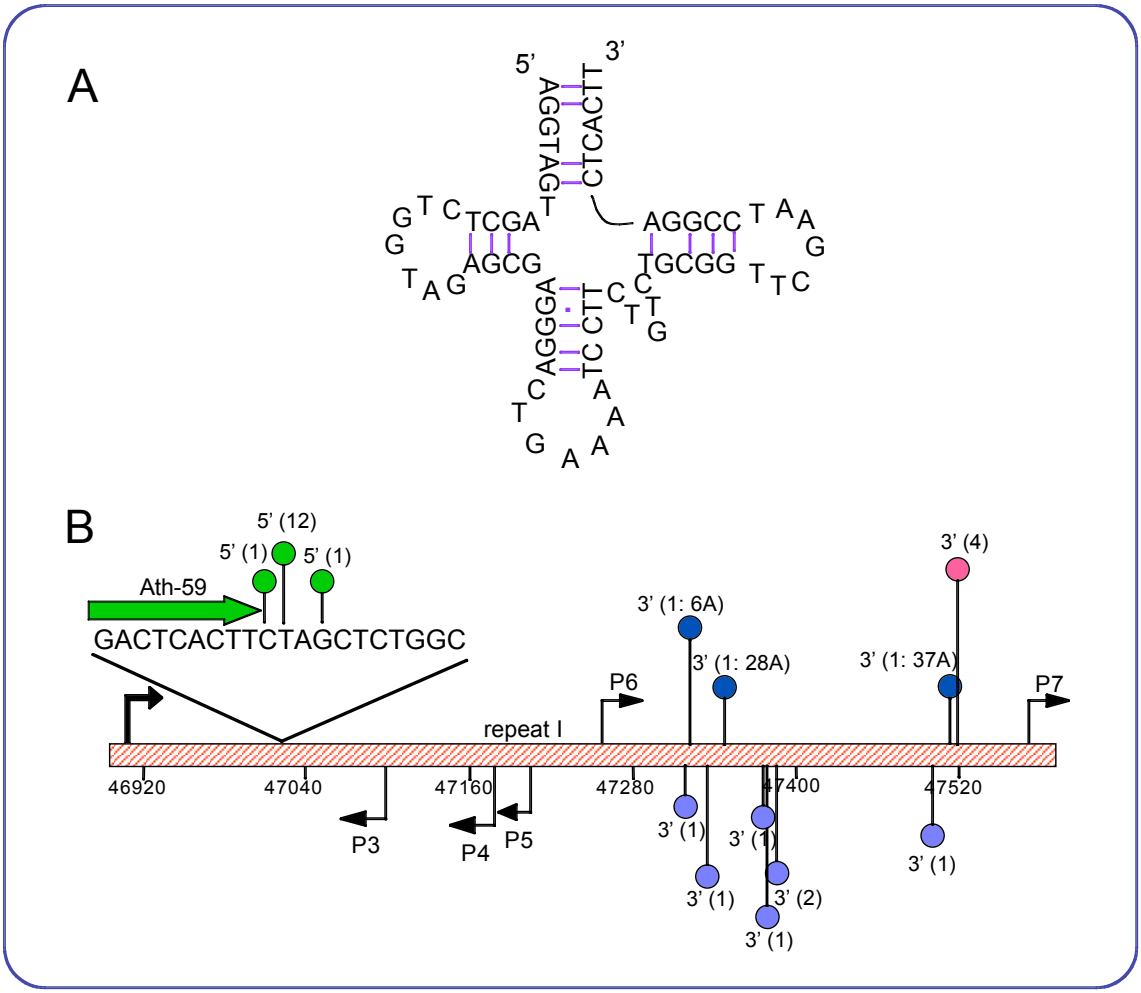


Figure 16 : **A**. The secondary structure of Ath-59 folds into a tRNA-like structure, with the $tRNA^{Phe}$ anticodon. **B**. 5' and 3' extremities of the *NCO* transcript for 14 clones of the cRT-PCR products. 5' extremities are indicated with a green bubble, 3' extremities are indicated with a blue bubble, faded and below the sequence when no poly(A) tail was observed. The pink bubble represents the major position of the 3' extremity. Position of the primers used for the cRT-PCR and of the small non messenger RNA Ath-59 are indicated. The number of clones are given in brackets for each position. The number of non-encoded adenosine in 3' of the sequence is indicated after the number of clones, except for the major position in pink, where several clones are ending.

The Ath-59 sequence was predicted to fold into a conventional tRNA structure (Laurence Maréchal-Drouard, personal communication, see Figure 16A). It is thus probable that processing of this tRNA like-element by RNase Z generates the

3' end of Ath-59 and, at the same time, the 5' extremity of the *NCO* transcript. Indeed, a recent work demonstrated that tRNA like-elements cleavage contribute to 5' and 3' end formation of mRNAs in plant mitochondria (Forner *et al*, 2007).

3' extremities of cRT-PCR clones were not all identical. 3' ends are spread from the position 47317 to the position 47521. However, a majority of 3' ends were located close to the position 47520. 4 clones at this position were clearly polyadenylated and carry 2, 15, 19, and 28 adenosines, respectively. These tails were mostly composed of As, except for one cytosine and one guanosine. The major 3' end position located 500nt downstream of the 5' extremity, in agreement with the length of the band detected by Northern blot (Figure 14). However, certain clones had their 3' extremity located upstream of the major position. Some of them were polyadenylated and likely represent degradation intermediates.

This analysis revealed that the novel NCO transcript generated from repeat I has precise 3' and 5' ends. Furthermore, this transcript is polyadenylated, and thus marked for degradation.

3. Maturation of NCO transcript is independent of PNPase and RNase II.

As described in the preceding chapter, *NCO* accumulates in absence of PNPase. This is demonstrated by the Northern blot analysis presented in Figure 17.

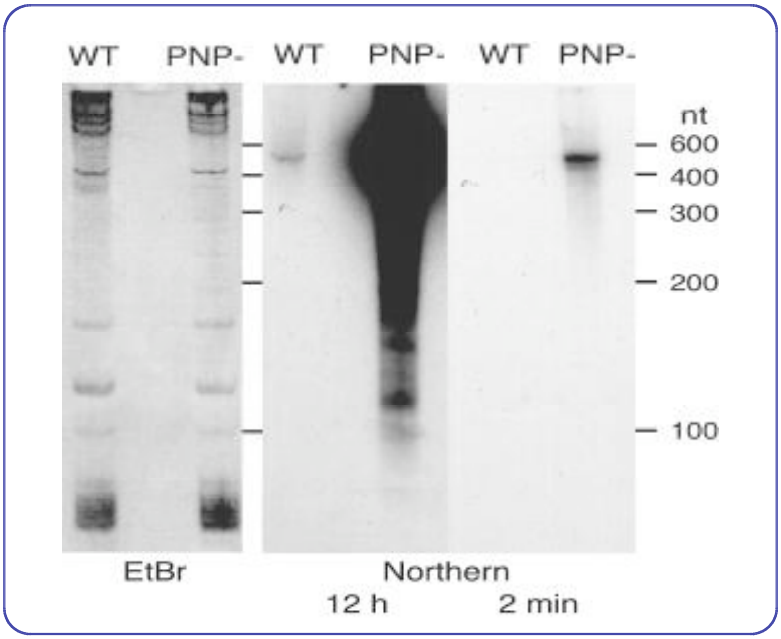


Figure 17 : Northern analysis of accumulation of the *NCO* transcript on total RNA from WT and PNP- plants. The probe used for detection is an *in vitro* transcribed RNA complementary to the position 47264 to 47395 of the mitochondrial genome. Two exposure times of the same experiment are presented, time is indicated at the bottom of each panel. A picture of the EtBr stained gel is showed for loading control.

The strong accumulation of *NCO* in absence of PNPase indicates that it is highly transcribed and that its degradation depends on PNPase. Interestingly, the size of the transcript is not affected by depletion of PNPase. Most of the RNA is detected as a single band of 500nt. Nevertheless, smaller RNA species, representing putative degradation intermediates are detected only upon long exposure times.

PNPase and RNase II exoribonucleases were proposed to act sequentially to process mitochondrial mRNAs such as *atp9* and *atp8*. PNPase could remove long 3' extensions and RNase II subsequently degrade short extensions (Perrin *et al*, 2004a). In chloroplast, both PNPase and RNase II are also involved in rRNA maturation (Walter *et al*, 2002; Bollenbach *et al*, 2005). To precisely map the extremities of *NCO* transcript in plants depleted for either PNPase or RNase II, I performed cRT-PCR experiments on RNA isolated from PNP- plants and from *rnasell* plants. Results are presented in the Figure 18.

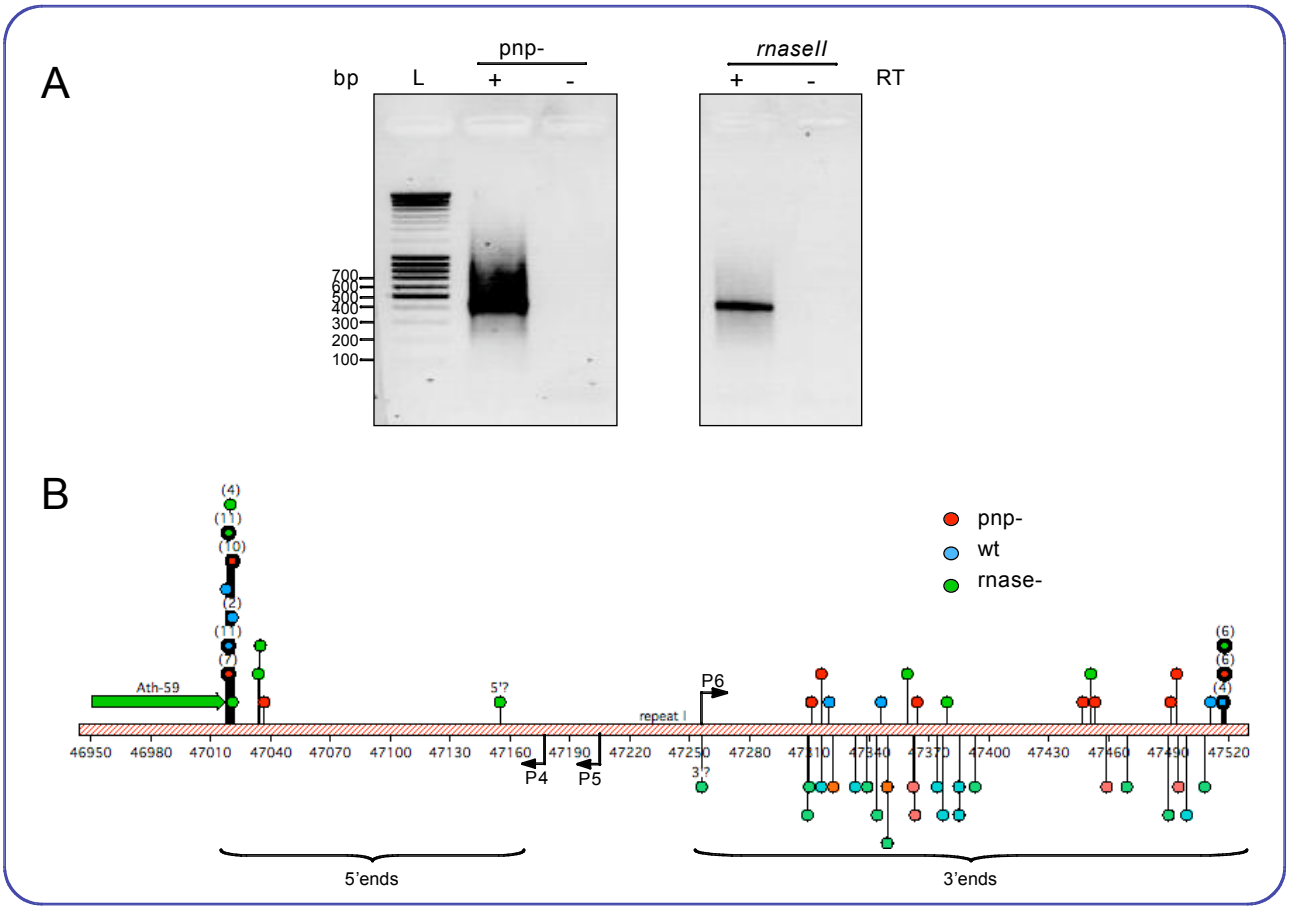


Figure 18 : *NCO* extremities in PNP- and *rnasell* backgrounds. **A**. cRT-PCR products separated on a 1,5% agarose gel. Presence and absence of reverse transcriptase are indicated by + and - RT at the top of the wells. 3µg of RNA are assayed for *rnasell* plants, 1,75µg for pnp- plants. Positions of the primers used on the genome are indicated in B. P5 was used for the reverse transcription, P4 and P6 were used for the PCR. **B**. Diagram of 5' and 3' extremities of cRT-PCR clones for pnp- (in red), *rnasell* (in green) plants and wt (from preceding analysis, in blue). Each bubble indicates an extremity for one clone, or more if mentioned in brackets. 3' extremities that lack any polyA tail are depicted below the sequence, in faded colors. ? represent possible experimental artifacts.

19 clones were analyzed here for PNP- and 20 for RNase II. 94,5% of PNP- clones and 84,2% of *rnase II* clones have their 5' extremity within 2 nucleotides after the 3' end of Ath-59. This resembles the situation in wild-type, indicating that 5' processing of *NCO* is not influenced by the lack of both RNase II and PNPase. 3' extremity positions varied between positions 47257 and 47521 of the genome. However a majority of polyadenylated clones ended at position 47520, representing a true 3' extremity. PolyA tails length of these clones varied between 8 and 30nt. Shorter clones harboring no or only few non-encoded adenosines could represent PCR artifacts that were preferentially cloned.

Processing of NCO is fast and efficient as it is properly matured in wild type plants despite its high turnover. These results show that maturation of NCO 3' end does not require PNPase or RNase II.

4. 3' end of the NCO transcripts is not generated by transcription termination.

5' ends of *NCO* are likely generated by an RNase Z activity that recognizes the tRNA-like structure of Ath-59. At the other end of the transcript, 3' ends maturation mechanisms of *NCO* transcript remain elusive. In theory, the mature 3' end of the *NCO* transcript could be generated either by transcription termination or by post-transcriptional processing.

To discriminate between termination of transcription and post-transcriptional processing, I measured the transcriptional activity in and after NCO coding sequence, by run-on experiments (Figure 19).

[^{32}P]-UTP radiolabeled transcription was performed *in organello* and products were hybridized to dot-blotted double stranded probes of 341 to 343bp, spanning NCO and *rrn26S* regions. 26S ribosomal RNA was used as an internal standard since this mitochondrial transcript is highly expressed similarly to *NCO*. The autoradiography of the dot-blot was quantified with a phosphor-imager to estimate the activity of transcription at the position of each probe. The evolution of transcription activity along the two regions investigated is presented in Figure 19.

For both regions, I observed a general decrease of the transcriptional activity in the 5' to 3' direction, that was already described for number of genes or gene clusters (Giegé *et al*, 2000). This decrease possibly reflects a general decrease of processivity of the RNA polymerase in these conditions or a bias of the labeling.

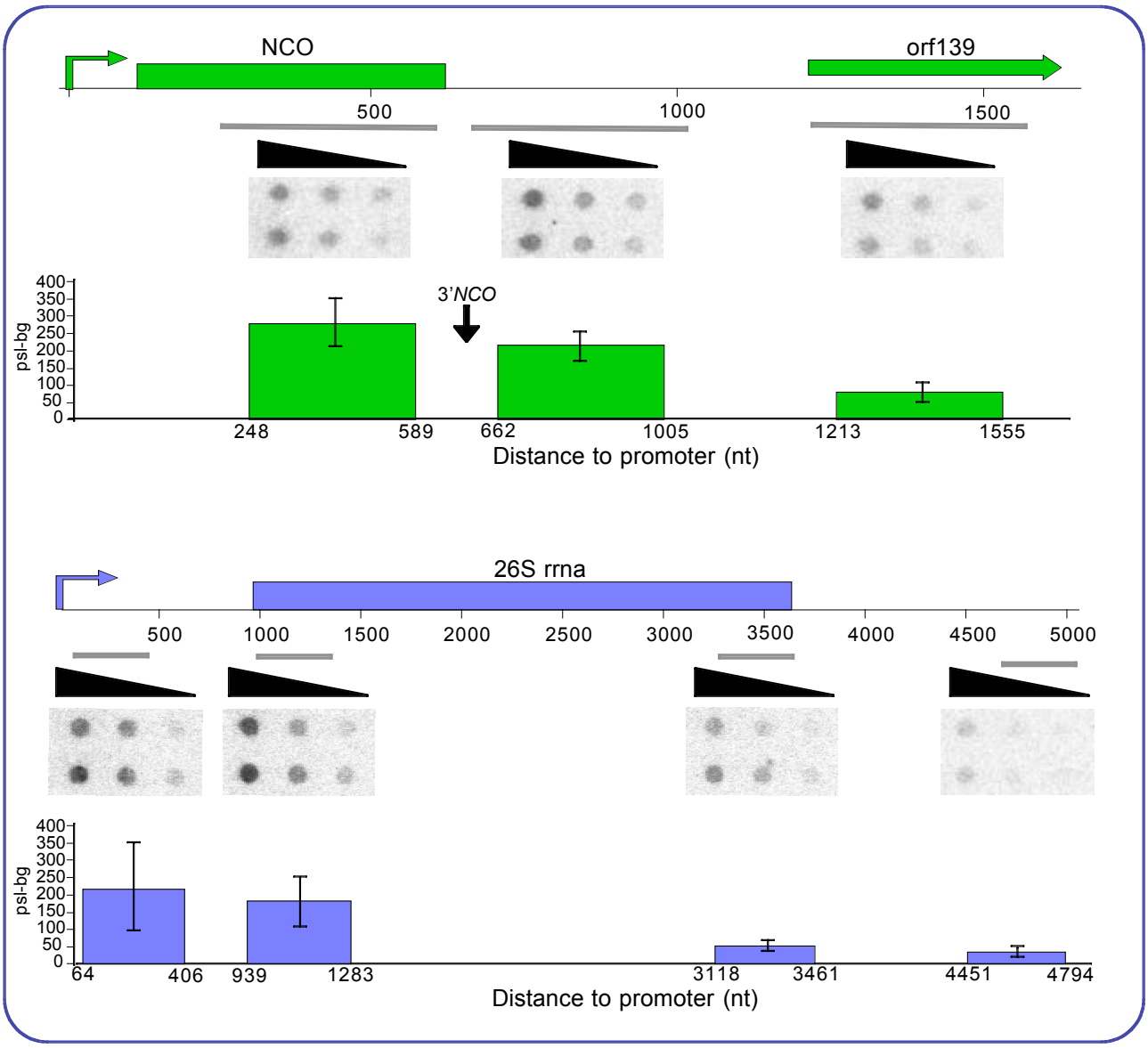


Figure 19 : Run-on assay to investigate a possible termination of transcription at the end of the *NCO* transcript. Probes are represented by gray bars, promoters are represented by bent arrows. Scales below the sequences represent the distance to the promoter given as nucleotides. For the run-on transcription, mitochondria from 200ml *Arabidopsis* cultured cells were incubated with 80µCi [^{32}P]-UTP and resulting transcripts were hybridized to duplicated dot-blots of 500ng, 150ng and 50ng of each probe. Signal were quantified by phospho-imager. Each point of the graph corresponds to the autoradiography shown above. Error bars are calculated from 3 independent experiments. psl-bg : photo-stimulated luminescence to which background has been subtracted.

3' ends of the 26S rRNA are not generated by transcription termination but by processing from large primary transcripts. Profile of transcription obtained for *NCO* appear similar to 26S rRNA transcription profile. In both cases, there is no clear drop of the transcription activity after the 3' end of the mature transcript.

This analysis provides evidence that the NCO 3' extremity is not generated by transcription termination.

5. What processes are involved in the generation of NCO 3' extremities?

As run-on assays showed that 3' extremities are not obtained by transcription termination, I wanted to characterize the processes involved in the maturation of the NCO 3' ends. A possible step of this maturation is an endonucleolytic cleavage. Thus following questions are raised : is there any endonucleolytic cleavage site in 3' of *NCO*? Can we map this site?

Two approaches are possible to address these questions : a biochemical approach, using an *in vitro* system or a molecular analysis aiming at characterizing intermediates generated by endonucleolytic cleavage.

5.1. Biochemical approach to test putative endonucleolytic processing.

In order to detect eventual endonucleolytic activity acting in *NCO* maturation, I performed *in vitro* processing assays (see Figure 20).

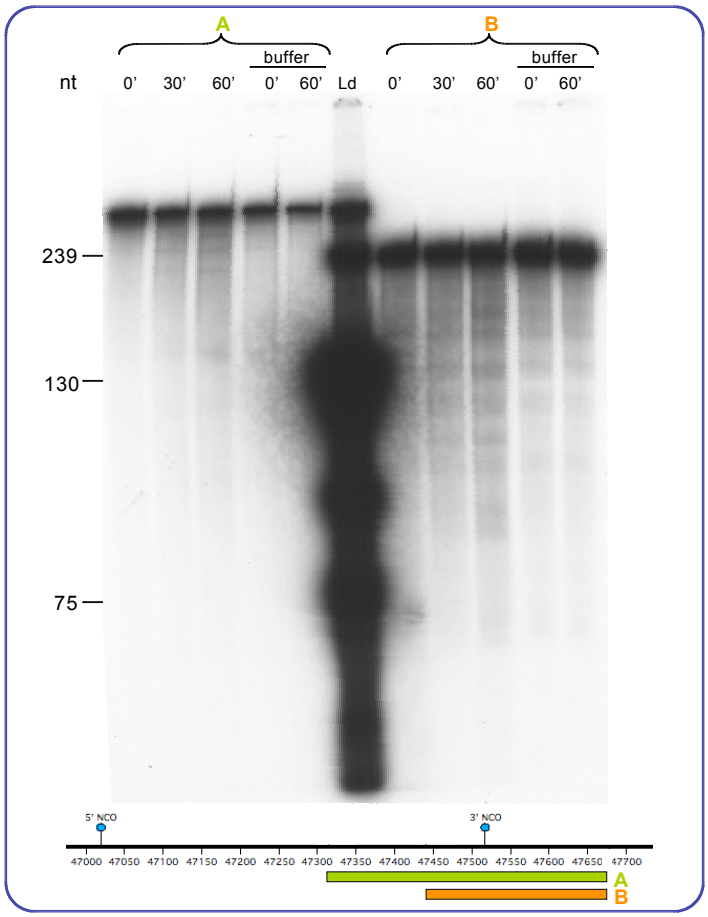


Figure 20 : *In vitro* processing assay. 0,7 mg mitochondrial proteins extracted from mitochondria of dark grown *Arabidopsis* cells were incubated with ^{32}P-[UDP] radiolabeled RNA probes A and B, positioned on the sequence on the scheme below the autoradiogram picture. Incubation times are mentioned in minutes above each line. Negative control was performed incubating probes with buffer without mitochondria.

Radiolabelled RNA probes were generated by *in vitro* transcription of T7 promoter containing PCR products spanning position 47314 to 47677 and 47441 to 47677 of the mitochondrial genome for probes A and B, respectively. These probes were incubated with mitochondrial protein extract for 0, 30 and 60 minutes. Resulting RNA species were separated by gel electrophoresis and detected by autoradiography. No changes were observed upon incubation with buffer alone. In contrast, weak bands smaller than full-length probes appeared after 30 and 60 minutes incubation with extract. These band could reflect degradation intermediates. However, a major band indicative of a at least semi-stable processing intermediate generated by a single endonucleolytic event was not observed. Neither was a significant ladder of bands suggesting an exoribonucleolytic activity.

In order to detect an eventual activity specific to *Arabidopsis thaliana*, or to the Brassicae, I repeated the assay using potato, cauliflower and *A.thaliana* mitochondrial extracts. However, this experiment did not reveal any clear difference from one extract to another, indicating that no specific activity could be detected.

5.2. Molecular approach to detect putative intermediates of endonucleolytic processing

First, I performed Northern blot experiments to analyze transcripts eventually generated from the region downstream of *NCO* (Figure 21).

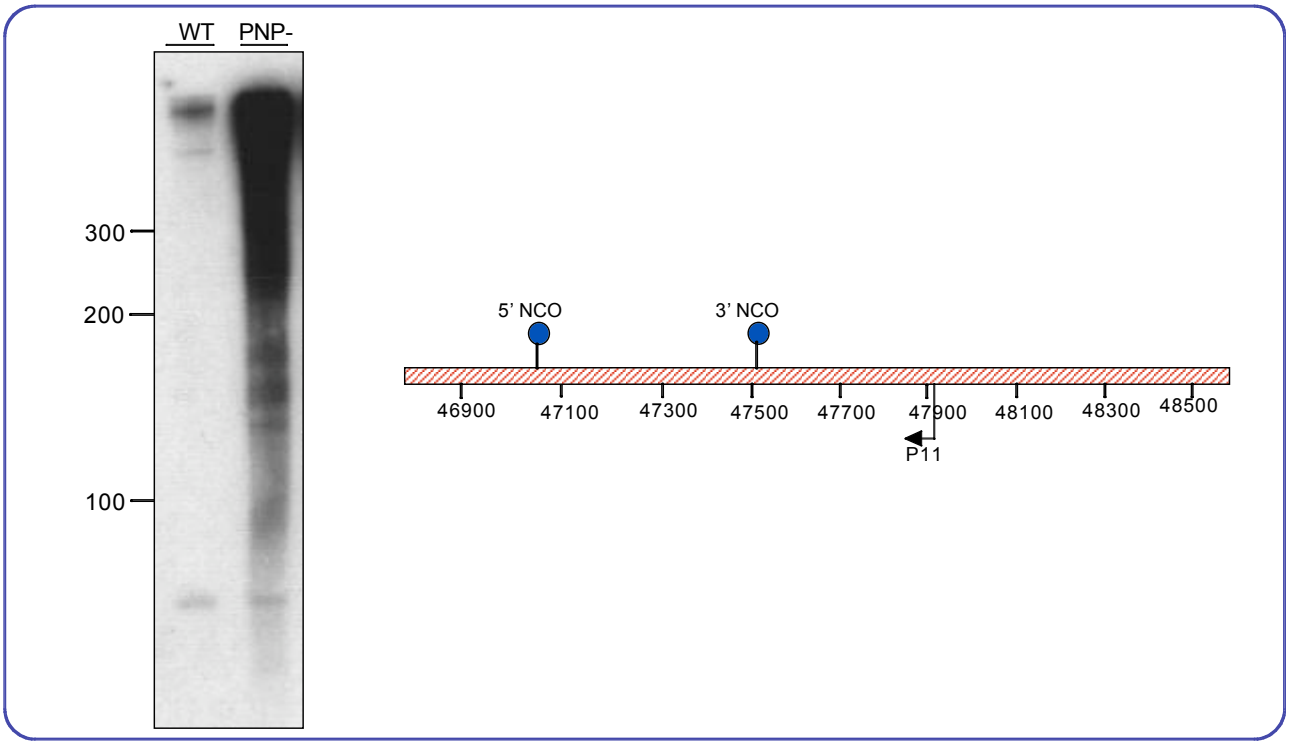


Figure 21 : Northern blot analysis on 5µg total RNA from wild-type (WT) and PNP- plants, using radiolabeled P11 primer (bent arrow on the scheme) to detect transcripts downstream *NCO*. Mature *NCO* extremities are depicted by blue bubbles on the scheme.

The Northern blot experiment using the P11 probe situated downstream of *NCO* detected a smear in PNP- plants, but not in wild-type plants. This result confirms that transcription does not terminate at *NCO* mature 3' end but continues beyond this point. However, PNPase efficiently degrades transcripts downstream of *NCO* as an important amount of RNAs accumulate in PNP- plants, as compared to wild-type plants. In contrast to the 500nt band detected by probes directed against *NCO*, the P11 probe detects a smear. This suggests that transcripts downstream of *NCO* do not accumulate as a single specie but as a population of transcripts of variable length.

Thus, transcripts corresponding to the region downstream of NCO exist, but apparently form a heterogeneous population without defined extremities.

As there is no termination of transcription at the *NCO* locus, the heterogeneity of the downstream transcripts could be due to different 3' ends or to different 5' ends, or both. I wanted to investigate the 5' extremity of transcripts downstream of *NCO* to possibly reveal an endonucleolytic cleavage after *NCO*. To this end, I used two techniques : ribonuclease protection assay (RPA) and 5' RACE (Rapid Amplification of cDNA Ends).

The ribonuclease protection assay (RPA) is presented in Figure 22. Total RNA from either wild-type (wt) or PNP- plants (p-) was hybridized to ^{32}P-[UDP]-labeled RNA probes prior to treatment with RNase A and T. Samples were analyzed by polyacrylamide gel electrophoresis (PAGE) and autoradiography. Probe A spans both the 3' end of *NCO* and the region downstream of *NCO* and should therefore hybridize to both precursor and mature *NCO* transcripts. Probe B corresponds to the sequence downstream of *NCO* and should therefore exclusively hybridize to precursor transcripts or to transcripts downstream of *NCO*. Both probes contain an additional sequence with no homology to the mitochondrial genome to allow discrimination between full-length and protected probes. As a negative control, probes were hybridized to yeast total RNA (y), which should not result in probe protection.

No bands were observed upon hybridization of probes A or probe B to wild-type RNA, indicating that levels of both precursors and mature NCO transcripts are below detection limits in these conditions. Upon hybridization of probe B with PNP- RNA, I observed a single band of 330nt corresponding to the protected full-length probe. This confirms the presence of *NCO* precursor and/or downstream transcripts in PNP- plants. However, no smaller band corresponding to a fragment downstream of *NCO* was observed. Incubation of RNA with probe A resulted in a major band of 500nt, representing the precursor transcript, and one band of 100nt, corresponding to the part of the probe A that overlap with the mature *NCO* transcript.

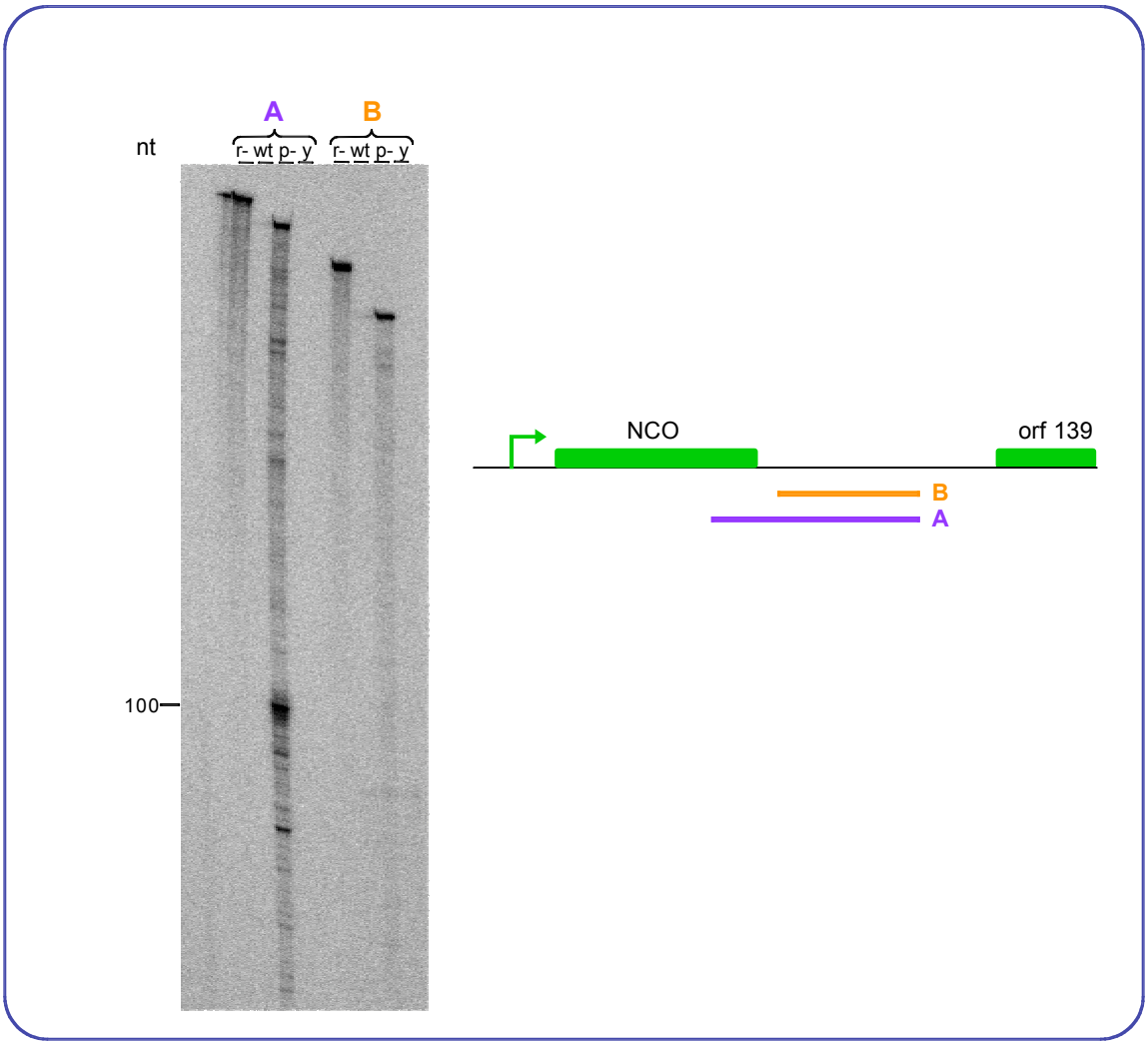


Figure 22 : Ribonuclease protection assay to determine 5' ends of transcripts downstream of the *NCO* transcript. r- : non-treated probe, wt : probe hybridized with wild-type RNA, p- : probe hybridized with RNA from PNP- plants, y : probe hybridized with yeast RNA. A and B probes are positioned on the sequence depicted on the right. They were obtained by cloning represented sequences in TOPO vectors and radiolabeled transcription to have non protected probes with a small portion not annealing to plant RNA, that is degraded during RNase treatment. Products of the ribonuclease assay are separated by acrylamide gel and revealed by exposure with a phosphorimager plate for 1 hour.

RPA detected NCO precursors or parts of NCO but no maturation intermediates that would have been generated by an endonucleolytic cleavage downstream of NCO.

In an attempt to characterize the species downstream of the *NCO* transcript with a more sensitive technique, I performed 5' Rapid Amplification of cDNA Ends (RACE) on the region downstream of the 3' extremity of the *NCO* transcript (Figure 23). Putative processing intermediates could be quickly degraded in wild-type plants. I therefore used PNP- plants to characterize expression of NCO and downstream sequence by 5' RACE using the SMART PCR cDNA synthesis kit (Clontech) and sequence-specific primers (Figure 23A). The two reverse primers upstream of

the 3' *NCO* extremity, P1 and P4, amplified a product corresponding to the *NCO* transcript (the smaller product detected by primer P4 could correspond to a degradation product or maybe is a PCR artifact due to overcycling). In contrast, primers spanning the region downstream of the *NCO* amplified a population of products, visible as a smear. However, a major species was amplified by P8, P9, P10 and P11 primers. These bands may correspond to transcripts that possibly share 5' ends with *NCO* but have not been processed at their 3' ends.

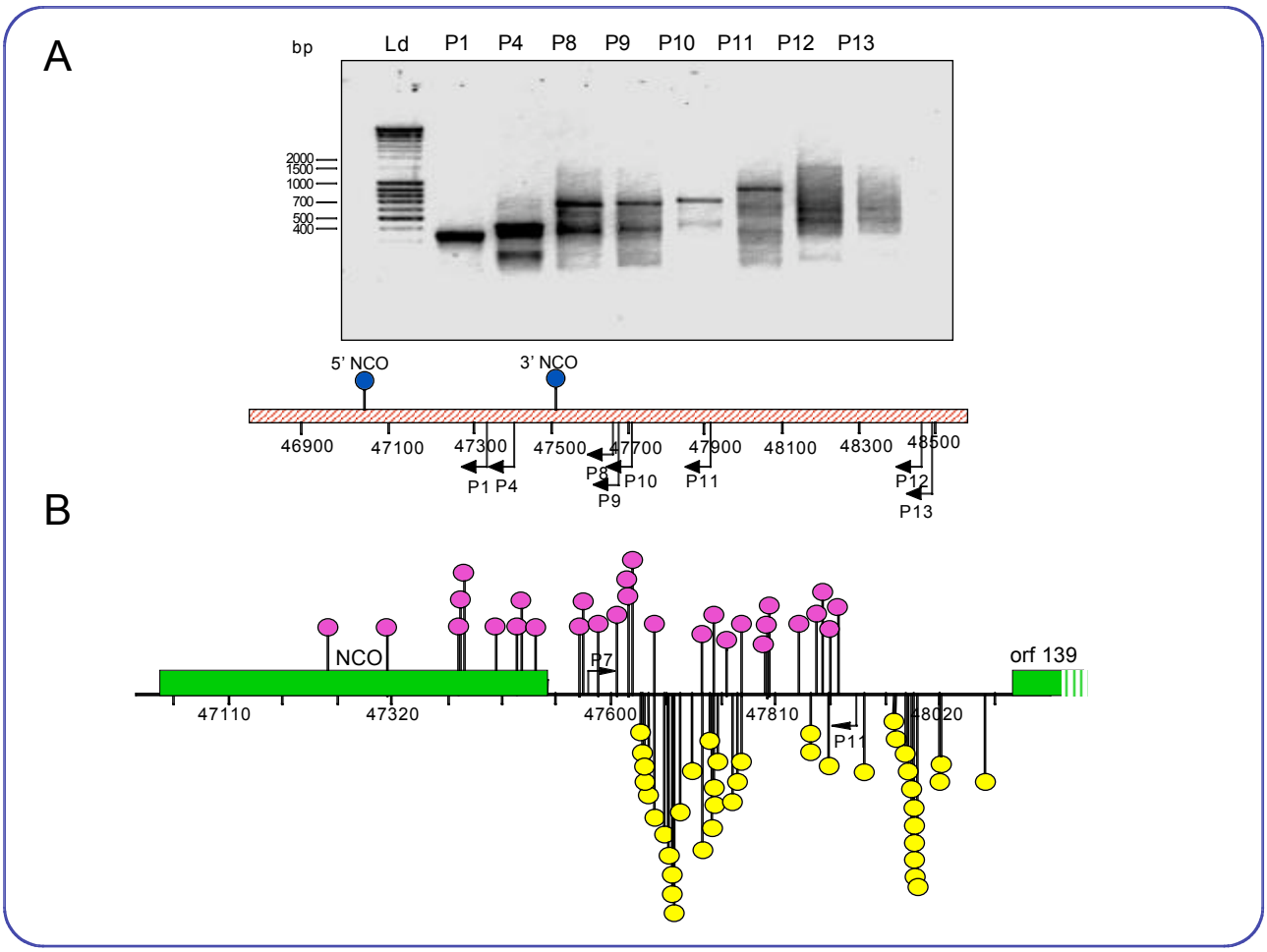


Figure 23 : RACE on cDNA from PNP- plants obtained with the SMART PCR cDNA synthesis kit (Clontech). **A**. Picture of electrophoresis-separated PCR products obtained from SMART cDNA using SMART-adapter primer and various reverse primers, positioned on the sequence represented below the gel picture. **B**. 5' and 3' RACE downstream of the *NCO* transcript on SMART cDNA from PNP- plants. *NCO* transcript and 5' part of orf 139 are depicted by green rectangles. 5' extremities of clones obtained with SMART-adapter primer and P11 primer for the 5' RACE are represented by pink bubbles. 3' extremities of clones obtained with oligo-dT and P7 primers for the 3' RACE are represented by yellow bubbles.

PCR products were cloned into TOPO vectors. Sequences of 29 clones resulting from 5' RACE experiments were analyzed. 5' ends of transcripts downstream to *NCO* were diverse and scattered between position 47240 and 47890 of the genome (Figure 23B, upper part). This result showed that 5' RACE did not reveal any product of an endonucleolytic cleavage. I therefore performed 3' RACE to investi-

gate the possibility of a conserved 3' extremity of transcripts downstream *NCO* (Figure 23B, lower part). The 40 clones resulting from 3' RACE presented 3' ends that did not map to a defined region, but were dispersed between position 47670 and 48080 of the mitochondrial genome.

Thus, I could not detect a maturation by-product that would have been produced by an endonucleolytic cleavage by Northern blot (Figure 21), RNase protection assay (Figure 22) or RACE (Figure 23). The random distribution of both 3' and 5' extremities did not allow to draw conclusions about number, nature or direction of ribonucleolytic activities that could be involved in NCO 3' end processsing.

Unexpectedly, some 5' extremities of the downstream transcripts were located upstream of the 3' end of the *NCO* transcript. This indicates that *NCO* precursor is a substrate for both processing and a minor endonucleolytic pathway, previously observed on tRNA maturation by-products.

Taken together, these results show that RNA species downstream of NCO exist as a heterogeneous population of transcripts that have undefined 3' and 5' ends. As this population exists in the absence of PNPase, they are generated by one or several other activities, but I could not show the existence of an endonucleolytic cleavage involved in 3' processing of NCO. However, there are evidences of a degrading activity acting on primary transcripts, unrelated to processing.

6. 3' end processing depends on the nuclear background.

While I was investigating the maturation of the *NCO* transcript, Matti Leino, in Kristina Glimelius' group at the Swedish University of Agricultural Science, analyzed CMS *Brassica napus* line that has a rearranged mitochondrial genome mostly inherited from *A. thaliana* (Leino et al, 2005). This CMS line was obtained by protoplast fusion between *A. thaliana*, ecotype Landsberg erecta (Ler) and *B. napus*, and successive backcrossing to *B. napus.* Descendants of CMS and cybrid lines were screened for a restored plant (Figure 24A). Interestingly, one of these lines accumulated transcripts of orf 139. This orf is located only 592nt downstream of NCO locus. Matti Leino confirmed that the genomic sequence of cybrid lines in this region is identical to the sequence in *A. thaliana*.

I was therefore interested in investigating NCO transcript maturation and abundance in both CMS and restored lines. Together with Matti Leino, I analyzed transcripts corresponding to the *NCO* region in different genetic backgrounds. To define 3' ends of these transcripts, we mapped polyA sites in *A. thaliana* and cybrid lines (Figure 24B and 24C). cDNA was synthesized using an oligodT-adapter primer, and 3' ends were amplified with reverse adapter primer and either primer

P14 located inside the NCO sequence, or primer P7 located just downstream of the *NCO* 3' end.

Using P14 primer, we obtained a band of 360nt corresponding to mature NCO in *A. thaliana*. No product was amplified by primer P7. This indicates that precursors longer than 500nt do not exist or are very low abundant in wild-type plants, in agreement with previous results. In contrast, both primer pairs amplified products of more than 1200nt (1800nt and 1400nt for P14 and P7, respectively) in cybrid lines, suggesting that primary transcripts are abundant and moreover, that 3' processing of NCO is absent in these lines.

This difference is not due to any variation in the genomic sequence or to differences in transcription as I have already shown that the 3' end of the NCO transcript is not generated by transcription termination (see paragraph 4).

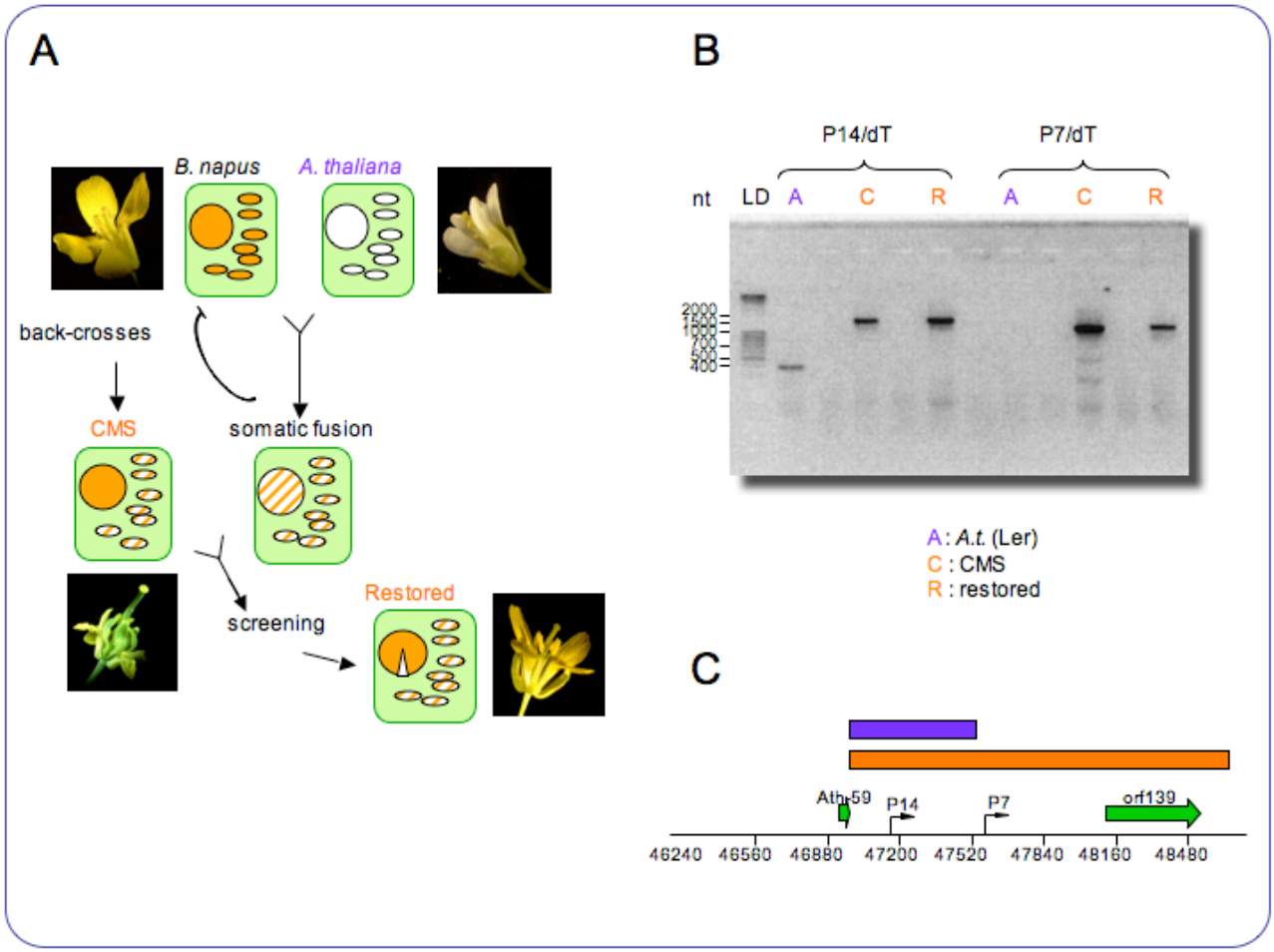


Figure 24 : Maturation of *NCO* transcripts is different between *Arabidopsis* and cybrid lines. **A.** Scheme showing generation of CMS and restored lines were obtained from somatic fusion between a *B. napus* and an *A. thaliana* cell (from Matti Leino, see Leino *et al*, 2003 and 2004). **B.** RT-PCR to map polyA sites of *NCO* transcript in *Arabidopsis* and cybrid lines. 2 µg plant RNA were reverse transcribed using an oligodT-adapter primer, and amplified with this primer and 2 different forward primers which position is shown in C. **C.** Transcripts at the NCO locus have different sizes in different backgrounds. Blue rectangle represent *NCO* transcript in A.t, orange rectangle represents the corresponding transcript in cybrid lines, as concluded from the RT-PCR. Green arrows represent Ath-59 and orf 139.

Hence, these experiments demonstrate that 3' processing of NCO depends on a nuclear factor that is present in A. thaliana but not in B. napus.

There are many examples of nuclear factors affecting processing of organellar transcript, often involving proteins of the PPR family. CMS systems provide good examples, such as the *Brassica napus* nap and pol CMS systems, where the nuclear restorer specifically removes the sequence from the 5' end of the sterility-associated transcript (Messana *et al*, 1999). Another example is the chloroplastic *clpP* preRNA which processing requires a PPR protein in *Physcomitrella patens* (Hattori *et al*, 2007).

7. Can we identify nuclear factor(s) responsible for the NCO 3' extremity processing?

I then used a genetic approach in the aim of identifying a nuclear factor(s) affects processing of the *NCO* transcript.

Molecular analysis of wild-type and PNP- plants did not allow to identify further factors that could be involved in 3' maturation of the *NCO* transcript. Therefore, I made use of natural *Arabidopsis* variants of the world wide collection of accession lines, displaying a high genetic diversity, that is available at INRA in Versailles (http://dbsgap.versailles.inra.fr/vnat/). The establishment of nested core collections allowed to maximize the genetic diversity, so that a selection of 48 accessions covers already most of the sequence variation present in the 265 accessions available (McKhann *et al*, 2004).

I analyzed 98 different ecotypes, including the 48 lines of the core collection, for polymorphisms in 3' end processing of the *NCO* transcript. RNA was extracted from 3-4 weeks old plantlets and 5µg were reverse transcribed with the oligo-dT primer. PCR, using the oligo-dT adapter primer and a forward primer in the NCO sequence or after the NCO sequence, were performed to search for an eventual polymorphism in the *NCO* length.

I did not observe any polymorphism concerning the 3' end of NCO.

This indicates that the nuclear factor(s) involved in NCO 3' maturation is expressed in all ecotypes tested and therefore cannot be identified by this genetic approach.

8. Conclusions.

I have identified a novel transcript generated from the repeated region I of the *Arabidopsis* mitochondrial genome. The genomic sequence coding for this *NCO* transcript is composed of recombined pieces of DNA. Thus, this transcript is likely not functional although this is difficult to prove experimentally.

Northern blot analysis and cRT experiments showed that *NCO* has defined ends both in wild-type background and in absence of PNPase. This indicates not only that *NCO* is processed at its 5' and 3' extremities but also that PNPase is not responsible for this maturation.

5' end processing is probably mediated by the cleavage of the upstream tRNA-like element. This hypothesis has to be further confirmed by *in vitro* and/or *in organello* processing assays (see model on Figure 25).

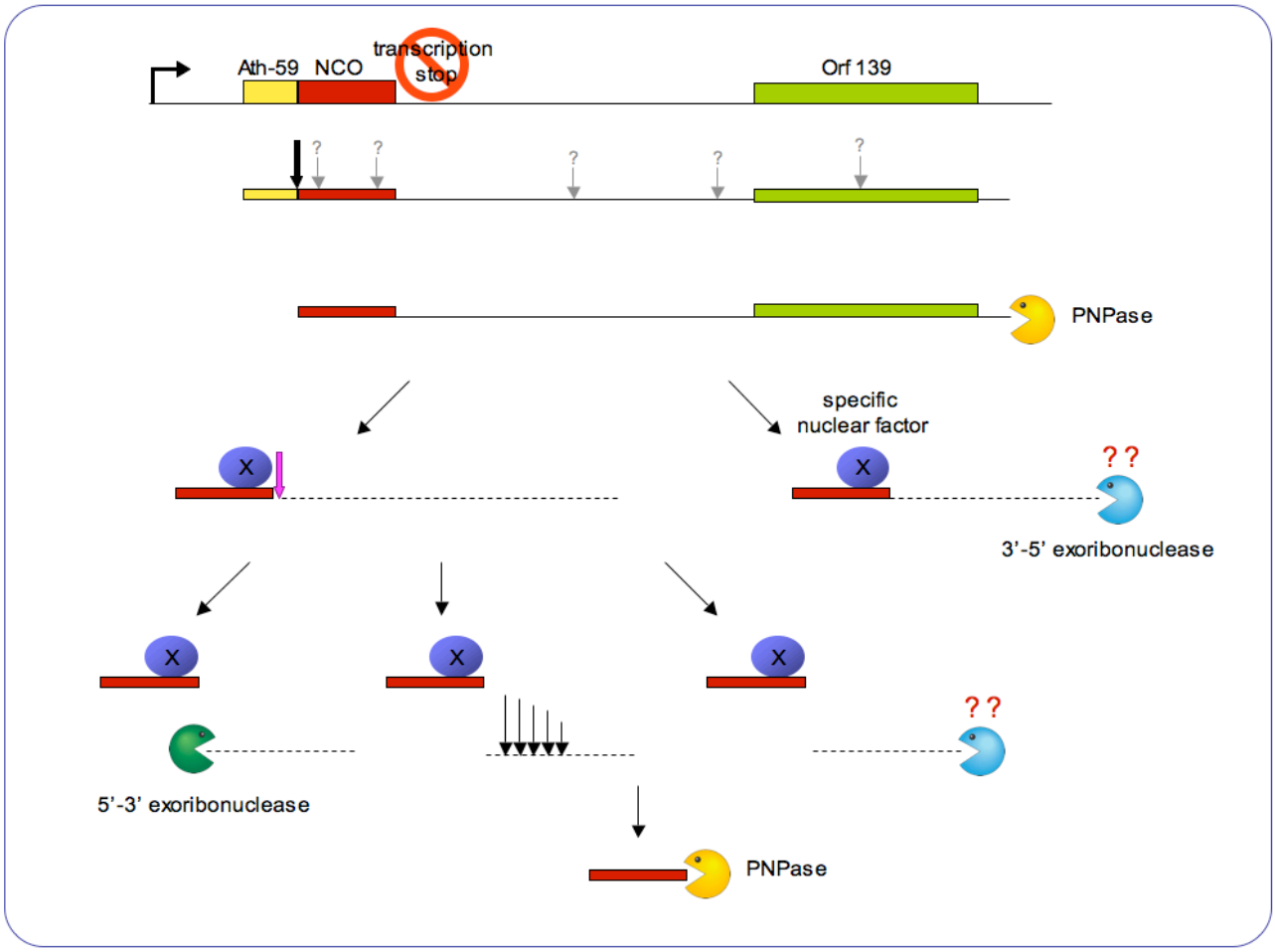


Figure 25 : Model of *NCO* transcription, processing and degradation. There is no transcription termination at the mature 3' end of *NCO,* and long primary transcripts are present. They are sensitive to random degradation activity, likely endonucleolytic (gray arrows with question mark). 5' processing (thick black arrow) is probably due to RNase Z cleavage of Ath-59. 3' end processing can be due to endonucleolytic cleavage (thick pink arrow) or to unidentified 3' to 5' exoribonuclease (blue pacman). In the case of an endonucleolytic cleavage, downstream product is rapidly degraded, either by a 5' to 3' exoribonucleolytic activity (green pacman), or by a wave of endonucleolytic cleavages, or by a 3' to 5' exoribonuclease. PNPase degrades mature and primary *NCO* transcripts.

Using run-on experiments, I could exclude that 3' ends of *NCO* are generated by transcription termination but rather are processed from a longer precursor. A long transcript of 1,8kb is not detectable in wild-type *Arabidopsis* but in *B. napus-A. thaliana* cybrid lines with identical mitochondrial genomic sequence, indicating that 3' end maturation of *NCO* involves a specificity factor encoded in the nucleus (Figure 25).

Northern analysis, RT-PCR and RACE experiments, as well as ribonuclease protection assays revealed that transcripts downstream of *NCO* form an heterogeneous population. Further, they are polyadenylated and accumulate in absence of PNPase, indicating that PNPase is responsible for their efficient decay.

The results of ribonuclease protection assay and 5' and 3' RACE experiments did not support the hypothesis that an endonucleolytic cleavage is involved in the generation of *NCO* 3' extremities. The different options to explain *NCO* 3' maturation are recapitulated in the model of the Figure 25. If there is an endonucleolytic cleavage, downstream product is rapidly degraded by other enzymes than PNPase. 3' products of the initial endonucleolytic cleavage could be degraded by a following wave of endonucleolytic cleavages as suggested for *atpB* RNA in Chlamydomonas chloroplast (Higgs *et al*, 1999 ; Hicks *et al*, 2002). Alternatively, degradation of the distal product of an endocleavage could be due to a 3' to 5' exoribonuclease or to a 5' to 3' exoribonuclease. This latter possibility is reinforced by the recent discovery of a bacterial 5' to 3' exoribonucleolytic activity, raising the possibility of an homolog acting in mitochondria (Mathy *et al*, 2007). In Chlamydomonas chloroplasts, transcripts are stabilized by factors in 5', indicating that 5' to 3' exoribonucleolytic activity exists (Raynaud *et al*, 2007).

Another possibility for *NCO* 3' maturation would be a 3' to 5' exonucleolytic trimming. However, PNPase and RNase II are the only 3' to 5' exoribonucleases identified so far in plant mitochondria. As cRT-PCR showed that they are not responsible for *NCO* 3' end maturation, processing could be the fact of another 3' to 5' exoribonucleolytic activity yet to be identified. However, this possibility is unlikely as this activity would be specific : it does not compensate for the lack of PNPase for all other RNA tested.

Unfortunately, neither genetic nor biochemical approaches led to identification of any of the factors involved. However, my results demonstrate that PNPase is responsible for degradation of both *NCO* and downstream transcripts.

More importantly, the example of *NCO* transcript shows for the first time that transcript 3' extremity maturation and stabilization can be uncoupled processes in plant mitochondria. This will be further developed in the Discussion & Perspectives chapter.

Chapter III
A novel degradation pathway in plant nucleus involving polyadenylation

RNA degradation mechanisms involving polyadenylation are not specific to mitochondria. Polyadenylation-dependent degradation was first discovered in bacteria and later in chloroplast, but also in archaea (Portnoy *et al*, 2006) suggesting that polyadenylation-mediated RNA degradation mechanisms probably derive from a common prokaryotic ancestor.

In contrast to the destabilizing role of poly(A) tails for bacterial, chloroplastic and mitochondrial transcripts, polyadenylation of nuclear encoded mRNAs is required for their export, stability and translatability. However, is has been shown in yeast that nuclear degradation of intergenic and non-coding RNAs is dependent on polyadenylation by a newly described complex called TRAMP (for Tfr4p/Air2p/Mtr4p Polyadenylation complex) (LaCava *et al*, 2005, Wyers *et al*, 2005, Vanacova *et al*, 2005). The degradative machinery is the nuclear exosome, comprising an hexameric ring of subunits homologous to *E. coli* RNase PH, associated with a 3' to 5' RNase II-like exoribonuclease and a specific 3' to 5' exoribonuclease of the RNase D type, named Rrp6p. The role of Rrp6p in degradation of polyadenylated rRNAs and snoRNAs/snRNAs was previously suggested in yeast nuclei (VanHoof *et al*, 2000, Kuai *et al*, 2004). In human cells, polyadenylated transcripts of ribosomal RNA, likely being degradation intermediates (Slomovic *et al*, 2006), and β-globin pre-mRNAs containing short A tails (West *et al*, 2006) have been reported. In addition, the low-abundant spacer 5S RNA transcript has been recently observed in a polyadenylated form in *Nicotiana* species (Fulnecek and Kovarik, 2007).

Interestingly, the small polyadenylated clones library we elaborated to investigate mitochondrial roles of PNPase contained some polyadenylated species corresponding to non-coding RNAs encoded in the nucleus.

Studies showed that eukaryotic polyadenylation could be involved in RNA degradation mechanisms. The presence of polyadenylated nuclear non-coding RNA in our library led us to address the possibility that RNA polyadenylation could trigger RNA degradation in plant nucleus.

1. Polyadenylated small RNAs are present in the plant nucleus.

Some of the polyadenylated RNAs that we cloned from total RNA isolated from PNP-plants did not correspond to mitochondrial DNA but mapped to sequences in the nuclear genome. 77 clones on 433, 18% of the library, originated from the nucleus. Most of them matched to sequences in messenger RNAs, likely representing fragments of a 5' to 3' mRNA degradation pathway, or products due to mechanical fragmentation. However, we also found polyadenylated sequences matching to non-coding RNAs or intergenic regions of the nuclear genome. 8 clones corresponded to ribosomal RNA or rRNA intergenic transcripts, 5 corresponded to small nucleolar RNAs (snoRNAs) and 1 to a previously characterized

small non-coding RNA, Ath-139 (Marker *et al*, 2002). The figure 26 presents these nuclear clones and shows the position of the clones found at the *rrn* genes locus.

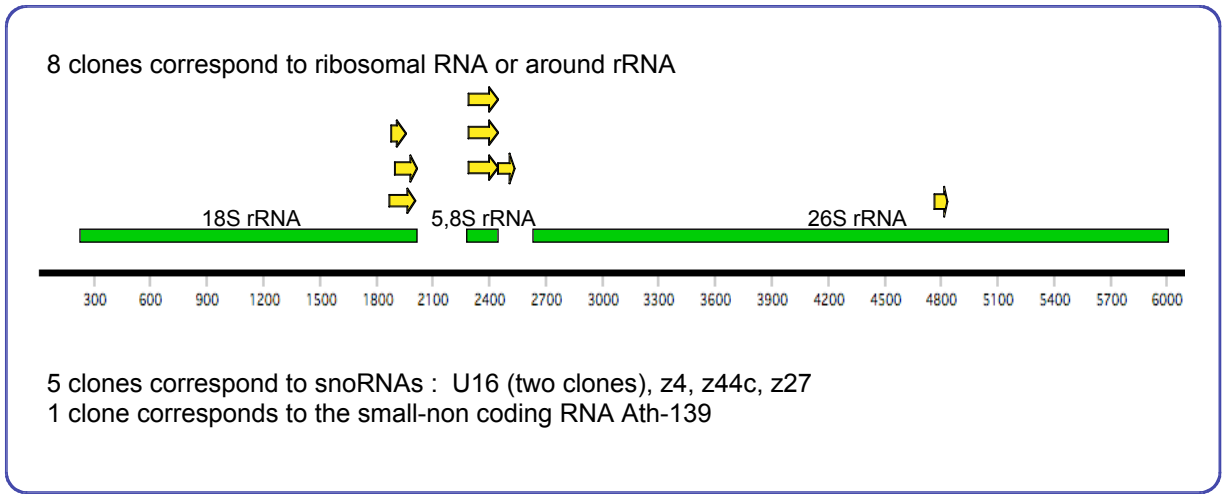


Figure 26 : Nuclear items of the small polyadenylated clones library. The scheme shows positions of the clones (yellow arrows) at the ribosomal genes (green rectangles) locus.

The existence of these clones corresponding to nuclear non-coding transcripts was the first hint that non-messenger nuclear transcripts can be polyadenylated in plants.

To analyze involvement of polyadenylation in nuclear non-coding RNAs degradation, expression and polyadenylation status of several snoRNAs and snRNAs were investigated. Heike Lange, post-doctoral fellow in the laboratory, performed oligo-dT primed RT-PCR on selected examples presented in figure 27.

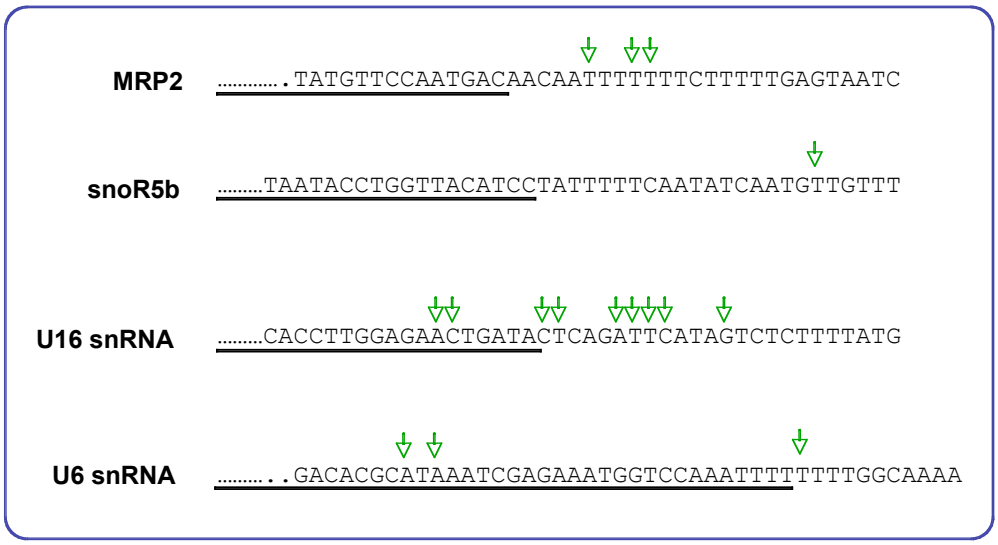


Figure 27 : Mapping of polyadenylation sites 3' regions of non-coding RNAs by RT-PCR. Total or size selected RNA isolated from *A.thaliana* was used for reverse transcription with an oligo-dT-adapter primer. PCR was performed with a forward primer specific for the indicated gene and a reverse primer binding the adapter sequence before cloning and sequencing. The underlined sequence represents the mature 3' end of the RNAs as deduced from genome annotation and the literature. Green arrows above the sequence represent polyadenylation sites.

This experiment confirmed that nuclear non-coding RNAs may be polyadenylated in *Arabidopsis thaliana.* Sites of polyadenylation were found into the RNA, at the 3' mature end and mostly downstream of this end. This indicates that transcripts precursors, as well as mature transcripts can be polyadenylated. Polyadenylation could apply to precursors that require processing or that are detected as aberrant by a quality-control system to target them for degradation. It could also apply to mature non-coding RNAs for their turnover if they are in excess.

These results raise the possibility that a degradation and/or processing pathway involving polyadenylation could act in plant nuclei.

2. A small family of yeast Rrp6p homologues in Arabidopsis.

During our analysis of polyadenylated nuclear non-coding RNAs, studies in yeast reported a nuclear polyadenylating complex named TRAMP, that is required to activate the RNA degradation activity of the nuclear exosome. It was demonstrated that a functional nuclear exosome, in complex with the associated Rrp6p protein, is indispensable for rRNA and snoRNA degradation, degradation of hypomodified tRNA and of intergenic transcripts (Kadaba *et al*, 2004, LaCava *et al*, 2005, Wyers *et al*, 2005, Vanacova *et al*, 2005). The involvement of Rrp6p in the degradation of polyadenylated RNAs was demonstrated by the accumulation of polyadenylated RNAs in yeast strains lacking Rrp6p (Briggs *et al,* 1998, Kuai *et al*, 2004). The cytoplasmic form of the exosome is not associated with Rrp6p in yeast.

As we were interested in nucleus-specific RNA degradation involving polyadenylation, we searched for homologs of Rrp6p in plants. A TAIR-BLAST analysis using the yeast and human Rrp6p protein sequence as queries detected three *Arabidopsis* sequences with good expectation values, namely At1g5440 (e-value of $2e^{-51}$), At5g35910 (e-value of $1e^{-53}$) and At2g32415 (e-value of $4e^{-50}$) that we named RRP6-like 1 (RRP6L1), RRP6-like 2 (RRP6L2) and RRP6-like 3 (RRP6L3), respectively. As full-length or fully spliced cDNAs were not available at the time of this work, we cloned and characterized cDNA of each gene by RT-PCR, using total RNA from wild-type plants as template. We also checked mRNAs extremities by 5' and 3' RACE experiments. The figure 28 shows the exon-intron structure of the 3 RRP6-like genes of *Arabidopsis thaliana*.

Public microarray data indicate that the three genes are expressed in all plant tissues, with a highest expression level in root tips and stem (www.genevestigator.ethz.ch). The general level of expression of the three RRP6-like mRNAs seems to be quite low.

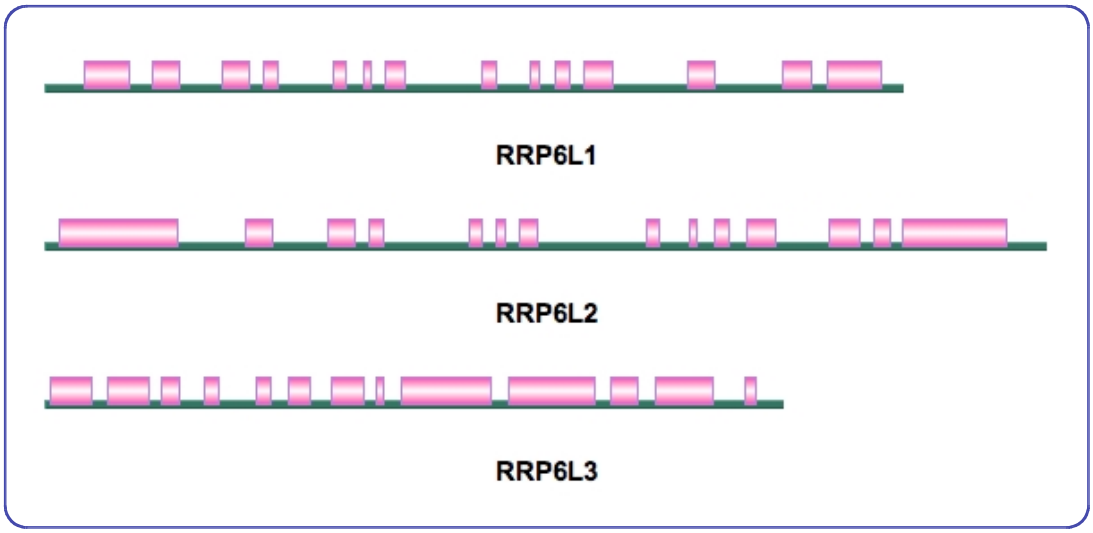


Figure 28 : Exon-intron structure of RRL6-like genes. Exons are depicted as pink boxes and introns as well as 5' and 3' UTRs as green line.

RRP6L1, RRP6L2 and RRP6L3 cDNAs are predicted to encode proteins of 637, 870 and 892 amino acids, respectively. RRP6L1 is significantly shorter than the others. As a result, RRP6L1 mRNA has a long 5' UTR (236nt) that is very similar to the 5' part of the RRP6L2 ORF sequence but is interrupted by a STOP codon. To determine whether this atypical mRNA can be translated *in vivo*, we used binary vectors containing full 5' UTR plus ORF sequence to express epitope-tagged versions of both RRP6L1 and RRP6L2 in *Nicotiana benthamiana* leaves. Western analysis with monoclonal antibodies directed against the epitope-tag revealed that both RRP6L1 and RRP6L2 are efficiently translated and that RRP6L1 is indeed shorter that RRP6L2 (Figure 29). Proteins had apparent molecular weights of 100 and 130 kDa respectively, which is larger than the calculated weight of the tagged proteins. However, we also observed a higher apparent mass for endogenous AtRRP6L2 using specific antibodies, suggesting that this unusual migration reflects an intrinsic property of the proteins (see figure 4E of the appendix 6).

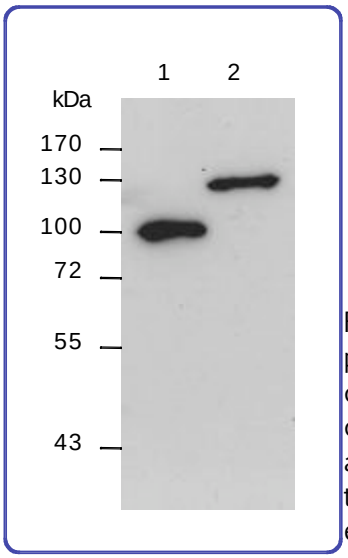
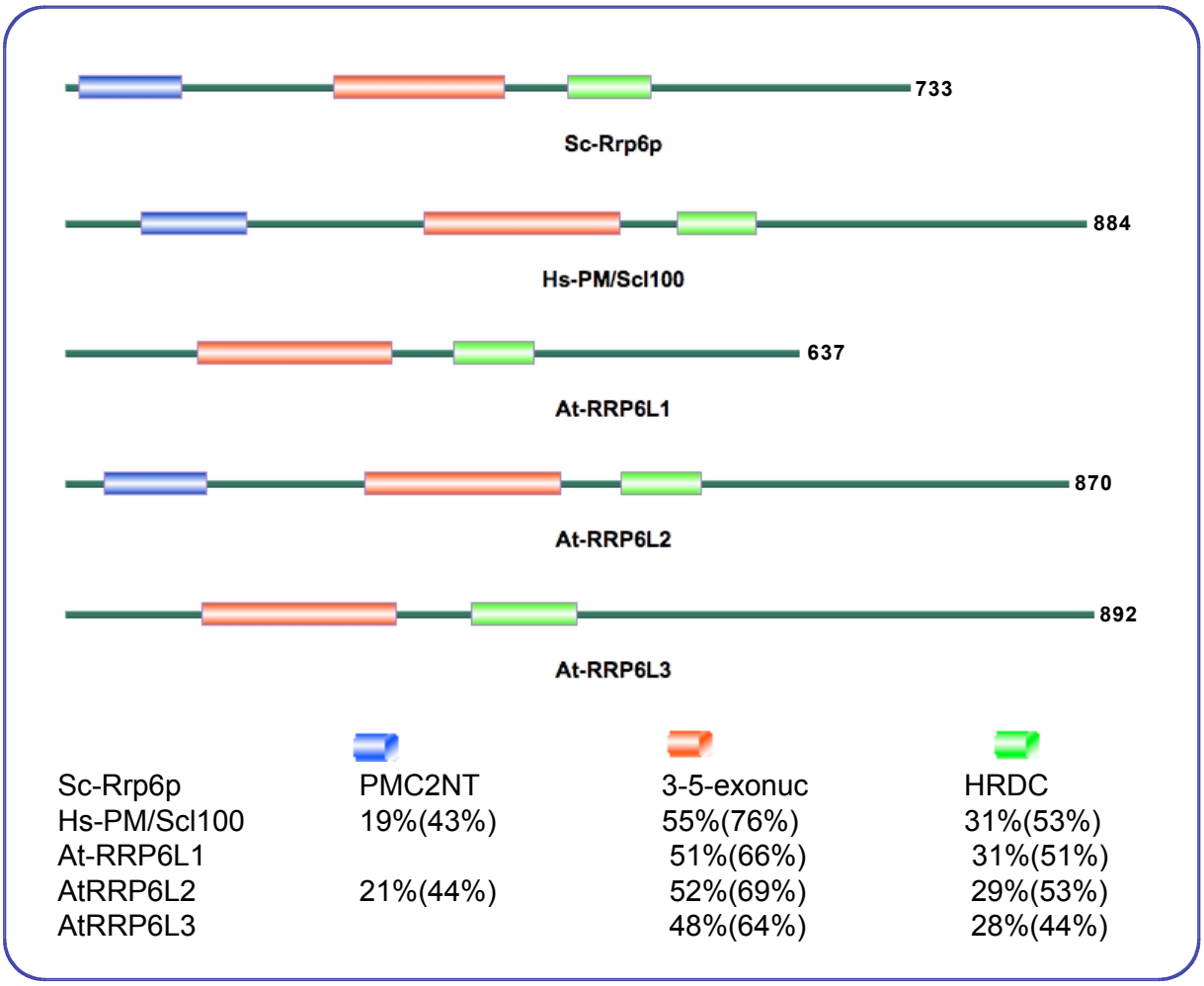


Figure 29 : Western blot analysis of RRP6L1 (1) and RRP6L2 (2) expression. 20µg proteins extracted from Nicotiana benthamiana overexpressing etag-CBP-6His tagged RRP6L1 or RRP6L2 are loaded on each lane. Proteins are detected with 1/2500e anti-etag primary antibody from mouse and 1/5000e goat anti mouse secondary antibody. Signals are revealed by the ECL method and the film was exposed for 1 minute.

The alignment of Rrp6 proteins from various organisms is presented in appendix 3. It highlights conserved regions. All proteins harbor two functional domains common to RNase D homologs : the 3' to 5' exonuclease (3-5-exonuc) (IPR002562) domain and the helicase and RNAse D C-terminal (HRDC) (IPR002121) motifs. In *Arabidopsis*, only the N-terminus of RRP6L2 shows weak homology to the PMC2NT domain present in the N-terminus of yeast Rrp6p. Figure 30 presents functional domains of yeast, human and *Arabidopsis* Rrp6 proteins. Percentages of similarity and identity to Sc-Rrp6p in these domains are indicated. Percentages are high for 3' to 5' exonuclease and HRDC domains, indicating a good conservation of the functional domains between related proteins.



	PMC2NT	3-5-exonuc	HRDC
Sc-Rrp6p			
Hs-PM/Scl100	19%(43%)	55%(76%)	31%(53%)
At-RRP6L1		51%(66%)	31%(51%)
AtRRP6L2	21%(44%)	52%(69%)	29%(53%)
AtRRP6L3		48%(64%)	28%(44%)

Figure 30 : Functional domains of Rrp6 proteins. Domains are depicted as colored boxes on the protein sequence represented as a green line. For each of these domains, percentages of similarity and, in brackets, of identity with Sc-Rrp6p are given. Number of amino acids of each proteins are showed. Sc : *Saccharomyces cerevisiae*, Hs : *Homo sapiens*, At : *Arabidopsis thaliana*.

The fact that the 5' UTR of RRP6L1 mRNA is similar to the region coding for the N-terminal part of RRP6L2 and the overall strong sequence homology between both proteins suggest a recent gene duplication event. To test this hypothesis, we searched for RRP6-like sequences in other plant genomes. Interestingly, both rice

and poplar contain also three genes encoding RRP6-like proteins. Phylogenetic analysis revealed that sequences form two groups represented by RRP6L1/2 on one hand and RRP6L3 on the other hand, the latter being absent from yeast, animal or human genomes (Figure 31). The presence of RRP6L3-like genes in rice, poplar and *Arabidopsis* indicates an early divergence in plants from a common ancestor.

The RRP6L1/2 group clusters with yeast and animal RRP6-like proteins. Despite the fact that this group consist of two members in plants, these two members cluster as a pair in each of the three species. It is therefore tempting to propose that RRP6L1/2 type genes have been duplicated independently in these three plant species rather than resulting from a gene duplication anterior to the speciation of rice, poplar and Arabidopsis.
Both rice and Arabidopsis have a short and a long version of the RRP6L protein of the RRP6L1/2 cluster.

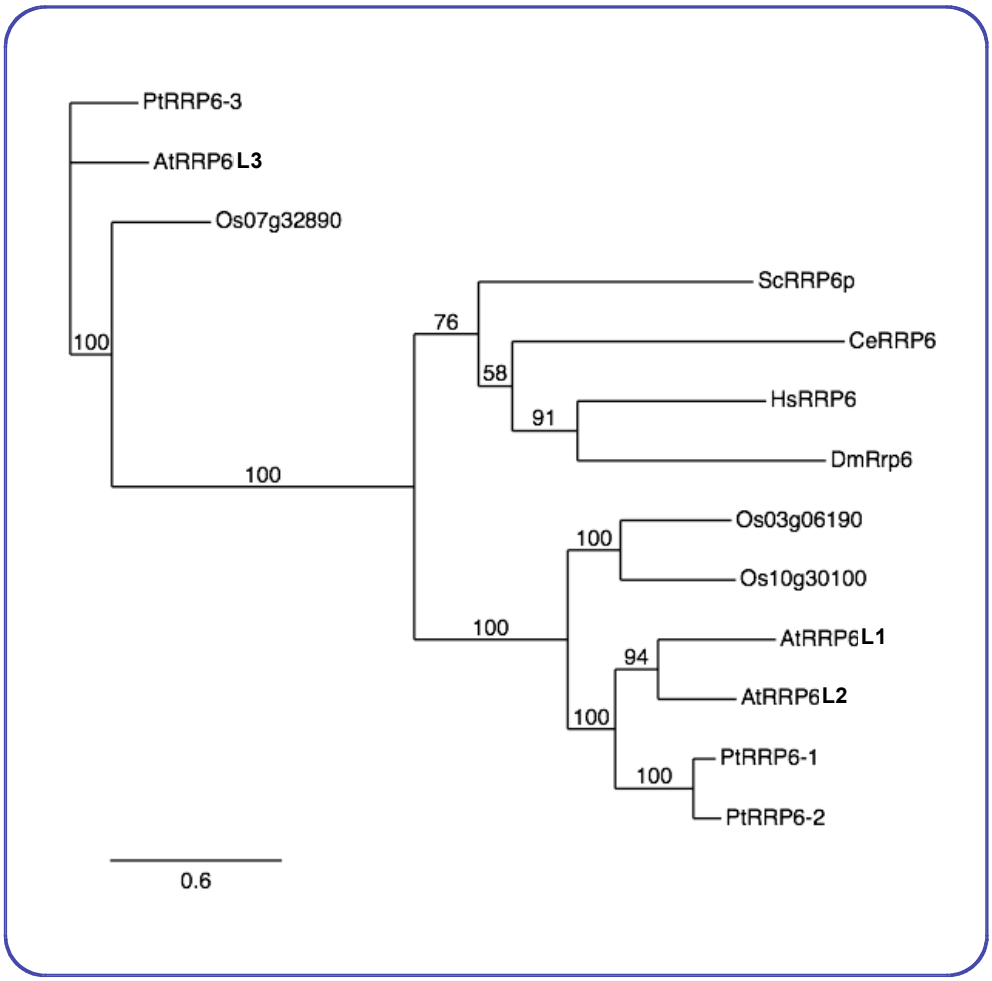


Figure 31 : Phylogenetical unrooted maximum likelyhood tree of RRP6 proteins. Bootstrap values above 70 out of 100 replicates, indicating a strong clade, are indicated along branches. The scale bar indicates the evolution distance (amino acid subtitutions per site). *Pt : Populus trichocarpa, At : Arabidopsis thaliana, Os : Oryza sativa, Sc : Saccharomyces cerevisiae, Ce : Caenorhabditis elegans, Hs : homo sapiens, Dm : Drosophila melanogaster.*

Thus, we have identified a family of RRP6-like proteins in Arabidopsis. This family is composed of 3 members : RRP6L1, RRP6L2 and RRP6L3. Three genes also encode RRP6-like proteins in rice and poplar. Two of these genes, represented by AtRRP6L1 and AtRRP6L2, are the closest homologues to yeast and human RRP6 whereas the AtRRP6L3 type seems to be specific to plants.

3. Accumulation of a ribosomal precursor in rrp6l2 mutant plants but not in rrp6l1 mutants.

To investigate the functions of RRP6L proteins in *Arabidopsis*, Heike Lange identified mutant plants obtained by T-DNA insertion, both in RRP6L1 and RRP6L2. Two independent mutant lines were obtained in each case. T-DNA was inserted in the first exon and in the sixth exon for *rrp6l1-2* and *rrp6l1-1* mutants, respectively. *rrp6l2* mutants harbored T-DNA insertion in the first exon : at the beginning of the exon for *rrp6l2-2* and at the end of the exon for *rrp6l2-1* (see Figure 32A).

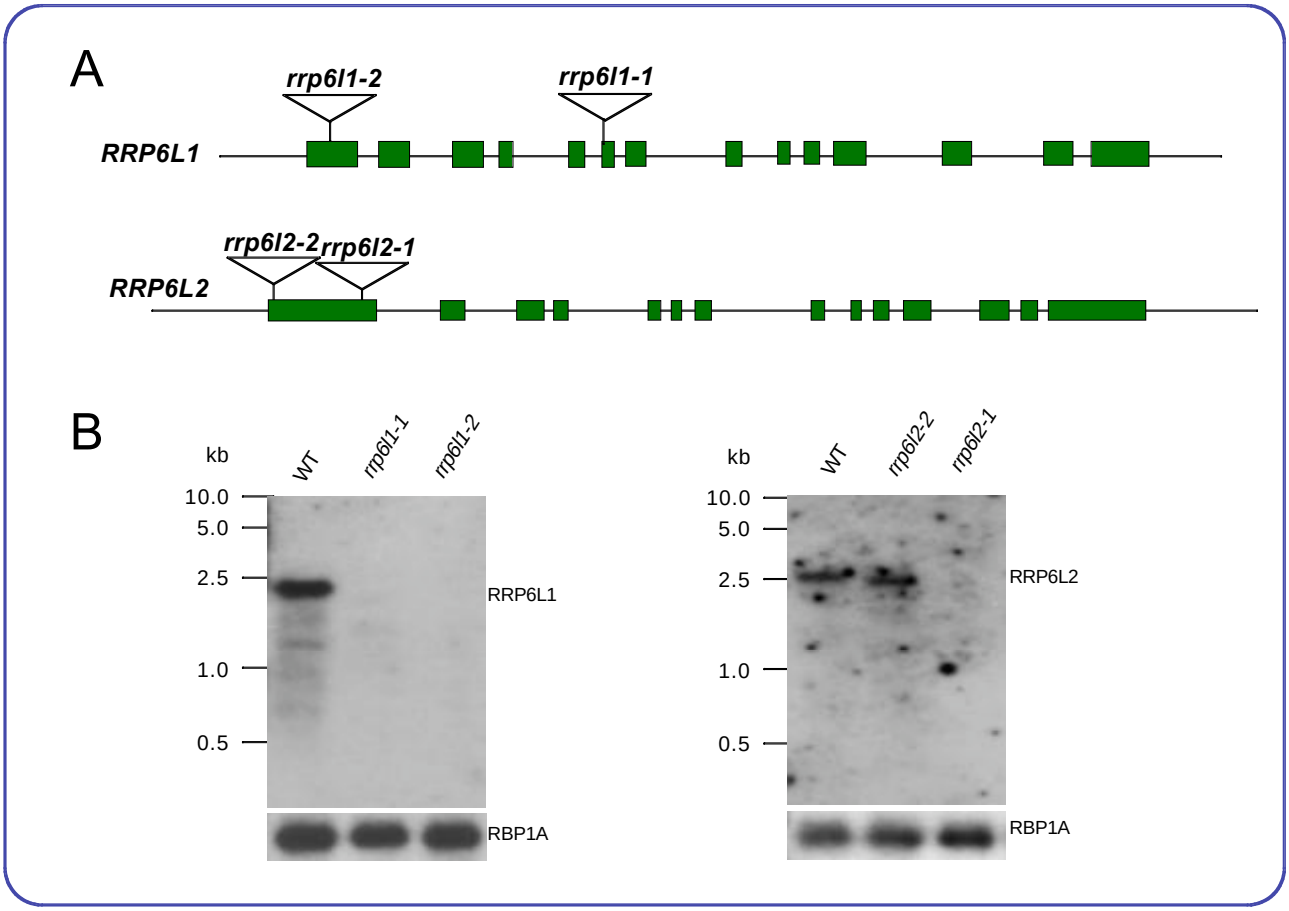


Figure 32 : Characterization of *rrp6l1* and *rrp6l2* mutant plants. **A**. Position of the T-DNA insertions in *RRP6L1* and *RRP6L2* genes. Names of the insertion lines are given above insertions. Exons are represented by green boxes. **B**. Virtual Northern blots showing expression of *RRP6L1* mRNA (left) and *RRP6L2* mRNA (right) in wild type and mutant plants. A probe recognising RBP1A mRNA was used as a loading control.

The level of expression of *RRP6L1* mRNA and *RRP6L2* mRNA in mutants lines and in wild-type plants was investigated by virtual Northern blot. We could not observe the 2,4 kb band corresponding to the *RRP6L1* mRNA in *rrp6l1-1* and *rrp6l1-2* mutant plants, indicating that T-DNA insertion indeed hampers *RRP6L1* expression. *RRP6L2* mRNA could be detected in wild-type plants but also in the *rrp6l2-2* line, with a reduced size. The mRNA was not detectable in the *rrp6l2-1* mutant. Western blot using antibodies raised against RRP6L2 showed that a truncated, potentially unfunctional protein is weakly expressed in *rrp6l2-2* and that the expression of the protein is bellow detection limit in *rrp6l2-1* (see figure 4E of the appendix 6).

All four mutant lines were undistinguishable from wild-type plants in terms of growth and development, under standard growth conditions. Homozygous *rrp6l1-2* plants were crossed with homozygous *rrp6l2-1* and *rrp6l2-2* plants but none of the homozygote *rrp6l1-2 rrp6l2-1* and *rrp6l1-2 rrp6l2-2* displayed any obvious growth or developmental phenotype.

Yeast *rrp6Δ* mutant strains present an accumulation of the ribosomal RNA maturation by-product 5'ETS (Allmang *et al*, 1999b). In *Arabidopsis* as in yeast, ribosomal 18S, 5.8S and 25S rDNA genes are cotranscribed as a single precursor containing internal (ITS1 and ITS2) and external transcribed spacers (5'- and 3'-ETS). The spacers are then removed by a series of endonucleolytic and exonucleolytic cleavage steps.

In yeast, the 5'ETS is cleaved off at A0 cleavage site taking place 90nt upstream of the 18S rRNA (see Figure 33). This site has not been characterized in *Arabidopsis* to date. However, a P cleavage site, absent from yeast, was characterized 560nt upstream of the 18S rRNA in Arabidopsis (Saez-Vasquez *et al*, 2004).

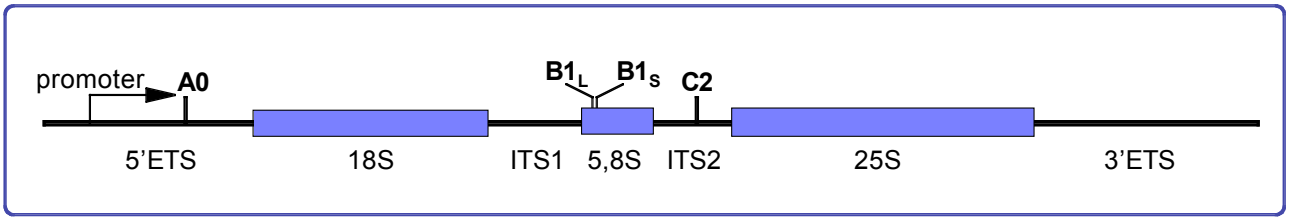


Figure 33 : Yeast ribosomal rDNA locus. rDNA genes are represented by blue boxes, cleavage sites are indicated in bold letters.

In a first approach, *rrp6l1* and *rrp6l2* mutants were tested for accumulation of the 5'ETS by circular RT-PCR experiments (figure 34A). Only cDNA from *rrp6l2* plants permitted amplification of the expected 5'ETS. Sequence analysis revealed that all 5' extremities mapped exactly to the P site. All of the 3' ends mapped about 80nt upstream of the 18S rRNA, suggesting the presence of an A0 site. Interestingly, 50% of the clones carried poly(A) tails, with an average length of 9-10 adenosines. As no oligo(dT) primer was used in the cRT-PCR experiments, these

results unambiguously show that this rRNA maturation by-product can be poly-adenylated.

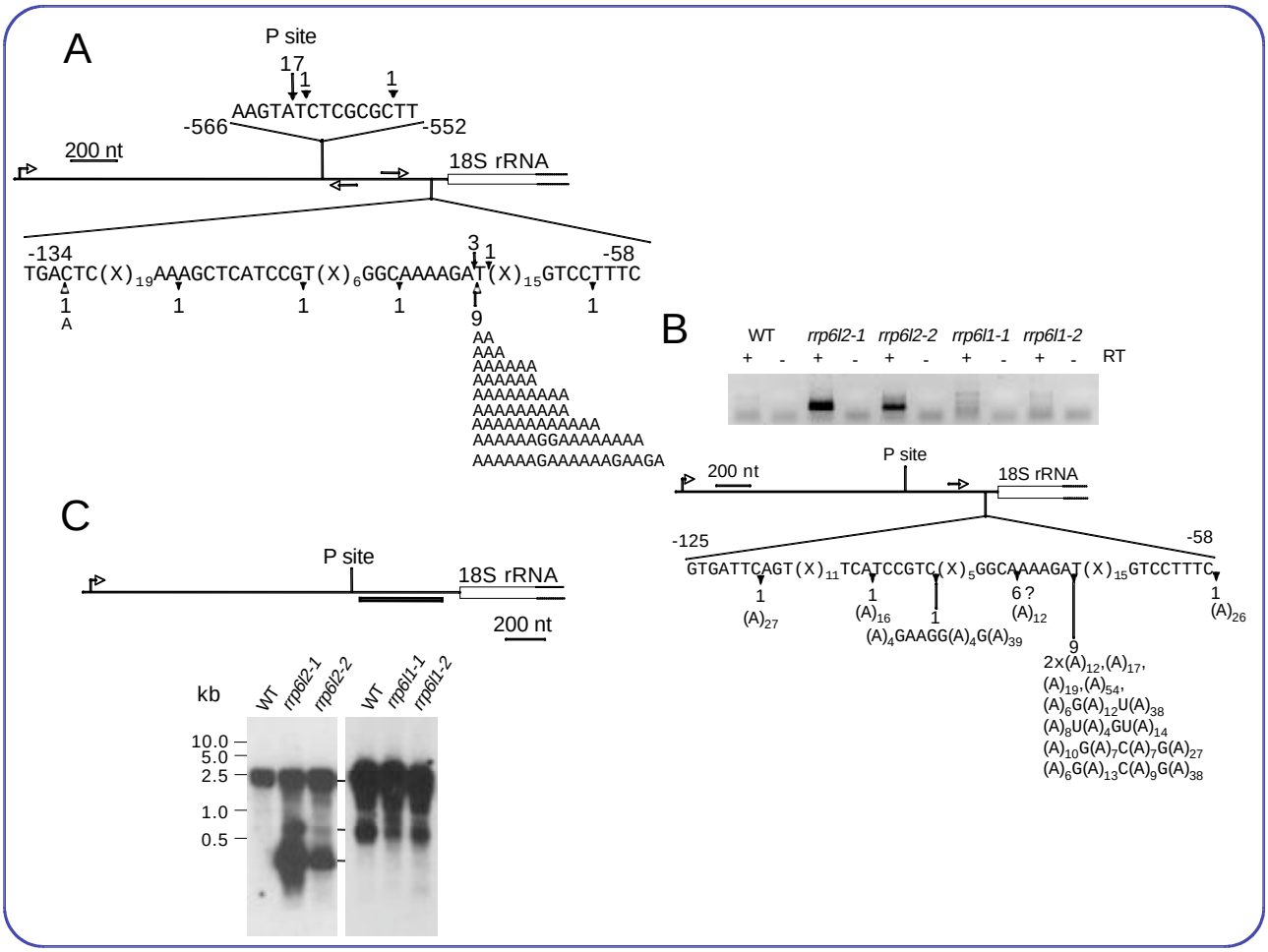


Figure 34 (extracted from submitted publication in appendix 6) : a polyadenylated rRNA maturation by-product accumulates in rrp6l2 mutants. **A.** Mapping of 5' and 3' ends of the excised 5' ETS accumulating in *rrp6l2* mutants by cRT-PCR. Total RNA from *rrp6l2* plants was self-ligated by T4 RNA ligase prior to cDNA synthesis with a gene specific reverse primer. In the following PCR reaction, the same primer was combined with a gene-specific forward primer to amplify joined 5' and 3' ends. Locations of primers are indicated in the scheme. 5' ends are indicated on the sequence above, 3' ends are indicated on the sequence below the scheme, and non-encoded nucleotides found at 3' ends of ETS transcripts are detailed. P-site, primary processing site of the rRNA precursor transcript (Saez-Velasquez, 2004). **B.** Detection of polyadenylated transcripts corresponding to the 5' ETS by 3' RACE. Total RNA from wild-type or mutant plants was used to synthesize cDNA with an oligo(dT)$_{12}$ primer. 3' ends were then amplified by PCR with a gene-specific forward primer (represented by an arrow above the scheme) and a reverse primer recognising the adapter sequence of the oligo(dT)$_{12}$ primer. PCR products were separated on 1,5 % agarose gel (top). PCR products obtained from *rrp6l2* samples were cloned and sequenced to map polyadenylation sites (bottom). Location and frequency of polyadenylation sites are indicated on the sequence, and sizes and nucleotide composition of poly(A) tails are detailed below. RT, reverse transcriptase **C.** Virtual Northern blots with samples from wild-type and *rrp6l2* (left) or wildtype and *rrp6l1* mutants (right) were probed for the 5' ETS, i.e. the RNA segment located between the primary processing site (P-site) and the 5' end of the 18S as indicated in the scheme of the 5' region of the primary rRNA transcript. The promoter is represented by a bent arrow, the probe is represented by a black bar.

To further analyze the polyadenylation status of the 5' ETS, 3' RACE experiment was performed on oligo(dT)-primed cDNA from wild type and *rrp6l2* plants (Figure 34B). Polyadenylated species were clearly detected in *rrp6l2-1* and *rrp6l2-2* while only a faint smear was observed for wild type and *rrp6l1* samples. Position and frequency of polyadenylation sites of the analyzed sequences for *rrp6l2-1* are presented in figure 34B. The polyadenylation sites are true polyadenylation sites since most of the clones harbor poly(A) tails larger than the oligo$(dT)_{12}$ used for the cDNA synthesis, and some sites were in regions devoid of adenosine. Most of the the sites were located at the 3' site determined by cRT-PCR. Several non-A nucleotides were detected indicating that these tails are slightly heteropolymeric.

Hence, we demonstrated by different approaches that the 5'ETS can by polyadenylated.

As 5' ETS was detected only in *rrp6l2* plants by cRT-PCR, accumulation of the 5' ETS was investigated by virtual Northern blot (Northern experiment relying on PCR amplification of cDNA reverse transcribed from polyadenylated RNA, before detection by radiolabeled probe) in wild type and mutant plants (figure 34C). Several RNA species were detected. The larger transcripts were present in all samples and could be rRNA maturation intermediates. An RNA of about 450nt was not detected in wild type plants but strongly accumulated in both *rrp6l2* mutants. This RNA corresponded in size to the 5' ETS, after endonucleolytic cleavage at the P site and putative A0 site.

Interestingly, the 450nt long RNA did not accumulate in any of the *rrp6l1* mutants. In addition, we did not observe an increased accumulation of the 5' ETS in double mutants. These results strongly suggest that the RRP6L2 but not RRP6L1 is involved in the degradation of the 5' ETS. Furthermore, the PMC2NT domain that is found only in RRP6L2, in Arabidopsis, was shown to mediate yeast Rrp6p interaction with Rrp47p, a RNA-binding protein (Stead *et al*, 2007). However, only RRP6L1 and not RRP6L2 complemented the yeast *rrp6*Δ growth defect (see submitted publication in appendix 6).

Together, these results demonstrate that RRP6L2, but not RRP6L1, is involved in degradation of an rRNA maturation by-product. This species can accumulate in a polyadenylated form, upon RRP6L2 knock-down.

These results support the existence of a nuclear RNA degradation pathway involving polyadenylation in plants.

4. Intracellular localization of Arabidopsis RRP6-like proteins.

To further investigate AtRRP6-like proteins functions, I wanted to determine the intracellular localization of the 3 members of the AtRRP6-like family.

As yeast Rrp6p is involved in ribosomal RNA processing and snoRNA/snRNA processing, the protein is expected to have a nuclear localization. Indeed, it localizes in the nucleus and is enriched in the nucleolus (Allmang *et al*, 1999b). In drosophila, RRP6 is predominantly, but not exclusively nuclear (Graham *et al*, 2006). Human Pm/Scl-100 is also predominantly nuclear but at least a small quantity could be detected in the cytoplasm of human cells (Brouwer *et al*, 2001). In trypanosome, even if RRP6 can be found independent from exosome association, RRP6 is an core subunit of the exosome and is detected in the cytoplasm and in the nucleus (Haile *et al*, 2006).

The WoLF PSORT program (Horton *et al*, 2007) predicts RRP6L1 and RRP6L2 to be nuclear but does not attribute clear localization to RRP6L3. RRP6L1 and RRP6L2 both have good probabilities to localize in the nucleus, as they exhibit two monopartite and two bipartite potential NLS. In contrast, RRP6L3 possess only two potential monopartite NLS. The potential nuclear localization signals (NLS) for the 3 proteins, determined with the same program, are represented on Figure 35.

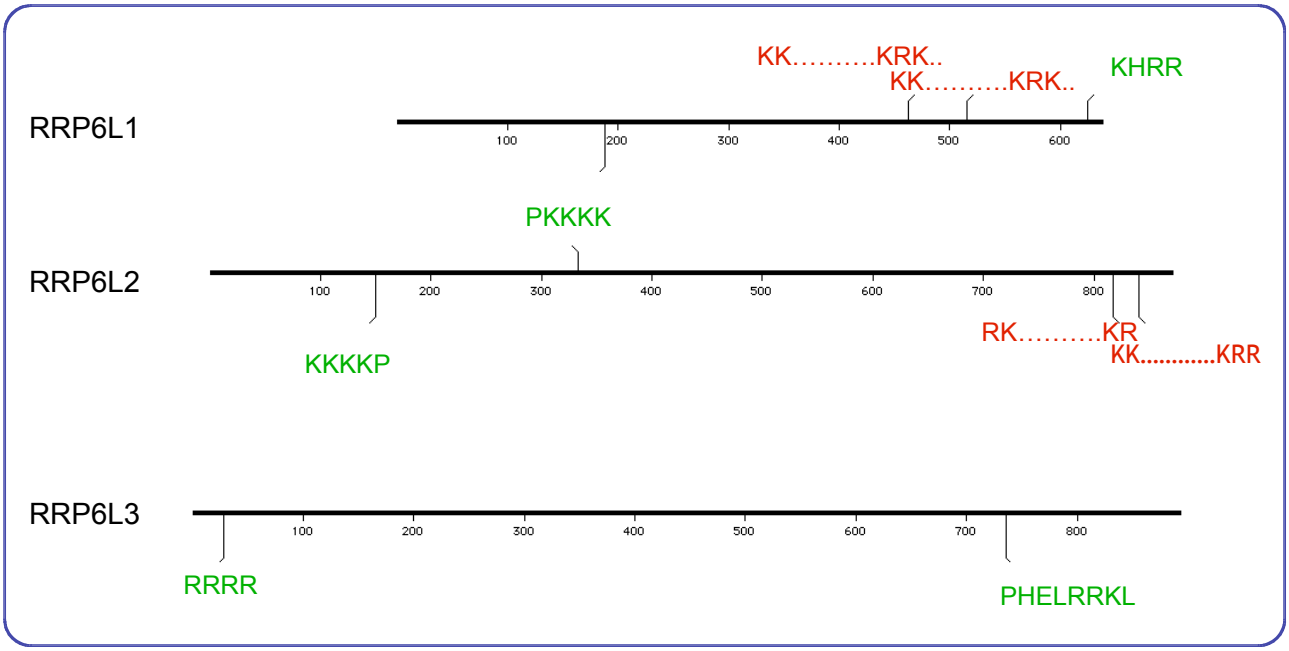


Figure 35 : Potential Nuclear Localization Sites (NLS) predicted by WoLF PSORT program for AtRRP6L proteins. Monopartite NLS are in green and bipartite NLS are in red. They are positioned on the protein sequence of each of the 3 AtRRP6L proteins.

To test these predictions, I performed GFP-fusion experiments. RRP6L1, RRP6L2 and RRP6L3 proteins were fused to an epitope-tagged version of the enhanced green fluorescent protein (GFP). N- and C-terminal fusions were elaborated to circumvent any artifact due to position of the fusion. I proceeded similarly with AtRRP4, a core subunit of the exosome, to mark the location of a core subunit of the exosome. These constructs were used to transiently transform tobacco BY2 cells by bombardment. A vector containing GFP alone was used as a negative control of GFP fluorescence localization and a vector containing GFP fused to glu-

curonidase (GUS), a construction excluded from the nucleus, was used as a non-nuclear control. Results were similar for both orientations of the fusions and representative examples are shown on Figure 36.

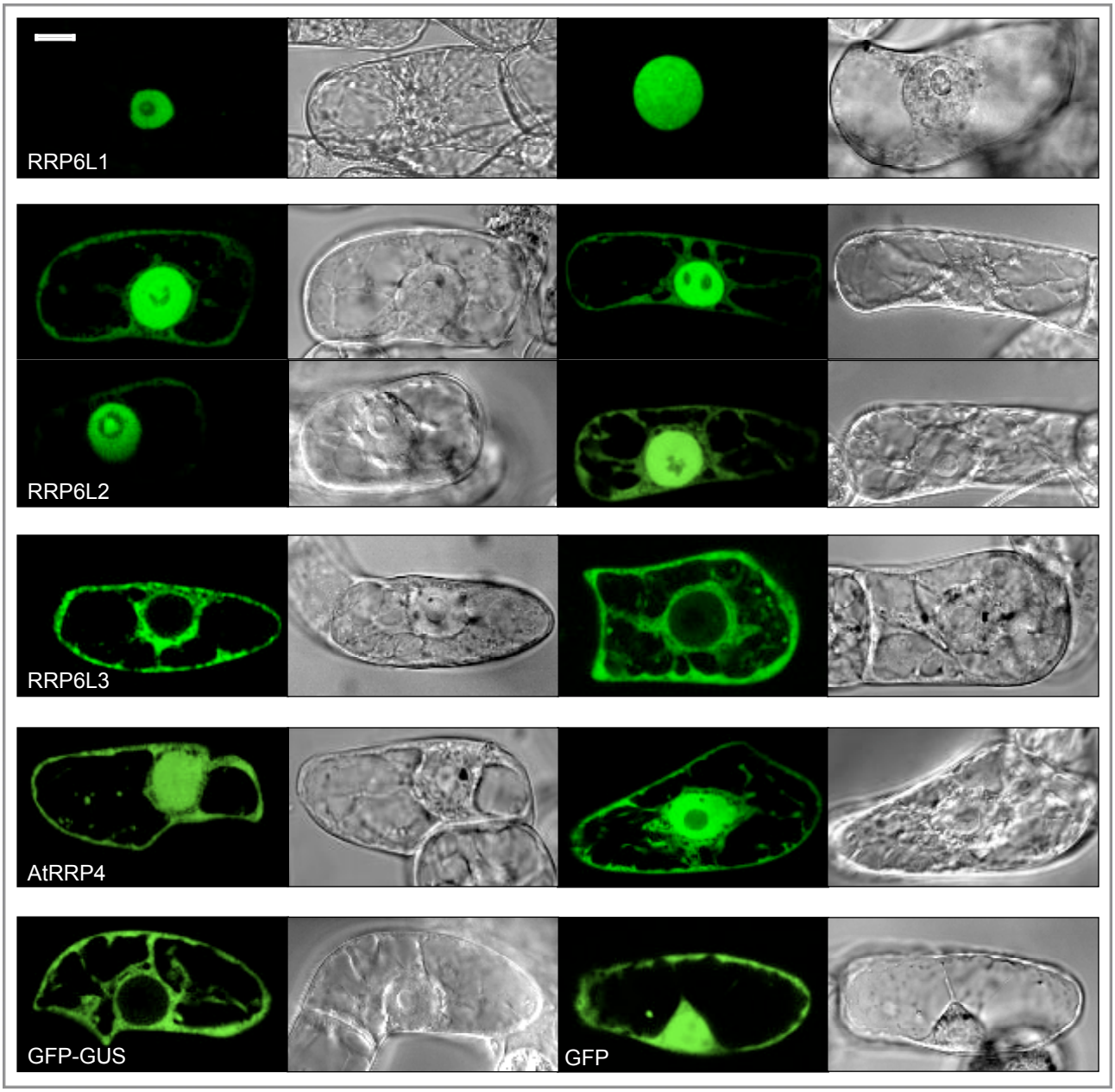


Figure 36 : confocal imaging of transitory expression of GFP-fused proteins by bombardment of 3 days-old BY2 cells. AtRRP4 marks localization of a core subunit of the exosome. GFP alone is the negative control, GFP-GUS is the nucleus excluded control. Patterns observed for N terminal and C terminal fusions were the same. The scale bar represents 10µm.

Transient expression of GFP-fused proteins in BY2 cells confirmed localizations predicted by the WoLF PSORT software : RRP6L1 and RRP6L2 fusion proteins had a nuclear localization, whereas RRP6L3 fusion protein was excluded from the nucleus.

RRP6L1-GFP was detected only in the nucleus. In most of the cases, both nucleoplasm and nucleolus showed fluorescence signal. However, in some cases,

only partial labeling of the nucleolus was observed (see picture of RRP6L1 on the left). I could not observe any cytoplasmic signal.

RRP6L2-GFP produced a strong nuclear signal. Additionally, a weak cytoplasmic labeling was observed. Figure 36 shows representative patterns of expression : fluorescence signal labeled the nucleoplasm and nucleolus, even if labeling in the nucleolus was not uniform. In some cases, parts of the nucleolus was more fluorescent than the nucleoplasm (top left and bottom right images of RRP6L2 pictures). In other cases, only foci of the nucleolus were labeled (top right image of RRP6L2 pictures) or only the center of the nucleole, likely the nucleolar vacuole, frequently observed in plants (bottom left image of RRP6L2 pictures).

RRP6L3-GFP, at the opposite of the two other constructs, showed exactly the same pattern than the GFP-GUS control construct, excluded from the nucleus. RRP6L3-GFP localization was thus strictly cytoplasmic.

The GFP constructs are under the control of a 35S promoter, driving a constitutive high expression. The use of such a promoter in the place of the endogenous promoter could induce artifactual localization of the expressed protein and results obtained with this method have to be carefully interpreted. However the distinct localizations of the three RRP6L proteins show that proteins features are sufficient to direct specific localization patterns. Nevertheless, localization patterns could be further confirmed using GFP constructs under the control of endogenous promoters, if these are active enough. Endogenous promoter of each of the RRP6L proteins could also be used to drive expression of glucuronidase (GUS) in order to investigate the tissue-specific expression of RRP6L proteins.

Transient expression of GFP-fused proteins in BY2 cells is a rapid method to have a first insight into protein localization. However, observation of stably expressed GFP-fused proteins in whole plants allows to observe localization patterns of proteins that have a weaker expression than transiently expressed proteins, thus minimizing artifactual localization. Furthermore, stable expression results in continuous expression, allowing time for the eventual association with the exosome. Correct expression of GFP-fused constructs was first checked by agrobacterium infiltration in *Nicotiana benthamiana* leaves, and the GFP-fused RRP6L proteins were then stably expressed in *Arabidopsis thaliana* plants.

Arabidopsis thaliana transformants were obtained by flower dip in agrobacterium suspension harboring the construct, and subsequent antibiotic selection of the offspring. I then observed fluorescence of root cells of 8 days *in vitro* grown plantlets using confocal microscopy (see Figure 37).

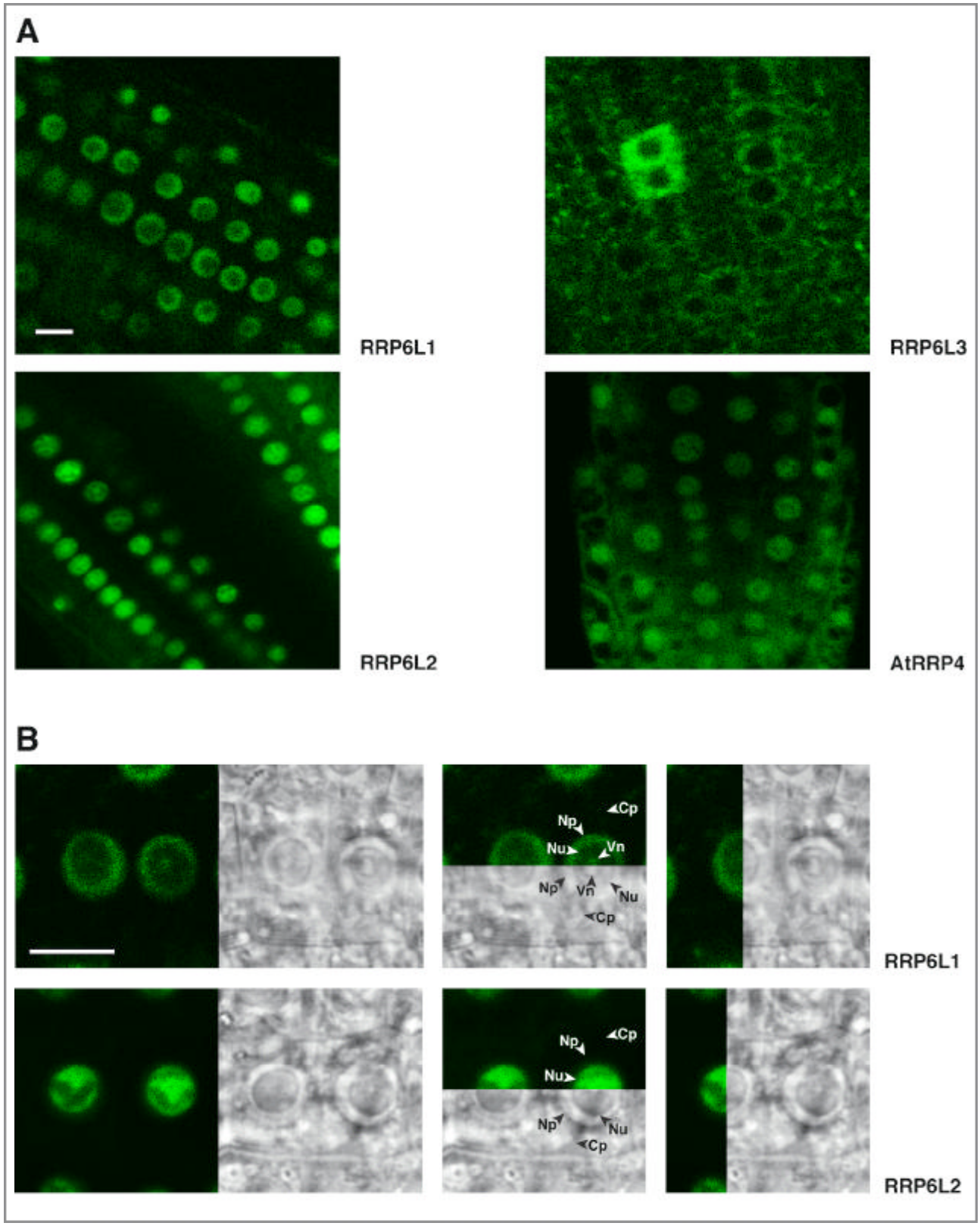


Figure 37 : Confocal images of Arabidopsis stable transformants roots expressing GFP-fused RRP6L proteins. **A.** Root apexes of 8 days old plantlets transformed with one of the RRP6L-GFP constructs or AtRRP4-GFP, as labeled. **B.** Magnification of images in A. for RRP6L1 and RRP6L2, showing nuclei. Cp : cytoplasm, Nu : nucleolus, Np : nucleoplasm, Vn : nucleolar vacuole Scale bar : 10µm

Stable expression of GFP-fusions in *Arabidopsis* root cells confirmed the localization patterns observed upon transient expression in BY2 cells : RRP6L1 localized in the nucleus, RRP6L2 was predominantly nuclear but also cytoplasmic, to a lesser extent. In contrast, RRP6L3 localized only in the cytoplasm. AtRRP4 pattern of expression was again nucleo-cytoplasmic. However, a closer examination of nuclei brought new informations : the localization patterns of RRP6L1 and RRP6L2 appeared to be mutually exclusive. RRP6L1-GFP fluorescence labeled the nucleoplasm and the nucleolar vacuole, whereas RRP6L2-GFP labeled predominantly the nucleolus, with a faint signal in the nucleoplasm. This is in agreement with virtual Northern results showing that RRP6L1 and RRP6L2 do not have overlapping functions in 5' ETS processing.

Thus, GFP-fusion experiments allowed me to define localization patterns of each member of the RRP6L family. RRLP6L1 and RRP6L2 are targeted to the nucleus whereas RRP6L3 is restricted to the cytoplasm. Thus, localization, as phylogeny, differentiates RRP6L3 from and RRP6L1 and RRP6L2.

In addition, RRP6L1 and RRP6L2 seem to have distinct subcompartimental localization in the nucleus.

5. Conclusions.

Thus we could demonstrate that non-coding nuclear RNAs are polyadenylated in plants. These results suggest that a putative RNA degradation pathway involving polyadenylation in the nucleus of plants, recalling the yeast pathway. snoRNA investigated were polyadenylated at internal sites and in downstream sequences, indicating that polyadenylation could be involved in their 3' end processing and degradation. The next step will be identification of other nuclear polyadenylated species and trying to determine whether polyadenylation is involved in processing and degradation functions. The factor responsible for polyadenylation has to be investigated in order to have a better idea on polyadenylation features. As the complex TRAMP was shown to be the yeast nucleus polyadenylating machinery, plant homologs of proteins belonging to the TRAMP complex are currently investigated in the laboratory.

In the purpose of identifying polyadenylated species in plant nuclei, we first focused on the characterization of the degrading factor to which polyadenylation targets substrates. *Arabidopsis* homologs of the yeast Rrp6p were good candidates because Rrp6p-associated exosome was shown to degrade polyadenylated intergenic and non-coding RNAs in yeast.

We found that a *Arabidopsis* family of three members is related to Rrp6p, and we named them RRP6L1, RRP6L2 and RRP6L3 respectively. Other plant species investigated also harbor a multi-component family homolog to Rrp6p, but this is not the case of other species, like human and drosophila. Thus the multiplication of Rrp6 genes could be a plant specificity. A phylogenetical analysis allowed

us to group RRP6L1 and RRP6L2 together, and set RRP6L3 apart. Whereas RRP6L3 appears to be plant specific, RRP6L1 and RRP6L2 are more closely related to the yeast and animal RRP6 proteins.

To gain access into RRP6L1 and RRP6L2 functions, we characterized plants down-regulated for these proteins. We detected a polyadenylated by-product of rRNA maturation accumulating in absence of functional RRP6L2. This species did not accumulate in *rrp6l1* mutants and no increase of this accumulation could be observed in *rrp6l1-rrp6l2* double mutants. These results show that the function of RRP6L2 in 5'ETS degradation is not shared by RRP6L1.

Intra-cellular localization analysis allowed to have insights into the function of the RRP6L proteins. Transitory expression of GFP-fusions gave localization patterns that were confirmed by stable expression in *Arabidopsis* plants. RRP6L3, in contrast to RRP6L1 and RRP6L2, was excluded from the nucleus and found strictly in the cytoplasm. RRP6L1 and RRP6L2 both presented a strong nuclear pattern, even if a small proportion of RRP6L2 seems to be cytoplasmic. Interestingly, a closer examination of RRP6L1 and RRP6L2 localization in the nucleus of *Arabidopsis* root cells showed that they seem to occupy different territories. This suggests that RRP6L1 and RRP6L2 do not have overlapping functions but rather have different substrates, localized in different parts of the nucleus and the nucleolus. The nucleolar localization of RRP6L2 is in agreement with a role in rRNA processing as rRNA biogenesis takes place in the nucleolus.

The cytoplasmic expression of RRP6L3 indicates that RRP6L3 has evolved to fulfill other functions than nuclear ones. This cytoplasmic expression of a protein related to Rrp6p is very interesting : it raises the possibility of a cytoplasmic pathway of RNA degradation, involving a Rrp6-like protein. It would be highly interesting to follow this possibility in further investigations.

Thus, our work presents good indications that polyadenylation has a conserved role of labeling RNAs for degradation in genetic systems ranging from bacteria to plant nucleus. Furthermore, we characterized a multi-component family of specialized RRP6-like proteins, suggesting an intriguing complexity of RNA degradation functions unique to plants.

DISCUSSION & PERSPECTIVES

1. A mitochondrial RNA surveillance pathway involving polyadenylation eliminates spurious RNAs.

Review in BBA-GSE

Plant mitochondria have large genomes which are mostly composed of non-coding DNA due to duplication and insertion events. Cryptic and potentially deleterious transcripts are massively and constantly produced from those intergenic regions. Production of these spurious transcripts can be explained by several factors such as duplication of promoters by recombination, relaxed promoter specificity and lack of transcription termination. Several deleterious effects of these cryptic transcripts are expected : depletion of factors involved in post-transcriptional processes by titration, disturbance of gene expression through antisense RNAs, production of "toxic" proteins, as examplified by CMS cases.

Illegitimate transcripts are polyadenylated and subsequently rapidly turned-over by PNPase. This statement raises the further question : how is the substrate recognized by the polyA polymerase?

A probable answer would be that there is no specific recognition of RNAs to be polyadenylated but rather a specific stabilization of functional RNAs. This stabilization is not conferred by a generic signal, as poly(A) tails for mitochondrial mRNAs of trypanosome and human. A few plant mitochondrial RNAs have a 3' stem-loop preventing exoribonucleolytic degradation, but individual stability is likely conferred by nuclear-encoded factors. Non-protected RNAs could be polyadenylated by default and thus targeted for degradation (see figure 1 of the following review). Such a system would allow plant mitochondria to tolerate the massive production of cryptic transcripts.

Such a surveillance pathway does not apply to defective transcripts commonly produced by inefficient post-transcriptional processes. Indeed, partially edited or unspliced transcripts are detected in polysomes, suggesting a lack of selection for functional transcripts at the translation level.

The ideas presented here are developed in the following review submitted for publication in a special issue on RNA quality-control and surveillance in Biochimica et Biophysica Acta-Gene Structure and Expression.

*** Manuscript**

Coping with cryptic and defective transcripts in plant mitochondria

Sarah Holec, Heike Lange, Jean Canaday and Dominique Gagliardi*

Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Unité Propre de Recherche 2357, Conventioné avec l'Université Louis Pasteur, 67000 Strasbourg, France.

*Corresponding author. Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Unité Propre de Recherche 2357, 12 rue du général Zimmer, 67000 Strasbourg, France.

Tel: +33 3 88 41 71 62

FAX: + 33 3 88 61 44 42

Email: dominique.gagliardi@ibmp-ulp.u-strasbg.fr

Key words: Plant mitochondria / polyadenylation / RNA degradation / PNPase

Abstract

Plant mitochondria are particularly prone to the production of both defective and cryptic transcripts as a result of the complex organisation and mode of expression of their genome. Cryptic transcripts are generated from intergenic regions due to a relaxed control of transcription. Certain intergenic regions are transcribed at higher rates than genuine genes and therefore, cryptic transcripts are abundantly produced in plant mitochondria. In addition, primary transcripts from genuine genes must go through complex post-transcriptional processes such as C to U editing and cis or trans splicing of group II introns. These post-transcriptional processes are rather inefficient and as a result, defective transcripts are constantly produced in plant mitochondria. In this review, we will describe the nature of cryptic and defective transcripts as well as their fate in plant mitochondria. Although RNA surveillance is crucial to establishing the final transcriptome by degrading cryptic transcripts, plant mitochondria are able to tolerate a surprising high level of defective transcripts.

Coping with cryptic and defective transcripts in plant mitochondria

Sarah Holec, Heike Lange, Jean Canaday, Dominique Gagliardi

Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2008, Vol. 1779, N° 9, Pages 566-573

Figure 1

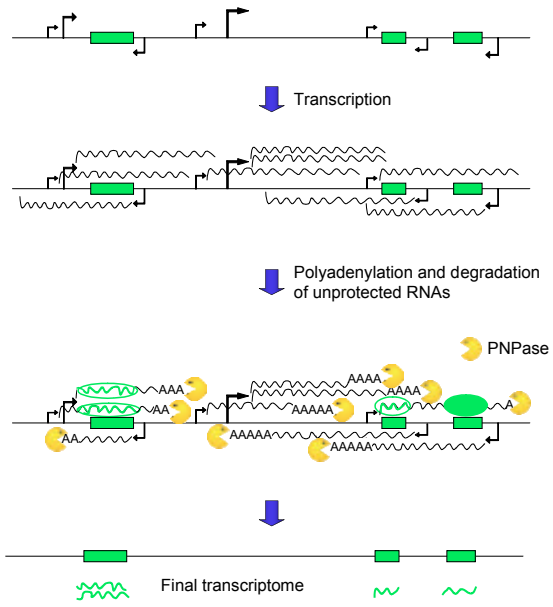


Figure 1

Figure 2

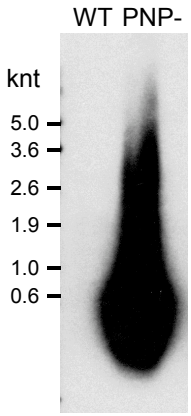


Figure 2

Figure 3

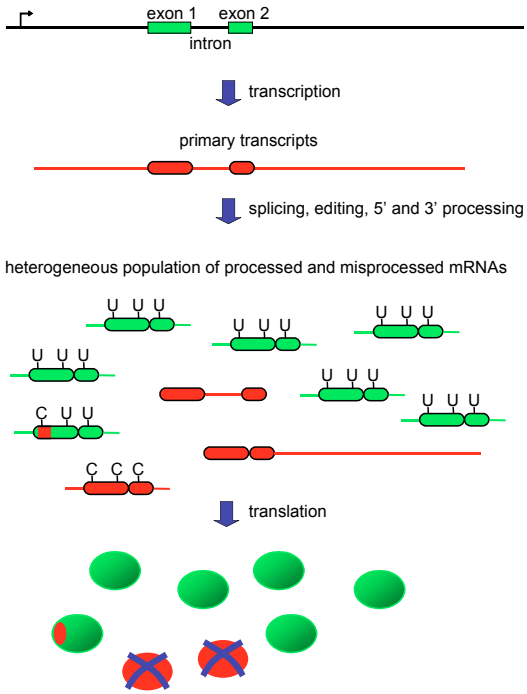


Figure 3

2. The mechanism of polyadenylation in plant mitochondria.

Control of RNA degradation in plant mitochondria can be viewed as a kinetic competition between the rate of fixation of stabilizing factors and the rate of polyadenylation triggering degradation. The idea of competition was first proposed by Li and colleagues to explain the mechanism of tRNA quality-control in *E. coli* : there would be a competition between processing enzymes and poly(A) polymerase at the level of tRNA precursor (Li *et al*, 2002). A kinetic competition is also the model proposed by Doma and Parker to explain the mechanisms of RNA degradation in the case of eukaryotic RNA quality-control. The kinetic competition occurs in the sense that "any defect leading to a delay in the normal forward reaction will trigger quality control". Either aberrant RNAs are defective in normal function or normal mRNAs promote normal function and thereby reduce quality control. Another possibility would be that aberrant ribonucleoproteins (RNPs) recruit quality-control machinery (Doma and Parker, 2007). In plant mitochondria, the model of polyadenylation directed specifically toward RNAs to be degraded is a possibility. However, an active recognition of cryptic transcripts would imply the existence of a common signature. This is unlikely given the diversity of sequence, structure and origin of mitochondrial transcripts. More likely, functional RNAs are taken over by stabilizing factors and escape polyadenylation whereas cryptic RNAs are not recognized by specific factors and are polyadenylated, probably by default.

To gain insights into the precise mechanism of polyadenylation, we will have to identify the mitochondrial poly(A) polymerase that is still unknown in plants. PNPase is the enzyme responsible for poly(A) polymerization in some systems as spinach chloroplast (Yehudai-Resheff *et al*, 2001) and *Streptomyces coelicolor* (Sohlberg *et al*, 2003). That is not the case in Arabidopsis chloroplast where polyadenylation is due to another poly(A) polymerase than PNPase as plants down-regulated for the chloroplastic PNPase showed an increase of polyadenylated chloroplastic RNAs (Walter *et al*, 2002). Similarly, PNPase is unlikely responsible for polyadenylation in Arabidopsis mitochondria as we detected accumulation of polyadenylated mitochondrial RNAs upon PNPase down-regulation. Nevertheless, we could still imagine that PNPase has an extremely efficient poly(A) polymerase activity and that a dim amount of PNPase is sufficient to take over the poly(A) polymerase activity but not the degrading activity. However, the enzyme responsible for polyadenylation of Arabidopsis mitochondrial RNAs is likely a poly(A) polymerase of the eubacterial type. Moreover, heteropolymeric tails are the mark of polyadenylation by PNPase (Slomovic *et al*, 2006) but the poly(A) tails detected in Arabidopsis mitochondria are mainly homopolymeric.

A novel class of poly(A) polymerase was described in several eukaryotes, such as Cid1-like proteins in the fission yeast *Schizosaccharomyces pombe* (Ste-

venson and Norbury, 2006), and Trf4p/Trf5p in *Saccharomyces cerevisiae* (Saitoh *et al*, 2002), GLD-2 in *Caenorhabditis elegans* (Wang *et al*, 2002), hmtPAP in human mitochondria (Tomecki *et al*, 2004) and kPAP2 in *Trypanosoma brucei* (Kao and Read, 2007). So far, a candidate for such a protein has not been reported in plant mitochondria. In contrast, three Arabidopsis genes were characterized as encoding putative poly(A) polymerase of the eubacterial-type, two of which are predicted to encode products targeted to mitochondria (Martin and Keller, 2004). But this bioinformatic study awaits confirmation by experimental data.

3. Is the Non Coding Overexpressed transcript really devoid of function?

The characterization of polyadenylated substrates of PNPase led us to the identification of the *NCO* transcript : a highly transcribed and rapidly turned-over transcript that is nevertheless processed at 5' and 3' ends, and edited. This transcript is generated from one of the large repetitions of the mitochondrial genome. This sequence is devoid of ORF of more than 30 codons and is therefore certainly not encoding any protein.

Several arguments can be advanced for the non-functionality of *NCO.* The first one is that the sequence coding for *NCO* is not conserved between various plant species. However, there are some cases of rapidly evolving RNAs that have a demonstrated functionality like *Xist* in mammals, showing that the consideration that rapid evolution is an evidence of lack of functionality may be incorrect. This is particularly true for non-coding DNA, and these sequences may in fact be simply able to drift easily because of different constraints and/or be subject to positive selection related to phenotypic variation (Pang *et al,* 2006). The proportion of non-coding sequences increases with the complexity of the organism (Taft *et al*, 2007) as, for example, it is estimated that 98% of the human genome is non-coding. The extensive amount of sequences (intronic and intergenic) not encoding any protein in eukaryotes has been regarded as accumulated evolutionary debris. But recent data show that a large proportion are actually transcribed : 60-70% at least of mammalian genome is transcribed on one or two strands and if most of these transcripts have unknown functions, there is no evidence that they do not have any (Mattick and Makunin, 2006).

There is evidence that transcription from upstream regions can affect the expression of the adjacent gene, leading to the hypothesis that the act of transcription is itself responsible for regulatory effects, by promoter interference or by altering chromatin structure and that the non-coding RNA would only be a by-product (Schmitt and Paro, 2004). Similarly, one could imagine that the function of *NCO* transcript is not in the expression of the genetic information it contains but rather in the act of its transcription itself. One could further hypothesize that the transcription of the repeated sequence generating the *NCO* transcript has a role in the mitochondrial genome replication or in sequence recombination, by

unwinding the double-stranded molecule of DNA, for instance. In agreement with this hypothesis, Shedge and colleagues suggested that strand-invasion events at the long repeats of plant mitochondrial genome are initiation sites for recombination-dependent replication (Shedge *et al,* 2007). Also, large repetitions of the mitochondrial genome have a long-known function in recombinations (Klein *et al,* 1994). To test an eventual function of *NCO* transcription in replication or recombination, experiments involving mutation of *NCO* promoter to deplete its function and detection of eventual defect in replication or recombination would be interesting to realize. However, this would necessitate the transformation of mitochondria and homologous recombination, a technique that is not feasible yet.

Thus, if NCO itself is likely devoid of function in genomic expression of plant mitochondria, we cannot exclude that its transcription could nevertheless have a role in replication or recombination.

4. NCO transcript is a model to study 3' extremity maturation of plant mitochondrial RNAs.

A combination of features distinguishes *NCO* from previously studied plant mitochondrial transcript :

- a high turnover
- defined extremities in the absence of PNPase
- a 3' maturation that depends on a specific nuclear factor.

For these reasons *NCO* is a good model to study 3' extremity maturation of plant mitochondrial transcripts.

In the light of the data obtained on *NCO* transcript, we can reexamine previous conclusions that were made in our laboratory on other plant mitochondrial transcripts. In the work of Perrin and colleagues, accumulation of long *atp9* precursors was detected in absence of PNPase. The processing of *atp9* mRNAs was proposed to be mainly exonucleolytic and due to PNPase (Perrin *et al*, 2004a). The absence of a maturation by-product generated by an endonucleolytic cleavage in PNP- plants reinforced this hypothesis. In addition, these plants had reduced levels of mature *apt9* mRNA. Whether these mature transcripts were produced before down-regulation of PNPase could not be known. But the example of *NCO* clearly shows that 3' extremity maturation can be independent of PNPase and raises the possibility that long *atp9* polyadenylated precursors observed in absence of PNPase are not intermediates of maturation but aberrant precursors targeted for degradation.

We could not detect an endonucleolytic cleavage generating the mature 3' extremity of *NCO*. The mechanism responsible for *NCO* 3' extremity maturation

remains elusive. If an endonucleolytic cleavage is responsible for *NCO* 3' maturation, the distal product of the cleavage is efficiently degraded in the absence of PNPase as we were unable to detect it by Northern blot analysis, ribonuclease protection assay or 5' RACE. However, further experiments can be tried to investigate efficient degradation of the putative distal product and try to distinguish between a 3' to 5' exoribonucleolytic activity or a 5' to 3' degrading activity- exoribonucleotidic or a succession of endocleavages (Figure d1).

As *in vitro* test conditions appear difficult to set up, an *in organello* degradation test has to be developed. Two different *in organello* expression systems exist to date, one based on electroporation (Farré and Arraya, 2001) and the other based on DNA uptake (Koulitchenko *et al,* 2003). Such a test would necessitate the *in organello* transcription of a DNA construct corresponding to the sequence downstream of NCO but containing a polyG to stop the degrading activity. Putative degradation products would subsequently be detected by sequence-directed RT-PCR. If the part in 5' of the polyG is detected, the activity proceeds in a 3' to 5' direction. If the other part is detected, the degrading activity acts from 5' to 3'. However, this experiment does not allow to distinguish between a wave of endonucleolytic cleavage and an exoribonuclease. This strategy relies on the fact that polyG would block the progression of RNases. This might not be the case as polyG sequences were previously observed as being degraded *in vitro* (Kuhn *et al*, 2001). A direct importation of a radiolabeled RNA probe would facilitate the setting up of such an *in organello* degradation test. Efforts to develop such an approach are in progress (Laurence Maréchal-Drouard, personal communication).

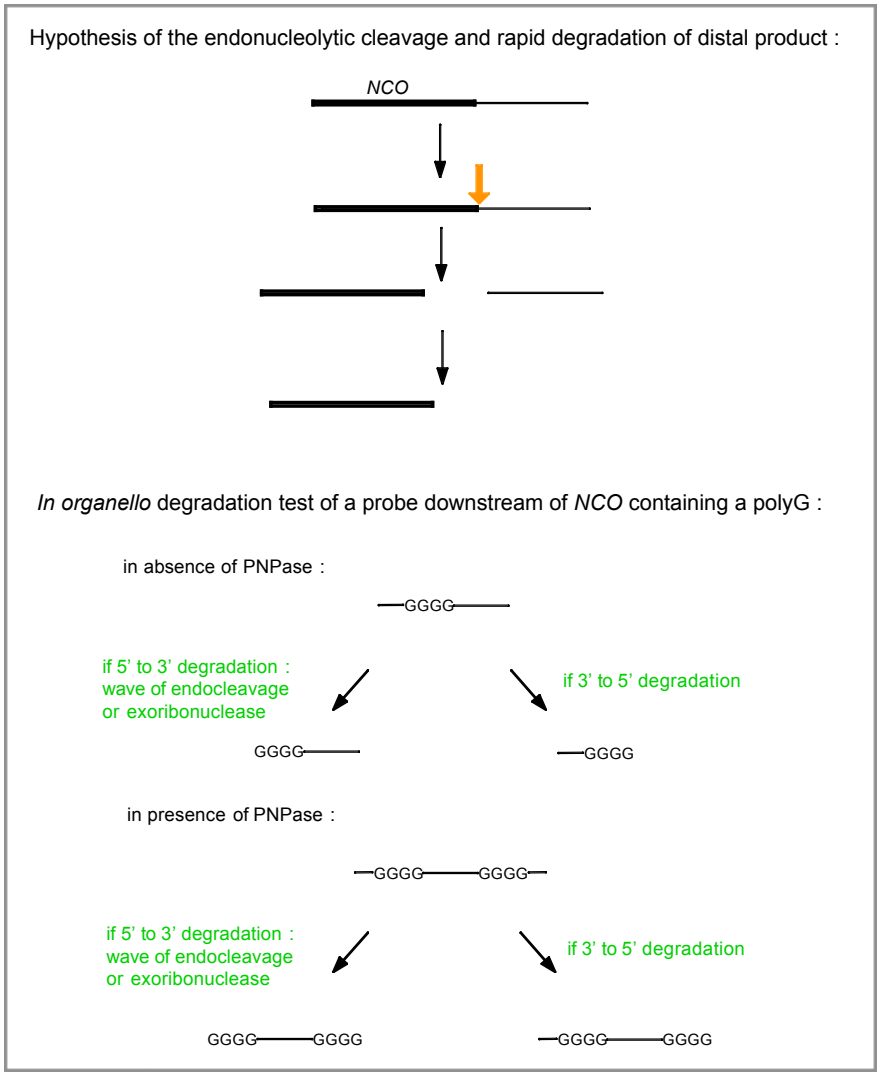


Figure d1 : In organello degradation test to determine the degrading activity putatively acting on distal endocleavage product. Upper part : Model of the 3' NCO maturation : after endonucleolytic cleavage (orange arrow), the distal product is rapidly degraded. Lower part : Expected results of an in organello degradation test using a probe containing a polyG.

Consecutive clones corresponding to tRNA maturation by-products in the library of polyadenylated RNAs are hints of the action of an endonucleolytic enzyme. Similarly, numerous 5' and 3' extremities were detected on transcripts of the *NCO* locus. These extremities are likely due to this same endonuclease and are probably not linked to any processing mechanism. The endonucleolytic activity is secondary as cleavage products are not easily detected and it does not compensate for PNPase down-regulation. The identification of this enzyme would however be interesting to have a global view of RNA degradation in plant mitochondria.

Thus, our work shows that *NCO* 3' processing depends on a specific *trans* factor. This transcript is matured in absence of PNPase. Interestingly, the *NCO* transcript can be efficiently processed in wild type plants but is not stable. This

situation raises the following question : are processing and stabilization interwoven or independent processes?
Originally, it was demonstrated that 3' stem-loops do not generate transcription termination in plant mitochondria but are rather processing signals (Dombrowski *et al*, 1997). Theses 3' stem-loop were further shown to act as stabilizing factors preventing RNA degradation (Kuhn *et al*, 2001). However, only few mitochondrial transcripts possess such a 3' stem-loop (Forner *et al*, 2007), indicating that another factor must act as stabilizing factor. The case of *NCO* shows that specific *trans* factor can be the missing player. *NCO* is matured but not stabilized. It demonstrates for the first time that processing and stabilization processes can be uncoupled and that maturation can occur independently of stabilization. Therefore, there is a possible competition at the 3' mature end of plant mitochondrial transcripts between polyadenylation and stabilizing factors (Figure d2). This proposition would explain why poly(A) tails are found mostly at mature 3' extremities in plant mitochondria.

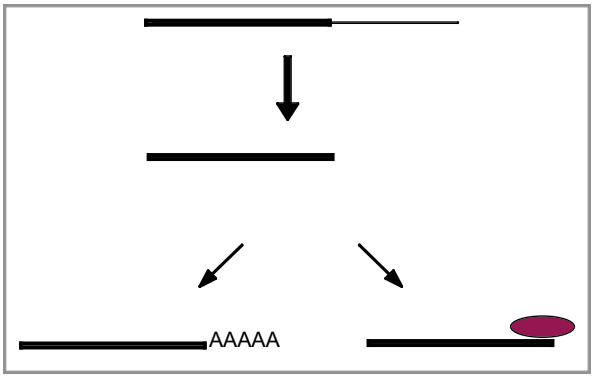


Figure d2 : 3' extremity maturation and stabilization of mitochondrial transcripts can be uncoupled processes. If the first step of maturation is distinct from stabilization, there could be competition beteween poladenylation and binding of a trans stabilizing factor at the mature 3' end of transcripts.

5. Three RRP6-like proteins perform specialized functions in plants?

We identified three RRP6-like proteins encoded by the genome of rice, poplar and *Arabidopsis thaliana* whereas a unique gene is found in yeast and animals investigated. Phylogenetic analysis clearly distinguished At-RRP6L1 and 2 - that group with human and yeast proteins - from At-RRP6L3 that seems to be a plant-specific gene. Could this situation, unique to plants, be the sign of a specialization of each RRP6L protein?

5.1. RRP6L3 : a plant cytoplasmic protein related to RRP6.

Interestingly, the localization pattern of RRP6L3 determined by GFP-fusion experiments in BY2 cells and in *Arabidopsis thaliana* is strictly cytoplasmic. Yeast Rrp6p was found to be strictly nuclear (Allmang *et al*, 1999b) but its counterpart was found to be nucleo-cytoplasmic in trypanosomes (Haile *et al*, 2007) and a small fraction of the protein was detected in the cytoplasm in human and drosophila (Brouwer *et al*, 2001, Graham *et al*, 2006). These data suggest a cytoplasmic function for RRP6-related proteins in higher eukaryotes including plants. It remains to be determined whether the biological role of RRP6L3 is related to the function of cytoplasmic RRP6 proteins in animals and humans or is a rather plant-specific function that could also be independent of the exosome. For this purpose, the association with cytoplasmic exosome must be tested, by co-immunoprecipitation or by purification of an eventual complex containing a tagged-RRP6L3.

To date, it is unknown whether polyadenylation triggers RNA degradation in the cytoplasm. Cytoplasmic non-canonical polyA polymerases have already been described in several eukaryotes, such as Cid1 and Cid13 in *Schizosaccharomyces pombe* and GLD2 in the nematode *Caenorhabditis elegans* and *Xenopus laevis* as well as human and mouse. Cytoplasmic polyadenylation has been characterized in animal oocytes and neuronal cells and in these cases, it is a rapid mechanism of translation upregulation (Stevenson and Norbury, 2006). However, an enzyme could be responsible for degradation of polyadenylated RNA in the cytoplasm of plants : AtRrp41p (Chekanova *et al*, 2000). AtRrp41p is a subunit of the exosome but the activity tested was the activity of the purified recombinant protein and it is unknown for the moment if the activity is retained in the context of the exosome. RRP6L3 would be another candidate for such an activity as Rrp6p-like proteins were shown to degrade polyadenylated substrates. To verify this hypothesis, we could try to monitor the accumulation of cytoplasmic polyadenylated RNAs destined for degradation in wild-type plants and in plants depleted for RRP6L3. Plant viruses with a life-cycle restricted to the cytoplasm, such as the Crinkle Turnip Virus (CTV), could be used as model substrates.

5.2 Nuclear RRP6L1 and RRP6L2 have non-overlapping functions.

RRP6L1 and RRP6L2 are both closely related to and share some functionality with yeast Rrp6p. RRP6L1 complements yeast *rrp6*Δ growth defect. On the contrary, only plants down-regulated for RRP6L2 accumulated the 5' ETS, a maturation by-product of rRNA synthesis and a classical substrate for yeast Rrp6p However, their distinct intra-nuclear localization and their substrate difference indicate that they have functions that do not completely overlap. The structural organization of RRP6L2 and its localization pattern support the role of RRP6L2 in the degradation of the 5'ETS. Indeed, among the three RRP6-like proteins in Arabidopsis, only RRP6L2 contains the N-terminal PMC2NT domain that mediates an interaction

of Rrp6p with the RNA binding protein Rrp47p in yeast (Stead *et al*, 2007). Absence of Rrp47p impairs both 3' processing of stable RNA and degradation of the 5' ETS (Mitchell *et al,* 2003). Thus, the N-terminal region of RRP6L2 could be necessary for degradation of the ETS in plants by interaction with the plant homolog of Rrp47p. In addition, the intra-nuclear distribution of the RRP6L2 fusion proteins is in agreement with the localization of the 5'ETS that was detected in pea root nucleoli (Shaw *et al*, 1995).

It is known that the nucleolus is not uniform. Subcomponents of the plant nucleolus are described. Plant cells have the particularity to display a nucleolar vacuole. Fibrillar centers are also observed as foci, surrounded by dense fibrillar components, themselves encircled by granular components. Transcription and processing of ribosomal RNA are not taking place uniformly in the nucleolus but that these events are spatially organized following ultrastructural regions. A probe to ETS sequence of the pre-rRNA transcripts labeled clearly demarcated regions corresponding to the dense fibrillar component (Shaw *et al*, 1995; Beven *et al* 1996). The level of resolution of our images do not allow to draw any conclusion on ultrastructural expression but it is still possible that the non uniform expression in the nucleolus has a biological significance.

The nucleolar vacuole is highly dynamic and its presence and structure depends on the cell cycle progression. Its role is unknown but small nuclear and nucleolar RNAs where shown to localize in this space (Shaw and Brown, 2004). A recent report showed by GFP-fusion that an *Arabidopsis* La protein, involved in small nuclear biogenesis, is expressed in the nucleoplasm and in some cases in the nucleolar vacuole (Fleurdépine et al, 2007). The pattern of expression is the same than the pattern seen in some cases for RRP6L2 transiently expressed in BY2 cells. In line with our results, expression of the GFP-fused AtRRP45a, a subunit of the plant core exosome, showed a pattern that we also observed for RRP6L2 transiently expressed in some BY2 cells : a nucleo-cytoplasmic expression with the nucleolus unlabeled except to some foci (Hooker *et al*, 2007).

However, the AtLa1 protein, as well as RRP6L1, RRP6L2, AtRRP4 and AtRRP45a, was not found in the Arabidopsis Nucleolar Protein Database (Brown et al, 2005). This could be explained by the fact that the database contains only 217 proteins and is certainly not exhaustive.

Further localization studies could help to better define localization of RRP6L1 and RRP6L2. For example, repeating GFP-fusion experiments using endogenous promoters would allow to test eventual localization bias due to the use of the strong constitutive 35S promoter. The use of another technique, such as immunolocalization, could also bring confirmation of the pattern observed with GFP fusions. To better investigate the territories occupied by RRP6L1 and RRP6L2, double-labeling experiments could be useful. Furthermore, as the nucleolar vacuole is highly dynamic during cell-cycle, time-laps imaging would be interesting to detect eventual cell-cycle-dependent localization.

Yeast Rrp6p is specifically associated with the exosome in the nucleus (Allmang *et al*, 1999b). In trypanosomes, RRP6 is a subunit of the cytoplasmic and

the nuclear exosome and but a fraction of RRP6 is also found independent from this complex (Haile *et al*, 2006). Are RRP6L1 and RRP6L2 both associated with the exosome in plants? If the two proteins have evolved to fulfill different roles in plants, it is possible that one or both of them is not found associated with the exosome. To answer this question, co-immunoprecipitation using antibodies raised against AtRRP4 and RRP6L1/2 proteins, or gradient purification of tagged RRP6L1/2, could be performed to detect eventual association to a high molecular weigh complex. Such experiments are in progress in the laboratory.

A crucial question for the comprehension of both RRP6L1 and RRP6L2 functions is the identity of the specific substrates of each of these proteins. To determine these substrates in an extensive manner, microarrays on plants depleted either in RRP6L1, or RRP6L2 or both would be very useful. Such mutants are available in the laboratory. However, 5' ETS accumulation is weak in *rrp6l2* mutants and in double-mutants as we can only detect it using polyadenylated RNAs. This indicates that there is a compensation from one or several other proteins. AtRRP44 and AtRRP41 could be good candidates for such a compensation. Nevertheless, accumulation of the 5'ETS is an observable phenotype using polyadenylated RNA, indicating that it is worthwhile performing microarrays on *rrp6l2* mutants.

The distinct subcellular localization and the initial functional characterization of RRP6L1 and RRP6L2 proteins suggest a specialization rather than functional redundancy among RRP6-like proteins in plants. Our first characterization of RRP6-like proteins in plants reveals an intriguing complexity as compared to other organisms studied to date.

6. Polyadenylation could trigger degradation in plant nucleus.

Interestingly, the 5'ETS accumulating upon knock-down of At-RRP6L2 is polyadenylated. This observation is consistent with recent discoveries in yeast, where several nuclear transcripts can be polyadenylated by the TRAMP complex and are subsequently degraded by Rrp6p (Wyers *et al*, 2005; La Cava *et al*, 2005). This is reminiscent of the bacterial polyadenylation-dependent RNA degradation that also occurs in chloroplast and plant mitochondria. Our results reinforce the idea that the ancient role of polyadenylation in targeting transcripts for degradation is conserved in most genetic systems and operates in all three genetic compartments of a plant cell (Bollenbach *et al*, 2004, Gagliardi *et al*, 2004, Doma and Parker 2007).

Recent discoveries have demonstrated promising interactions with other pathways gene expression control. In drosophila cells, the two siRNA-mediated cleavage products are rapidly degraded exoribonucleolytically, without undergoing decapping and deadenylation : the 5' fragment from its 3' end by the exosome, and

the 3' fragment from its 5' extremity by XRN1 (Orban and Izaurralde, 2005). RNA induced silencing complex (RISC)-cleavage 5' products were shown to be oligoadenylated by a non canonical poly(A) polymerase in *Chlamydomonas reinhardti,* possibly for degradation by the exosome (Ibrahim *et al*, 2006). Further, it appears that polyadenylation-triggered RNA degradation plays a role in gene silencing and in genome integrity. In fission yeast *Schizosaccharomymces pombe*, the exosome was shown to contribute to heterochromatin silencing of centromeric repeats for proper kinetochore formation and establishment of kinetochore-microtubule interaction (Murakami *et al*, 2007). Later, polyadenylation by a TRAMP-like complex was described to contribute to robust silencing of heterochromatic genes via recruitment of the exosome and/or RNAi machinery (Bühler *et al,* 2007). Wang and colleagues further assessed the role of polyadenylation-assisted nuclear RNA turnover in eliminating a variety of RNA targets involved in controlling diverse processes such as heterochromatic gene silencing, meiotic differentiation, and maintenance of genomic integrity (Wang *et al*, 2007). In the budding yeast *Saccharomyces cerevisiae* strains lacking the non canonical poly(A) polymerase Trf4, cryptic transcripts were detected from regions of repressed chromatin at telomeres and in the rDNA intergenic spacer region (Houseley *et al*, 2007). Furthermore, this work gave indications that Trf4 is involved in rDNA copy number control. Very recently, Camblong and colleagues demonstrated that even if the budding yeast is devoid of core RNAi machinery, gene silencing can occur if antisense transcripts are stabilized by exosome impairment, via the recruitment of histone deacetylases (Camblong *et al*, 2007).

Thus, the function of polyadenylated-assisted RNA degradation in the nucleus and cytoplasm of eukaryotes is an exploding field of research. Plants constitute an interseting model in which polyadenylation-mediated RNA degradation poerates in all three genetic compartments : nucleus, chloroplast and mitochondria.

General conclusion

PNPase was previously shown to be essential for degradation of plant mitochondrial mRNAs and rRNAs precursors and maturation intermediates. All substrates of PNPase were polyadenylated. However, the function of PNPase was investigated only for a few chosen RNAs. Therefore, the aim of this work was to obtain a more comprehensive identification of substrate RNAs through analysis of a library of polyadenylated RNAs from plants lacking mitochondrial PNPase. The general objective was to determine biological roles of PNPase in plant mitochondria. The degrading role of PNPase could be generalized in removing maturation by-products of rRNAs and tRNAs. Moreover, we determined a novel role of PNPase in degrading spurious transcripts generated by a relaxed control of transcription. We thus described an essential role of PNPase in modeling the plant mitochondrial final transcriptome. PNPase eliminates potentially deleterious RNAs : RNAs that could titer editing machinery, antisense RNAs and chimeric ORFs. Thus, in addition to its role in turnover of mitochondrial transcripts and maturation by-products, PNPase is involved in a general surveillance pathway in plant mitochondria.

We also characterized the existence of plant mitochondrial transcripts that were not only polyadenylated but also extended by heteropolymeric extensions, some of them complementary to the encoded 3' sequence, could anneal to that sequence and from a double stranded RNA stretch. The enzyme(s) responsible for the extensions could be a phage-type polymerase but this is still to be demonstrated. If these extensions have any function or not remains to be determined.

The analysis of polyadenylated RNAs allowed us to identify a mitochondrial transcript that we used as a model to study a 3' processing event that is independent of PNPase. This *NCO* transcript is highly transcribed but is rapidly turned-over by PNPase. It has defined mature ends. We showed that 3' mature extremity was not generated by termination of transcription but by processing. This processing is independent from PNPase or RNase II but we could not reveal any endonucleolytic cleavage generating the mature 3' extremity. The mechanism and enzymes responsible for *NCO* 3' extremity maturation remain to be identified but we demonstrated that a nucleus-encoded specificity factor is necessary and that processing and stabilization processes are uncoupled.

The presence of sequences corresponding to nuclear RNAs in our library of polyadenylated clones prompted us to investigate the possibility of a degradation pathway involving polyadenylation in plant nuclei. We identified 3 genes encoding RRP6L1, RRP6L2 and RRP6L3 proteins respectively. These proteins are homolog to yeast Rrp6p, a nuclear subunit of the exosome involved in degradation and processing of nuclear polyadenylated non-coding RNAs. RRP6L3 appears to be a plant-specific protein and is located in the cytoplasm, whereas RRP6L1 and RRP6L2, that occupy different territories of the nucleus, are more closely related to yeast

Rrp6p. RRP6L1 complements yeast *rrp6Δ* growth defect and RRP6L2 has a role in a polyadenylated ribosomal RNA maturation by-product degradation.

Thus, our results suggest that a degradation pathway involving polyadenylation takes place in plant nucleus, supporting the idea of a conserved role of polyadenylation in triggering RNA degradation. Furthermore, we identified a plant-specific multiple family of RRP6L proteins. The functions of each members have to be better defined but they appear to play different roles, indicating a complexity that would be unique to plants.

MATERIAL & METHODS

I. MATERIAL.

1. Plant material.

1.1. *Arabidopsis thaliana.*

Arabidopsis thaliana of the Columbia 0 ecotype (Col0) was used for the work presented here. Plants were cultivated either in greenhouses under the following conditions : day/night photoperiod of 12/12 hrs, 20°C with 50% humidity and at least 10 kilolux luminosity, or in growth chambers under the following conditions : day/night photoperiod of 12/12 hrs, 20/18°C, with 80% humidity and at least 10 kilolux luminosity.

1.2. *Nicotiana benthamiana.*

Nicotiana benthamiana plants of 15 days were used to perform infiltration experiments for transient expression of GFP-fused proteins. Plants were cultiveted in IBPM greenhouses under the following conditions : day/night photoperiod of 12/12 hrs, 20°C with 50% humidity and at least 10 kilolux luminosity.

1.3. *Arabidopsis thaliana* seedlings grown *in vitro.*

In vitro culture of *Arabidopsis thaliana* is performed in Petri dishes containing the following culture medium. The Murashige and Skoog (MS) medium for *in vitro* culture contains macro, micro MES and vitamins (MS255 Duchefa) and is supplemented with saccharose (BDH AnalarR, 10g/L) and agar (Agar-agar, Merck, 8g/L); pH 5,7. Specific antibiotics for transgene selection were added after sterilization of the medium. After vernalization of the seeds 3 days at 4°C in obscurity, Petri dishes were placed in growth chamber with following conditions : 2500 lux, day/night photoperiod of 16/8 hrs, 20-22°C.

1.4. *Arabidopsis thaliana* cell culture.

Arabidopsis thaliana cells were used for mitochondria purification. *A.t.* cells were cultivated in Gamborg's B5 medium (Duchefa), containing saccharose 20g/L, 2,4D 1 mg/L, pH5.7. Cultures were placed under agitation (125 rpm, 25°C), in obscurity for etiolated cells. 100 mL fresh medium was inoculated under sterile conditions every week with 10mL. Mitochondria were purified from 5 days old cells.

1.5. *Nicotiana tabacum* cv. Bright Yellow-2 (BY-2) cells.

Transient expression for localization studies were made using BY2 cells. BY2 cells were cultivated in BY2 medium containing micro- and macro-elements of MS221 (Duchefa) 4.3g/L, thiamin 1mg/L, KH_2PO_4 200mg/L, 2,4D 0.2mg/L, myo-inositol 100mg/L, saccharose 3%, pH 5.8. Cultures were maintained in obscurity, under 110 rpm agitation, at 25°C. 100 mL fresh medium was inoculated under sterile conditions every week with 10mL. 3 days old cells were used for transient expression.

2. Bacterial strains.

2.1. *Escherichia coli* strains.

The *Escherichia coli* strains used derive from *E. coli* K12. RecA-strains were used to avoid any risk of recombination during cloning. It allows the α-complementation and selection of recombinant clones through colony colour in presence of IPTG, a substrate for ß-galactosidase.

- TOP10F' (Invitrogen) :F'[lacI^q Tn10(tet^R)] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ-
- DH5α : F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deo^R nupG Φ80d*lacZ*ΔM15 Δ (*lacZYA-argF*)U169, hsdR17(r_K- m_K+), λ–
- XL1-Blue : endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacI^q Δ(lacZ)M15] hsdR17(r_K- m_K+)

2.2. *Agrobacterium tumefaciens* GV3101 (pM90) strain.

The GV3101 strain contains the Rifampicin resistance gene on its chromosome. This strain contains a Ti plasmid devoid of oncogenic (*ONC*) genes and T-DNA border sequences but carrying the virulence (*VIR*) genes. This allows the transfer, in *trans*, of a DNA sequence introduced between the artificial border sequences of a second vector such as pBIN (binary vector).

2.3. Culture medium.

The culture medium used for *E. coli* and *A. tumefaciens* culture was Luria Bertoni (LB) medium (Bactotryptone 1%, yeast extract 0.5%, NaCl 0.5%, pH 7.4). *E. coli* and *A. tumefaciens* cultures were incubated overnight in shakers at 37°C and 28°C respectively.

3. Vectors.

3.1. pCRII-TOPO® (Invitrogen).

pCRII-TOPO® is a 3.9 kb vector, provided as linearized form, with 3'T-overhangs and an activated topoisomerase I, to readily accept PCR products with 3'A-overhangs. It carries the Ampicillin and Kanamycin resistance genes and EcoRI sites flanking the PCR product insertion site for easy removal of the insert. The cloning cassette is flanked by SP6 and T7 phage polymerase promoters to allow transcription of the sequence inserted in the cassette. The PCR product insertion site is located within LacZα gene, which allows the screening of recombinant colonies by α-complementation. M13 forward and reverse priming sites can be used for PCR screening and sequencing.

3.2. pBINPLUS.

pBINPLUS (van Engelen *et al*, 1995) is a 12.3 kb vector derived from pBiN19 vector used to obtain transgenic cells. It is a binary vector as its origin of replication RK2 allows plasmid production in *E. coli* and *A. tumefasciens* cells. The Col E1 origin of replication allows to multiply by 4 the number of copies in *E.coli* compared to the RK2 replicon. NPTIII an NPTII genes confer kanamycin resistance. NPTIII gene allows selection of transformed bacteria whereas NPTII gene allows selection of the transformed plant cells. This vector has the β-galactosidase gene allowing bacteria selection through α-complementation. The pBINH vector is a vector derived from pBINPLUS vector, where the kanamycin resistance genes have been replaced by hygromycin resistance gene.

3.3. pUCAP.

This plasmid is a pUC vector with rare *AscI* and *PacI* restriction sites flanking the cloning cassette. These sites are also present in the pBINPLUS plasmid. pUCAP is a smaller vector than pBINPLUS (2.7 kb) and is a high-copy number vector. It is used to easily realize construction before transfer to the pBINPLUS vector by cloning in *AscI* and *PacI* sites.

3.4. pNEG-X1.

The pNEG-X1 (Niwa-Evrard-X1, 4.2 kb) vector was constructed by Dr Jean-Luc Evrard from the Dr Yasuo Niwa pTH2 vector containing the sequence coding for eGFP (enhanced GFP) downstream of the CaMV 35S promoter and upstream of the Nos terminator. It contains the β-lactamase gene for ampicilin resistance in *E. coli*. The sequence coding for the Etag peptide GAPVPYPDPLEPR upstream of the eGFP sequence allows expression of the Etag-eGFP fusion protein. Etag peptide is recognized with a high specificity by mAb701 (Amersham) antibody. In addition, multiple restriction sites are flanking the *Etag-eGFP* sequence.

4. Oligonucleotides.

Oligonucleotides were ordered from Sigma-Aldrich company (https://www.sigmaaldrich.com). The melting temperature is estimated with the following formula : Tm(°C) = 2x (A+T) + 4 (C+G).

Following primers have been used for the work presented in this manuscript :

Universal primers for sequencing of insert in TOPO vector	
pF	GTAAAACGACGGCCAG
pR	CAGGAAACAGCTATGAC
Primers for reverse transcription and amplification of cDNA from polyadenylated RNA	
oligodT	GAATTCATGTCGACGGTCTCATTTTTTTTTTTT
oligodT adapter	GAATTCATGTCGACGGTCTCA
Primers for SMART cDNA synthesis from BD SMART kit (BD Biosciences)	
A Oligonucleotide	AAGCAGTGGTATCAACGCAGAGTACGCGGG
CDS Primer II A	AAGCAGTGGTATCAACGCAGAGTACGCGGG5(T)30
PCR Primer II A	AAGCAGTGGTATCAACGCAGAGT
Primers for analysis of the NCO transcript : Northern analysis, cRT-PCR, 5' and 3' RACE, run-on	
P1	ACGTGATGCGAGAACCTACT
P2	GATAATCCTCAAGTGATAGAAAGGTT
P3	CAGCCTTTACCTAACCAAC
P4	AAAAAGGATTTCAAGTCCCTG
P5	GAAAAGGCACTTCCAAAGAG
P6	GCAATAACGGAAGGTAAC
P7	CAAAGAGTTCCTGTCCAAC
P8	GCACGATGTTCTCACAAC
P9	ACGTGATGCGAGAACCTACT
P10	GCACACCCGTTGATGATCC
P11	CACTGAAGACCTAACCTTGAAG
P12	GTGGAACTGGCATTTTGG
P13	ATTATGCTGCTTGCTGGG
P14	CAGGGACTTGAAATCCTTTTT
P15	CGAGTAACTTGTGCTGAGGTG
Primers for PNPase miRNA construction	
PmiR1	TAACTCGAGATGAGAGAGTCCCTTTGATC
PmiR2	GAGTCCCTTTGATCATGTGAAGTCGATAATAGATC
PmiR3	GTCGATAATAGATCTTACCTGACCACACACG
PmiR4	ATTACTAGTACGAGAGAGTACTGAGATCA
PmiR5	GAGTACTGAGATCATGTGGAGTCGATCATAGATA
PmiR6	GTCGATCATAGATAATCTAGAGAGAATAATG
Primers for substrate construction for in vitro RNA polymerization assay	
PS1.1	AAGCAGTGGTATCAAGAAGAAAGATCGT
PS1.2	ACTCGTAAAGGCAAAGAAAAAAGGGATCCG
PS2.1	AAGCAGTGGTATCAATGCCTGCCAAGCCAATAGGCGA
PS2.2	CCGGGGTTGTTACCCCGGAGAGCTTTGAGTTGCT
PS3.1	AAGCAGTGGTATCAACATTCTGTGGCT
PS3.2	CCGGAAGGAATGCAAATAAAAG
Primers for 26S probe construction	
P26S.1	GCTACTTAGATGTTTCAGTTCG
P26S.2	TTCTTTGCCTTTACGAGTTG
P26S.3	CTGTATTCGGAGTTTCCCTG
P26S.4	TGGGTCAGCGAGGAAATG
P26S.5	GGTGCGGCAACTAAAGAG
P26S.6	GTTCCTATCTACCGTTGGTG
P26S.7	CTACGGACCAACAAAGATTG
P26S.8	TGGTATTAGAGTAAGACAGCCAAC
Primers for substrate construction for in vitro degradation assay	
PA	atgtcTAATACGACTCACTATAgggCTGGGAGCTCAACCA
PB	atgtcTAATACGACTCACTATAgggAAGTAGGTTCCCGCATC
PC	TTGATCATTTACGAATTGCGTAGATAATCCTCAAG
Primers for probe construction for Ribonuclease Protection Assay	

PRPA.A	GTAGCACAGAAGCCATCATC
PRPA.B	CAAAGAGTTCCTGTCCAAC
PRPA.C	CACTGAAGACCTAACCTTGAAG
	Primers to clone and sequence RRP6L1 cDNA
PR6L1.1	CGATTCTAAAACCCTAAGCGAG
PR6L1.2	TTCTCTCTCCAAGTCCGACGAG
PR6L1.3	CAGACTGGAGAATAAGACCCC
PR6L1.4	TTGGCATCTCCTTGGCTATG
PR6L1.5	GATACCGACCAAAACATTGC
PR6L1.6	CTATTGTCTGAAGATACCGACC
PR6L1.7	GTTAGACGATGATGATGACTCG
PR6L1.8	CGGAAAGCAGAGAAGATGAAG
PR6L1.9	TCTCAGCCACTTTCACATCG
PR6L1.10	TCACTCTCAGCCACTTTCAC
PR6L1.11	AAACGGAACCTCTGCCTTGC
	Primers to clone and sequence RRP6L2 cDNA
PR6L2.1	GTGGTTACTGGCTCGGCA
PR6L2.2	TGTTGAAGAGTTTGCGGTTG
PR6L2.3	GATGACACTGACCACCGAATC
PR6L2.4	GGAGATGCCTGATAGTGTTGG
PR6L2.5	CGAAACTGGATTCTTCTGTAGC
PR6L2.6	GAAAAAAGAAAGCCCACTGC
PR6L2.7	CCCCCGACTAACACATAACG
	Primers to clone and sequence RRP6L3 cDNA
PR6L3.1	TAGCCCCGATGACAGATTC
PR6L3.2	TGAGGAAGACGCAGCACTAC
PR6L3.3	CCGTTACCGAATAATACCGTC
PR6L3.4	CCGTAGCGATTTGTTTCTTG
PR6L3.5	CCCTCACGAAAGAAGAAAACTC
PR6L3.6	GCAAACTATGCTGCTCAGTGTC
PR6L3.7	CGAACAGTTTTTCGAGGTTAGAAG
PR6L3.8	TGACGAAACGGTAGAGATGG
	Primers for RRP6L proteins and AtRRP4 fusion to GFP
PR6L1MluI	gtacaaACGCGTGAGATTTGATGATCCCA
PR6L1NotI	agccggGCGGCCGCCTAAATGTTAAGAAAGCCA
PR6L1KpnI	gccccgGGTACCgaAATGTTAAGAAAGCCACGT
PR6L1SalI	tacaGTCGACGACTTGACTTCAATGGCGA
PR6L2SalI	AAAACTCGAGTAATGTTCCTAAAACCCTAGCCA
PR6L2KpnI	AAGGTACCATATTTTTAAAGGACATACTTCGGTTC
PR6L2MluI	AAACGCGTAATGAGCGACGGTAACATGG
PR6L2EagI	TTCGGCCGTCAATTTTTAAAGGACATACTTCG
PR6L3SalI	AAAAGTCGACATGGAGCTCAAAGAAAAGGC
PR6L3KpnI	TTTGGTACCTCGGTGGAGAGTCTCTTTGT
PR6L3MluI	AAAACGCGTCATGGAGCTCAAAGAAAAGGC
PR6L3EagI	TTTCGGCCGCTACTCGGTGGAGAGTCTCTTTG
PR4SalI	TACAGTCGACAAGGGAACTAACGTCAAGG
PR4KpnI	cggggcGGTACCacCTTTTTCCTCTTTGTACGTC
PR4MluI	GTACAAACGCGTCATGGTGATGAGAAAGCTACA
PR4NotI	AGCCGGGCGGCCGCTCACTTTTTCCTCTTTGTAC
	Primers on GFP gene to sequence fusion constructs
PGFPC	GCTAACTTGTGGCCGTTTACGTCG
PGFPN	GTCCTGCTGGAGTTCGTGACCGCCGCC

5. Informatics tools.

5.1. Sequence analyses.

Sequences comparison, alignments and restriction analysis have been performed using MacVector software, version 7.2.3 to 9.5.2.

5.2. Sequence databases.

Sequence researches were realized using mainly following web sites :
NCBI BLAST : http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
Arabidopsis data base TAIR : http://www.arabidopsis.org/index.jsp
Rice data base TIGR : http://www.tigr.org/tdb/e2k1/osa1/index.shtml

5.3. RNA analyses.

Secondary structure prediction :
BiBiserv : http://bibiserv.techfak.uni-bielefeld.de/rnashapes/submission.html
MFOLD : http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html
Exon-intron structure research EBI : http://www.ebi.ac.uk/Wise2/
Expression profile Genevestigator : https://www.genevestigator.ethz.ch/

5.4. Protein analyses.

Sub-cellular localization prediction WOLF-PSORT : http://wolfpsort.org/
Motif database InterproScan : http://www.ebi.ac.uk/InterProScan/
Family and domain database Pfam : http://www.sanger.ac.uk/Software/Pfam/

5.5 Phylogenetic analyses.

Alignments were performed with Jalview v.2 multiple alignment editor (Clamp *et al*, 2004) using the Muscle v.3.0 program (Edgar, 2004). Phylogenetic trees were performed by PhyML method (http://atgc.lirmm.fr/phyml/) and designed with Treedyn tree editor (Chevenet *et al*, 2006).

5.6. Image processing.

ImageJ v.1.37 and Adobe Photoshop CS2 software were used for confocal images processing.

II. METHODS.

1. Arabidopsis mitochondria purification.

Protoplasts were prepared from 5 days old suspension-cultured Arabidopsis cells. Typically, 60 g of cells (FW) were collected from 400 mL of culture and incubated 3 h at 25°C under 45 rpm agitation in 250 mL of digestion buffer (0.4 M mannitol, 3.5 mM MES-KOH, pH 5.7) containing 1% (w/v) cellulase RS (Yakult, Tokyo), 0.1% (w/v) pectolyase Y-23 (Seishin Pharmaceutical, Tokyo). Protoplasts were harvested by centrifugation at 800xg for 10 min at 4°C, washed three times with an equal volume of digestion buffer without the enzymes. Protoplasts were resuspended in 50 ml of extraction buffer (0.4 M saccharose, 30 mM potassium phosphate, pH 7.5, 2 mM EDTA, 2 mM DTT and 0.1% (w/v) BSA) and ruptured by forcing them successively through 90-µm, 45-µm and 25-µm nylon meshes. Cell debris and nuclei and chloroplasts were sedimented by centrifugation at 2000xg for 10 min at 4°C. The supernatant was centrifuged at 16.000xg for 10 min at 4°C to pellet mitochondria, which were then resuspended in a small amount of cold washing buffer (300 mM saccharose, 30 mM potassium phosphate, pH 7.5, 1 mM EDTA and 0.1% (w/v) BSA) with a pre-cold Dounce homogenizer, on ice. The motichondria-enriched fraction was loaded onto a 18%/23%/40% (v/v) Percoll step gradient and centrifuged at 20.000xg for 15 min, at 4°C, without break. Intact mitochondria were collected at the 18%/23% interface, washed three times by successive pellet dilution in washing buffer and centrifugation at 20.000xg for 10 min.

2. Nucleic acid analyses.

2.1. Cloning techniques.

2.1.1. Plasmid DNA isolation.

Plasmid DNA was isolated from bacteria using a “mini-prep” DNA Quick Pure kit (MachereyNagel) for 0.7-2mL culture, and a “midi-prep” DNA NucleoBond PC kit (MachereyNagel) for 50mL culture, and according to the manufacturer’s protocol.

2.1.2. DNA amplification by PCR (Polymerase Chain Reaction).

A 50 µL reaction medium contains :

-1 to 100 ng template DNA.
-50 pmoles of each oligodesoxyribonucleotide primer
-200 µM of each dNTP
-2.5 mM of MgCl2
-5 µL of 10x PCR buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl; Invitrogen)
-2.5 U of Taq B DNA polymerase (Invitrogen)
-water to 50 µL

A basic PCR was (i) an initial denaturation step of 1 min at 95°C, (ii) a denaturation step of 30 sec at 95°C, (iii) an annealing step of 30 sec at 50°C, (iv) an elongation step of 30 sec at 72°C (steps (ii) to (iv) were repeated 30 times). A final elongation step of 1 min 30 sec at 72°C was performed. Cycles conditions were optimized for each couple of primers (annealing temperature) and length of the desired products (elongation time). PCRs were performed in a thermal cycler (Gene Cycle, Biorad, T3000 thermocycler, Biometra or Robocycler gradient 96, Stratagene).

2.1.3. Analysis and purification of DNA fragments on agarose gels.

DNA fragments were separated by electrophoresis in horizontal 1% agarose gels. The gels were prepared by boiling 1g of agarose with 100mL of 0.5X TAE (Tris Acetate EDTA) buffer (Euromedex) and adding 10µL (10ng/mL) of ethidium bromide. DNA samples were prepared by adding 5 µL of DNA loading buffer (50% (v/v) glycerol, 1% (w/v) SDS, 1 mM EDTA, 0.1% (w/v) cyanol xylen, 0.1% (w/v) bromophenol blue; Fermentas) and loaded. The gel was run in 0.5X TAE buffer at 100V. Analytical gels were photographed on a short wavelength UV trans-illuminator (gelDoc 1000, BioRad). For isolation of DNA from agarose gels, DNA was visualized on a long wavelength UV trans-illuminator (TM40, UVP) and the desired fragment was excised using a razor blade. Fragments were purified using a DNA extraction kit (Nucleospin Extract, Macherey Nagel).

2.1.4. Sub-cloning in pCRII-TOPO vector.

Cloning in pCRII-TOPO vector was made according to the manufacturer's instructions (Invitrogen).

2.1.5. DNA digestion by restriction enzymes.

1-2 µg of DNA was generally digested in a 20µL reaction containing 1-5µL of DNA, 2µL of 10X buffer -adapted to the enzyme-, 10U of restriction enzyme and water to 20µL. Reactions were generally carried out 1-2 hrs at 37°C and stopped by a 10 min incubation at 65°C if necessary.

DNA was purified by phenol/chloroform extraction (addition of 1 volume). Following 20.000xg centrifugation for 5 min at 4°C, aqueous phase is taken and mixed with 0.1 volume 3M sodium acetate, pH 5.2 and 2.5 volume 100% EtOH, and placed at -20°C for at least 20 min for precipitation. DNA is pelleted by 15 min centrifugation at 20.000xg at 4°C and washed by repeating centrifugation of the pellet in 1mL 70% EtOH. Pellet is dissolved in water.

2.1.6. Ligation.

Ligation of purified linearized vector (100ng) to DNA fragment/insert (in a 1:1 or 1:3 stoichiometric ratio), were performed with the Rapid Ligation kit (Fermentas) according to the manufacturer's instructions, except that the ligation was performed for 1 h at room temperature. T4 DNA ligase was deactivated by incubation ligation reaction tube at 70°C for 2 min. Ligation products were possibly ex-

tracted with phenol/chloroform, precipitated with ethanol and resuspended in 10µL of water or 0.8-5µl of the ligation were directly used for bacteria transformation.

2.1.7. Bacterial transformation.

2.1.7.1. Heat-shock transformation of *E. coli* cells.

25-50µL of competent cells were mixed with 0.8-5µL ligation mix (or 1-10ng of plasmid DNA) and incubated on ice for 20 min. The cells were heat-shocked at 42°C for 30 seconds and chilled on ice for 2 minutes. The cells were then mixed with 400µL of SOC medium (tryptone (pancreatic digest of casein) 2% (w/v), yeast extract 0.5% (w/v), NaCl 8.6 mM, KCl 2.5 mM, MgSO4 20 mM, Glucose 20 mM) and incubated for 1hr at 37°C. 100-200µL of this transformation mix was spread onto LB-agar (LB medium supplemented with 1.5% (w/v) agar) plates supplemented with the appropriate antibiotic for selection of transformants. For α-complementation selection, 30 µL IPTG at 50 mg/mL and 30 µL X-Gal at 50 mg/mL are added. Following overnight culture at 37°C, colonies with recombinant plasmid are white.

2.1.7.2. Electroporation of *E. coli* and *Agrobacterium* cells.

Electroporation was performed using a cell-electroporator and voltage booster (Biorad), and pre-chilled cuvettes. 1µL of plasmid DNA (10ng-100ng) was added to 40µL of electro-competent cells. The cell-DNA mixture was carefully suspended between the electrodes in the cuvette, and the cuvette placed in the electroporator. Electroporation of *E. coli* cells was performed with the following settings: capacitance: 25µF, resistance: 200Ω, voltage: 2.5V. After electroporation, cells were mixed with 500µL of SOC and incubated at 37°C for 1 hr before spreading on LB-agar plates supplemented with appropriate antibiotics or substrates.

Agrobacterium cells were transformed following the same protocol except that electroporation settings were the following : 25µF, resistance: 400Ω, voltage: 2.5V. Transformed cells were selected on LB-agar plates supplemented with the appropriate antibiotics after 48 hrs at 28°C.

2.1.8. Screening by PCR directly on colony.

This technique allows verification of the presence of DNA insert of interest directly on the bacterial colony. Bacteria of a white colony are transfered in 30 µL water with a tooth-pick and resuspended. 1 µL of the solution is used as DNA template for a PCR as described in paragraph 2.1.2. The first step of 1 min denaturation at 95°C is sufficient for bacteria lysis and DNA freeing.

2.1.9 DNA Sequencing.

DNA sequencing is realized by M. Alioua (sequencing service of IBMP) with the "Applied Biosystems 373 DNA sequencer" (Perkin Elmer). The method used is based on the Sanger technique (Sanger *et al*, 1977).

2.2. Northern analyses.

2.2.1. RNA extraction.

100mg of plant tissue were ground under liquid nitrogen. The resulting powder was homogenized in 1mL of TRIZOL® Reagent (Invitrogen), a mono-phasic solution of phenol and guanidine isothiocyanate. 200µL of chloroform was added, the mixture thoroughly mixed and let at room temperature (RT) for 2 minutes (min). The aqueous and organic phase were separated by centrifugation at 4°C and at maximum speed (12.000xg) for 15 min. The upper aqueous phase (approximately 600µL) was transferred to a new tube and mixed with 500µL isopropanol. The mixture was incubated at RT for 5 minutes. A second centrifugation step (12.000g, 10 min, at 4°C) was carried out to collect the RNA precipitate. The supernatant was discarded and the pellet washed with 70% ethanol, then allowed to air-dry at RT for 5 min. The pellet was resuspended in 20-40µL sterile water by repeated pipetting. 220-300nm spectrum was recorded with a spectrophometer (Genesys2, Spectronic) at a 1/100 dilution in sterile water.

2.2.2. High molecular weight (HMW) RNA blots.

A denaturing agarose gel (1.5%) was prepared by melting 1.5g of agarose in 77mL of deionised water. 5mL of 20X MOPS buffer (200mM MOPS pH 7; 50mM $NaCH_3COO$ EDTA 10mM) added and the solution allowed to cool down at RT. 18mL of 2.2M formaldehyde was added, the solution mixed and poured to set the gel. About 3 volumes of RNA loading buffer (for 200µL : 42.5µL of water, 10.5µL of 20X MOPS, 105µL of de-ionized formamide, 42.5µL of formaldehyde) and one volume of STOP-Blue (15% (v/v) glycerol, de-ionized formamide, 0.1% (w/v) cyanol xylen, 0.1% (w/v) bromophenol blue) were added to one volume of each RNA sample to test (5-10µg of total RNAs). The samples were denatured at 70°C for 10 min and chilled on ice for 5 min. After 5 min pre-run of the gel at 70V, samples were loaded. The gel was run at 50V until bromophenol blue was at 1cm to the bottom of the gel (5-6 hours) in 1X MOPS buffer. The gel was washed three times 10 min in RNase-free water and twice 30 min in 20X SSC (NaCl 3M, trisodic citrate 0.3M, pH7). RNAs were transferred by capillarity, overnight (o/n) in 20X SSC, on a neutral nylon membrane Hybond™-NX (Amersham). The membrane was rinsed in 2X SSC, put on a Whatmann paper soaked with 2X SSC and UV cross-linked in Stratalinker® (120 kilojoules). The membrane was then rinsed in RNase-free water, briefly incubated in methylen blue 0.3M Na acetate, pH5.2 and rinced again in RNase-free water. The membrane is destained in 100mL 0.1X SCC, 0.1% SDS and 5mL 20% SDS after scanning.

2.2.3. Low molecular weigh (LMW) RNA blots.

For 6% denaturing polyacrylamide gels preparation, 12mL of 7M urea, 1X TBE (Tris Borate EDTA), 6% acrylamide/bis-acrylamide 19:1 (Euromedex) were poured between glass slides. 100µL of 10% fresh APS and 10µL of TEMED were added just before pouring the mixture. Samples of 5-20µg RNA were mixed to 5-

15µL loading buffer (95% (v/v) formamide, 20 mM EDTA, xylen cyanol 0.05% (p/v), bromophenol blue 0.05% (p/v)) denatured at 70˚C for 4 min and put on ice for 5-10 min, then. Wells were washed with 1X TBE buffer and RNA samples loaded after a 30 min pre-run of the gel at 13mA. The gel was run in 1X TBE buffer at 20mA for 2-3 hrs until the blue dye gets out of the gel. Equal loading was checked by observation under UV light ethidium bromide staining. RNAs were transfered on a nylon membrane Hybond™-N+ (Amersham) by a semi-dry method using 2X 3 Whatmann papers saturated with 0.5X TBE buffer in a V10-SDB semi-dry blotter for 1 hr at 80mA (0.8mA/cm^2) at RT. The membrane was then UV cross-linked in Stratalinker® (120 kilojoules).

2.2.4. Hybridization of RNA blots.

2.2.4.1. Probe preparation.

RNA blots were hybridized with end-labelled oligonucleotide DNA probes, or with *in vitro* transcribed radiolabelled RNA probes :

· End-labelling reaction

3µL of oligonucleotide at 1µM (3pmol) were mixed with 4.5µL of water and denatured 3 min at 70˚C and then incubated on ice. 1.5µL of 10X T4-Polynucleotide Kinase (PNK) buffer (Fermentas), 5µL of [^{32}P]-γ-dATP (50µCi specific activity) and 1µL of T4-PNK (10U/µL) were added to a final volume of 15µL. The reaction was incubated at 37˚C for 45 min. Non-incorporated nucleotides were removed by loading the reaction supplemented with water to 50µL onto a G50 sephadex column (prepared by centrifugation of 120µL G-50 sephadex at 1.000rpm for 30 sec in an eppendorf tube) and performing a centrifugation step at 1.000rpm for 30 sec. The purified probe was denatured at 100˚C for 5 min and incubated on ice for 5 min before hybridization.

· *In vitro* transcription of radiolabelled RNA probes

50-100ng DNA template containing the sequence of the T7 promoter are supplemented with RNase-free water to 7µL and mixed with 4µL 5mM rNTPs (ATP, CTP, GTP) and 4µL 5X transcription buffer (Promega : 400mM HEPES-KOH (pH 7.5), 120mM $MgCl_2$, 10mM spermidine, 200mM DTT), 4µL [α^{32}P]-UTP (40µCi) and 1µL T7 enzyme mix (Promega). The reaction was incubated at 37˚C for 45 min. 1µL DNAse RQ1 was added and the reaction was further incubated at 37˚C for 15 min. Non-incorporated nucleotides were removed by G50 filtration as described before.

2.2.4.2. Hybridization.

The membranes were pre-hybridized for at least 30 min in hybridization buffer (0.5M sodium phosphate, pH 7.2; 7%SDS). After probe addition, hybridizations were carried out with 3mL hybridization buffer o/n at 65˚C and 45˚C for RNA probes and oligonucleotide probes, respectively.

2.2.4.3. Washes.

- Membranes hybridized with oligonucleotide probes were washed twice for 15 min at 45°C in a low stringency washing buffer (2X SSC, 0.1% SDS) and, if necessary, another 15 min in a higher stringency buffer (0.5X SSC, 0.1% SDS).
- Membranes hybridized with RNA probes were washed twice for 10 min at 65°C in 2X SSC, 0.1% SDS, 10 min in 1X SSC, 0.1% SDS, 10 min in 0.5X SSC, 0.1% SDS, 10 min in 0.1X SSC, 0.1% SDS and 10 min in 1X SSC.

2.2.4.4. Signal revealing.

The washed membrane was exposed to a phosphorimager plate (Fuji) for developing by a phosphorimager system (Fuji Bas 1000) or to a X-ray film and enhancing screen for autoradiography. Films were developed in a X-OMAT 2000 processor (Kodack).

2.3. Run-on assay.

2.3.1. *In organello* run-on transcription.

Intact mitochondria (corresponding to 2 mg proteins) were added to a reaction mix containing 60 µL of R buffer (17 mM $MgCl_2$, 66 mM KCl, 23 mM HEPES pH 7.9), 10 µL of 10x nucleotide mix (0.3 mM UTP, 5 mM ATP, CTP and GTP each) and 8µL (80 µCi) of [α-^{32}P]UTP. The reaction mix was gently pipetted up and down six times to disrupt mitochondria. The solution was incubated for 8 min at 25°C and the reaction stopped by adding 10 µL of lysis buffer (50 mg/ml N-laurylsarcosine, 50 mM Tris–HCl pH 8.0, 25 mM EDTA pH 8.0). Run-on transcripts were extracted twice with phenol/chloroform/isoamyl alcohol (25:25:1), precipitated with NaAc and EtOH and purified on a G-50 sephadex column.

2.3.2. Nylon filter preparation.

DNA samples, i.e. PCR products were spotted manually on Hybond N+ nylon filters (Amersham). Each sample was spotted at three different dilutions, i.e. 500, 150, and 50ng of DNA per 100 µL spot. DNA spotted on the membrane were denatured 1 min on a Whatmann paper soaked with 0.5M NaOH, 1.5M NaCl. Denaturing solution was then neutralized 1 min on a Whatmann paper soaked with 1M Tris-HCl, 1.5M NaCl. The membrane was washed 30 sec in 2X SSC and UV cross-linked in Stratalinker® (120 kilojoules).

2.3.3. Hybridization and washes.

The membrane was pre-hybridized in hybridization buffer (as before) for 30 min and incubated o/n with run-on transcripts and 3mL hybridization buffer at 65°C. After hybridization, membrane was rinsed in 2X SSC, 0.1%SDS and washed 15 min in 1X SSC, 0.1%SDS and twice 10 min in 0.2X SSC, 0.1%SDS.

2.3.4. Quantification.

The autoradiogram signals were quantified with a phosphorimager using a video densitometric system. Each experiment was repeated three times with independently labeled RNA obtained from separate preparations of mitochondria. The three experiments were averaged and standard deviations were calculated.

2.4. *In vitro* degradation assay.

^{32}P-[UDP] radiolabelled RNA probes were generated by *in vitro* transcription using T7 promoter-containing PCR products (Promega RiboMax™ Large Scale RNA production system-T7 kit) spanning position 47314 to 47677 and 47441 to 47677 of the mitochondrial genome for probes A and B, respectively. Mitochondrial protein were extracted from *Arabidopsis thaliana* mitochondria corresponding to 4mg proteins dissolved and carefully homogenized in 250µL extraction buffer (20mM MOPS pH 7.5, 0.1M DEPC-treated KCl, 0.5% Triton X100 and 10µL protease inhibitor cocktail (Roche)), incubated on ice 15 min and centrifuged 15 min at 20.000xg at 4°C. 65µL supernatant or 65µL non centrifuged extraction medium were supplemented with 1mM DTT, 1mM rATP, 1mM $MgCl_2$ and 2U RNAseOUT and incubated with 4µL A and B probes for 0, 30 and 60 minutes. Negative control was performed incubating probes with buffer without mitochondria. Resulting RNA species were extracted with 1 volume phenol/chloroform, separated by 6% acrylamide gel electrophoresis and detected by autoradiography.

2.5. *In vitro* RNA-dependent RNA polymerization assay.

^{32}P-[UDP] radiolabelled S1, S2 and S3 RNA substrates were generated by *in vitro* transcription of T7 promoter containing PCR products (Ambion Maxiscript kit) and purified by gel-excision and elution for 9 hrs at 33°C in 350µL elution buffer (0.5M NH_4OAc, 1mM EDTA, 0.1%SDS). These substrates were precipitated in 2.5 volume EtOH with 1µL glycogen and dissolved in water to have a 50ng/µL solution. 1, 2 or 3µL of each substrate were then incubated with 0.08pmol overexpressed RpoTm polymerase, 0.02pmol RpoTmp RNA polymerase, or 0.3U T7 polymerase (Ambion) for 50 min at 30°C in transcription buffer (10mM Tris-HCl pH7.9, 10mM KCl, 10mM $MgCl_2$, 1mM DTT and 0.1mg/mL BSA, 0.4mM ATP, 0.4mM CTP, 0.4mM GTP, 0.02mM UTP) supplemented with 1U RNasin. Resulting RNA products were extracted in 115µL extraction buffer (6M urea, 360mM NaCl, 20mM EDTA, 10mM Tris-HCl, pH 8, 1%SDS) supplemented with 20µL 3M NaOAc and 150µL phenol/chloroform. After precipitation with 375µL EtOH and 1µL glycogen, RNAs were separated by 10% acrylamide gel electrophoresis, along with unpurified and purified substrates alone and detected by autoradiography.

2.6. Ribonuclease Protection assay.

The ribonuclease protection assay was performed using the Ambion RPAIII kit, according to the manufacturing's instruction.

For the RPA, radiolabeled RNA probes A and B, complementary to 3' part of *NCO* transcript and downstream sequence (position 47407 to 47658 of Arabidopsis

mitochondrial genome), and solely downstream sequence (position 47571 to 47658), respectively, were synthesized *in vitro* (Promega RiboMax™ Large Scale RNA production system-T7 kit) and purified on gel. 1 µL of the pure labeled probes were incubated with RNA samples (from PNP- plants, from WT plants or with yeast RNA for negative control) under conditions that favor hybridization of complementary sequences. After o/n hybridization at 42°C, the mixture was treated with ribonuclease to degrade unhybridized probe. Labeled probe that was hybridized to complementary RNA from the sample was protected from ribonuclease digestion, and was separated on a long 6% polyacrylamide gel and visualized by autoradiography.

2.7. 3' Rapid Amplification of cDNA Ends (RACE), circular RT-PCR and SMART cDNA library construction.

Chapter in Methods in Enzymology

The following chapter on "Polyadenylation-mediated RNA degradation in plant mitochondria" is to be published in Methods in Enzymology.

3' Rapid Amplification of cDNA Ends (RACE), circular RT-PCR and SMART cDNA library construction techniques I used for the work presented in this manuscript are described in details and commented in this review.

The only difference is for the cDNA library construction : in the review, a variation of the SMART cDNA library is described where an oligo$(dT)_{12}$-anchor primer is used to synthesize polyadenylated cDNA in the place of the oligo(dT)-anchor primer provided with the SMART kit. For the polyadenylated cDNA library constructed described in this manuscript, the primer provided in the SMART kit was used.

Polyadenylation-mediated RNA degradation in plant mitochondria

Sarah Holec, Heike Lange, André Dietrich and Dominique Gagliardi*

Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Unité Propre de Recherche 2357, Conventionné avec l'Université Louis Pasteur, Strasbourg, France.

*Corresponding author. Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Unité Propre de Recherche 2357, 12 rue du général Zimmer, 67000 Strasbourg, France.

Tel: +33 3 88 41 71 62

FAX: + 33 3 88 61 44 42

Email: dominique.gagliardi@ibmp-ulp.u-strasbg.fr

Running title: RNA degradation in plant mitochondria

Polyadenylation-mediated RNA degradation in plant mitochondria

Sarah Holec, Heike Lange, André Dietrich, Dominique Gagliardi

in Methods in Enzymology. Vol. 447, RNA Turnover in Prokaryotes, Archae and Organelles (chap. 21). [SI] : Elsevier, 2008

FIGURE LEGENDS

Figure 1 : Mapping of polyadenylation sites by 3' RACE. RNA and DNA strands are shown as thin and bold lines, respectively. The anchor and gene-specific primers are represented by dark and light grey rectangles, respectively.

Figure 2 : Mapping of polyadenylation sites by cRT-PCR. RNA is shown as thin lines, cDNA as broken bold lines and other DNA as bold lines. Primer P1 is used for cDNA synthesis, primers P2 and P3 for PCR amplification. Positions corresponding to initial 5' and 3' extremities of the RNA are indicated throughout by dotted lines.

Figure 3 : cDNA library construction using the BD SMART™ PCR cDNA synthesis kit (BD Biosciences Clontech). RNA and DNA strands are shown as thin and bold lines, respectively. The anchor and SMART primers are represented by dark and light grey rectangles, respectively. M, RNA size marker ; MMLV-RT, Moloney Murine Leukemia Virus - Reverse Transcriptase.

Figure 4 : Flow-chart of the mitochondrial lysate preparation for the *in vitro* RNA degradation assay.

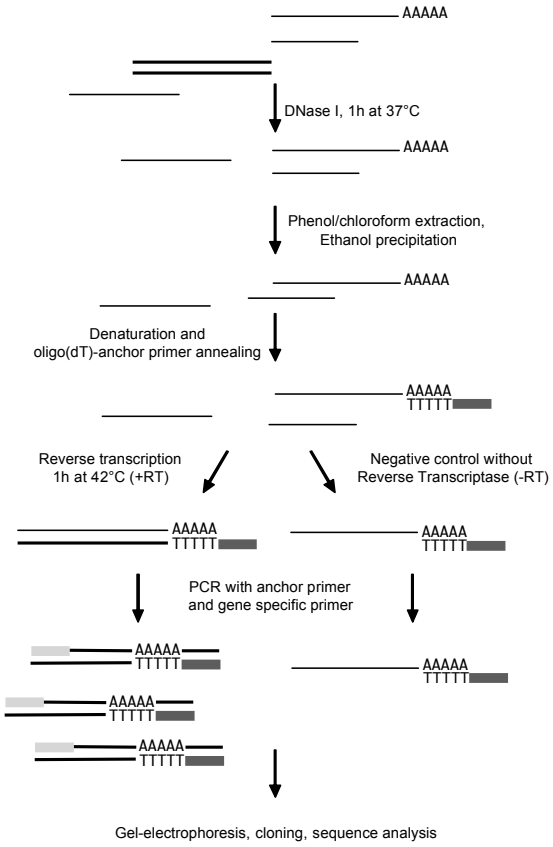

AAAAA
DNase I, 1h at 37°C
AAAAA
Phenol/chloroform extraction,
Ethanol precipitation
AAAAA
Denaturation and
oligo(dT)-anchor primer annealing
AAAAA
TTTTT
Reverse transcription
1h at 42°C (+RT)
Negative control without
Reverse Transcriptase (-RT)
AAAAA
TTTTT
AAAAA
TTTTT
PCR with anchor primer
and gene specific primer
AAAAA
TTTTT
AAAAA
TTTTT
AAAAA
TTTTT
AAAAA
TTTTT
Gel-electrophoresis, cloning, sequence analysis

Figure 1

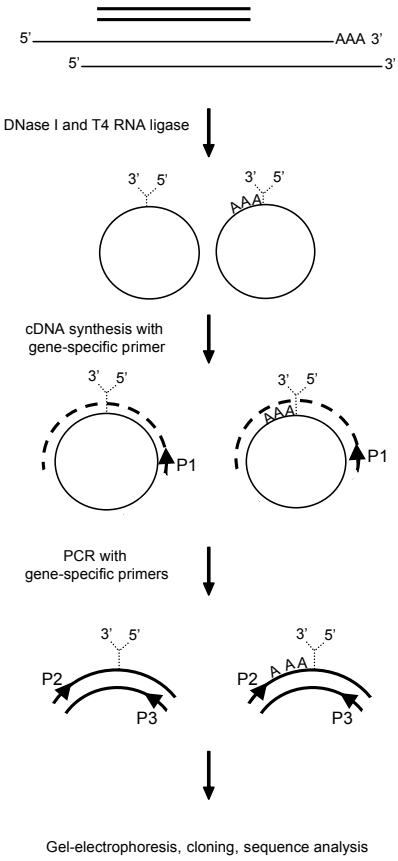

5'
AAA 3'
5'
3'
DNase I and T4 RNA ligase
3' 5'
3' 5'
AAA
cDNA synthesis with
gene-specific primer
3' 5'
P1
3' 5'
AAA
P1
PCR with
gene-specific primers
3' 5'
P2
P3
3' 5'
P2
AAA
P3
Gel-electrophoresis, cloning, sequence analysis

Figure 2

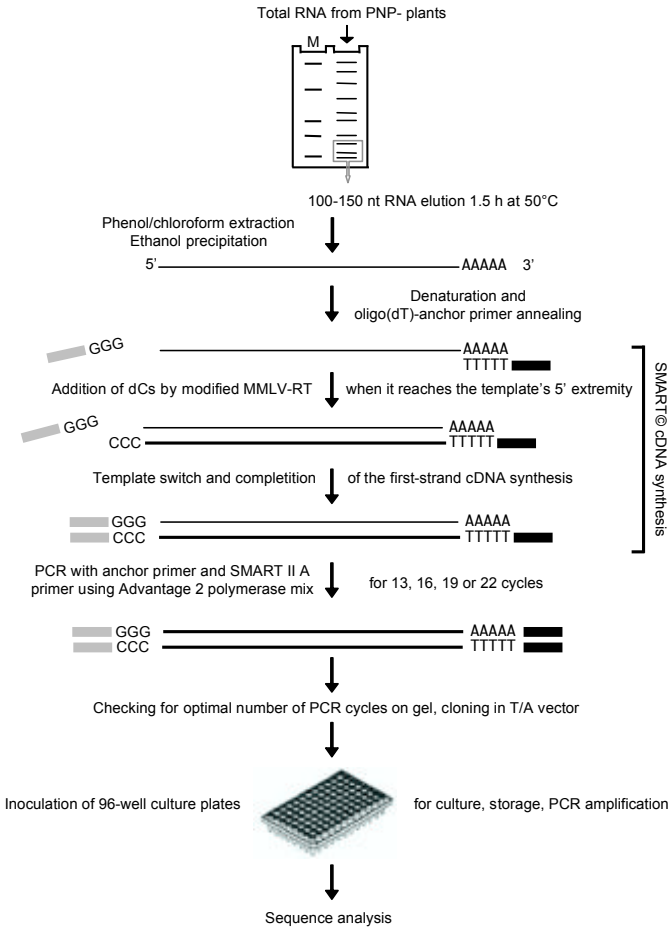

Total RNA from PNP- plants
M
100-150 nt RNA elution 1.5 h at 50°C
Phenol/chloroform extraction
Ethanol precipitation
5'
AAAAA 3'
Denaturation and
oligo(dT)-anchor primer annealing
GGG
AAAAA
TTTTT
Addition of dCs by modified MMLV-RT
when it reaches the template's 5' extremity
CCC
Template switch and completition
of the first-strand cDNA synthesis
SMART© cDNA synthesis
PCR with anchor primer and SMART II A
primer using Advantage 2 polymerase mix
for 13, 16, 19 or 22 cycles
Checking for optimal number of PCR cycles on gel, cloning in T/A vector
Inoculation of 96-well culture plates
for culture, storage, PCR amplification
Sequence analysis

Figure 3

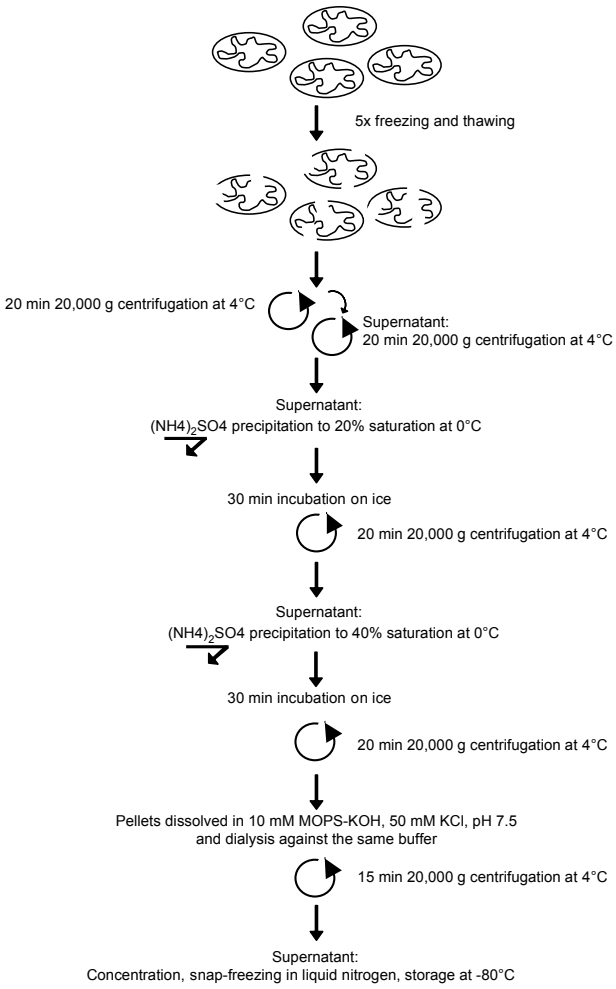

5x freezing and thawing
20 min 20,000 g centrifugation at 4°C
Supernatant:
20 min 20,000 g centrifugation at 4°C
Supernatant:
(NH4)2SO4 precipitation to 20% saturation at 0°C
30 min incubation on ice
20 min 20,000 g centrifugation at 4°C
Supernatant:
(NH4)2SO4 precipitation to 40% saturation at 0°C
30 min incubation on ice
20 min 20,000 g centrifugation at 4°C
Pellets dissolved in 10 mM MOPS-KOH, 50 mM KCl, pH 7.5
and dialysis against the same buffer
15 min 20,000 g centrifugation at 4°C
Supernatant:
Concentration, snap-freezing in liquid nitrogen, storage at -80°C

Figure 4

3. Protein analyses.

3.1. Protein extraction.

Plant tissue was ground in liquid nitrogen with mortar and pestle and 100-200mg were mixed with 2 volumes of 2X denaturing buffer (8M urea, 160mM Tris-HCl pH 6.8, 200mM DTT, 20% glycerol, 4% SDS). The mixture was vortexed, denatured at 95°C for 5 min, then centrifuged at 13.000xg to recover the supernatant.

3.2. Western blot analysis.

3.2.1. Gel preparation.

Stacking and resolving denaturing gels were prepared as follows :

12% acrylamide resolving gel :	40% acryl/bis-acryl 37.5:1	3mL
	1.5M TrisHCl pH 8.8	2.5mL
	H_2O	4.5mL
	APS 10%	130µL
	Temed	8.5µL
4% acrylamide stacking gel :	40% acryl/bis-acryl 37.5:1	0.4mL
	1.5M TrisHCl pH 6.8	1mL
	H_2O	2.6mL
	APS 10%	70µL
	Temed	3.5µL

3.2.2. Gel migration and protein transfer.

5-20µL of protein extracts were loaded after denaturation, along with a protein ladder (Euromedex). The gel was run at 10mA for about 2.5 hrs in a 1X Tris-Glycine (TG)/SDS 0.1% "running" buffer (25mM Tris, 250mM glycine pH 8.3, 0.1%SDS). Proteins were transfered in 1X TG "transfer" buffer (480mM Tris, 390mM glycine, 0.375% SDS) on a Immobilon™-P polyvinylidene fluoride (PVDF) microporous membrane (Millipore) -that was prewet in methanol and then water and then transfer buffer - by a semi-dry method : gel and membrane were put between 2X 3 Whatmann papers saturated with transfer buffer in a V10-SDB semi-dry blotter for 1 hr at 80mA (0.8mA/cm^2) at RT.

3.2.3. Incubation with antibodies.

The membrane was blocked with 1X Tris Buffered Saline (TBS) (20X TBS : 500mM TrisHcL, 3M NaCl pH 7.6) supplemented with 5% milk and 0.2% Tween 20 for 1 hr at RT or o/n at 4°C. It was then incubated with the primary antibody (in general diluted at 1/10000) for 1 hr at RT and washed 5 times for 5 min in 1X TBS/0.2% Tween 20. It was then incubated with the secondary antibody (in general diluted at 1/10000) coupled with HRP (horseradish peroxidase, Santa-Cruz) for 1hr at RT and washed again as described above. Proteins of interest were re-

vealed by chemiluminescence with ECL Western Blotting System (Amersham) by the exposition of a film. After revelation, the membranes were stained in a Coomassie blue solution (10% acetate, 25% ethanol, 0.5g Coomassie blue) during 1 min and subsequently washed in 10% acetate with 25% ethanol until correct destaining of the membrane to estimate the loading of total proteins.

4. Plant transformation.

4.1. Arabidopsis transformation using the floral dip method.

This protocol was as previously described (Clough and Bent, 1998) with some slight modifications. *Agrobacterium* cells were grown overnight at 28°C in 10mL of LB medium with the appropriate antibiotics. 2 mL of overnight culture were inoculated into 250mL of LB medium with appropriate antibiotics for overnight incubation at 28°C. *Agrobacterium* cells were then collected by centrifugation at 5000 rpm for 15 min and resuspended in an equal volume of the following infiltration medium:

	per 1 liter:
½ Murashige and Skoog (MS) 255 medium	2.165g
5% sucrose	50g
Silvet L-77	500µl
Acetosyringone 200mM	1 ml

Flowering Arabidopsis plants were dipped into the mixture for 90 sec. High humidity was maintained during 48 hours.

Seeds were sterilized by overnight incubation in a vacuum vessel with 3mL HCl added to 100mL of a solution of 1 volume sodium hypochlorite and 1 volume water. Seeds were sown on MS-agar containing specific antibiotic for transgene selection, and vernalized for 3 days at 4°C before grown for 3 wks *in vitro* in a growth chamber. Selected plants were bedded out to obtain descendants.

4.2. Nicotiana tabacum agro-infiltration.

Agro-infiltrations were carried out in *Nicotiana benthamiana* plants. *Agrobacterium* cultures were grown to an OD_{600}=1, pelleted by a 5 min centrifugation step at 4000 rpm. Cells were resuspended in equal volume of 10mM magnesium chloride (MgCl2) and incubated for 3 hrs with 200µM of acetosyringone to induce the virulence gene (Vir) expression. *Nicotiana benthamiana* leaves were wounded and agro-infiltrated using a syringe without needle. GFP expression was monitored 2 to 4 days post-infiltration (dpi) under UV light.

4.3. BY2 cells particle bombardment.

Biolostic transformation of BY2 cells allows to introduce exogenous DNA in plant cell nucleus via tungsten particles thanks to a compressed air particle delivery system. The expression of plasmidic constructs allows the analysis of fluorescent protein localization.

4.3.1. Particle preparation.

For each trial, two firings of 0.5 mg 1.1mm tungsten particles (Bio-Rad) coated with DNA are performed. Appropriate amount of particles are sterilized in 1 mL absolute alcohol for 20 min and resuspended in 50% glycerol. For each trial, 33µL particles-glycerol were mixed with 10 µg plasmid DNA (1µg/µL) supplemented with 33µL 2M $CaCl_2$ and 15µL 0.5M spermidine. Coated particles are successively washed with 70%EtOH and 100%EtOH and resuspended in absolute EtOH (6 µL/firing).

4.3.2. Cells preparation.

3 days old BY2 cells are used for biolistic transient espression. 5 mL harvested cells were filtered onto Whatmann discs and placed on 0.5% agar Murashige-Skoog (MS) media plates. The firing distance was 11 cm and helium pressure 7 bars. After bombardment, cells are incubated in the dark for 3-6hrs.

5. GFP observation.

Roots of *in vitro* selected Arabidopsis plantlet transformed via agrobacteria, discs of inflitrated Nicotiana benthamiana leaves or biolistically transformed BY2 are put between slide and cover-slip for confocal laser scanning microscopy observations. Observation are made with a LSM 510-405 microscope equipped with a X63 water corrected objective (numerical aperture 1.2) and a Zeiss AIM V.4.1 software package (Zeiss, Germany). Excitation/emission wavelenght were 488nm/505-550nm for eGFP fluorochrome. Image processing was carried out using ImageJ and Photoshop v.7 softwares.

BIBLIOGRAPHY

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content : gene loss and transfer to the nucleus. *Molecular Phylogenetics and evolution* **29** : 380-395

Allmang, C, Kufel, J, Chanfreau, G, Mitchell, P, Petfalski, E and Tollervey, D. (1999a) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *Embo J*, **18**: 5399-5410.

Allmang, C, Petfalski, E, Podtelejnikov, A, Mann, M, Tollervey, D and Mitchell, P. (1999b) The yeast exosome and human PM-Scl are related complexes of 3' to 5' exonucleases. *Genes Dev*, **13**: 2148-2158.

André C, Levy A and Walbot V (1992) Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet*, **8** :128-132

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

Aloy P, Ciccarelli FD, Leutwein C, Gavin AC, Superti-Furga G, Bork P, Bottcher B, Russell RB. (2002) A complex prediction: three-dimensional model of the yeast exosome. *EMBO Rep.* **3** : 628-635.

Araki Y, Takahashi S, Kobayashi T, Kajiho H, Hoshino S, Katada T. (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast. *EMBO J.* **20** : 4684-4693.

Backert S, Nielsen BL, Börner T (1997) The mystery of the rings : structure and replication of the mitochondrial genomes from higher plants. *Trends in Plant Science,* **2** : 477-483

Backert S, Börner T (2000) Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant Chenopodium album (L.) *Current Genetics* **37** : 304-14.

Baginsky S, Shteiman-Kotler A, Liveanu V, Yehudai-Resheff S, Bellaoui M, Settlage RE, Shabanowitz J, Hunt DF, Schuster G, Gruissem W. (2001) Chloroplast PNPase exists as a homo-multimer enzyme complex that is distinct from the Escherichia coli degradosome. *RNA.* **7** : 1464-1475.

Bellaoui M, Pelletier G, 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.

Beran RK, Simons RW. (2001) Cold-temperature induction of Escherichia coli polynucleotide phosphorylase occurs by reversal of its autoregulation. *Mol Microbiol.* **39** : 112-125.

Bernstein JA, Lin PH, Cohen SN, Lin-Chao S. (2004) Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci U S A.* **101** : 2758-2763.

Binder S, Brennicke A (1993a) Transcription initiation sites in mitochondria of Oenothera berteriana J Biol Chem **268** : 7849-7855

Binder S, Brennicke A (1993b) A tRNA gene transcription initiation site is similar to mRNA and rRNA promoters in plant mitochondria *Nucleic Acids Res.* **21** : 5012–5019.

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

Binder S, Brennicke A. (2003) Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci.* **358** : 181-189.

Bollenbach TJ, 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

Bollenbach TJ, Schuster G, Stern DB. (2004) Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. *Prog Nucleic Acid Res Mol Biol.* **78** : 305-337.

Bollenbach TJ, Stern DB. (2003) Secondary structures common to chloroplast mRNA 3'-untranslated regions direct cleavage by CSP41, an endoribonuclease belonging to the short chain dehydrogenase/reductase superfamily. *J Biol Chem.* **278** : 25832-25838.

Bonawitz N, Clayton D, Shadel G (2006) Initiation and Beyond: Multiple Functions of the Human Mitochondrial Transcription Machinery. *Molecular Cell* **24** : 813-825

Bousquet-Antonelli C, Presutti C, Tollervey D. (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell.* **102** : 765-775.

Brandt P, Sünkel S, Unseld M, Brennicke A, Knoop V. (1992) The nad4L gene is encoded between exon c of nad5 and orf25 in the Arabidopsis mitochondrial genome. *Molecular and General Genetics* **236** :33-8.

Brangeon J, Sabar M, Gutierres S, Combettes B, Bove J, Gendy C, Chetrit P, Des Francs-Small CC, Pla M, Vedel F, De Paepe R. (2000) Defective splicing of the first nad4 intron is associated with lack of several complex I subunits in the Nicotiana sylvestris NMS1 nuclear mutant. *Plant J.* **21** : 269-280.

Briggs MW, Burkard KT, Butler JS. (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J Biol Chem.* **273** : 13255-13263.

Brodersen P, Voinnet O. (2006) The diversity of RNA silencing pathways in plants. *Trends Genet.* **22** : 268-280.

Brown GG, Auchincloss AH, Covello PS, Gray MW, Menassa R, 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.

Brouwer, R, Allmang, C, Raijmakers, R, van Aarssen, Y, Egberts, WV, Petfalski, E, van Venrooij, WJ, Tollervey, D and Pruijn, GJ. (2001) Three novel components of the human exosome. *J Biol Chem*, **276**: 6177-6184.

Buchanan BB, Gruissem W, Jones RL (2000) Biochemitry & Molecular Biology of Plants. *American Society of Plant Physiologists*.

Bühler M, Haas W, Gygi SP, Moazed D. (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell.* **129** : 707-721.

Burger G, Gray MW, Lang BF (2003) Mitochondrial genomes : anything goes. *Trends in Genetics* **19** : 709-716

Butler JS. (2002) The yin and yang of the exosome. *Trends Cell Biol.* **12** : 90-96.

Butow RA, Zhu H, Perlman P, Conrad-Webb H. (1989) The role of a conserved dodecamer sequence in yeast mitochondrial gene expression. *Genome.* **31** : 757-760.

Büttner K, Wenig K, Hopfner KP. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol Cell.* **20** : 461-471.

Callaghan AJ, Marcaida MJ, Stead JA, McDowall KJ, Scott WG, Luisi BF. (2005) Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature.* **437** : 1187-1191

Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell.* **131** : 706-717.

Carpousis AJ. (2007) The RNA Degradosome of Escherichia coli: A Multiprotein mRNA-Degrading Machine Assembled on RNase E. *Annu Rev Microbiol.* **61** : 71-87

Cavalier-Smith T. (1991) Archamoebae: the ancestral eukaryotes? *Biosystems.* **25** : 25-38.

Celesnik H, Deana A, Belasco JG. (2007) Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal.*Mol Cell.* **27** : 79-90.

Chapdelaine Y, Bonen L. (1991) The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. *Cell.* **65** : 465-472.

Chapman EJ, Carrington JC. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet.* **8** : 884-896.

Chateigner-Boutin,A.L. and Hanson,M.R. (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared ciselements. *Mol. Cell. Biol.,* **22**, 8448–8456.

Chekanova JA, Dutko JA, Mian IS, Belostotsky DA.(2002) Arabidopsis thaliana exosome subunit AtRrp4p is a hydrolytic 3'-->5' exonuclease containing S1 and KH RNA-binding domains. *Nucleic Acids Res.* **30** : 695-700.

Chekanova JA, Shaw RJ, Wills MA, Belostotsky DA. (2000) Poly(A) tail-dependent exonuclease AtRrp41p from Arabidopsis thaliana rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells. *J Biol Chem.* **275** : 33158-33166.

Chen C., Deutscher M.P. (2005) Elevation of RNase R in response to multiple stress conditions. *J. Biol. Chem.* **280** : 34393-34396.

Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, Pruijn GJ, Stoecklin G, Moroni C, Mann M, Karin M. (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell.* **107** : 451-464.

Chen HW, Rainey RN, Balatoni CE, Dawson DW, Troke JJ, Wasiak S, Hong JS, McBride HM, Koehler CM, Teitell MA, French SW. (2006) Mammalian polynucleotide phosphorylase is an intermembrane space RNase that maintains mitochondrial homeostasis. *Mol Cell Biol.* **26** : 8475-8487.

Cheng ZF, Deutscher MP. (2005) An important role for RNase R in mRNA decay. *Mol Cell.* **17** : 313-318.

Chevenet F., Brun C., Banuls A.L., Jacq B. and R. Christen (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees *BMC Bioinformatics* **7** : 439.

Christianson T, Rabinowitz M. (1983) Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. *J Biol Chem.* **258** : 14025-14033.

Choi JM, Park EY, Kim JH, Chang SK, Cho Y. (2004) Probing the functional importance of the hexameric ring structure of RNase PH. *J Biol Chem.* **279** : 755-764.

Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004) The Jalview Java Alignment Editor. Bioinformatics. 20 : 426-427.

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

Condon C. (2006) Shutdown decay of mRNA. *Mol Microbiol.* **61** : 573-583.

Condon C. (2007) Maturation and degradation of RNA in bacteria. *Current Opinion in Microbiology.* **10** : 271-278.

Covello PS, Gray MW. (1989) RNA editing in plant mitochondria. *Nature.* **341** : 662-666.

Daley DO, Adams KL, Clifton R, Qualmann S, Millar AH, Palmer JD, Pratje E, Whelan J. (2002) Gene transfer from mitochondrion to nucleus: novel mechanisms for gene activation from Cox2. *Plant J.* **30** : 11-21.

Dez C, Houseley J, Tollervey D. (2006) Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of Saccharomyces cerevisiae. *EMBO J.* **25** : 1534-1546.

Ding SW, Voinnet O. (2007) Antiviral immunity directed by small RNAs. *Cell.* **130** : 413-426

Doma MK, Parker R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature.* **440** : 561-564.

Doma MK, Parker R. (2007) RNA quality control in eukaryotes.*Cell.* **131** : 660-668.

Dombrowski S, Brennicke A and Binder S (1997) 3'-inverted repeats in plant mitochondrial mRNAs are processing signal rather than transcription terminators *EBMO J* **16** : 5069-5076

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

Dong FG, Wilson KG, Makaroff CA. (1998) The radish (Raphanus sativus L.) mitochondrial cox2 gene contains an ACG at the predicted translation initiation site. *Curr Genet. 34 : 79-87.*

Drager RG, Higgs DC, Kindle KL, Stern DB. (1999) 5' to 3' exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. *Plant J.* **19** : 521-531.

Dziembowski A, Piwowarski J, Hoser R, Minczuk M, Dmochowska A, Siep M, van der Spek H, Grivell L, Stepien PP. (2003) The yeast mitochondrial degradosome. Its composition, interplay between RNA helicase and RNase activities and the role in mitochondrial RNA metabolism. *J Biol Chem.* **278** : 1603-1611.

Edgar R.C. (2004) MUSCLE : a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5** : 113.

Emanuel C, Weihe A, Graner A, Hess WR, Börner T (2004) Chloroplast development affects expression of phage-type RNA polymerases in barley leaves. *Plant Journal* **38** : 460-472

Embley TM, van der Giezen M, Horner DS, Dyal PL, Foster P. (2003) Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**:191-201

Falcon de Longevialle A, Meyer EH, Andrés C, Taylor NL, Lurin C, Millar AH, Small ID. (2007) The Pentatricopeptide Repeat Gene OTP43 Is Required for trans-Splicing of the Mitochondrial nad1 Intron 1 in Arabidopsis thaliana. *Plant Cell* .In press.

Farré JC, Araya A. (1999) The mat-r open reading frame is transcribed from a non-canonical promoter and contains an internal promoter to co-transcribe exons nad1e and nad5III in wheat mitochondria. *Plant Mol Biol.* **40** : 959-967.

Farré JC, Araya A. (2001) Gene expression in isolated plant mitochondria: high fidelity of transcription, splicing and editing of a transgene product in electroporated organelles. *Nucleic Acids Res.* **29** : 2484-2491.

Farré JC, Leon G, Jordana X, Araya A. (2001) cis Recognition elements in plant mitochondrion RNA editing. *Mol Cell Biol.* **21** : 6731-6737.

Fey J and Maréchal-Drouard L (1999) Compilation and analysis of plant mitochondrial promoter sequences : an illustration of divergent evolution between monocot and dicot mitochondria. *Biochem Biophys Res Commun,* **256** : 409-414

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

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

Fulnecek J, Kovarik A. (2007) Low abundant spacer 5S rRNA transcripts are frequently polyadenylated in Nicotiana. *Mol Genet Genomics.* **278**: 565-573.

Forner J, Weber B, Thuss S, Wildum S, Binder S. (2007) Mapping of mitochondrial mRNA termini in Arabidopsis thaliana: t-elements contribute to 5' and 3' end formation. *Nucleic Acids Res.* **35** : 3676-3692.

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

Gagliardi D. & Binder S. (2007) Expression of the plant mitochondrial genome, in Plant mitochondria, Editor : D. C Logan, Blackwell publishing (Oxford) *Annual Plant Reviews*, **31** : 50-96.

Gagliardi D, Leaver CJ. (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO J.* **18** : 3757-3766.

Gagliardi D, Perrin R, Marechal-Drouard L, Grienenberger JM, Leaver CJ. (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** : 43541-4347.

Gagliardi D, Stepien PP, Temperley RJ, Lightowlers RN, Chrzanowska-Lightowlers ZM. (2004) Messenger RNA stability in mitochondria: different means to an end. *Trends in Genetics.* **20** : 260-267.

Gatfield D, Izaurralde E. (2004) Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. *Nature.* **429** : 575-578.

Giegé P, Brennicke A. (1999) RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A.* **96** : 15324-15329.

Giegé P, Hoffmann M, Binder S and Brennicke A (2000) RNA degradation buffers asymmetries of transcription in *Arabidopsis* mitochondria. *EMBO Rep* 1 : 164-170

Graham, AC, Kiss, DL and Andrulis, ED. (2006) Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci. *Mol Biol Cell*, **17**: 1399-1409.

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science* **283**:1476–1481

Gualberto JM, Lamattina L, Bonnard G, Weil JH, Grienenberger JM. (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature.* **341** : 660-662.

Hanson MR, Bentolila S. (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell.* **16** : 154-169.

Haile S, Cristodero M, Clayton C, Estévez AM. (2007) The subcellular localisation of trypanosome RRP6 and its association with the exosome. *Mol Biochem Parasitol.* **151** : 52-58.

Hajnsdorf E, Braun F, Haugel-Nielsen J, Regnier P. (1995) Polyadenylylation destabilizes the rpsO mRNA of Escherichia coli. *Proc Natl Acad Sci U S A.* **92** : 3973-3977.

Hattori M, Miyake H, Sugita M. (2007) A Pentatricopeptide repeat protein is required for RNA processing of clpP Pre-mRNA in moss chloroplasts. *J Biol Chem.* **282** : 10773-10782.

Hatzack F, Dombrowski S, Brennicke A, Binder S (1998) Characterization of DNA-binding proteins from pea mitochondria. *Plant Physiol.* **116** : 519-528

Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J and Millar AH (2004) Experimental analysis of the *Arabidopsis* proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell*, **16** : 241-256

Hedtke B, Meixner M, Gillandt S, Richter E, Borner T, Weihe A. (1999) Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. *Plant J.* **17** : 557-561.

Hedtke B, Börner T, Weihe A. (2000) One RNA polymerase serving two genomes. *Embo Report* **1** : 435-40.

Hegeman C, Halter C, Owens T and Hanson M (2005) Expression of complementary RNA from chloroplast transgenes affects editing efficiency of transgene and endogenous chloroplast transcripts. *Nucleic Acids Research,* **33,** 1454–1464

Hicks A, Drager RG, Higgs DC, Stern DB. (2002) An mRNA 3' processing site targets downstream sequences for rapid degradation in Chlamydomonas chloroplasts. *J Biol Chem.* **277** : 3325-3333.

Higgs DC, Shapiro RS, Kindle KL, Stern DB. (1999) Small cis-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. *Mol Cell Biol.* **19** : 8479-8491

Hiesel R, Wissinger B, Schuster W, Brennicke A. (1989) RNA editing in plant mitochondria. *Science.* **246** : 1632-1634.

Hoffmann M, Dombrowski S, Guha C, Binder S. (1999) Cotranscription of the rpl5-rps14-cob gene cluster in pea mitochondria. *Mol Gen Genet.* **261** : 537-545.

Hoffmann M, Binder S. (2002) Functional importance of nucleotide identities within the pea atp9 mitochondrial promoter sequence. *J Mol Biol.* **320** : 943-950.

Holec S, Lange H, Kühn K, Alioua M, Börner T, 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.

Holec S, Gagliardi D. (2006) Importance de la surveillance des ARN dans les mitochondries d'*Arabidopsis thaliana Med Sci* **22** : 681-682.

Houseley J, Tollervey D. (2006) Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep.* **7** : 205-211.

Houseley J, Kotovic K, El Hage A, Tollervey D. (2007) Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO J.* **26** : 4996-5006.

Ibrahim F, Rohr J, Jeong WJ, Hesson J, Cerutti H. (2006) Untemplated oligoadenylation promotes degradation of RISC-cleaved transcripts. *Science.* **314** : 1893.

Ikeda TM, Gray MW (1999) Characterization of a DNA-binding protein implicated in trancription in wheat mitochondria *Mol Cel Biol* **19** : 8113-8122

Jiang T, Altman S. (2002) A protein subunit of human RNase P, Rpp14, and its interacting partner, OIP2, have 3'-->5' exoribonuclease activity. *Proc Natl Acad Sci U S A.* **99** : 5295-5300

Kao C-Y and Read L.K. (2005) Opposing Effects of Polyadenylation on the Stability of Edited and Unedited Mitochondrial RNAs in *Trypanosoma brucei Mol. Cel. Biol.* **25** : 1634-1644.

Kao CY, Read LK.(2007) Targeted depletion of a mitochondrial nucleotidyltransferase suggests the presence of multiple enzymes that polymerize mRNA 3' tails in Trypanosoma brucei mitochondria. *Mol Biochem Parasitol.* **154** : 158-169.

Klein M, Eckert-Ossenkopp U, Schmiedeberg I, Brandt P, Unseld M, Brennicke A, Schuster W. (1994) Physical mapping of the mitochondrial genome of Arabidopsis thaliana by cosmid and YAC clones. *Plant J.* **6** : 447-455.

Kobayashi Y, Matsuo M, Sakamoto K, Wakasugi T, Yamada K, Obokata J. (2007) Two RNA editing sites with cis-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. *Nucleic Acids Res.* In press.

Komine Y, Kikis E. A., Schuster G. and Stern D. B. (2002) In vivo modulation of chloroplast RNA stability by 3' UTR homopolymeric tails in Chlamydomonas reinhardtii *Proc. Natl. Acad. Sci. USA* **99**: 4085-4090.

Kotera E, Tasaka M, Shikanai T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature.* **433** : 326-330.

Koulintchenko M, Konstantinov Y, Dietrich A. (2003) Plant mitochondria actively import DNA via the permeability transition pore complex. *EMBO J.* **22** :1245-1254

Kuai,L., Fang,F., Butler,J.S. and Sherman,F. (2004) Polyadenylation of rRNA in Saccharomyces cerevisiae. *Proc. Natl Acad. Sci. USA.* **101** : 8581–8586.

Kuhn J, Tengler U, 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.

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

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

Kühn K, Bohne A-V, Liere K, Weihe A and Börner T (2007) *Arabidopsis* phage-type RNA polymerases : accurate *in Vitro* transcription of organellar genes. *The Plant Cell* **19** : 959-971

Kunzmann A, Brennicke A, 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.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell.* 121 : 713-724.

Leino M, Teixeira R, Landgren M, Glimelius K. (2003) Brassica napus lines with rearranged Arabidopsis mitochondria display CMS and a range of developmental aberrations. *Theor Appl Genet.***106** : 1156-1163.

Leino M, Thyselius S, Landgren M, Glimelius K. (2004) Arabidopsis thaliana chromosome III restores fertility in a cytoplasmic male-sterile Brassica napus line with A. thaliana mitochondrial DNA. *Theor Appl Genet.* **109** : 272-279.

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

Lelandais C, Gutierres S, Mathieu C, Vedel F, Remacle C, Maréchal-Drouard L, Brennicke A , Binder S, Chétrit P. (1996) A promoter element active in run-off transcription controls the expression of two cistrons of nad and rps genes in Nicotiana sylvestris mitochondria. *Nucleic Acids Research* **24** :4798-804.

Levinger L, Mörl M, Florentz C. (2004) Mitochondrial tRNA 3' end metabolism and human disease. *Nucleic Acids Res.* **32** : 5430-5441

Li H, Zassenhaus HP. (1999) Purification and characterization of an RNA dodecamer sequence binding protein from mitochondria of Saccharomyces cerevisiae. *Biochem Biophys Res Commun.* **261** : 740-745.

Li Z, Reimers S, Pandit S, Deutscher MP. (2002) RNA quality control: degradation of defective transfer RNA. *EMBO J.* **21** : 1132-1138.

Li-Pook-Than J, Carrillo C, Bonen L. (2004) Variation in mitochondrial transcript profiles of protein-coding genes during early germination and seedling development in wheat. *Curr Genet.* **46** : 374-380.

Lin-Chao S, Chiou NT, Schuster G. (2007) The PNPase, exosome and RNA helicases as the building components of evolutionarily-conserved RNA degradation machines. *J Biomed Sci.* **14** : 523-532.

Linder T, Park CB, Asin-Cayuela J, Pellegrini M, Larsson N-G, Falkenberg M, Samuelsson T, Gustavsson CM (2005) A family of putative transcription termination factors shared amongst metazoan and plants *Curr Genet* **48** : 265-269

Lisitsky I, Klaff P and Schuster G. (1996) Addition of destabilizing poly (A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA. *Proc Natl Acad Sci U S A.* **93** : 13398-133403.

Lisitsky I, Klaff P, and Schuster G. (1997). Blocking polyadenylation of mRNA in the chloroplast inhibits its degradation. *Plant J.* **12**: 1173–1178.

Liu Q, Greimann JC, Lima CD. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127** : 1223-1237

erratum in Liu Q, Greimann JC, Lima CD. (2007) *Cell* **131**, 188-189

Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat Struct Mol Biol.* **12** : 575-581.

Lorentzen E, Dziembowski A, Lindner D, Seraphin B, Conti E. (2007) RNA channelling by the archaeal exosome. *EMBO Rep.* **8** : 470-476.

Lu B, Hanson MR. (1996) Fully edited and partially edited nad9 transcripts differ in size and both are associated with polysomes in potato mitochondria. *Nucleic Acids Res.* **24** : 1369-1374.

Lupold DS, Caoile AG, Stern DB. (1999) Polyadenylation occurs at multiple sites in maize mitochondrial cox2 mRNA and is independent of editing status. *Plant Cell.* **11** : 1565-1578.

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

Mackenzie SA (2007) The unique biology of mitochondrial genome instability in plants. in Plant mitochondria, Editor : D. C Logan, Blackwell publishing (Oxford) *Annual Plant Reviews,* **31** :36-49

Makaroff CA, Apel IJ, Palmer JD. (1989) The atp6 coding region has been disrupted and a novel reading frame generated in the mitochondrial genome of cytoplasmic male-sterile radish. *J Biol Chem.* **264** : 11706-11713.

Manchekar M, Scissum-Gunna K, Songb D, Khazia F,McLeana S.L.,Nielsena B.L. (2006) DNA Recombination Activity in Soybean Mitochondria *Journal of Molecular Biology*, **356** : 288-299

Marchfelder A, Brennicke A, Binder S. (1996) RNA editing is required for efficient excision of tRNA(Phe) from precursors in plant mitochondria. *J Biol Chem.* **271** : 1898-1903.

Marchfelder A, Schuster W, Brennicke A.(1990) In vitro processing of mitochondrial and plastid derived tRNA precursors in a plant mitochondrial extract. *Nucleic Acids Res.* **18** : 1401-1406.

Maréchal-Drouard L, Kumar R, Remacle C, Small I. (1996) RNA editing of larch mitochondrial tRNA(His) precursors is a prerequisite for processing. *Nucleic Acids Res.* **24** : 3229-3234.

Marienfeld J, Unseld M, Brennike A (1999) The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. *Trends in Plant Science* **4** : 495:502

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

Martin G, Keller W (2004) Sequence motifs that distinguish ATP(CTP):tRNA nucleotidyl transferases from eubacterial poly(A) polymerases. *RNA.* **10** : 899-906.

Marujo PE, Hajnsdorf E, Le Derout J, Andrade R, Arraiano CM, Regnier P. (2000) RNase II removes the oligo(A) tails that destabilize the rpsO mRNA of Escherichia coli. *RNA.* **6** : 1185-1193.

Mathy N, Benard L, Pellegrini O, Daou R, Wen T, Condon C. (2007) 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell*.**129** : 681-692.

Mattick JS, Makunin IV. (2006) Non-coding RNA. *Hum Mol Genet*. **15** Spec No 1:R17-29.

McCulloch V, Seidel-Rogol BL, Shadel GS. (2002) A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. *Mol Cell Biol.* **22** : 1116-1125.

McKhann HI, Camilleri C, Bérard A, Bataillon T, David JL, Reboud X, Le Corre V, Caloustian C, Gut IG, Brunel D. (2004) Nested core collections maximizing genetic diversity in Arabidopsis thaliana. *Plant J.* **38** : 193-202.

Menassa R, L'Homme Y, Brown GG. (1999) Post-transcriptional and developmental regulation of a CMS-associated mitochondrial gene region by a nuclear restorer gene. *Plant J.* **17** : 491-499.

Militello KT, Read LK. (1999) Coordination of kRNA editing and polyadenylation in Trypanosoma brucei mitochondria: complete editing is not required for long poly(A) tract addition. *Nucleic Acids Res* **27** : 1377-1385.

Millar A.H., Heazlewood J.L., Kristensen B.K., Braun H.-P., Moller I.M. (2005) The plant mitochondrial proteome. *Trends in Plant Science*, **10** : 36-43.

Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell.* **91** : 457-466.

Mitchell, P, Petfalski, E, Houalla, R, Podtelejnikov, A, Mann, M and Tollervey, D. (2003) Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol Cell Biol*, **23**: 6982-6992.

Mitchell P, Tollervey D. (2003) An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. *Mol Cell.* **11** : 1405-1413.

Mohanty BK, Kushner SR. (2000) Polynucleotide phosphorylase functions both as a 3'-5' exonuclease and a poly(A) polymerase in Escherichia coli. *Proc Natl Acad Sci U S A.* **97** : 11966-11971.

Mohanty BK, Kushner SR. (2003) Genomic analysis in Escherichia coli demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay. *Mol Microbiol.* **50** : 645-658.

Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijn G, Lutz CS, Wilusz J. (2002) The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J.* **21** : 165-174.

Mulligan RM, Lau GT, Walbot V (1988) Numerous transcription initiation sites exist for the maize mitochondrial genes for subunit 9 of the ATP synthase and subunit 3 of cytochrome oxidase. *Proc. Natl. Acad. Sci USA* **85** :7998-8002

Mulligan RM, Leon P, Walbot V. (1991) Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression. *Mol Cell Biol.* **11** : 533-543.

Murakami H, Goto DB, Toda T, Chen ES, Grewal SI, Martienssen RA, Yanagida M. (2007) Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS ONE*. **2** : e317

Nakagawa N, Sakurai N. (2006) A mutation in At-nMat1a, which encodes a nuclear gene having high similarity to group II intron maturase, causes impaired splicing of mitochondrial NAD4 transcript and altered carbon metabolism in Arabidopsis thaliana. *Plant Cell Physiol.* **47** : 772-783.

Nicholson AW. (1999) Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol Rev.* **23** : 371-390.

Nishimura Y, Kikis EA, Zimmer SL, Komine Y, Stern DB. (2004) Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in chlamydomonas chloroplasts. *Plant Cell.* **16** : 2849-2869.

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

O'Hara EB, Chekanova JA, Ingle CA, Kushner ZR, Peters E, Kushner SR. (1995) Polyadenylylation helps regulate mRNA decay in Escherichia coli. *Proc Natl Acad Sci U S A.* **92** : 1807-1811.

Okuda K, Myouga F, Motohashi R, Shinozaki K, Shikanai T. (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A.* **104** : 8178-8183.

Orban TI, Izaurralde E. (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA*. **11** : 459-469.

Osaki H, Nakamura H, Nomura K, Matsumoto N, Yoshida K. (2005) Nucleotide sequence of a mitochondrial RNA virus from the plant pathogenic fungus, Helicobasidium mompa Tanaka. *Virus Res.* **107** : 39-46.

Ow MC, Liu Q, Kushner SR. (2000) Analysis of mRNA decay and rRNA processing in Escherichia coli in the absence of RNase E-based degradosome assembly. *Mol Microbiol.* **38** : 854-866.

Palmer JD and Herbon LA (1987). Unicircular structure of the *Brassica hirta* mitochondrial genome. Curr Genet **11**:565–570.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci U S A*. **97** : 6960-6966.

Pang KC, Frith MC, Mattick JS. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* **22** :1-5.

Parker R, Song H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol.* **11** : 121-127.

Perrin R, Meyer EH, Zaepfel M, Kim YJ, Mache R, Grienenberger JM, Gualberto JM, Gagliardi D. (2004a) Two exoribonucleases act sequentially to process mature 3'-ends of atp9 mRNAs in Arabidopsis mitochondria. *J Biol Chem.* **279** : 25440-25446.

Perrin R, Lange H, Grienenberger JM, Gagliardi D. (2004b) AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria *J Biol Chem.* **279** : 25440-25446.

Perrotta G, Grienenberger JM, Gualberto JM. (2002) Plant mitochondrial rps2 genes code for proteins with a C-terminal extension that is processed. *Plant Mol Biol.* **50** : 523-533.

Phreaner CG, Williams MA, Mulligan RM. (1996) Incomplete editing of rps12 transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. *Plant Cell.* **8** : 107-117.

Placido A, Gagliardi D, Gallerani R, Grienenberger JM, Marechal-Drouard L. (2005) Fate of a larch unedited tRNA precursor expressed in potato mitochondria. *J Biol Chem.* **280** : 33573-33579

Portnoy V, Evguenieva-Hackenberg E, Klein F, Walter P, Lorentzen E, Klug G, Schuster G. (2005) RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus. *EMBO Rep.* **6** : 1188-1193.

Portnoy V, Schuster G. (2006) RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. Nucleic Acids Res. 34 : 5923-5931.

Rapp WD, Lupold DS, Mack S, Stern DB. (1993) Architecture of the maize mitochondrial atp1 promoter as determined by linker-scanning and point mutagenesis.*Mol Cell Biol.* **13** : 7232-7238.

Raynaud C, Loiselay C, Wostrikoff K, Kuras R, Girard-Bascou J, Wollman FA, Choquet Y. (2007) Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. *Proc Natl Acad Sci U S A.* **104** : 9093-9098.

Remacle C, Maréchal-Drouard L. (1996) Characterization of the potato mitochondrial transcription unit containing 'native' trnS (GCU), trnF (GAA) and trnP (UGG). *Plant Mol Biol.* **30** : 553-563.

Rogowska AT, Puchta O, Czarnecka AM, Kaniak A, Stepien PP, Golik P. (2006) Balance between transcription and RNA degradation is vital for Saccharomyces cerevisiae mitochondria: reduced transcription rescues the phenotype of deficient RNA degradation *Mol Biol Cell.* **17** : 1184-1193.

Ryan CM, Militello KT, Read LK. (2003) Polyadenylation regulates the stability of Trypanosoma brucei mitochondrial RNAs. *J Biol Chem.* **278** : 32753-32762.

Ryan CM, Read LK. (2005) UTP-dependent turnover of Trypanosoma brucei mitochondrial mRNA requires UTP polymerization and involves the RET1 TUTase. *RNA.* **11** : 763-773.

Sáez-Vasquez J, Caparros-Ruiz D, Barneche F, Echeverría M. (2004) Characterization of a crucifer plant pre-rRNA processing complex.*Biochem Soc Trans.* **32** : 578-580.

Saitoh S, Chabes A, McDonald WH, Thelander L, Yates JR, Russell P. (2002) Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell.* **109** : 563-573.

Salinas T, Duchêne AM, Delage L, Nilsson S, Glaser E, Zaepfel M, Maréchal-Drouard L.(2006) The voltage-dependent anion channel, a major component of the tRNA import machinery in plant mitochondria. *Proc Natl Acad Sci U S A.* **103** : 18362-18367.

Sanger F, Nicklen S, Coulson A.R. (1977) DNA Sequencing with Chain-Terminating Inhibitors *Proc Natl Acad Sci U S A* **74** : 5463-5467.

Sarria R, Lyznik A, Vallejos CE, Mackenzie SA. (1998) A cytoplasmic male sterility-associated mitochondrial peptide in common bean is post-translationally regulated. *Plant Cell.* **10** : 1217-1228.

Shadel GS, Clayton DA. (1993) Mitochondrial transcription initiation. Variation and conservation. *J Biol Chem.* **268** : 16083-16086.

Schäfer B. (2005) RNA maturation in mitochondria of S. cerevisiae and S. pombe. *Gene.* **354** : 80-85.

Shaw, PJ, Highett, MI, Beven, AF and Jordan, EG. (1995) The nucleolar architecture of polymerase I transcription and processing. *Embo J*, **14**: 2896-2906.

Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell.* **19** : 1251-1264.

Schmitt S, Paro R. (2004) Gene regulation: a reason for reading nonsense. *Nature.* **429** : 510-511.

Schmitz-Linneweber C, Williams-Carrier R, 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.

Schonfeld C, Wobbe L, Borgstadt R, Kienast A, Nixon PJ, Kruse O (2004) The nucleus-encoded protein MOC1 is essential for mitochondrial light acclimatation in Chlamydomonas Reinhardtii *J Biol Chem.* **279** :50366-50374

Schuster J, Binder S. (2005) The mitochondrial branched-chain aminotransferase (AtBCAT-1) is capable to initiate degradation of leucine, isoleucine and valine in almost all tissues in Arabidopsis thaliana. *Plant Mol Biol.* **57** : 241-254.

Schuster W, Brennicke A (1989) Conserved sequence elements at putative processing sites in plant mitochondria. *Curr Genet.* **15** : 187-92.

Siculella L, Pacoda D, Treglia S, Gallerani R, Ceci LR. (1996) GTG as translation initiation codon in the apocytochrome b gene of sunflower mitochondria. *DNA Seq.* **6** : 365-369.

Simpson L, Sbicego S, Aphasizhev R. (2003) Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA.* **9** : 265-276.

Singh M, Brown GG (1991). Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell.* **3** : 1349-62.

Slomovic S, Laufer D, Geiger D, Schuster G. (2005) Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark.*Mol Cell Biol.* **25**: 6427-6435.

Slomovic S, Laufer D, Geiger D, Schuster G. (2006) Polyadenylation of ribosomal RNA in human cells. *Nucleic Acids Res.* **34** : 2966-2975.

Slomovic S, Portnoy V, Liveanu V, Schuster G. (2006) RNA polyadenylation in prokaryotes and organelles; different tails tell different tales. *Crit. Rev. Plant Sci.* **25** : 65–77.

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

Sohlberg B, Huang J, Cohen SN. (2003) The Streptomyces coelicolor polynucleotide phosphorylase homologue, and not the putative poly(A) polymerase, can polyadenylate RNA. *J Bacteriol.* **185** : 7273-7278.

Stead JA, Costello JL, Livingstone MJ, Mitchell P. (2007) The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. *Nucleic Acids Res.* **35** : 5556-5567.

Stern DB, Gruissem W. (1987) Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell.* **51** : 1145-1157.

Stern D, Kindle K (1993) 3' end maturation of the *Chlamydomonas reinhardtii* atpB mRNA is a two-step process *Molecular and Cellular Biol.* **13** : 2277-2285

Stevenson AL, Norbury CJ. (2006) The Cid1 family of non-canonical poly(A) polymerases. *Yeast.* **23** : 991-1000.

Symmons MF, Jones GH, Luisi BF. (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure.* **8** : 1215-1226.

Taft RJ, Pheasant M, Mattick JS. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays.* **29** : 288-299

Takenaka M, Neuwirt J, Brennicke A. (2004) Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.* **32** : 4137-4144.

Temperley RJ, Seneca SH, Tonska K, Bartnik E, Bindoff LA, Lightowlers RN, Chrzanowska-Lightowlers ZM. (2003) Investigation of a pathogenic mtDNA microdeletion reveals a translation-dependent deadenylation decay pathway in human mitochondria. *Hum Mol Genet.* **12** : 2341-2348.

Tillich M, Lehwark P, Morton BR, Maier UG. (2006) The evolution of chloroplast RNA editing. *Mol Biol Evol.* **10** : 1912-1921.

Timmis JN, Ayliffe MA, Huang CY, Martin W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* **5** : 123-135.

Tomecki R, Dmochowska A, Gewartowski K, Dziembowski A, Stepien PP. (2004) Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Res.* **32** : 6001-6014.

Vaistij FE, Boudreau E, Lemaire SD, Goldschmidt-Clermont M, Rochaix JD. (2000) Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast psbB/psbT/psbH gene cluster in Chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A.* **97** : 14813-14818.

Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* **3** : e189.

van Engelen FA, Molthoff JW, Conner AJ, Nap JP, Pereira A, Stiekema WJ. (1995) pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Res.* **4** : 288-290.

van Hoof A, Frischmeyer PA, Dietz HC, Parker R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science.* **295** : 2262-2264.

van Hoof A, Lennertz P, Parker R. (2000) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol.* **20** : 441-452.

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

Wang HW, Wang J, Ding F, Callahan K, Bratkowski MA, Butler JS, Nogales E, Ke A. (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing. *Proc Natl Acad Sci U S A.* **104** : 16844-16849

Wang L, Eckmann CR, Kadyk LC, Wickens M, Kimble J. (2002) A regulatory cytoplasmic poly(A) polymerase in Caenorhabditis elegans. *Nature.* **419** : 312-316

Wang SW, Stevenson AL, Kearsey SE, Watt S, Bähler J. (2007) Global role for polyadenylation-assisted nuclear RNA degradation in post-transcriptional gene silencing. *Mol Cell Biol.* In press.

Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, 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** : 676-687.

Ward B.L., Anderson,R.S. and Bendich,A.J. (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell*, **25** : 793-803.

West S, Gromak N, Norbury CJ, Proudfoot NJ. (2006) Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol Cell.* **21**: 437-443.

Williams MA, Johzuka Y, Mulligan RM. (2000) Addition of non-genomically encoded nucleotides to the 3'-terminus of maize mitochondrial mRNAs: truncated rps12 mRNAs frequently terminate with CCA. *Nucleic Acids Res.* **28** : 4444-4451.

Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B, Libri D, Jacquier A. (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell.* **121** : 725-737.

Yehudai-Resheff S, Hirsh M, Schuster G. (2001) Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts. *Mol Cell Biol.* **16** : 5408-5416.

Yehudai-Resheff S, Portnoy V, Yogev S, Adir N, Schuster G. (2003) Domain analysis of the chloroplast polynucleotide phosphorylase reveals discrete functions in RNA degradation, polyadenylation, and sequence homology with exosome proteins. *Plant Cell.* **15** : 2003-2019.

Yokobori S, Paabo S. (1997) Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA(Tyr). *J Mol Biol.* **265** : 95-99.

Yue D, Maizels N, and Weiner A.M. (1996) CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily : Characterization of the CCA-adding enzyme from the archaeal hyperthermophile Sulfolobus shibatae. *RNA* **2**: 895–908.

Zaegel V, Guermann B, Le Ret M, Andrés C, Meyer D, Erhardt M, Canaday J, Gualberto J.M , Imbault P (2006) The Plant-Specific ssDNA Binding Protein OSB1 Is Involved in the Stoichiometric Transmission of Mitochondrial DNA in *Arabidopsis. Plant Cell* **18** : 3548-63

Zandueta-Criado A, Bock R. (2004) Surprising features of plastid ndhD transcripts: addition of non-encoded nucleotides and polysome association of mRNAs with an unedited start codon. *Nucleic Acids Res.* **32** : 542-550.

APPENDIX

Appendix 1 :

Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phoshorylase.

Sarah Holec, Heike Lange, Kristina Kühn, Malek Alioua
Thomas Börner and Dominique Gagliardi.

Molecular and Cellular Biology, 2006

Relaxed Transcription in Arabidopsis Mitochondria Is Counterbalanced by RNA Stability Control Mediated by Polyadenylation and Polynucleotide Phosphorylase

Sarah Holec, Heike Lange, Kristina Kühn, Malek Alioua, Thomas Börner, Dominique Gagliardi

Molecular and cellular biology, 2006, vol. 26, N°7, Pages 2869-2876

Appendix 2 :
Table 1 (supplementary data to Holec *et al*, 2006)

Table1

Stable RNA

	Query Id	Location	Length	Align. length	Ext. length	q. start	q. end	s. start	s. end	Strand	Mism.	Duplicates
rRNA	mIVF8R	rrn26	127	128	-1	1	127	8944	9071	-	0	
	mIVE6R	rrn26	131	131	0	1	131	10161	10290	-	0	
	mTIB3R	rrn26	130	130	0	1	130	10162	10290	-	0	
	mTIF7R	rrn26	158	151	7	5	155	10404	10553	-	3	
	mIIIG5F	rrn26	147	128	19	1	128	10607	10734	-	0	
	mIVE4R	rrn26	128	113	15	1	113	11183	11295	-	0	
	mIIE9R	rrn26	105	89	16	1	89	11283	11371	-	0	
	mIIIC1F	5' rrn26/rrn26	135	134	1	2	135	11357	11490	-	1	
	mTIF3F	rrn18	106	105	1	1	105	361400	361504	-	0	
	mIIC7R	rrn18	145	144	1	1	144	361658	361801	-	0	
	mVF12R	rrn18	45	45	0	1	45	362420	362464	-	0	
	mTIC11F	rrn18	153	153	0	1	153	362981	363133	-	0	
	mIIIF11F	rrn18	115	115	0	1	115	363019	363133	-	0	
	mIIIH10R	rrn18	152	150	2	1	149	363134	363283	-	0	
	mIIIC6R	rrn18	153	150	3	1	149	363134	363283	-	0	
	mIVA6F	rrn18	131	131	0	1	131	363136	363266	-	0	
	mVD3R	rrn18	144	144	0	1	144	363137	363280	-	0	
mRNA	mIVF12R	5' nad5 exon d	158	157	1	1	157	22299	22454	-	0	
	mIVC3F	5' nad5 exon d	145	135	10	13	145	22495	22629	-	0	
	mIVH8R	5' nad5 exon d	148	145	3	1	143	22565	22706	-	4	
	mIIG8R	5' nad5 exon d	167	167	0	1	167	22699	22864	-	1	
	mIVB3R	5'rlp16/rpl16	123	123	0	1	123	25500	25621	-	2	
	mIVF3R	ccmFc	150	145	5	1	145	51741	51885	-	0	
	mIIIG9F	5'ccmFc/ ccmFc	141	133	8	1	133	53491	53623	-	1	
	mB11R	5'ccmFc/ ccmFc	63	63	0	1	63	53552	52614	-	3	
	mIIIG7F	5'ccmFc/ ccmFc	128	127	1	1	127	53607	53733	-	0	
	mIVC6F	cob	155	155	0	1	155	60294	60448	+	1	
	mVD4F	atp6-1	119	110	9	8	117	111975	112084	+	0	
	mB8F	5' of nad7 exon1/nad7 exon1	149	149	0	1	149	132013	132161	+	4	
	mTIH12Rb	nad5 exon b	17	13	4	1	13	141555	141567	+	0	
	mTIC7F	matR	89	89	0	1	89	146379	146467	-	0	
	mIVD12F	rpl2 exon 1	64	53	11	1	53	155564	155616	+	0	
	mIIIH12R	atp1	108	106	2	1	106	302604	302709	-	0	
	mIIB10F	atp1	95	86	9	2	87	302985	302900	+	9	
	mTIE5R	3' nad1exon3	138	135	3	2	136	317657	317791	-	1	
	mIVF2F	nad2 exon e	106	106	0	1	106	328742	328846	-	1	
	mVB3R	cox1	23	23	0	1	23	350189	350211	+	0	

Flanking regions < 1.5kb

	Query Id	Location	Length	Align. length	Ext. length	q. start	q. end	s. start	s. end	Strand	Mism.	Duplicates
rRNA	mIIID8R	3'rrn26	9	9	0	1	9	6605	6613	+	0	
	mTIH1R	5' rrn26	142	142	0	1	142	11842	11983	-	0	
	mVG8R	5' rrn26	115	113	2	1	113	11948	12060	-	0	
	mIIB7F	5' rrn26	99	99	0	1	99	11986	12084	-	1	
	mVG11Ra	5' rrn26	57	54	3	1	54	12093	12146	-	0	
	mIIIE3Ra	5' rrn26	52	50	2	1	50	12097	12146	-	0	
	mIVC7Ra	5' rrn26	47	47	0	1	47	12100	12146	-	0	
	mIIIH3Fa	5' rrn26	37	37	0	1	37	12110	12146	-	0	
	mVA2Ra	5' rrn26	61	61	0	1	61	12160	12220	-	1	
	mIIE3Ra	5' rrn26	56	56	0	1	56	12163	12218	-	0	
	mIIA8Ra	5' rrn26	85	85	0	1	85	12224	12308	-	0	
	mIIIH9Ra	5' rrn26	84	78	6	1	78	12224	12301	-	0	
	mIIIA1Ra	5' rrn26	93	80	13	1	80	12229	12308	-	0	
	mTID2Ra	5' rrn26	77	77	0	1	77	12232	12308	-	1	
	mTID4Ra	5' rrn26	77	77	0	1	77	12232	12308	-	0	IVE1F-IVC4R-IIIG10R
	mIIC3Fa	5' rrn26	68	68	0	1	68	12232	12299	-	0	IIIC8F
	mIIIE6Fa	5' rrn26	76	77	-1	1	76	12232	12308	-	0	
	mIIIC8Fa	5' rrn26	68	68	0	1	68	12232	12299	-	0	IIC3F
	mIVC4Ra	5' rrn26	77	77	0	1	77	12232	12308	-	0	TID4R-IVE1F-IIIG10R
	mIVE9Ra	5' rrn26	99	70	29	30	99	12232	12301	-	0	
	mIIF11Fa	5' rrn26	77	75	2	1	75	12233	12307	-	1	
	mIIIE2Fa	5' rrn26	76	76	0	1	76	12233	12308	-	0	B29F
	mB29Fa	5' rrn26	76	75	1	1	75	12234	12308	-	0	IIIE2F
	mB17Fa	5' rrn26	75	75	0	1	75	12234	12308	-	0	
	mIIB11Fa	5' rrn26	50	29	21	16	44	12235	12263	-	0	
	mIVH1Ra	5' rrn26	62	61	1	1	61	12240	12300	-	1	
	mIIIG10Ra	5' rrn26	77	77	0	1	77	12308	12232	+	0	TID4R-IVE1F-IVC4R
	mIVE1Fa	5' rrn26	77	77	0	1	77	12308	12232	+	0	TID4R-IVC4R-IIIG10R
	mIIIA8R	3' rrn5	119	119	0	1	119	359811	359929	-	0	
	mIIC6R	3' rrn5	103	102	1	2	103	360060	360161	-	1	
	mIVC11F	3' rrn5	106	103	3	1	103	360060	360162	-	1	
	mB1R	3' rrn5	22	22	0	1	22	360226	360247	-	0	
	mIIIF7F	3' rrn5	142	139	3	1	139	360237	360375	-	1	
	mVB5R	3' rrn5	112	112	0	1	112	360323	360434	-	1	
	mIVF4R	3' rrn5	148	148	0	1	148	360532	360678	-	0	
	mIIH11R	3' rrn5	68	60	8	9	68	360559	360618	-	0	
	mIIA9R	3' rrn5	87	85	2	1	85	360897	360981	-	0	
	mIID6R	3' rrn5	151	151	0	1	151	361030	361180	-	1	
	mVB8F	IR rrn18-rrn5	54	54	0	1	54	361196	361249	-	0	
	mTIA2R	IR rrn18-rrn5	134	133	1	2	134	361199	361331	-	0	
	mIVE7R	IR rrn18-rrn5	164	164	0	1	164	361201	361364	-	0	
	mIIIC4Rb	5'rrn18	54	52	2	1	52	363378	363429	-	0	
	mVF8Rb	5'rrn18	47	46	1	1	46	363383	363428	-	0	
tRNA	mVD8Fa	3' trnK	90	87	3	1	87	28809	28895	-	0	
	mB4R	5' trnK	96	96	0	1	96	29503	29597	-	7	
	mVC5F	5 trnG	31	29	2	1	29	53955	53983	-	0	
	mIID2R	trnS-trnfM IR	139	139	0	1	139	67561	67698	-	0	
	mIIIH8F	5'trnfM	91	91	0	1	91	71356	71446	+	0	
	mVA8F	5'trnfM	96	96	0	1	96	71356	71451	+	0	
	mIID10F	5'trnfM	67	67	0	1	67	71359	71425	+	0	
	mTIH2R	5'trnfM	86	82	4	5	86	71360	71441	+	0	
	mIVD6R	5'trnfM	86	86	0	1	86	71441	71356	-	0	
	meIID5R	trnfM/3 'trnfM	146	146	0	1	146	71546	71691	+	1	
	mIIF3F	3'trnfM	91	91	0	1	91	71820	71910	+	0	
	mVH3F	3'trnfM	90	90	0	1	90	71838	71927	+	0	
	mIVE8Fa	trnS-trnY IR/repeat II	163	161	2	1	161	103914	104074	+	1	
	mIIIF1Fa	trnS-trnY IR/repeat II	173	117	56	18	134	103926	104041	+	11	
	mIVC12Ra	trnS-trnY IR/repeat II	163	163	0	1	163	103931	104093	+	0	
	mVA4Fa	trnS-trnY IR/repeat II	81	81	0	1	81	103933	104013	+	0	
	mIIIH10Ra	trnS-trnY IR/repeat II	151	150	1	1	150	103967	104116	+	0	
	mIVB4Fa	trnS-trnY IR/repeat II	78	74	4	1	74	103975	104048	+	0	
	mVA5Ra	trnS-trnY IR/repeat II	114	114	0	1	114	103980	104093	+	0	
	mIVF10Ra	trnS-trnY IR/repeat II	65	65	0	1	65	103984	104048	+	0	
	mVA12Ra	trnS-trnY IR/repeat II	116	95	21	1	95	103985	104079	+	0	
	mIIA10Ra	trnS-trnY IR/repeat II	138	138	0	1	138	104081	104218	+	0	
	mVG7Ra	trnS-trnY IR/repeat II	66	66	0	1	66	104081	104146	+	1	
	mVD5Fa	trnS-trnY IR/repeat II	99	96	3	1	96	104120	104215	+	1	
	mIVE3Ra	trnS-trnY IR/repeat II	44	44	0	1	44	104122	104165	+	0	
	mIID8F	trnY-trnP IR/ repeat II/3' repeat II	133	133	0	1	133	104296	104428	+	0	

	mIVH7F	trnY-trnP IR/ repeat II/3' repeat II	130	130	0	1	130	104296	104425	+	0	
	mVE4F	trnY-trnP IR/ repeat II/3' repeat II	163	144	19	1	144	104296	104439	+	0	
	mVF10R	trnY-trnP IR/ repeat II/3' repeat II	130	130	0	1	130	104298	104427	+	0	
	meIIA6F	trnY-trnP IR/ repeat II/3' repeat II	130	130	0	1	130	104299	104428	+	1	
	mIVG6R	trnY-trnP IR/ repeat II/3' repeat II	133	126	7	1	126	104300	104424	+	0	
	mIIIE12R	trnY-trnP IR/ repeat II/3' repeat II	145	142	3	1	142	104306	104447	+	0	
	mTIE4R	trnY-trnP IR/ repeat II/3' repeat II	112	112	0	1	112	104314	104425	+	0	
	mTIC3R	trnP-trnC IR	61	61	0	1	61	104533	104592	+	0	
	mTIE1R	trnP-trnC IR	60	60	0	1	60	104533	104592	+	1	
	mVC1F	trnP-trnC IR	66	65	1	1	65	104536	104600	+	1	
	mIIIH8R	trnP-trnC IR	105	104	1	2	105	104537	104640	+	0	
	mIIID6F	trnP-trnC IR	122	122	0	1	122	104598	104719	+	0	
	mVB1F	trnP-trnC IR	127	125	2	1	125	104599	104723	+	0	
	mTIC12R	trnP-trnC IR	153	153	0	1	153	104601	104753	+	0	
	mTIG2R	trnP-trnC IR	147	138	9	1	138	104749	104884	+	0	
	mIIE7F	trnC/3'trnC	147	139	8	1	139	104957	105095	+	0	
	mVD11F	trnC/3'trnC	94	76	18	1	76	104986	105061	+	0	
	mIIIA4R	trnC-trnN IR	130	130	0	1	130	105021	105150	+	0	
	mIVG9F	trnC-trnN IR	89	89	0	1	89	105021	105109	+	0	
	mVE1F	trnC-trnN IR	97	88	9	1	88	105064	105151	+	0	
	mIIID11R	trnC-trnN IR	82	82	0	1	82	105121	105202	+	0	
	m+cIIIH7R	trnC-trnN IR	133	132	1	2	133	105309	105440	+	1	
	mIIIE10F	trnC-trnN IR	144	143	1	2	144	105649	105791	+	0	
	mIVF6R	trnC-trnN IR	84	82	2	1	82	105793	105874	+	10	
	mIVG7F	trnN-trnY IR	129	129	0	1	129	105955	106083	+	0	
	mVD12F	trnN-trnY IR	124	98	26	27	124	105988	106085	+	1	
	mTID6R	trnN-trnY IR	45	45	0	1	45	106029	106073	+	1	
	mIIA5R	trnN-trnY IR	61	60	1	1	60	106034	106093	+	1	
	mVB2R	trnN-trnY IR	67	67	0	1	67	106034	106100	+	0	
	mIVH12F	trnN-trnY IR	63	63	0	1	63	106035	106097	+	0	
	mTIF1R	trnN-trnY IR	89	83	6	1	83	106106	106188	+	2	
	mIVC5R	trnN-trnY IR	150	150	0	1	150	106221	106370	+	0	
	mIVD5R	trnN-trnY IR	154	154	0	1	154	106221	106374	+	0	
	mVC4R	trnN-trnY IR	136	130	6	1	130	106226	106355	+	0	
	mIIB1R	trnN-trnY IR	47	47	0	1	47	106243	106289	+	0	
	mIVE10F	trnN-trnY IR	135	136	-1	1	135	106535	106670	+	0	
	mIIIA6F	trnN-trnY IR	101	102	-1	1	101	106693	106794	+	0	
	mIIIF5F	trnN-trnY IR	100	101	-1	1	100	106693	106793	+	0	
	mVH5F	trnN-trnY IR	54	55	-1	1	54	106693	106747	+	0	
	mIIIB7R	trnN-trnY IR	50	50	0	1	49	106699	106748	+	0	
	mB1F	trnN-trnY IR	47	47	0	1	47	106701	106748	+	0	
	mIIF8R	3'trnY	78	77	1	1	77	106882	106956	+	0	
	mIVC9Fb	3'trnY	104	96	8	9	104	107005	107099	+	3	
	mIIIE11F	3'trnY	124	107	17	13	119	107861	107967	+	5	
	mB30R	3'trnY	105	105	0	1	105	107874	107978	+	0	B18R
	mB18R	3'trnY	105	105	0	1	105	107874	107978	+	0	B30R
	mIVC1R	3'trnY	94	93	1	1	93	107886	107978	+	0	
	mIVD1R	3'trnY	94	93	1	1	93	107886	107978	+	2	
	mIIID7Fa	3' trnS/repeat III	158	156	2	1	156	113831	113986	+	0	
	mTIG11R	3' trnI IR	152	124	28	29	152	125431	125553	-	1	
	mIVD4R	3' trnI IR	114	112	2	1	112	125578	125689	-	0	
	mIVG12R	3' trnI IR	110	110	0	1	110	125580	125689	-	0	
	mIIIG1F	3' trnI IR	95	95	0	1	95	125884	125790	+	0	
	mIIIC12F	5'of trnI	145	144	1	1	144	127098	127241	-	0	
	mIIIA7R	5'of trnI	153	152	1	1	152	127099	127250	-	0	
	mVG6R	5'of trnI	141	140	1	1	140	127099	127238	-	0	
	mIIF12R	5'of trnI	21	21	0	1	21	127126	127146	-	0	IIH12R
	mIIIH12R	5'of trnI	21	21	0	1	21	127126	127146	-	0	IIF12R
	mVB12F	5'of trnI	62	62	0	1	62	127177	127238	-	1	
	mIIB12R	5'of trnI	39	39	0	1	39	127183	127221	-	0	
	mIVE8Fb	trnS-trnY IR/repeat II	163	161	2	1	161	227196	227356	+	1	
	mIIIF1Fb	trnS-trnY IR/repeat II	173	117	56	18	134	227208	227323	+	11	
	mIVC12Rb	trnS-trnY IR/repeat II	163	163	0	1	163	227213	227375	+	0	
	mVA4Fb	trnS-trnY IR/repeat II	81	81	0	1	81	227215	227295	+	0	
	mIIIH10Rb	trnS-trnY IR/repeat II	151	150	1	1	150	227249	227398	+	0	
	mIVB4Fb	trnS-trnY IR/repeat II	78	74	4	1	74	227257	227330	+	0	
	mVA5Rb	trnS-trnY IR/repeat II	114	114	0	1	114	227262	227375	+	0	
	mIVF10Rb	trnS-trnY IR/repeat II	65	65	0	1	65	227266	227330	+	0	
	mVA12Rb	trnS-trnY IR/repeat II	116	95	21	1	95	227267	227361	+	0	
	mIIA10Rb	trnS-trnY IR/repeat II	138	138	0	1	138	227363	227500	+	0	
	mVG7Rb	trnS-trnY IR/repeat II	66	66	0	1	66	227363	227428	+	1	
	mVD5Fb	trnS-trnY IR/repeat II	99	96	3	1	96	227402	227497	+	1	
	mIVE3Rb	trnS-trnY IR/repeat II	44	44	0	1	44	227404	227447	+	0	
	mVH7F	3'trnE	66	64	2	1	64	227690	227753	+	0	
	mIIG6F	3'trnE	171	170	1	1	170	227730	227899	+	0	
	mTID9F	3'trnE	118	118	0	1	118	228026	228142	+	0	
	mB7F	3'trnE	147	147	0	1	147	228157	228303	+	0	
	mIIA3R	3'trnE	133	133	0	1	133	228728	228860	+	0	
	mIIID1R	3' trnQ	105	103	2	3	105	254299	254401	-	1	
	mIVH2F	5' trnQ	156	108	48	50	156	254799	254905	-	7	
	mVC3R	5' trnQ	110	110	0	1	110	254838	254947	-	0	
	mIIC3R	5' trnQ	15	15	0	1	15	255126	255140	-	0	
	mIVG5F	5' trnQ	143	143	0	1	143	255330	255472	-	0	
	mIIID7Fb	3' trnS/repeat III	158	156	2	1	156	295741	295896	-	0	
	mIIIA2R	5' trnM	159	158	1	1	158	336603	336760	+	1	
	mTIE10F	5' trnM	121	121	0	1	121	336802	336918	+	1	
	mVC6F	5' trnM	165	130	35	1	130	336868	336997	+	0	
	mIIIG12F	5' trnM	121	118	3	1	118	337436	337553	+	0	
mRNA	mIIIB9F	5' rps4	157	153	4	1	152	84284	84436	-	2	
	mVG1F	5' of nad7 exon1	104	95	9	1	95	131817	131911	+	0	
	TIH12Ra	3' of rpl2 exon 1	17	13	4	1	13	155728	155740	+	0	
	mIID4R	orfX-nad4 exon1 IR	26	26	0	1	26	160531	160556	+	0	
	mVH2R	5' nad4L	137	137	0	1	137	189663	189799	-	1	
	mTID10R	3' cox3 exon2	92	77	15	8	84	219423	219499	+	0	
	mTIB10R	3'rpsl2	115	112	3	1	112	259957	260067	-	0	
	mIID3R	3' apt9	117	117	0	1	117	279226	279329	+	10	
	mIIIF12F	3' apt9	100	100	0	1	100	279229	279320	+	0	
	mIVC9Fa	3' apt9	104	90	14	2	91	279251	279340	+	1	
	mVH10F	5' nad1	113	111	2	1	111	289424	289534	-	1	
	mVA9R	5' cox1	148	147	1	1	147	351508	351654	-	0	

Intergenic regions

Query id	Location	Length	Align. length	Ext. length	q. start	q. end	s. start	s. end	Strand	Mism.	Duplicates
mIVF7F	3' orf 315	121	120	1	1	120	15032	15151	-	0	
mTIF2R	3' orf 315	113	112	1	1	112	16117	16228	-	1	
mIIIC7R	3' orf 315	110	100	10	1	100	16614	16712	-	0	
mIIIB4F	orf 315	164	161	3	1	161	17251	17411	-	0	
mIVE12R	orf 315	107	102	5	1	102	17561	17662	-	0	
mIIIB11R	5' orf 315	149	129	20	1	129	18063	18191	-	0	
mB14Fa	5' orf 315	90	89	1	1	89	18290	18378	-	1	
mIVB6Fa	repeat I	124	122	2	1	122	44928	45049	+	0	
mIIIA11Ra	repeat I	152	148	4	1	148	45393	45540	+	0	
mVF7R	repeat I	139	139	0	1	139	45462	45600	+	1	
mVE5Ra	repeat I	150	144	6	1	144	46098	46241	+	0	
mIVA2Fa	repeat I	123	123	0	1	123	46371	46493	+	0	
mIIIB3Ra	repeat I	73	73	0	1	73	46655	46726	+	0	
mB33Fa	repeat I	94	93	1	1	93	46848	46940	+	0	B21F
mB21Fa	repeat I	94	93	1	1	93	46848	46940	+	0	B33F

mIIE3Rc	repeat I	56	55	1	2	56	46949	47003	+	6	
mVA11Ra	repeat I	121	115	6	1	115	47045	47159	+	0	
mIVF11Ra	repeat I	159	155	4	1	155	47058	47212	+	0	
mIVA10Fa	repeat I	152	152	0	1	152	47072	47223	+	0	
mVB10Fa	repeat I	118	118	0	1	118	47117	47234	+	0	
mVH6Fa	repeat I	88	88	0	1	88	47156	47243	+	0	
mIIIH11Fa	repeat I	154	141	13	1	141	47238	47378	+	1	
mIVH6Ra	repeat I	110	110	0	1	110	47244	47353	+	3	
mIVA8Fa	repeat I	136	134	2	1	134	47245	47378	+	2	
mIVA3Fa	repeat I	93	91	2	1	91	47259	47349	+	2	
mTIH3Ra	repeat I	92	92	0	1	92	47262	47353	+	1	
mIIE8FASa	repeat I	133	132	1	1	132	47264	47395	-	0	
mTIE2Ra	repeat I	123	123	0	1	123	47268	47390	+	1	
meIIH1Fa	repeat I	61	60	1	1	60	47274	47333	+	1	
mB5Fa	repeat I	98	98	0	1	98	47277	47374	+	1	
mTID12Fa	repeat I	118	98	20	21	118	47277	47374	+	2	
mVD10Fa	repeat I	169	167	2	1	167	47317	47483	+	0	
mTID3Ra	repeat I	157	156	1	1	156	47329	47484	+	1	
mIIH5Ra	repeat I	120	97	23	1	97	47342	47434	+	1	
mIIG11Ra	repeat I	170	153	17	1	153	47361	47513	+	0	
TIB12Ra	repeat I	134	134	0	1	134	47385	47518	+	0	
mIIE2Fa	repeat I	137	129	8	1	129	47385	47513	+	0	
mB23Fa	repeat I	126	126	0	1	126	47394	47518	+	0	
mB35Fa	repeat I	125	125	0	1	125	47394	47518	+	0	
mIVD3Fa	repeat I	123	123	0	1	123	47396	47518	+	0	
mIIF6Fa	repeat I	115	115	0	1	115	47404	47518	+	0	
mVE12Fa	repeat I	117	111	6	1	111	47409	47519	+	0	
mIIIC11Ra	repeat I	107	107	0	1	107	47412	47518	+	0	
mIIIG8Ra	repeat I	49	47	2	1	47	47446	47492	+	0	
mIID9Ra	repeat I	116	113	3	1	113	47525	47637	+	0	
mIID7Fa	repeat I	138	131	7	1	131	47545	47675	+	0	
mIIF2Fa	repeat I	137	130	7	1	130	47546	47675	+	0	
mIIG12Ra	repeat I	98	101	-3	1	96	47644	47744	+	0	
mIVE2R	repeat I/3'repeat I	84	82	2	1	82	48837	48918	+	0	IVF1R
mIVF1R	repeat I/3'repeat I	83	82	1	1	82	48837	48918	+	0	IVE2R
mVC2F	3' repeat I	145	145	0	1	145	49219	49363	+	0	
mIIIE11F	3' repeat I	160	160	0	1	160	49347	49506	+	0	
mIVB2R	3' repeat I	128	128	0	1	128	49925	50052	+	0	
mVA10F	3' repeat I	156	156	0	1	156	50054	50209	+	0	
mIIIB2F	rps4 - orf 145a IR	94	94	0	1	94	87663	87755	-	0	
mB29Fb	orf154 - repeat II IR	76	75	1	1	75	98946	99020	+	0	
mB17Fb	orf154 - repeat II IR	75	75	0	1	75	98946	99020	+	0	
mTID2Rb	orf154 - repeat II IR	77	77	0	1	77	98946	99022	+	1	
mTID4Rb	orf154 - repeat II IR	77	77	0	1	77	98946	99022	+	0	
mIIA8Rb	orf154 - repeat II IR	85	85	0	1	85	98946	99030	+	0	
mIIIG10Rb	orf154 - repeat II IR	77	77	0	1	77	98946	99022	+	0	
mIIIE6Fb	orf154 - repeat II IR	76	77	-1	1	76	98946	99022	+	0	
mIIIA1Rb	orf154 - repeat II IR	93	80	13	1	80	98946	99025	+	0	
mIIE2Fb	orf154 - repeat II IR	76	76	0	1	76	98946	99021	+	0	
mIVC4Rb	orf154 - repeat II IR	77	77	0	1	77	98946	99022	+	0	
mIVE1Fb	orf154 - repeat II IR	77	77	0	1	77	98946	99022	+	0	
mIIF11Fb	orf154 - repeat II IR	77	75	2	1	75	98947	99021	+	1	
mIIIH9Rb	orf154 - repeat II IR	84	78	6	1	78	98953	99030	+	0	
mIVE9Rb	orf154 - repeat II IR	99	70	29	30	99	98953	99022	+	0	
mIVH1Rb	orf154 - repeat II IR	62	61	1	1	61	98954	99014	+	1	
mIIC3Fb	orf154 - repeat II IR	68	68	0	1	68	98955	99022	+	0	
mIIIC8Fb	orf154 - repeat II IR	68	68	0	1	68	98955	99022	+	0	
mIIB11Fb	orf154 - repeat II IR	50	29	21	16	44	98991	99019	+	0	
mVA2Rb	orf154 - repeat II IR	61	61	0	1	61	99034	99094	+	1	
mIIE3Rb	orf154 - repeat II IR	56	56	0	1	56	99036	99091	+	0	
mIIIH3Fb	orf154 - repeat II IR	37	37	0	1	37	99108	99144	+	0	
mIIIE3Rb	orf154 - repeat II IR	52	50	2	1	50	99108	99157	+	0	
mIVC7Rb	orf154 - repeat II IR	47	47	0	1	47	99108	99154	+	0	
mVG11Rb	orf154 - repeat II IR	57	54	3	1	54	99161	99108	-	0	
mIIA12F	orf154 - repeat II IR	149	48	101	1	48	99167	99214	+	0	
mIVB12F	orf154 - repeat II IR	150	150	0	1	150	99167	99316	+	0	
mVD9F	orf154 - repeat II IR	149	149	0	1	149	99167	99315	+	0	
mIIG10F	orf154 - repeat II IR	149	149	0	1	149	99168	99316	+	2	
mTIF12R	orf154 - repeat II IR	146	146	0	1	146	99171	99316	+	0	TIB11R
mTIB11R	orf154 - repeat II IR	146	146	0	1	146	99171	99316	+	0	TIF12R
mTIA3F	orf154 - repeat II IR	142	141	1	2	142	99191	99327	+	1	
mB15R	orf154 - repeat II IR	128	127	1	2	128	99402	99528	+	1	
mIIIA5R	orf154 - repeat II IR	148	147	1	1	147	99438	99584	+	0	IVD2R
mIVD2R	orf154 - repeat II IR	148	147	1	1	147	99438	99584	+	0	IIIA5R
mIVC2R	orf154 - repeat II IR	152	146	6	5	150	99442	99584	+	8	
mIIIH2R	orf154 - repeat II IR	94	93	1	1	93	99498	99590	+	1	
mB28F	IR 3' repeat II	105	104	1	1	104	100830	100933	+	0	TIC1F
mB16F	IR 3' repeat II	104	104	0	1	104	100830	100933	+	0	
mTIC1F	IR 3' repeat II	105	104	1	1	104	100830	100933	+	0	B28F
mIVE11F	IR 3' repeat III	22	19	3	1	19	122345	122363	+	1	
mB14Fb	IR 3' repeat III	90	89	1	1	89	122367	122455	+	1	
m+nVH9F	IR 3' repeat III	104	103	1	1	103	122650	122752	+	0	
m+nVF9F	IR 3' repeat III	101	101	0	1	101	122652	122752	+	0	
mIIIF5F	IR 3' repeat III	75	75	0	1	75	122678	122751	+	1	
mIIIF3R	5 'orf 161 IR	95	95	0	1	95	178655	178749	-	1	
mIVB6Fb	repeat I	124	122	2	1	122	179093	179214	+	0	
mIIIA11Rb	repeat I	152	148	4	1	148	179558	179705	+	0	
mVF7R	repeat I	139	139	0	1	139	179627	179765	+	1	
mVE5Rb	repeat I	150	144	6	1	144	180263	180406	+	0	
mIVA2Fb	repeat I	123	123	0	1	123	180536	180658	+	0	
mIIB3Rb	repeat I	73	73	0	1	73	180820	180891	+	0	
mB33Fb	repeat I	94	93	1	1	93	181013	181105	+	0	
mB21Fb	repeat I	94	93	1	1	93	181013	181105	+	0	
mIIE3Rb	repeat I	56	55	1	2	56	181114	181168	+	6	
mVA11Rb	repeat I	121	115	6	1	115	181210	181324	+	0	
mIVF11Rb	repeat I	159	155	4	1	155	181223	181377	+	0	
mIVA10Fb	repeat I	152	152	0	1	152	181237	181388	+	0	
mVB10Fb	repeat I	118	118	0	1	118	181282	181399	+	0	
mVH6Fb	repeat I	88	88	0	1	88	181321	181408	+	0	
mIIIH11Fb	repeat I	154	141	13	1	141	181403	181543	+	1	
mIVH6Rb	repeat I	110	110	0	1	110	181409	181518	+	3	
mIVA8Fb	repeat I	136	134	2	1	134	181410	181543	+	2	
mIVA3Fb	repeat I	93	91	2	1	91	181424	181514	+	2	
mTIH3Rb	repeat I	92	92	0	1	92	181427	181518	+	1	
mIIE8FASb	repeat I	133	132	1	1	132	181429	181560	-	0	
mTIE2Rb	repeat I	123	123	0	1	123	181433	181555	+	1	
mIIH1Fb	repeat I	61	60	1	1	60	181439	181498	+	1	
mB5Fb	repeat I	98	98	0	1	98	181442	181539	+	1	
mTID12Fb	repeat I	118	98	20	21	118	181442	181539	+	2	
mVD10Fb	repeat I	169	167	2	1	167	181482	181648	+	0	
mTID3Rb	repeat I	157	156	1	1	156	181494	181649	+	1	
mIIH5Rb	repeat I	120	97	23	1	97	181507	181599	+	1	
mIIG11Rb	repeat I	170	153	17	1	153	181526	181678	+	0	
mTIB12Rb	repeat I	134	134	0	1	134	181550	181683	+	0	
mIIE2Fb	repeat I	137	129	8	1	129	181550	181678	+	0	
mB23Fb	repeat I	126	126	0	1	126	181559	181683	+	0	

mB35Fb	repeat I	125	125	0	1	125	181559	181683	+	0	
mIVD3Fb	repeat I	123	123	0	1	123	181561	181683	+	0	
mIIF6Fb	repeat I	115	115	0	1	115	181569	181683	+	0	
mVE12Fb	repeat I	117	111	6	1	111	181574	181684	+	0	
mIIIC11Rb	repeat I	107	107	0	1	107	181577	181683	+	0	
mIIIG8Rb	repeat I	49	47	2	1	47	181611	181657	+	0	
mIID9Rb	repeat I	116	113	3	1	113	181690	181802	+	0	
mIID7Fb	repeat I	138	131	7	1	131	181710	181840	+	0	
mIIF2Fb	repeat I	137	130	7	1	130	181711	181840	+	0	
mIIG12Rb	repeat I	98	101	-3	1	96	181809	181909	+	0	
mB6F	5' orf 149	112	110	2	1	110	190650	190759	-	0	
mIIIA3F	orf 275	144	134	10	1	125	191632	191765	-	1	
mIIID3F	orf 275	130	130	0	1	130	191683	191812	-	0	
mTIC10F	orf 275	152	117	35	42	152	191769	191885	-	4	
mB32R	orf 275	96	96	0	1	96	191771	191866	-	3	
mB20R	orf 275	96	96	0	1	96	191771	191866	-	3	
mIIIG6F	orf 275	157	104	53	60	157	191782	191885	-	4	
mIIIE1F	orf 275	88	88	0	1	88	191782	191869	-	1	
mVE11F	5'orf 275	148	148	0	1	148	192032	192179	-	0	
mIIE1R	5'orf 275	144	144	0	1	144	192034	192177	-	0	
mIVA11R	3' orf 240a, antisense	166	162	4	1	161	203275	203436	+	0	
mIVD8Fb	5' orf240a/orf240a	90	87	3	1	87	204251	204337	-	0	
mTIH11R	orf 240a - 120 IR	84	84	0	1	84	205210	205293	-	0	
mIIIG4F	5' repeat II	135	120	15	1	120	226799	226918	+	0	
mVF1F	orf 215a - 215b IR	137	137	0	1	137	244672	244808	-	1	
mIIIC5R	3' orf 215b	143	143	0	1	143	248881	249023	-	0	
mVF5F	orf 114	119	119	0	1	119	260966	261084	-	3	
mTIA7R	orf 135b	144	144	0	1	144	261832	261975	-	2	
mIIIB1F	orf 135b	100	97	3	4	100	261893	261989	-	0	
mIVD9F	5' orf 135b	101	101	0	1	101	262217	262317	-	1	
mTIA6R	5' orf 135b	72	62	10	11	72	262240	262301	-	0	
mIIIC4Ra	5' orf 135b	54	52	2	1	52	262265	262311	-	0	
mVF8Ra	5' orf 135b	47	46	1	1	46	262270	262310	-	0	
mIVH9F	orf 106e - orf 107f IR	110	102	8	1	102	267556	267657	-	0	
mIIIC9F	orf 106e - orf 107f IR	138	136	2	1	136	268653	268788	-	0	
mVE9R	orf 106e - orf 107f IR	92	92	0	1	92	269157	269248	-	1	
mIIID4R	3' orf 139	85	84	1	1	84	272606	272689	-	0	
mIVG10F	3' orf107g	160	126	34	1	126	276057	276182	+	0	
mIIIA10R	orf294	106	106	0	1	106	304332	304437	-	0	
mIIC10F	orf294	56	48	8	1	48	304424	304471	-	0	
mTIH10R	orf113 - orf145b IR	154	152	2	1	152	306558	306709	+	0	
mTIA4R	5' orf136a/orf136a, antisense	91	90	1	1	90	325249	325338	-	0	
mIIH7F	3' trnM	154	154	0	1	154	339623	339776	+	0	
mIVA12R	orf 145c	41	41	0	1	41	347110	347150	+	0	
mIIC9R	cox1 - orf111d IR	139	139	0	1	139	357008	357146	-	0	
mVC8R	cox1 - orf111d IR	147	147	0	1	147	357091	357237	-	0	
mB27R	cox1 - orf111d IR	91	91	0	1	91	358196	358286	-	0	
mVA3R	cox1 - orf111d IR	105	103	2	1	103	359265	359367	-	1	

Arabidopsis Landsberg mt genome accession number D84192

Query Id	Location	Length	Align. length	Ext. length	q. start	q. end	s. start	s. end	Strand	Mism.	Duplicates
uVC7R	3' rps3	74	74	0	1	74	1870	1943	+	0	

Table 1: Degradation tags were sorted by categories: Stable RNA (including rRNA and mRNA); Flanking regions to known genes (up to 1.5 kbp on each side); Intergenic regions (IR). Query ID: clone identifier; Location: position to closest gene when applicable; Length: length of query in nt; Align. length: number of nt matching the *Arabidopsis* mt genome; Ext. length: number of nt non-coded by the *Arabidopsis* mt genome; q. start: position in query of the first matching nt; q. end: position in query of the last matching nt; s. start: position in the *Arabidopsis* mt genome of the first matching nt; s. end: position in the *Arabidopsis* mt genome (NC_001284) of the last matching nt; Strand + or -: indicates the sense or antisense strand as in NC_001284; Mism.: number of mismatches between query and the *Arabidopsis* mt genome; Duplicates: indicate identical clone identifier in case of duplicated sequences. For clones matching repeated sequences in the *Arabidopsis* mt genome, both location are given and the clone identifier is followed by a or b. m+n and m+c preceding clone identifier indicate a possible mt or nuclear origin or mt or chloroplastic origin, respectively.

Appendix 3 :
Importance de la surveillance des ARN dans les mitochondries d'*Arabidopsis thaliana*,
Sarah Holec, Dominique Gagliardi, Médecine/Sciences 2006

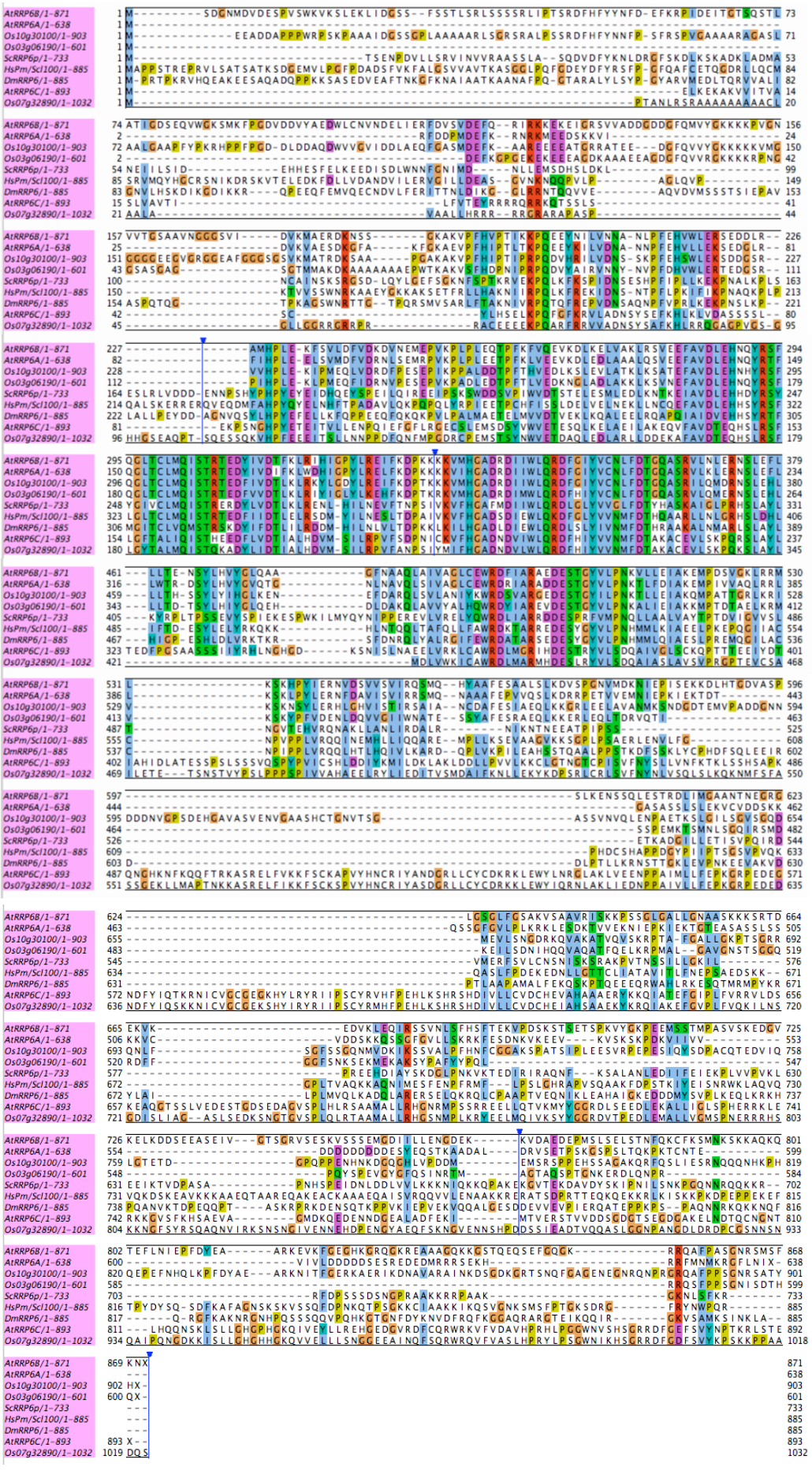
Importance de la surveillance des ARN dans les mitochondries d'Arabidopsis thaliana

Sarah Holec, Dominique Gagliardi

Médecine/Sciences, 2006, Vol. 22, N° 8-9, pages 681-682

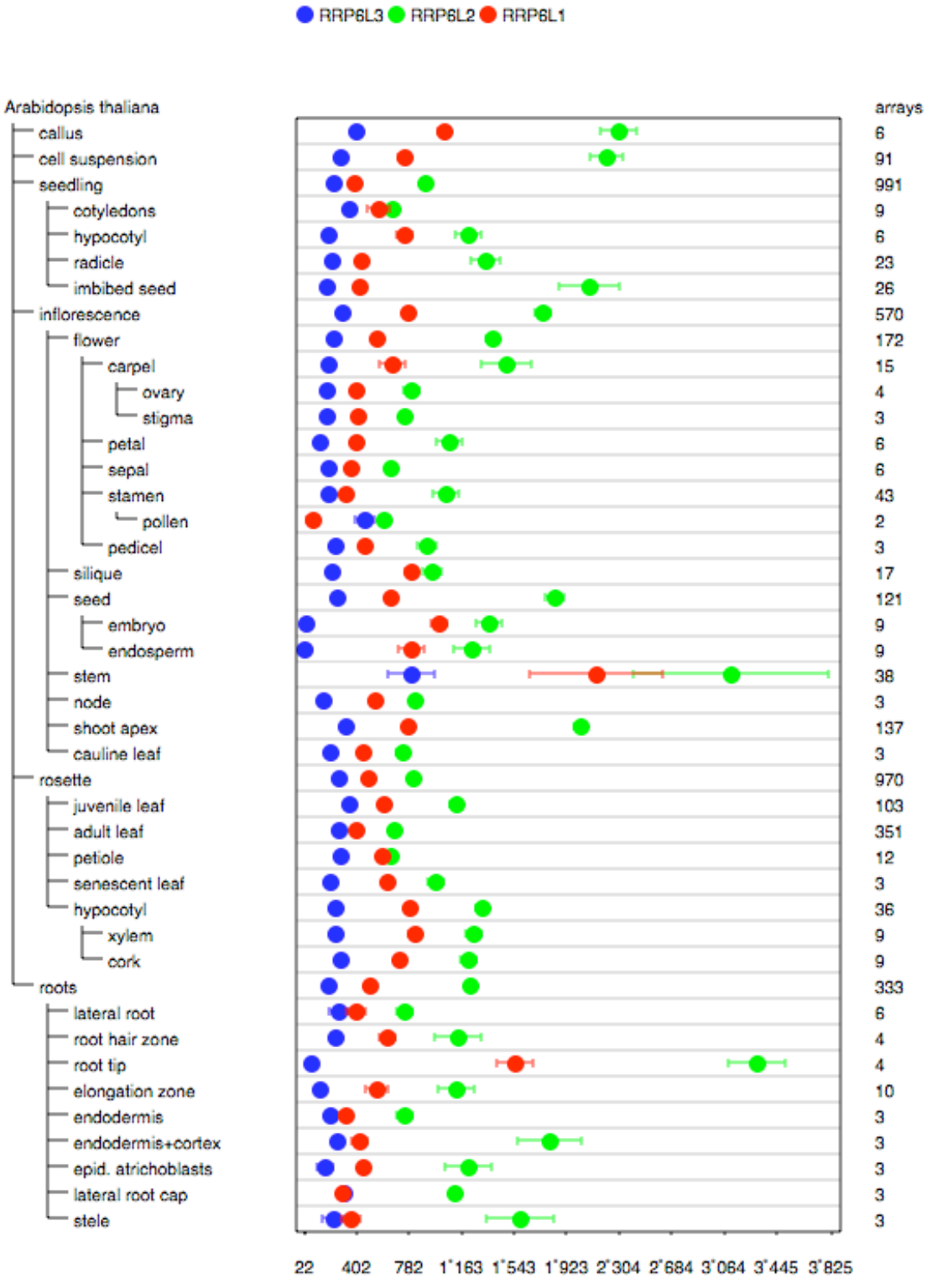
Appendix 4 : Sequence alignment of RRP6 proteins.

Sequence of RRP6 proteins from *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Saccharomyces cerevisae* (Cs), *Homo sapiens* (Hs) and *Drosophila melanogaster* (Dm) were aligned. Alignment was performed using Muscle v.3.52 program and visualized using Jalview editor following removing of poorly aligned positions (blue arrows).


AtRRP68/1-871
AtRRP6A/1-638
Os10g30100/1-903
Os03g06190/1-601
ScRRP6p/1-733
HsPm/Scl100/1-885
DmRRP6/1-885
AtRRP6C/1-893
Os07g32890/1-1032

Appendix 5 : Genevestigator tissue-specific expression profile of RRP6L mRNAs

(Meta profile analysis performed on http://www.genevestigator.ethz.ch/)


RRP6L3 RRP6L2 RRP6L1
Arabidopsis thaliana
callus
cell suspension
seedling
cotyledons
hypocotyl
radicle
imbibed seed
inflorescence
flower
carpel
ovary
stigma
petal
sepal
stamen
pollen
pedicel
silique
seed
embryo
endosperm
stem
node
shoot apex
cauline leaf
rosette
juvenile leaf
adult leaf
petiole
senescent leaf
hypocotyl
xylem
cork
roots
lateral root
root hair zone
root tip
elongation zone
endodermis
endodermis+cortex
epid. atrichoblasts
lateral root cap
stele
arrays
6
91
991
9
6
23
26
570
172
15
4
3
6
6
43
2
3
17
121
9
9
38
3
137
3
970
103
351
12
3
36
9
9
333
6
4
4
10
3
3
3
3
3
22
402
782
1˙163
1˙543
1˙923
2˙304
2˙684
3˙064
3˙445
3˙825

Appendix 6 :

Degradation of a polyadenylated rRNA maturation by-product involves one of the three RRP6 like proteins in *Arabidopsis thaliana*

Heike Lange, Sarah Holec, Valérie Cognat, Laurent Pieuchot, Jean Canaday and Dominique Gagliardi.

submitted to Molecular and Cellular Biology

Degradation of a polyadenylated rRNA maturation by-product involves one of the three RRP6-like proteins in *Arabidopsis thaliana*

Heike Lange, Sarah Holec, Valérie Cognat, Laurent Pieuchot, Monique Le Ret, Jean Canaday and Dominique Gagliardi*

Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Unité Propre de Recherche 2357, Conventionné avec l'Université Louis Pasteur, 67000 Strasbourg, France.

*Corresponding author. Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Unité Propre de Recherche 2357, 12 rue du général Zimmer, 67000 Strasbourg, France.

Tel: +33 3 88 41 71 62

FAX: + 33 3 88 61 44 42

Email: dominique.gagliardi@ibmp-ulp.u-strasbg.fr

Running title: Specialization of RRP6-like proteins in Arabidopsis

Keywords: RRP6 / exosome / polyadenylation / RNA degradation / nucleolus

Manuscript statistics

Abstract: 143 words

Material and Methods: 554 words

Introduction, Results and Discussion: 3112 words

ABSTRACT

Yeast Rrp6p and its human counterpart PM/Scl100 are exosome-associated proteins involved in processing of non-coding RNAs and degradation of aberrant transcripts. Activity of yeast Rrp6p is stimulated by the polyadenylation of its RNA substrates. We identified three RRP6-like proteins in *Arabidopsis thaliana*: At-RRP6L3 is restricted to the cytoplasm, whereas At-RRP6L1 and 2 have different intra-nuclear localizations. Both nuclear RRP6L proteins share functional homology with yeast Rrp6p: At-RRP6L1 complements the temperature sensitive phenotype of a yeast *rrp6Δ* strain and mutation of At-RRP6L2 leads to accumulation of an rRNA maturation by-product. This by-product accumulates as a polyadenylated transcript, suggesting that the basic mechanism of poly(A)-mediated RNA degradation is conserved in plant nuclei. Interestingly, the rRNA maturation by-product is a substrate of At-RRP6L2 but not of AtRRP6L1. This result and the distinctive subcellular distribution of AtRRP6L1 to 3 indicate a specialization of RRP6-like proteins in Arabidopsis.

INTRODUCTION

The exoribonucleolytic activity of the exosome is fundamental to many aspects of RNA metabolism. The exosome is involved in mRNA turn-over, degradation of defective and cryptic transcripts, but also in processing of 3' extremities of a variety of non-coding RNA and elimination of RISC-cleaved mRNA (for recent reviews see 25, 44). While most exosome functions were initially characterized in the yeast *Saccharomyces cerevisiae*, related complexes are present in all higher eukaryotes investigated and in several Archaea (4, 9, 10, 16, 24, 26, 32, 33).

Eukaryotic exosomes are composed of a core complex to which cytoplasmic- and nuclear-specific subunits associate (reviewed in 35). One of these subunits, Rrp6p, a member of the RNase D family, associates with the nuclear exosome in yeast (2). Its human counterpart, PM/Scl-100, is predominantly nuclear but is also detected in the cytoplasm (7). Rrp6p plays a role in nuclear mRNA surveillance and in the degradation of rRNA maturation by-products or intergenic transcripts (27, 30, 45, 48). In addition, Rrp6p is involved in the final steps in processing several non-coding RNAs (1, 6).

In yeast, the TRAMP complex polyadenylates RNA substrates, which triggers their degradation by the nuclear exosome (27, 45, 48). In higher eukaryotes, there is little data on polyadenylation of nuclear transcripts destined for degradation. Short poly(A) tails were detected upon co-transcriptional cleavage of human ß-globin and murine serum albumin pre-mRNA (47). Human ribosomal RNA can be also polyadenylated at putative sites of endonucleolytic cleavage (41). In plants, polyadenylation of nuclear non-coding RNA may also occur as polyadenylated transcripts of the low abundant 5S rRNA spacer were reported recently in *Nicotiana* (18).

We show here that three RRP6-like proteins (RRP6L1-3) are encoded by the *Arabidopsis thaliana* genome and form two distinct subfamilies, one of which is specific to plants.

Furthermore, each At-RRP6L protein has a specific subcellular distribution. Yeast complementation and analysis of mutant plants demonstrates that both nuclear isoforms are functional homologs of Rrp6p but have non-overlapping roles in plants. Interestingly, we also observe polyadenylation of an rRNA maturation by-product that accumulates upon knock-down of RRP6L2, indicating that polyadenylation can mark a transcript destined for degradation in plant nuclei.

MATERIAL AND METHODS

Plant material. All *Arabidopsis thaliana* plants used in this study were of Columbia ecotype (Col-0). Insertion mutant information was obtained from the SIGnAL website at http://signal.salk.edu. T-DNA insertion lines were generated in the context of Gabi-KAT and Salk T-DNA programs and provided by Bernd Weisshaar (MPI for Plant Breeding Research; Cologne, Germany) or retrieved from NASC (http://arabidopsis.info/), respectively (3, 36, 38). Salk_004432 and Gabi_344G09 correspond to T-DNA insertion lines in At1g54440 and were named *rrp6l1-1* and *rrp6l1-2*, respectively. Gabi_825G09 and Salk_113786 correspond to T-DNA insertion lines in At5g35910 and were named *rrp6l2-1* and *rrp6l2-2*, respectively.

Sequence analysis. Rice (*Oryza sativa*) and poplar (*Populus trichocarpa*) sequences were identified by BlastP and retrieved from TIGR Rice Genome (http://www.tigr.org/tdb/e2k1/osa1) and the Doe Joint Genome Institute (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html), respectively. To obtain full-length sequences for poplar proteins, predictions were improved using FGENESH+ software (www.softberry.com; data available upon request). Arabidopsis sequences initially identified in TAIR (http://www.arabidopsis.org) were determined by translation of cloned cDNA sequences (Genbank accession numbers EU240662-4). PFAM (17) was used to determine protein

domains. Sequence alignment was done with Muscle v3.52 program (15). Phylogenetic analysis was performed using Phyml (21), following removal of poorly aligned and overly divergent positions using Gblocks program (8). The unrooted tree was drawn with Treedyn (11).

Yeast complementation. cDNAs for At-*RRP6L1*, *2* , *3* and yeast *RRP6* were cloned into pRS 426-TDH (40) and transformed into wild type (BY4742) and *rrp6*Δ (Euroscarf Y11777) yeast strains using the lithium acetate method. Following selection at 24°C, cells were tested for growth at 24° and 37°C.

Expression of GFP fusion proteins. Full–length cDNAs of At*RRP6L1*, *2*, and *3* were obtained by RT-PCR and fused to E-tagged eGFP in N- and C-terminal orientation in pBinH, a pBinPLUS derivative (43) containing CaMV 35S promoter and terminator sequences and a hygromycin resistance cassette. Biolistic transformation of tobacco BY2 cells and analysis by confocal microscopy was as described previously (46). For stable transformants, Arabidopsis plants were transformed by the floral dip method (12). T1 and T2 plants were selected on agar plates containing MS salts, 0.5% sucrose, and 50 μg/ml hygromycin (MSH). Stable transformants expressing GFP-fusion proteins were sown on cover slips coated with MSH agar and root tips and examined 8 days after germination by confocal microscopy.

RNA extraction and analysis. RNA was extracted using the TRI-reagent (Molecular Research Center) according to the manufacturer's manual. Polyadenylation sites were mapped by amplification of 3’ ends as described previously (34). 5' and 3' ends of the ETS were determined by circular RT-PCR (cRT-PCR) as described previously (34). A complete list of primers used in this study is available upon request.

5

Smart cDNA and virtual northern blots. Full length cDNAs were synthesized using the cDNA SMART synthesis kit (Clontech) following the manufacturer's protocol except that different sequence tags at both 3' and 5' ends of polyadenylated RNA were introduced during cDNA synthesis. Total cDNAs were amplified by PCR using the PCR Advantage II kit (Clontech) for 16 cycles (30s at 94°C, 30s at 60°C, 6min at 68°C). Diluted SMART PCR products (representing total full-length polyadenylated RNA) were used as a template for 3' and 5' RACE PCR. For virtual northern blots, SMART PCR products were separated on a 1.2 % agarose gel/ 0.5 x TAE, blotted on Hybond N+ membrane and hybridized to [α-^{32}P]dCTP labelled DNA probes.

RESULTS

A small multigene family encodes RRP6-like proteins in plants

In a previous study, aimed at identifying polyadenylated mitochondrial RNA degradation intermediates, we constructed a oligo-dT primed library from size-selected RNA isolated from *Arabidopsis thaliana* seedlings (23). Interestingly, some clones in this library matched to fragments of the polycistronic rRNA precursor encoded in the nuclear genome. This raised the interesting possibility that nuclear non-mRNA transcripts could be polyadenylated in *Arabidopsis thaliana*, a situation reminiscent of the one recently described in yeast. Degradation of yeast nuclear non-coding RNAs depends on polyadenylation by the TRAMP complex and involves RRP6, a hydrolytic exoribonuclease associated with the nuclear exosome (27, 45, 48). While in yeast and humans, RRP6 is encoded by a single gene, the Arabidopsis genome contains three genes encoding RRP6-like proteins: At1g54440, At5g35910 and At2g32415, which we refer to as At-*RRP6L1*, At-*RRP6L2* and At-*RRP6L3*, respectively. At-*RRP6L1-3* encode proteins of 637, 870 and 892 amino acids, respectively.

These three proteins are the only Arabidopsis proteins containing both the IPR002562 3'–5' exoribonuclease and the IPR002121 HDRC motifs that characterize RRP6 proteins (Fig. 1A). The PMC2NT domain, present in the N-terminus of *S. cerevisiae* Rrp6p, is only found in At-RRP6L2 (Fig. 1A).

We searched for RRP6-like sequences in other plant genomes and found that both rice and poplar contain also three genes encoding RRP6-like proteins. Phylogenetic analysis reveals that sequences form two strongly supported groups (Fig. 1B). The first group contains At-RRP6L1 and At-RRP6L2 and the second group contains At-RRP6L3. The presence of At-*RRP6L3*-like genes in rice, poplar and Arabidopsis indicates an early divergence in plants from a common ancestor. While RRP6L3 related proteins form a plant-specific subgroup, the At-RRP6L1/2 group clusters with yeast and animal RRP6-like proteins. In each of the three plants species, RRP6L1 and RRP6L2 cluster as a pair (Fig. 1B). In addition, positions of introns are clearly conserved in an intra- rather than inter-specific manner (data not shown). Therefore, we propose that the At-*RRP6L1/2* gene duplication occurred independently in these three plant species rather than anterior to the speciation of rice, poplar and Arabidopsis.

It is interesting to note that both rice and Arabidopsis genomes encode RRP6-like proteins containing or lacking the N-terminal PMC2NT domain.

In conclusion, three RRP6-like proteins are encoded in the genomes of rice, poplar and Arabidopsis. Two of these genes, represented by At-*RRP6L1* and *2*, are the closest homologues to yeast and human *RRP6* whereas the At-*RRP6L3* type appears to be specific to plants.

Subcellular distribution of RRP6-like proteins in plants

Both At-RRP6L1 and At-RRP6L2 are predicted to be nuclear by the PSORT programme (http://psort.ims.u-tokyo.ac.jp) while no specific localization is predicted for At-RRP6L3. To

address this issue experimentally, we expressed N- and C- terminal GFP-fusion proteins of At-RRP6L1, 2 and 3 in tobacco BY2 cells. Results were identical for both orientations of fusion proteins. At-RRP6L3 fusion proteins were excluded from the nucleus and located exclusively in the cytosol (Fig. 2A). In contrast, both At-RRP6L1 and 2 fused to GFP accumulated predominantly in the nucleus, although a faint cytoplasmic signal is observed for At-RRP6L2 (Fig. 2A). These results indicate that, in contrast to At-RRP6L3, both At-RRP6L1 and 2 can be targeted to the nucleus. To further study the distribution of At-RRP6-like proteins, we produced stably transformed Arabidopsis plants. The results obtained confirmed the cytoplasmic localization of At-RRP6L3 and the nuclear localization of both At-RRP6L1 and 2 (Fig. 2B). In root tip cells where nucleoli are particular large, the intra-nuclear distribution of RRP6L1 and RRP6L2 could be distinguished. The At-RRP6L1 fusion protein accumulated in both the nucleoplasm and the nucleolar vacuole (Fig. 2C top). In contrast, At-RRP6L2 accumulated predominantly in nucleoli and the signal was weaker in the nucleoplasm (Fig. 2C bottom).

Hence, At-RRP6L1 and 2 are targeted to the nucleus whereas At-RRP6L3 is restricted to the cytoplasm. The subcellular distribution thus again differentiates At-RRP6L3 from At-RRP6L1 and 2. In addition, At-RRP6L1 and 2 each have a distinct intra-nuclear distribution.

At-RRP6L1 complements the growth defect of a yeast *rrp6Δ* strain

To determine whether Arabidopsis RRP6-like proteins can functionally replace *S. cerevisiae* Rrp6p, we expressed At-RRP6L1, 2 and 3 in the temperature-sensitive *rrp6Δ* yeast strain (2). Expression of transgenes was confirmed by RT-PCR (not shown). At-RRP6L3, cytoplasmic in plants, did not support growth of *rrp6Δ* at the non-permissive temperature (Fig. 3). In contrast, At-RRP6L1 was able to complement the thermo-sensitive growth phenotype of yeast *rrp6Δ* (Fig. 3), as did a truncated Rrp6p lacking the N-terminal domain (42). This

demonstrates that At-RRP6L1 is a functional protein that can perform at least one of the biological roles of Rrp6p in yeast. Surprisingly, expression of the nuclear/nucleolar protein At-RRP6L2 (that alone shares the N-terminal PMC2NT domain with its yeast counterpart) did not restore the growth phenotype of *rrp6Δ*, although expression and stability of the protein were sufficient as determined by western analysis (not shown). One possible explanation is that At-RRP6L2 fails to associate properly with nuclear exosome complexes in yeast due to its longer C-and N-terminal domains. Nonetheless, RRP6L2 is a functional homologue of yeast Rrp6p as determined below by mutant analysis.

Characterization of mutant plants

Since RRP6L1 and 2 are putatively nuclear proteins, we wanted to test whether they could be involved in the turnover of polyadenylated non-coding RNAs transcribed in the nucleus. To this end, two independent T-DNA insertion lines for At-*RRP6L1* (*rrp6l1-1*, *rrp6l1-2*) and for At-*RRP6L2* (*rrp6l2-1*, *rrp6l2-2*) were characterized. In *rrp6l1-1,* the T-DNA is inserted into the first exon of *RRP6L1* (Fig. 4A). This insertion event led to a deletion of 29 bp of the coding region. Line *rrp6l1-2* harbors a T-DNA in the sixth exon of *RRP6L1*. No *RRP6L1* mRNA was detected by virtual northern blots in either *rrp6l1* mutant (Fig. 4B), indicating that both T-DNA insertions severely affect *RRP6L1* expression. In both *rrp6l2-1* and *rrp6l2-2*, the T-DNA is inserted in the first exon of *RRP6L2* (Fig. 4C). No *RRP6L2* mRNA was detected in *rrp6l2-1*. In *rrp6l2-2*, mRNA was slightly reduced in size but accumulated to wild-type level (Fig. 4D). We therefore mapped mRNA extremities by 5' and 3' RACE. This analysis revealed that *rrp6l2-2* mRNA consists of the entire *RRP6L2* mRNA in which the complete 5' UTR plus 29 nt of *RRP6L2* ORF is replaced by 16 nt encoded by the T-DNA. The resulting chimeric *rrp6l2-2* mRNA is 2839 nt, as compared to 2936 nt for wild-type

RRP6L2 mRNA. The *rrp6l2-2* mRNA potentially codes for a protein lacking the 86 N-terminal aminoacids as compared to wild-type RRP6L2.

Both *rrp6l2* mutants were further characterized by western blots using antibodies raised against RRP6L2. We did not detect any RRP6L2 protein in *rrp6l2-1* and observed a faint signal that could correspond to the truncated protein in *rrp6l2-2* (Fig. 4E). We therefore cannot exclude that low levels of a truncated RRP6L2 can be produced from *rrp6l2-2* mRNA. However, these results indicate that *RRP6L2* expression is severely decreased in *rrp6l2-2* and probably abolished in *rrp6l2-1*.

All four single mutants were indistinguishable from WT plants in terms of growth and development under standard growth conditions. Similarly, both homozygous double mutants, *rrp6l1-2 rrp6l2-1* and *rrp6l1-2 rrp6l2-2*, and did not show any obvious growth or developmental phenotype.

An rRNA maturation by-product accumulates in *rrp6l2* but not in *rrp6l1* mutants

We tested these insertion mutants for the accumulation of an rRNA maturation by-product, which is a typical molecular phenotype of yeast *rrp6* mutant strains (2). One of the earliest processing events of rRNA maturation is an endonucleolytic cleavage in the 5' external transcribed sequence (5' ETS). In *S. cerevisiae*, this cleavage takes place at site A0, located about 90 nt upstream of the 18S rRNA. The excised ETS accumulates in absence of functional Rrp6p. In Arabidopsis, the complete 5' ETS is relatively large (about 1,8 kb). The primary processing site (P-site) is located 1275 nt downstream of the transcription initiation site and about 560 nt upstream of the 5' extremity of the 18S rRNA (37). To our knowledge, a second processing site, equivalent to the A0 site in yeast, has not been characterized in Arabidopsis. Cleavage at such a site will release a plant rRNA maturation by-product akin to the yeast 5' ETS. We will refer to this product as 5'pETS (5' proximal ETS) to distinguish it

from the distal 5' ETS corresponding to the sequence between the transcription initiation site and the P site (Fig. 5A).

To characterize the putative 5'pETS in *Arabidopsis thaliana*, we performed circular RT-PCR (cRT-PCR) to determine both 5' and 3' extremities. Briefly, 3' and 5' ends were ligated prior to cDNA synthesis and PCR amplification of the region spanning the joined ends was done. The expected 5'pETS was amplified only when using cDNA from *rrp6l2* plants. All 5' extremities mapped exactly to the P site, and 3' ends mapped about 80 nt upstream of the 18S rRNA (Fig. 5B). Interestingly, non-encoded adenosines were present in 55% of the clones analyzed. Poly(A) tails were up to 18 nt in length with an average size of 9-10 nt. These lengths do not necessarily reflect the in vivo RNA population, since cRT-PCR may preferentially amplify RNA with shorter poly(A) tails. However, these results unequivocally show that this rRNA maturation by-product can be polyadenylated in *rrp6bl2* plants, since oligo(dT) primers were not used in cRT-PCR.

We further analyzed the polyadenylation status of the 5'pETS by 3' RACE. Oligo(dT)-primed cDNA was produced from WT and *rrp6* RNA, followed by 3' RACE with a sense primer located 210 nt upstream of the 18S rRNA. We clearly detected polyadenylated 5'pETS in *rrp6l2-1* and to a lesser extent in *rrp6l2-2*, while only a faint smear was observed for WT and *rrp6l1* samples (Fig.. 5C). Sequence analysis of the cloned PCR products from the WT sample revealed that most were PCR artefacts; only one in fifty clones actually corresponded to the polyadenylated 5'pETS. This suggests that the 5'pETS can be polyadenylated in WT plants but the level is extremely low. In contrast, all sequences from *rrp6l2* mutants corresponded to polyadenylated 5'pETS. Position and frequency of polyadenylation sites are presented for *rrp6l2-1* (Fig. 5C). Some clones mapped to regions devoid of A's and had poly(A) tails longer than the oligo$(dT)_{12}$ used for cDNA synthesis. These clones represent unambiguous polyadenylation sites. For six clones (indicated by a question mark in Fig. 5C)

we cannot exclude artificial priming by a stretch of 4 As. However, nine clones mapped precisely to the 3' site previously determined by cRT-PCR, immediately downstream of this A-rich region (AAAAG). The fact that all of these clones contained the G nucleotide indicates that cDNA synthesis was not primed by the genomically encoded A-stretch. Remarkably, poly(A) tails were slightly heteropolymeric, i.e. they contained several non-A nucleotides. These data confirm that the 5'pETS is polyadenylated in *rrp6l2* mutants.

To determine the abundance of polyadenylated 5'pETS, we performed virtual northern blots with a probe corresponding to the 5'pETS (Fig. 5D). Several RNA species were detected. The larger transcripts are rRNA precursor transcripts that are present in all samples. The smaller RNA, which is the size of the 5' pETS (450 nt), accumulates in both *rrp6l2* mutants (Fig. 5D), indicating that At-RRP6L2 is involved in degradation of the 5'pETS. Accumulation of this RNA was not observed in either of the *rrp6l1* mutants or in WT. This result indicates that the functions of RRP6L1 and RRP6L2 are distinct or, at least, do not completely overlap. This idea was reinforced by analysis of At-RRP6L1/2 double mutants. If RRP6L1 is involved in the degradation of the 5'pETS, we would expect an increased accumulation of the 5'pETS in double mutants as compared to single *rrp6l2* mutants. No increase in accumulation was observed in *rrp6l1-2 rrp6l2-1* and *rrp6l1-2 rrp6l2-2* (data not shown), strongly suggesting that the 5'pETS is a substrate for At-RRP6L2 but not for At-RRP6L1.

DISCUSSION

While composition and function of exosome complexes in yeast and humans have been investigated in depth, only a few components of plant exosomes and associated factors have been characterized to date. Arabidopsis RRP4 and RRP41 were shown to be associated in a 500 kDa complex containing (9, 10) and RR45 counterparts have been recently identified (24). These initial data reveal some interesting features. First, At-RRP41, a component of the

core exosome, is a true phosphorolytic subunit as judged by the conservation of essential residues and its *in vitro* activity (10, 14). If RRP41 retains its activity upon complex assembly, the plant exosome core complex would therefore have a similar exoribonucleolytic activity as archaeal exosomes (29), while yeast and human core subunits are catalytic inactive (14, 28). Second, at least two components of plant exosomes are not encoded by single genes: RRP44 and RRP45. Although *RRP44* genes have not been investigated to date, a recent study has now identified the two Arabidopsis RRP45-like proteins. In this case, single loss-of-function mutants were viable while simultaneous down regulation of both proteins was lethal, indicating a certain degree of redundancy (24). However, loss of At-*RRP45b* but not of At-*RRP45a* specifically affected cuticular wax biosynthesis, suggesting that their functional overlap is only partial. These results illustrate that isoforms of core components of the exosome can particularly impact a specific cellular process. In the present study, we show that plants as diverse as Arabidopsis, poplar or rice contain 3 RRP6-like proteins in contrast to single RRP6 proteins in yeast, Drosophila, trypanosomes and human. Although their association with the exosome remains to be investigated, our results show that at least two of these proteins are functional homologues of yeast RRP6. Importantly, RRP6L proteins show distinctive subcellular distribution which suggests a specialization of RRP6L proteins in Arabidopsis.

Phylogenetic analysis clearly distinguished At-*RRP6L1* and *2*, which group with human and yeast proteins, from At-*RRP6L3* that seems to be a plant-specific gene. Interestingly, AtRRP6L3-GFP fusion proteins were exclusively found in the cytoplasm in both tobacco BY2 cells and *Arabidopsis thaliana*. While yeast Rrp6p is restricted to the nucleus (2), the protein was found in both cytoplasm and nucleus of trypanosomes (22). Interestingly, in human and Drosophila, a small fraction of PM/Scl-100 and dRRP6, respectively, was also detected in the cytoplasm (7, 20). Taken together, these results suggest that RRP6-related

proteins could have a function in the cytoplasm in higher eukaryotes including plants. It remains to be determined whether the biological role of RRP6L3 is similar to that of cytoplasmic RRP6 proteins in animals and humans or whether it has a plant-specific function that could also be independent of the exosome.

In contrast to RRP6L3, both RRP6L1 and RRP6L2 share some functions with yeast Rrp6p. As indicated by their different intra-nuclear distribution and their effect on the 5'pETS, these two proteins have non-overlapping functions in plants. Only plant mutants down-regulated for AtRRP6L2 accumulated the 5'pETS, a maturation by-product of rRNA synthesis and a classical substrate for yeast Rrp6p. A role of At-RRP6L2 in the degradation of plant ETS is supported by its domain architecture and its subcellular distribution. At-RRP6L2 is the only RRP6-like protein in Arabidopsis, which contains the N-terminal PMC2NT domain. This domain mediates interaction of yeast Rrp6p with the RNA binding protein Rrp47 (42). The absence of Rrp47p impairs both 3' processing of stable RNA and degradation of the 5' ETS (31). Thus, the N-terminal region of RRP6L2 could be necessary for degradation of the ETS in plants by interaction with a cofactor analogous to Rrp47. In fact, a protein with significant homology to Rrp47p is encoded by At5g25080. In addition, the intra-nuclear distribution of the At-RRP6L2 fusion proteins coincides with the localization of the 5'ETS in nucleoli (39).

The fact that only RRP6L2 is able to degrade 5'pETS combined with the distinct subcellular distribution of RRP6L proteins suggests that RRP6-like proteins are specialized or at least not fully redundant in plants. These results reveal an intriguing complexity as compared to other organisms studied to date.

In addition, we show that the 5'pETS accumulating upon knock-down of At-RRP6L2 is polyadenylated. This observation is consistent with recent results in yeast, which show that several nuclear transcripts can be polyadenylated by the TRAMP complex and are subsequently degraded by Rrp6p (27, 45, 48). The fact that polyadenylation can trigger RNA

degradation was initially discovered in bacteria, and a similar pathway is present in both chloroplast and plant mitochondria (5, 13, 19). Our results now give further support to the idea that the ancient role of polyadenylation in marking transcripts destined for degradation is conserved in most genetic systems and operates in all three genetic compartments of a plant cell.

ACKNOWLEDGEMENTS

This work was funded by the Centre National de la Recherche Scientifique (CNRS, France), a fellowship of the Deutsche Forschungsgemeinschaft to H. L. and a PhD fellowship from the French Ministère de l'Education Nationale, de l'Enseignement supérieur et de la Recherche (MNER, France) to S. H.

We thank Uli Mühlenhoff (Marburg) and Jean-Luc Evrard (Strasbourg) for kindly providing the yeast and GFP cloning vectors, respectively.

REFERENCES

1. **Allmang, C., J. Kufel, G. Chanfreau, P. Mitchell, E. Petfalski, and D. Tollervey.** 1999. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J **18:**5399-410.
2. **Allmang, C., E. Petfalski, A. Podtelejnikov, M. Mann, D. Tollervey, and P. Mitchell.** 1999. The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. Genes Dev **13:**2148-58.
3. **Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen, P. Shinn, D. K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, C. Gadrinab, C. Heller, A. Jeske, E. Koesema, C. C. Meyers, H. Parker, L. Prednis, Y. Ansari, N. Choy, H. Deen, M. Geralt, N. Hazari, E. Hom, M. Karnes, C. Mulholland, R. Ndubaku, I.**

Schmidt, P. Guzman, L. Aguilar-Henonin, M. Schmid, D. Weigel, D. E. Carter, T. Marchand, E. Risseeuw, D. Brogden, A. Zeko, W. L. Crosby, C. C. Berry, and J. R. Ecker. 2003. Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301:**653-7.

4. **Andrulis, E. D., J. Werner, A. Nazarian, H. Erdjument-Bromage, P. Tempst, and J. T. Lis.** 2002. The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature **420:**837-41.
5. **Bollenbach, T. J., G. Schuster, and D. B. Stern.** 2004. Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. Prog Nucleic Acid Res Mol Biol **78:**305-37.
6. **Briggs, M. W., K. T. Burkard, and J. S. Butler.** 1998. Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. J Biol Chem **273:**13255-63.
7. **Brouwer, R., C. Allmang, R. Raijmakers, Y. van Aarssen, W. V. Egberts, E. Petfalski, W. J. van Venrooij, D. Tollervey, and G. J. Pruijn.** 2001. Three novel components of the human exosome. J Biol Chem **276:**6177-84.
8. **Castresana, J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17:**540-52.
9. **Chekanova, J. A., J. A. Dutko, I. S. Mian, and D. A. Belostotsky.** 2002. Arabidopsis thaliana exosome subunit AtRrp4p is a hydrolytic 3'-->5' exonuclease containing S1 and KH RNA-binding domains. Nucleic Acids Res **30:**695-700.
10. **Chekanova, J. A., R. J. Shaw, M. A. Wills, and D. A. Belostotsky.** 2000. Poly(A) tail-dependent exonuclease AtRrp41p from Arabidopsis thaliana rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells. J Biol Chem **275:**33158-66.

11. **Chevenet, F., C. Brun, A. L. Banuls, B. Jacq, and R. Christen.** 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics **7:**439.

12. **Clough, S. J., and A. F. Bent.** 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16:**735-43.

13. **Dreyfus, M., and P. Regnier.** 2002. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. Cell **111:**611-3.

14. **Dziembowski, A., E. Lorentzen, E. Conti, and B. Seraphin.** 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol **14:**15-22.

15. **Edgar, R. C.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:**113.

16. **Estevez, A. M., T. Kempf, and C. Clayton.** 2001. The exosome of Trypanosoma brucei. EMBO J **20:**3831-9.

17. **Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman.** 2006. Pfam: clans, web tools and services. Nucleic Acids Res **34:**D247-51.

18. **Fulnecek, J., and A. Kovarik.** 2007. Low abundant spacer 5S rRNA transcripts are frequently polyadenylated in Nicotiana. Mol Genet Genomics **278:**565-73.

19. **Gagliardi, D., P. P. Stepien, R. J. Temperley, R. N. Lightowlers, and Z. M. Chrzanowska-Lightowlers.** 2004. Messenger RNA stability in mitochondria: different means to an end. Trends Genet **20:**260-7.

20. **Graham, A. C., D. L. Kiss, and E. D. Andrulis.** 2006. Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci. Mol Biol Cell **17:**1399-409.

21. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:**696-704.

22. **Haile, S., M. Cristodero, C. Clayton, and A. M. Estevez.** 2007. The subcellular localisation of trypanosome RRP6 and its association with the exosome. Mol Biochem Parasitol **151:**52-8.

23. **Holec, S., H. Lange, K. Kuhn, M. Alioua, T. Borner, and D. Gagliardi.** 2006. Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. Mol Cell Biol **26:**2869-76.

24. **Hooker, T. S., P. Lam, H. Zheng, and L. Kunst.** 2007. A core subunit of the RNA-processing/degrading exosome specifically influences cuticular wax biosynthesis in Arabidopsis. Plant Cell **19:**904-13.

25. **Houseley, J., J. LaCava, and D. Tollervey.** 2006. RNA-quality control by the exosome. Nat Rev Mol Cell Biol **7:**529-39.

26. **Koonin, E. V., Y. I. Wolf, and L. Aravind.** 2001. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. Genome Res **11:**240-52.

27. **LaCava, J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, and D. Tollervey.** 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell **121:**713-24.

28. **Liu, Q., J. C. Greimann, and C. D. Lima.** 2006. Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell **127:**1223-37.

29. **Lorentzen, E., P. Walter, S. Fribourg, E. Evguenieva-Hackenberg, G. Klug, and E. Conti.** 2005. The archaeal exosome core is a hexameric ring structure with three catalytic subunits. Nat Struct Mol Biol **12:**575-81.

30. **Milligan, L., C. Torchet, C. Allmang, T. Shipman, and D. Tollervey.** 2005. A nuclear surveillance pathway for mRNAs with defective polyadenylation. Mol Cell Biol **25:**9996-10004.

31. **Mitchell, P., E. Petfalski, R. Houalla, A. Podtelejnikov, M. Mann, and D. Tollervey.** 2003. Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. Mol Cell Biol **23:**6982-92.

32. **Mitchell, P., E. Petfalski, A. Shevchenko, M. Mann, and D. Tollervey.** 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. Cell **91:**457-66.

33. **Orban, T. I., and E. Izaurralde.** 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA **11:**459-69.

34. **Perrin, R., H. Lange, J. M. Grienenberger, and D. Gagliardi.** 2004. AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria. Nucleic Acids Res **32:**5174-82.

35. **Raijmakers, R., G. Schilders, and G. J. Pruijn.** 2004. The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm. Eur J Cell Biol **83:**175-83.

36. **Rosso, M. G., Y. Li, N. Strizhov, B. Reiss, K. Dekker, and B. Weisshaar.** 2003. An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol **53:**247-59.

37. **Saez-Vasquez, J., D. Caparros-Ruiz, F. Barneche, and M. Echeverria.** 2004. A plant snoRNP complex containing snoRNAs, fibrillarin, and nucleolin-like proteins is

competent for both rRNA gene binding and pre-rRNA processing in vitro. Mol Cell Biol **24:**7284-97.

38. **Scholl, R. L., S. T. May, and D. H. Ware.** 2000. Seed and molecular resources for Arabidopsis. Plant Physiol **124:**1477-80.

39. **Shaw, P. J., M. I. Highett, A. F. Beven, and E. G. Jordan.** 1995. The nucleolar architecture of polymerase I transcription and processing. EMBO J **14:**2896-906.

40. **Sikorski, R. S., and P. Hieter.** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122:**19-27.

41. **Slomovic, S., D. Laufer, D. Geiger, and G. Schuster.** 2006. Polyadenylation of ribosomal RNA in human cells. Nucleic Acids Res **34:**2966-75.

42. **Stead, J. A., J. L. Costello, M. J. Livingstone, and P. Mitchell.** 2007. The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. Nucleic Acids Res **35:**5556-67.

43. **van Engelen, F. A., J. W. Molthoff, A. J. Conner, J. P. Nap, A. Pereira, and W. J. Stiekema.** 1995. pBINPLUS: an improved plant transformation vector based on pBIN19. Transgenic Res **4:**288-90.

44. **Vanacova, S., and R. Stefl.** 2007. The exosome and RNA quality control in the nucleus. EMBO Rep **8:**651-7.

45. **Vanacova, S., J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, and W. Keller.** 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol **3:**e189.

46. **Vetter, G., J. M. Hily, E. Klein, L. Schmidlin, M. Haas, T. Merkle, and D. Gilmer.** 2004. Nucleo-cytoplasmic shuttling of the beet necrotic yellow vein virus RNA-3-encoded p25 protein. J Gen Virol **85:**2459-69.

47. **West, S., N. Gromak, C. J. Norbury, and N. J. Proudfoot.** 2006. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. Mol Cell **21:**437-43.

48. **Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri, and A. Jacquier.** 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell **121:**725-37.

FIGURE LEGENDS

Fig. 1. Analysis of RRP6-like proteins. (A) Conservation of functional domains in RRP6 and RRP6-like proteins. Comparison of *S. cerevisiae* (Sc Rrp6p) and *H. sapiens* (Hs PM/Scl-100) and the three *Arabidopsis thaliana* RRP6-like proteins (At RRP6L1, At RRP6L2 and At RRP6L3). Percent identity and, in parentheses, similarity with Sc Rrp6p are given below each domain drawn as boxes. (B) Phylogenetic analysis of RRP6-like proteins presented as an unrooted maximum likelihood tree. Bootstrap values above 70 (using 100 replications) are indicated along branches. The scale bar indicates the evolutionary distance (amino acid substitutions per site). Pt, *Populus trichocarpa*; At, *Arabidopsis thaliana*; Os, *Oriza sativa*; Sc, *Saccharomyces cerevisiae*; Ce, *Caenorhabditis elegans*; Hs *Homo sapiens*; Dm, *Drosophila melanogaster*.

Fig. 2. Subcellular distribution of RRP6-like proteins. (A) GFP-Fluorescence (left panel) and Nomarski (right panel) of tobacco BY2 cells transiently expressing eGFP and eGFP-fusion proteins. GFP-GUS encodes a large protein that cannot enter the nuclear compartment by passive diffusion. (B) Root tips of stably transformed Arabidopsis plants expressing RRP6L-eGFP fusion proteins. (C) Enlarged view of root tips of transformed Arabidopsis plants showing the intra-nuclear distribution of fusion proteins. Comparison of fluorescence and Nomarski panel shows that RRP6L1-eGFP (top) is mainly in the nucleoplasm and in the nucleolar vacuole. RRP6L2-eGFP is detected mainly inside the nucleolus (bottom). Cy, Cytoplasm; Np, Nucleoplasm; Nu, Nucleolus. Size bar = 10μm.

Fig. 3. At-RRP6L1 complements yeast *rrp6Δ*. Growth at the non-permissive temperature (37°C) of wild type (WT) and *rrp6Δ* yeast strains harbouring empty vector (vec), or vectors

encoding either *S. cerevisiae* Rrp6p (Sc-*RRP6*) or the indicated Arabidopsis RRP6-like proteins.

Fig. 4. Characterization of *rrpl1* and *2* TDNA-insertion mutants. (A) Diagram of the intron-exon structure of *RRP6L1*. Exons in black, introns in white. 5' and 3' UTRs are drawn as grey blocks. T-DNA-insertion sites for *rrp6l1-1* and *2* are shown. In *rrp6l1-2*, the T-DNA insertion removed 29 nt from the genomic sequence. (B) Virtual northern blot of samples from wildtype and mutant plants probed for RRP6L1 mRNA. A probe for RBP1A mRNA was used as a loading control. (C) Organization and T-DNA insertion sites in *RRP6L2*. (D) Virtual northern blot analysis of RRP6L2 mRNA in wild-type and mutant plants shows a size difference between transcripts in wild-type and *rrp6b-2* mutants. (E) Western blot of proteins extracted from wild-type, *rrp6l21-1* or *rrp6l2-2* seedlings were probed with antibodies against At-RRP6L2 (left). The truncated protein encoded by the mutant *rrp6b-2* allele is indicated by an arrow. As a loading control, the membrane was stained with Coomassie blue (right).

Fig. 5. A polyadenylated rRNA maturation by-product accumulates in *rrp6l2* mutants (A) Diagram showing the 5' region of the polycistronic rRNA transcript of *A. thaliana*. Distal (5'dETS) and proximal (5'pETS) RNA segments and processing sites P and P', respectively, are indicated. The promoter is shown by a bent arrow. (B) Mapping of 5' and 3' ends of the 5'pETS by cRT-PCR. Total RNA from *rrp6l2-1* plants was self-ligated by T4 RNA ligase and cDNA was synthesized using a gene specific reverse primer. The same primer was combined with a gene-specific forward primer to amplify joined 5' and 3' ends by PCR. Primers are indicated on the diagram. 5' ends are shown above the diagram, 3' ends are shown below, and non-encoded nucleotides at 3' ends of 5' pETS transcripts are indicated. (C) Characterization of the 5' pETS by 3' RACE. Oligo$(dT)_{12}$-primed cDNA was synthesized

from total RNA from wild-type or mutant plants. 3' ends were then amplified by PCR using a gene-specific forward primer (arrow above the diagram) and a reverse primer specific for the oligo(dT) primer adapter sequence. PCR products were analysed by electrophoresis (top). PCR products obtained from *rrp6l2* samples were cloned and sequenced to map polyadenylation sites (bottom). Location and frequency of polyadenylation sites are indicated on the sequence, and sizes and nucleotide composition of poly(A) tails are given below. RT, reverse transcriptase. (D) Virtual northern blots of wild-type and *rrp6l2* (left) or wild type and *rrp6l1* mutants (right) hybridized with a DNA probe corresponding to the 5' pETS.

24

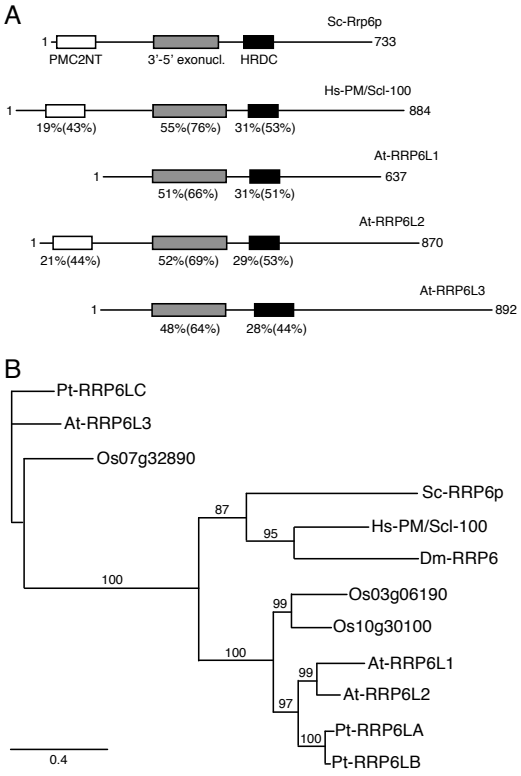

A
Sc-Rrp6p
1
733
PMC2NT
3'-5' exonucl.
HRDC
Hs-PM/Scl-100
1
884
19%(43%)
55%(76%)
31%(53%)
At-RRP6L1
1
637
51%(66%)
31%(51%)
At-RRP6L2
1
870
21%(44%)
52%(69%)
29%(53%)
At-RRP6L3
1
892
48%(64%)
28%(44%)
B
Pt-RRP6LC
At-RRP6L3
Os07g32890
Sc-RRP6p
Hs-PM/Scl-100
Dm-RRP6
Os03g06190
Os10g30100
At-RRP6L1
At-RRP6L2
Pt-RRP6LA
Pt-RRP6LB
87
95
100
99
100
99
97
100
0.4

Figure 1

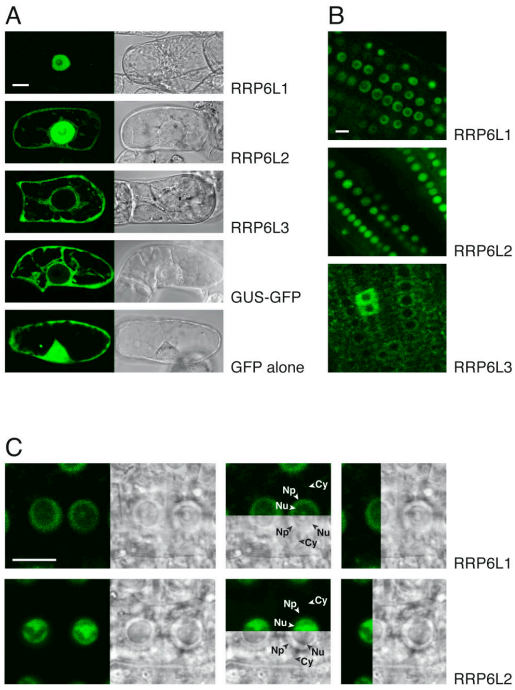

A
B
RRP6L1
RRP6L2
RRP6L3
GUS-GFP
GFP alone
RRP6L1
RRP6L2
RRP6L3
C
Np
Cy
Nu
RRP6L1
RRP6L2

Figure 2

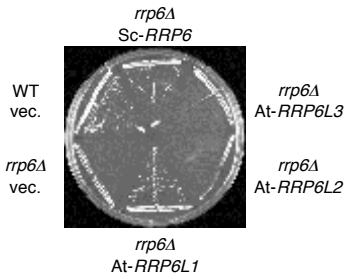

rrp6Δ
Sc-RRP6
WT
vec.
rrp6Δ
At-RRP6L3
rrp6Δ
vec.
rrp6Δ
At-RRP6L2
rrp6Δ
At-RRP6L1

Figure 3

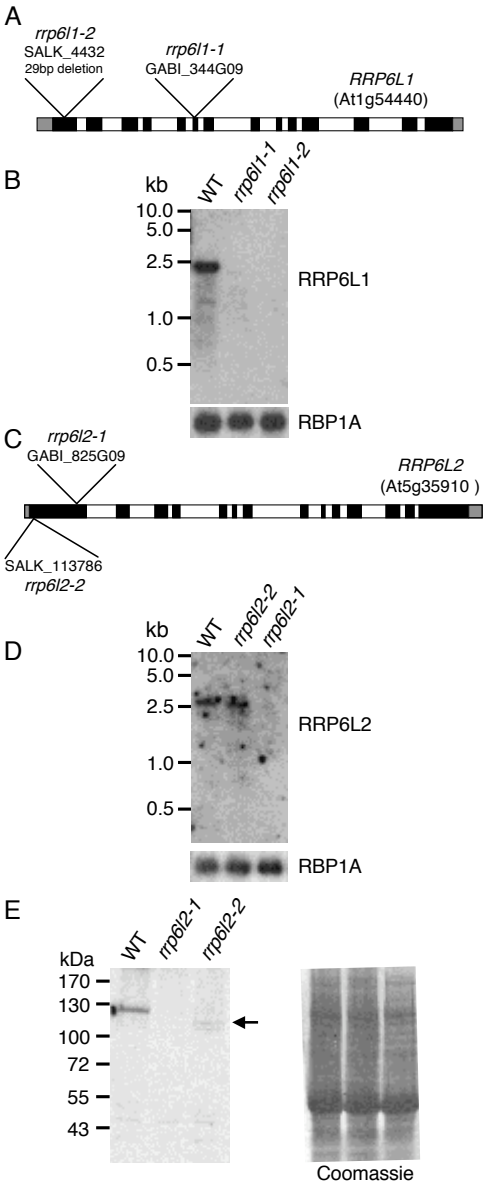

A
rrp6l1-2
SALK_4432
29bp deletion
rrp6l1-1
GABI_344G09
RRP6L1
(At1g54440)
B
kb
WT
rrp6l1-1
rrp6l1-2
10.0
5.0
2.5
1.0
0.5
RRP6L1
RBP1A
C
rrp6l2-1
GABI_825G09
RRP6L2
(At5g35910)
SALK_113786
rrp6l2-2
D
kb
WT
rrp6l2-2
rrp6l2-1
10.0
5.0
2.5
1.0
0.5
RRP6L2
RBP1A
E
kDa
WT
rrp6l2-1
rrp6l2-2
170
130
100
72
55
43
Coomassie

Figure 4

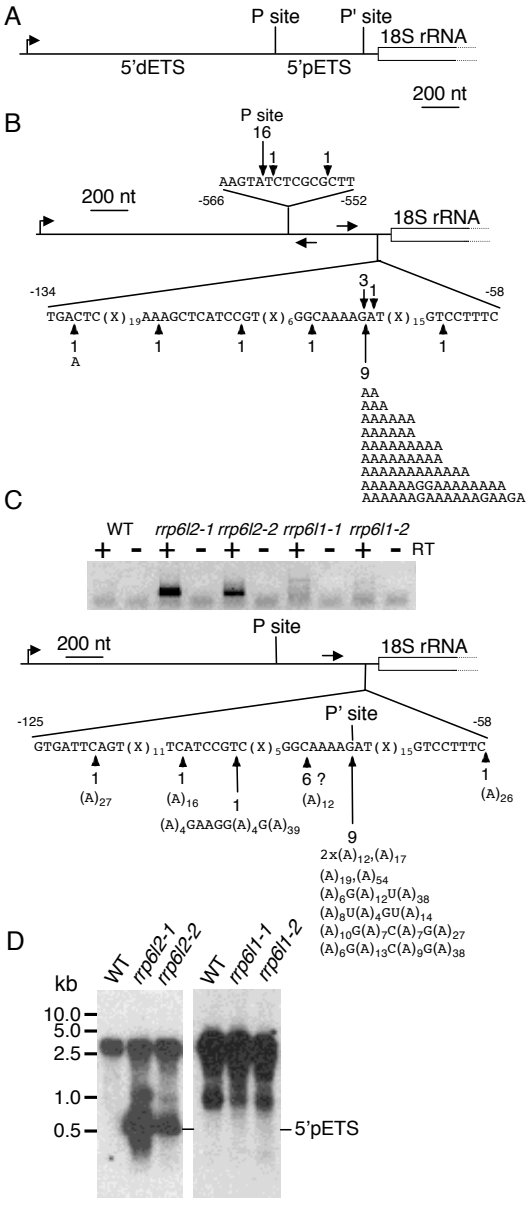

A
P site
P' site
18S rRNA
5'dETS
5'pETS
200 nt
B
P site
16
1
1
AAGTATCTCGCGCTT
-566
-552
200 nt
18S rRNA
-134
-58
3
1
TGACTC(X)19AAAGCTCATCCGT(X)6GGCAAAAGAT(X)15GTCCTTTC
1
A
1
1
1
1
9
AA
AAA
AAAAAA
AAAAAA
AAAAAAAAA
AAAAAAAAA
AAAAAAAAAAAA
AAAAAAGGAAAAAAAA
AAAAAAGAAAAAAGAAGA
C
WT
rrp6l2-1 rrp6l2-2 rrp6l1-1 rrp6l1-2
+ - + - + - + - + - RT
P site
200 nt
18S rRNA
-125
P' site
-58
GTGATTCAGT(X)11TCATCCGTC(X)5GGCAAAAGAT(X)15GTCCTTTC
1
(A)27
1
(A)16
1
(A)4GAAGG(A)4G(A)39
6 ?
(A)12
9
2x(A)12,(A)17
(A)19,(A)54
(A)6G(A)12U(A)38
(A)8U(A)4GU(A)14
(A)10G(A)7C(A)7G(A)27
(A)6G(A)13C(A)9G(A)38
1
(A)26
D
kb
WT
rrp6l2-1
rrp6l2-2
WT
rrp6l1-1
rrp6l1-2
10.0
5.0
2.5
1.0
0.5
5'pETS

Figure 5

Polyadenylation and RNA degradation : from mitochondria to the nucleus of Arabidopsis thaliana.

PNPase was previously shown to be essential for degradation of plant mitochondrial mRNAs and rRNAs precursors and maturation intermediates. All substrates of PNPase are polyadenylated. The aim of this work was to obtain a more comprehensive identification of substrate RNAs through analysis of a library of polyadenylated RNAs from plants lacking mitochondrial PNPase. The degrading role of PNPase could be generalized in removing maturation by-products of rRNAs and tRNAs. Moreover, we determined a novel role of PNPase in degrading spurious transcripts generated by a relaxed control of transcription in the context of a general surveillance pathway in plant mitochondria. The analysis of polyadenylated RNAs allowed us to identify a mitochondrial transcript that we used as a model to study a 3' processing event that is independent of PNPase. We demonstrated that a nucleus-encoded specificity *trans* factor is necessary for this *NCO* transcript 3' maturation and that processing and stabilization processes can be uncoupled in plant mitochondria. The presence of sequences corresponding to nuclear RNAs in our library of polyadenylated clones prompted us to investigate the possibility of a degradation pathway involving polyadenylation in plant nuclei. We identified 3 genes encoding RRP6L1, RRP6L2 and RRP6L3 proteins respectively. RRP6L3 appears to be a plant-specific protein and is located in the cytoplasm, whereas RRP6L1 and RRP6L2, that occupy different territories of the nucleus, are more closely related to yeast Rrp6p, a nuclear subunit of the exosome. The accumulation of a polyadenylated rRNA maturation by-product upon down-regulation of RRP6L2 supports the existence of a polyadenylation mediated RNA degradation pathway in the plant nucleus.

Polyadénylation et dégradation de l'ARN : des mitochondries au noyau d'Arabidopsis thaliana.

Le rôle de la PolyNucleotide Phosphorylase (PNPase) mitochondriale de plante dans la dégradation de précurseurs d'ARNm et ARNr et sous-produits de leur maturation avait été montré précédemment. Tous les substrats de la PNPase sont polyadénylés. Le but de ce travail était d'avoir une idée plus exhaustive de l'identité des substrats de la PNPase par l'analyse d'une banque de cDNAs polyadénylés à partir de plantes déplétées pour la PNPase mitochondriale. Ainsi, nous avons pu généraliser le rôle de la PNPase dans l'élimination des sous-produits de la maturation des ARNr et ARNt. De plus, nous avons décrit un nouveau rôle de la PNPase dans la dégradation de transcrits illégitimes générés par un contrôle relâché de la transcription, dans le contexte d'une surveillance par défaut dans les mitochondries de plantes. L'analyse des ARN polyadénylés nous a permis d'identifier un transcrit modèle pour l'étude de la maturation en 3', indépendamment de la PNPase. Nous avons démontré qu'un *trans* facteur spécifique était nécessaire à la maturation en 3' de ce transcrit, appelé *NCO* et que les processus de maturation et de stabilisation pouvaient être découplés dans la mitochondrie de plante. La présence de séquences correspondant à des ARN nucléaires dans la banque de clones polyadénylés nous a poussé à étudier la possibilité de l'existence d'une voie de dégradation des ARNs impliquant la polyadénylation dans les noyaux de plantes. Nous avons identifié 3 genes codant pour les protéines RRP6L1, RRP6L2 et RRP6L3, respectivement. RRP6L3 semble être spécifique aux plantes et est localisée dans le cytoplasme, tandis que RRP6L1 et RRP6L2, qui occupent des territoires distincts du noyau, sont plus proches de la protéine Rrp6p de levure, une sous-unité nucléaire de l'exosome. L'accumulation d'un sous-produit polyadénylé de la maturation d'un ARNr lors de la sous-expression de RRP6L2 corrobore l'existence d'une voie de dégradation des ARNs assistée par la polyadénylation dans le noyau des plantes.