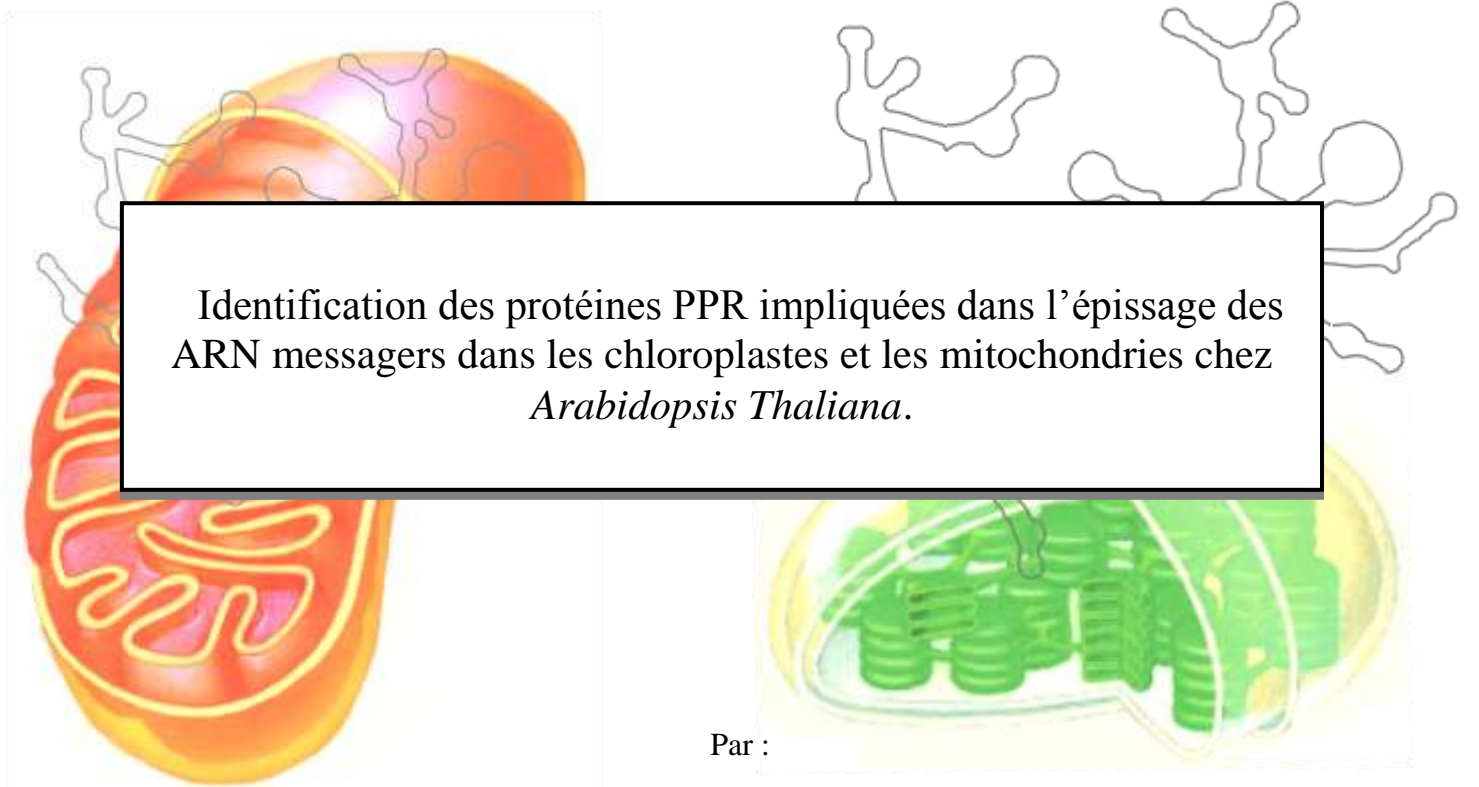


THESE EN COTUTELLE

Présentée pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE D'EVRY VAL D'ESSONNE

Et le grade de
DOCTEUR DE L'UNIVERSITE DE WESTERN AUSTRALIA

Spécialité : Biologie cellulaire et moléculaire



Identification des protéines PPR impliquées dans l'épissage des
ARN messagers dans les chloroplastes et les mitochondries chez
Arabidopsis Thaliana.

Par :

Andeol Alexis FALCON DE LONGEVIALLE

Le jury est composé de:

Claire LURIN

Ian SMALL

Francis-André WOLLMAN

José GUALBERTO

Christian SCHMITZ-LINNEWEBER

Bénédicte STURBOIS

Dominique GAGLIARDI

Martin CRESPI

Directrice de thèse

Directeur de thèse

Rapporteur UEVE et UWA

Rapporteur UEVE

Examineur

Examinatrice

Rapporteur UWA

Rapporteur UWA

Préambule :

Cette thèse est le résultat d'un cursus peu habituel. En effet, une grande partie des travaux qui sont présentés ont été réalisés alors que je n'étais pas encore inscrit en thèse.

Après l'obtention de mon diplôme de DESS intitulé « Productivité Végétale-Biotechnologie-Génome » réalisé à l'université Denis Diderot PARIS VII, j'ai intégré l'équipe de Ian Small et de Claire Lurin en tant qu'Ingénieur d'étude sur le projet européen AGRIKOLA. C'est à la fin de ces trois années que j'ai commencé peu à peu à m'immerger dans le monde des PPR. Puis j'ai suivi Ian Small qui est parti en Australie diriger le centre Plant Energy Biology à Perth. C'est sous un contrat de Post-doctorant mais sans thèse... que j'ai effectué mes recherches sur le rôle des protéines PPR dans l'épissage des ARNm. Suite à ces deux années enrichissantes, j'ai réintégré l'équipe de Claire Lurin à Evry sous un nouveau contrat européen, AGRONOMICS, sur la mise en place d'une technique d'interaction protéine-protéine in vivo chez arabidopsis, le Split-DHFR.

Dans l'optique de valoriser le travail que j'avais réalisé en Australie, j'ai donc décidé de m'inscrire en thèse en co-tutelle entre l'université d'Evry Val d'Essonne et l'université de Western Australia. Ainsi pendant les deux dernières années, j'ai donc concilié à la fois le travail sur le Split-DHFR et mon travail de thèse sur les protéines PPR.

Cette thèse est donc le fruit d'un peu plus de quatre ans dans le monde des PPR, deux ans en Australie et les deux dernières en France dans le laboratoire de Claire Lurin.

Remerciements :

Je tiens tout d'abord à remercier Claire et Ian pour m'avoir donné chacun l'opportunité de travailler sur ces protéines PPR, de m'avoir encadré tout en me donnant une grande liberté de recherche.

Je remercie également tous les membres du jury qui ont pris le temps de lire ce manuscrit.

En grand Merci à ma chère et tendre épouse, Maud qui m'a accompagné et soutenu pendant ces quatre années, même si pour elle ce que je fais reste en grande partie incompréhensible.

Un autre merci à Paloma, ma première fille et à ma toute dernière Bianca. Sans elles, la rédaction de ce manuscrit n'aurait pas été autant un challenge...

Merci à Anne-Laure, l'experte en édition !

Merci à Etienne Delannoy pour toutes nos discussions enrichissantes sur le monde des PPR et son aide précieuse. Welcome Back !

Merci à Etienne Meyer, mon partenaire de squash, pour son expertise et son aide sur le complexe I.

Merci à Charles pour son aide au début de la caractérisation des premiers mutants.

Merci également à Cathie qui est devenue l'experte en Epissage, Anne, Kristina, Fred, Sandra, Nick Taylor et Nick O'Toole à Perth. Et sans oublier Jenny et Jude et leurs agréables sourires !!!

Un grand merci aussi à Alex et Laure, qui ont été mes binômes pendant un moment et qui sont toujours disponibles et prêtes à me rendre service.

Merci aussi à Dario, mon binôme du split que j'ai laissé un peu tomber au moment de la rédaction...

Merci à Clément, à Richard, à Sandrine et à tout le reste de l'équipe Génomique fonctionnelle.

Merci à Geert De Jaeger et Mike Bevan pour leur collaboration dans ce travail.

Enfin un dernier merci à toutes les personnes qui de loin ou de près m'ont aidé pendant cette thèse et que je n'ai pas cité ici.

SOMMAIRE

| | |
|--|-----------|
| Sommaire..... | 3 |
| Principales Abréviations :..... | 7 |
| CHAPITRE I : INTRODUCTION | 8 |
| I.1 L'épissage des ARN messagers | 8 |
| I.1.1 Définition..... | 8 |
| I.1.2 Origine de l'épissage | 8 |
| I.1.3 Rôle de l'épissage | 9 |
| I.1.4 L'épissage des ARN messagers dans les organites | 9 |
| <i>Article 1 : de Longevialle A.F., Small I.D. and Lurin C. (2010) Nuclearly Encoded Splicing Factors Implicated in RNA Splicing in Higher Plant Organelles. Molecular Plant, sous presse. ...</i> | <i>9</i> |
| I.2 La famille des protéines PPR (Pentatricopeptide Repeat) | 25 |
| I.2.1 Caractéristiques des PPRs | 25 |
| I.2.1.1 Description des motifs PPRs | 25 |
| I.2.1.2 Composition de la famille | 26 |
| I.2.1.3 Distribution de la famille..... | 27 |
| I.2.1.4 Données de localisation subcellulaire | 29 |
| I.2.2 Rôle des protéines PPR..... | 29 |
| I.2.2.1 Les protéines PPR dans l'initiation de la transcription | 32 |
| I.2.2.2 Les protéines PPR dans la maturation des ARNm | 33 |
| I.2.2.2.1 <i>Maturation en 5' et 3' des ARNm.....</i> | <i>33</i> |
| I.2.2.2.2 <i>L'édition des ARNm.....</i> | <i>34</i> |
| I.2.2.2.3 <i>L'épissage des ARNm.....</i> | <i>36</i> |
| I.2.2.3 Les PPRs dans la stabilisation des ARNm et la traduction | 36 |
| I.2.2.4 Les protéines PPR impliquées dans la restauration de la fertilité mâle cytoplasmique (CMS) | 38 |
| Présentation du sujet de thèse :..... | 40 |
| CHAPITRE II : RESULTATS..... | 41 |
| II.1 Identification de protéines PPR impliquées dans l'épissage des ARNm dans les chloroplastes..... | 41 |
| II.1.1 OTP51..... | 41 |
| <i>Article 2 : de Longevialle A.F., Hendrickson L., Taylor N.L., Delannoy E., Lurin C., Badger M., Millar A.H. and Small I.D. (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. The Plant Journal, 56: 157 – 168.</i> | <i>41</i> |
| II.1.2 Identification de défauts d'épissage en utilisant la PCR quantitative. | 62 |
| II.1.2 Identification de défauts d'épissage en utilisant la PCR quantitative. | 62 |
| II.1.2.1 OTP52 (At3g06430, emb2750)..... | 64 |

| | | |
|------------|--|------------|
| II.1.2.2 | Profil d'épissage des mutants affectés dans la transcription de type PEP66 | |
| II.2. | Identification de protéines PPR impliquées dans l'épissage des ARNm dans les mitochondries..... | 69 |
| II.2.1 | OTP43..... | 69 |
| | <i>Article 3 : de Longevialle A.F., Meyer E.H., Andrés C., Taylor N.L., Lurin C., Millar A.H. and Small I.D. (2007) The Pentatricopeptide Repeat Gene OTP43 Is Required for trans-Splicing of the Mitochondrial nad1 Intron 1 in Arabidopsis thaliana. The Plant Cell, 19:3256-3265.....</i> | 69 |
| II.2.2 | TANG2 | 84 |
| | <i>Article 4 : Li Y., Smith C., Colas des Francs-Small C., de Longevialle A.F., Corke F., Zheng L., Tanz S., Small I.D. and Bevan M.W. (2010) TANG2, a Glucose-Responsive Gene, Encodes a Pentatricopeptide Repeat Protein Required for Splicing of Mitochondrial nad5 Transcripts, in submission.....</i> | 84 |
| II.2.3 | PPR40 | 110 |
| II.2.4 | Comment identifier de nouvelles PPRs impliquées dans l'épissage des introns nad ? | 112 |
| II.2.4.1 | Co-régulation des PPRs..... | 112 |
| II.2.4.2 | Les nouveaux facteurs d'épissage identifiés | 113 |
| II.2.4.2.1 | « Pure » PPRs (OTP439)..... | 115 |
| II.2.4.2.2 | Influence de l'édition dans l'épissage..... | 117 |
| II.2.4.2.3 | Effet du sucre sur l'épissage | 120 |
| II.3 | L'épissage alternatif des PPR..... | 122 |
| II.3.1 | Epissage alternatif identifié dans les banques d'EST | 124 |
| II.3.2 | Analyse préliminaire des 48 PPRs « du laboratoire » | 128 |
| II.3.3 | Cas d'OTP43 et de ses différentes isoformes | 132 |
| II.3.2.1 | Identification des isoformes par RT-PCR | 132 |
| II.3.2.2 | Preuves de traduction des différents transcrits : | 134 |
| II.3.2.3 | Complémentation fonctionnelle avec les isoformes | 137 |
| II.4 | Identification de partenaires des protéines PPR | 140 |
| II.4.1 | Méthode utilisée : TAP-TAG | 141 |
| II.4.2 | Partenaires d'OTP43 | 142 |
| II.4.3 | Partenaires de DYW1 | 144 |
| II.4.4 | Partenaires de CLB19 | 144 |
| | CHAPITRE III : CONCLUSION et DISCUSSION GENERALE..... | 148 |
| III.1 | Les protéines PPR, des facteurs d'épissage intron spécifique : | 148 |
| III.2 | L'épissage alternatif, un moyen d'augmenter le nombre de protéines PPR : | 151 |
| III.3 | Rôle de l'épissage alternatif : une nouvelle hypothèse sur la reconnaissance de multi-sites d'édition par une seule protéine PPR ? | 152 |
| III.4 | Action des protéines PPR, une action seule ou en partenariat : | 154 |
| III.5 | Rôle et régulation des protéines PPR | 155 |

| | | |
|---|--|------------|
| III.5.1 | Régulation transcriptionnelle et post-transcriptionnelle : | 155 |
| III.5.2 | Régulation post-traductionnelle : | 156 |
| CHAPITRE IV : MATERIELS ET METHODES..... | | 159 |
| IV.1 | Matériels..... | 159 |
| IV.1.1 | Souches bactériennes | 159 |
| IV.1.2 | Oligonucleotides utilisés au cours de ma thèse et non répertoriés dans les articles. 159 | |
| IV.1.2 | Lignées d'insertion..... | 161 |
| IV.2 | Méthodes | 162 |
| IV.2.1 | Méthodes relatives à la culture d'Arabidopsis | 162 |
| IV.2.1.1 | Culture en terre | 162 |
| IV.2.1.2 | Culture <i>in vitro</i> | 162 |
| IV.2.1.3 | Préparation des graines d'Arabidopsis | 162 |
| | <i>IV.2.1.3.1 Vernalisation.....</i> | <i>162</i> |
| | <i>IV.2.1.3.2 Stérilisation des graines.....</i> | <i>162</i> |
| IV.2.2 | Protocoles relatifs à l'ADN | 163 |
| IV.2.2.1 | Extraction rapide d'ADN génomique | 163 |
| IV.2.2.2 | Extraction d'ADN génomique (Dneasy 96 Plant kit QIAGEN)..... | 163 |
| IV.2.2.3 | Amplification d'ADN par PCR..... | 164 |
| | <i>IV.2.2.3.1 Méthode de PCR « classique ».....</i> | <i>164</i> |
| | <i>IV.2.2.3.2 Méthode de PCR « clonage ».....</i> | <i>165</i> |
| IV.2.2.3 | PCR quantitative..... | 165 |
| IV.2.2.3 | Electrophorèse | 165 |
| IV.2.2.4 | Purification de produits PCR en vue d'un clonage ou d'un séquençage 166 | |
| | <i>IV.2.2.4.1 Utilisation du kit QIAquick (Qiagen).....</i> | <i>166</i> |
| | <i>IV.2.2.4.2 Purification au PEG.....</i> | <i>166</i> |
| IV.2.2.5 | Clonage de produit PCR..... | 166 |
| | <i>IV.2.2.5.1 Clonage dans le vecteur pGEMT-Easy (Promega).....</i> | <i>166</i> |
| | <i>IV.2.2.5.2 Technologie de clonage GatewayTM (Invitrogen)</i> | <i>166</i> |
| | <i>IV.2.2.5.2.1 Principe du clonage Gateway.....</i> | <i>166</i> |
| | <i>IV.2.2.5.2.2 PCR de clonage.....</i> | <i>166</i> |
| | <i>IV.2.2.5.2.3 Mutagenèse dirigée par PCR en vue d'un clonage dans un vecteur d'entrée du système Gateway 167</i> | |
| | <i>IV.2.2.5.2.4 Réaction de BP clonase</i> | <i>168</i> |
| | <i>IV.2.2.5.2.5 Réaction de LR clonase</i> | <i>168</i> |
| IV.2.2.5.3 | Transformation bactérienne, sélection et test des colonies transformées. 168 | |

INTRODUCTION

| | | |
|--------------|---|------------|
| IV.2.2.5.3.1 | Protocole de préparation de cellules DH5a Thermo-compétentes..... | 168 |
| IV.2.2.5.3.2 | Protocole de transformation par choc thermique de cellules DH5a compétentes 169 | |
| IV.2.2.5.4 | Transformation d'Agrobactéries, sélection et test des colonies transformées 169 | |
| IV.2.2.5.4.1 | Protocole de préparation d'agrobactéries thermocompétentes | 169 |
| IV.2.2.5.4.2 | Protocole de transformation par choc thermique de cellules C58C1 compétentes 169 | |
| IV.2.2.6 | Extraction et purification de plasmides (minipréparation d'ADN)..... | 170 |
| IV.2.2.7 | Séquençage d'ADN | 170 |
| IV.2.3 | Méthodes relatives à l'ARN..... | 170 |
| IV.2.3.1 | Extraction d'ARN total..... | 170 |
| IV.2.3.2 | Traitement de décontamination de l'ADN contaminant..... | 170 |
| IV.2.3.3 | Réaction de transcription « inverse » | 170 |
| IV.2.3.4 | Northern blot..... | 170 |
| IV.2.3.5 | Protocole pour l'analyse PPE | 171 |
| IV.2.4 | Méthodes propres à Arabidopsis..... | 171 |
| IV.2.4.1 | Transformation d'Arabidopsis par « floral deeping » | 171 |
| IV.2.4.2 | Etude du phénotype | 171 |
| | Références Bibliographiques | 172 |
| | ANNEXES :..... | 182 |

PRINCIPALES ABREVIATIONS :

ADN : Acide Desoxyribonucléique
ADNc : ADN complémentaire
ADNg : ADN génomique
AGRIKOLA : Arabidopsis Genomic RNAi Knock-out Line Analysis
ARN : Acide Ribonucléique
ARNm : ARN messenger
ARNi : ARN interférence
ARNr : ARN ribosomique
ARNt : ARN de transfert
BET : Bromure d'ETHidium
CMS : "Cytoplasmic Male Sterility", stérilité mâle cytoplasmique
dNTP : désoxyribonucléotide
EST : Expressed Sequence Tag
GFP : Green Fluorescent Protein
GST : Gene-specific Sequence Tag
IPTG : Isopropyl-Beta-Thio-Galactoside
MS (1/2MS) : milieu Murashige et Skoog
NEP : Nuclear Encoded Polymerase
NMD : nonsense mediated mRNA decay
ORF : Open Reading Frame
PCR : "Polymerase chain reaction", réaction en chaîne par polymérase
PEP : "Plastidial Encoded Polymerase", polymérase chloroplastique codée par le génome
PLS : sous-famille de protéines PPR
PPE : Poisoned Primer Extension
PPR : Pentatricopeptide Repeat
PTC : Premature Termination Codon
qPCR : PCR quantitative
RIP-CHIP : RNA ImmunoPrecipitation and DNA CHIP hybridization
RRM : RNA Recognition Motifs (motif de reconnaissance de l'ARN)
RT-PCR : réaction de transcription inverse suivie d'une PCR
RUST : Regulation by Unproductive Splicing and Translation mechanism
snRNP : Small Nuclear Ribonucleoprotein
TAIR : The Arabidopsis Information Resource
UTR : UnTranslated Region (région non traduite)

Nomenclature :

Pour un gène nucléaire, la nomenclature suivante est utilisée :

Gène : *NUC*, mutant : *nuc*, protéine : NUC

Pour un gène cytoplasmique, la nomenclature bactérienne est adoptée :

Gène : *cyT*, protéine : CyT

CHAPITRE I : INTRODUCTION

I.1 L'épissage des ARN messagers

I.1.1 Définition

L'épissage est un processus de maturation post-transcriptionnelle par lequel les pré-ARN transcrits à partir de l'ADN génomique peuvent subir des étapes de coupure et ligature conduisant à la suppression de certaines régions dans l'ARN mature. Ce phénomène permet l'assemblage des séquences codantes, les exons, mais aussi l'assemblage de séquences non codantes comme des transcrits d'ARNt contenant des introns.

L'épissage est assuré par un ensemble de complexes ribonucléoprotéiques, appelés les spliceosomes. Chaque complexe snRNP pour Small Nuclear Ribonucleoprotein, contient un ARN et plusieurs protéines. Cet ensemble de protéines reconnaît des séquences spécifiques et permet la mise en place d'une structure secondaire optimale de l'ARN pour l'épissage.

Il existe également des introns appelés auto-épissables ou auto-catalytiques, c'est-à-dire capable de s'exciser sans intervention d'un spliceosome, ceux-ci ont été décrits dans les mitochondries, les plastes et certaines bactéries. Cependant, ce n'est pas la règle générale dans les mitochondries et les chloroplastes, chez lesquels la majorité des introns nécessitent l'intervention de protéines codées par le génome nucléaire et importées dans les organites.

Par analogie avec le fonctionnement du ribosome, il a été proposé que, dans le spliceosome, c'est l'ARN qui est catalytique et donc que le spliceosome est un ribozyme. Le mécanisme catalytique exact du spliceosome est encore inconnu.

I.1.2 Origine de l'épissage

Dès la découverte des introns en 1977, trois questions ont été immédiatement posées : Quelle est l'origine de ces introns ? Quel est leur rôle dans l'évolution ? Mais aussi quelle est leur fonction ?

Deux hypothèses antagonistes s'opposent concernant l'origine des introns au cours de l'évolution : l'hypothèse « des introns précoces » propose que les introns existaient avant la divergence des procaryotes et des eucaryotes (Gilbert 1978). Cependant cette hypothèse ne peut expliquer l'absence d'introns chez les procaryotes et, de ce fait, une explication a été émise, proposant qu'il y ait eu une pression de sélection sur les introns chez les procaryotes et que ceux-ci aient été éliminés de leur génome. En effet, la plus grande pression que subit le monde procaryote est de devoir se multiplier très rapidement, de ce fait toute séquence inutile est éliminée pour aboutir à un génome minimal.

En revanche, du fait de l'absence de spliceosome chez les procaryotes, une deuxième hypothèse a été formulée, dite de « l'hypothèse des introns tardifs » qui propose que les introns aient été insérés dans les gènes eucaryotes après cette divergence et que les espèces n'ayant pas d'introns n'aient jamais eu d'introns (Doolittle and Stoltzfus 1993). Certains éléments sont plus en faveur de la première hypothèse et certains de la deuxième et il est donc difficile de conclure sur l'une ou l'autre. De plus, la génomique comparative n'est toujours pas capable de dissocier entre ces deux hypothèses et il semblerait plutôt que l'origine des introns se situe entre ces deux hypothèses, « existence d'introns tardifs » et « d'introns précoces » (Koonin 2006).

Du fait de la caractéristique des introns nucléaires et du mécanisme nécessaire pour leur épissage, il a été proposé que les introns de group II auto-épissables sont à l'origine des introns nucléaires, et qu'ils auraient envahi massivement le génome nucléaire au cours de l'endosymbiose entre une α -proteobactérie et une archebactérie.

Quelle que soit l'origine des introns, ils ont par la suite évolué au sein des eucaryotes. Ainsi, par exemple, le mécanisme d'épissage des introns nucléaires chez les plantes diffère des mammifères, le site de reconnaissance en 3' est enrichi en AU dans les introns des plantes, alors que chez les mammifères, celui-ci est enrichi en poly pyrimidine (Wiebauer, Herrero et al. 1988). Par ailleurs, les introns de group II dits auto-épissables ne le sont dans la majorité des cas plus dans les organites et nécessitent désormais différents facteurs d'épissage comme il en sera discuté dans cette thèse.

I.1.3 Rôle de l'épissage

L'épissage de manière générale permet de joindre les exons entre eux en excisant toutes les parties non codantes afin d'obtenir un ARNm mature qui va pouvoir être traduit. Cette étape, d'une grande complexité au niveau nucléaire mais également dans les organites, fait partie des différentes étapes de maturation post-transcriptionnelle des ARN pré-messagers et est donc importante dans la régulation de l'expression des gènes et peut être contrôlée grâce à l'implication de différents facteurs d'épissage. Cependant d'autres fonctions peuvent lui être attribuées. En effet, l'épissage sous sa forme dite alternative permet d'établir différents isoformes dont les fonctions varient.

L'épissage alternatif est décrit dans le noyau mais n'est pas présent dans les organites et se définit sous quatre formes. 1. L'excision d'exons ; 2. La rétention d'introns ; 3. Site d'épissage donneur alternatif ; 4. Ou site accepteur d'épissage alternatif. Ce phénomène qui sera plus longuement discuté dans le chapitre sur l'épissage alternatif, permet soit de réguler l'expression des gènes, soit de synthétiser de nouvelles protéines à partir d'un même gène, conférant ainsi une adaptation et une évolution.

I.1.4 L'épissage des ARN messagers dans les organites

Article 1 : de Longevialle A.F., Small I.D. and Lurin C. (2010) Nuclearly Encoded Splicing Factors Implicated in RNA Splicing in Higher Plant Organelles. Molecular Plant, sous presse.

I.2 La famille des protéines PPR (Pentatricopeptide Repeat)

Le séquençage du génome d'*Arabidopsis* en décembre 2000 a permis de révéler la présence de milliers de gènes jusque là encore insoupçonnés et dont la fonction était inconnue. Parmi ces gènes, une nouvelle famille, la famille des gènes PPR (pour Pentatricopeptide Repeat) a été mise en évidence de façon indépendante par deux équipes françaises (Aubourg, Boudet et al. 2000; Small and Peeters 2000). Deux grands points caractérisent cette famille, tout d'abord les deux tiers de ses membres sont prédits localisés soit dans les mitochondries, soit dans les chloroplastes en utilisant le logiciel de prédiction Predotar (Small and Peeters 2000; Small, Peeters et al. 2004) et le deuxième point concerne leur étrange abondance dans le génome d'*Arabidopsis* (Aubourg, Boudet et al. 2000). En effet, cette famille, comportant plus de 450 membres chez *Arabidopsis* représente près de 2% du génome, ce qui a fait d'elle la plus grande famille pour laquelle la fonction n'était pas connue. Au moment de leur découverte, peu de données permettaient de leur attribuer un rôle fonctionnel, mais, grâce à la caractérisation de trois protéines de ce type dans d'autres espèces, une idée plus précise de leur fonction avait pu être proposée. Ainsi, chez le maïs, le gène *CRP1* code pour une protéine PPR impliquée dans la traduction de certains transcrits chloroplastiques (Barkan, Walker et al. 1994) ; la protéine PET309 chez *Saccharomyces cerevisiae* est également impliquée dans la traduction mais aussi dans la stabilité du transcrit *cox1* dans les mitochondries (Manthey and McEwen 1995) ; et le gène *cya-5* chez *Neurospora crassa* code pour une protéine PPR, homologue de PET309 et impliquée également dans la maturation post-transcriptionnelle du transcrit *cox1* (Coffin, Dhillon et al. 1997). Ces premières données suggéraient donc une fonction des protéines PPR dans la régulation post-transcriptionnelle des ARN messagers des organites. Mais quelles sont leurs caractéristiques, leurs fonctions et leur distribution dans le monde vivant ?

I.2.1 Caractéristiques des PPRs

I.2.1.1 Description des motifs PPRs

Les protéines PPR sont caractérisées par la présence d'un motif de 35 acides aminés répété en tandem, entre 2 et 26 fois (Lurin, Andres et al. 2004). Ce motif est assez proche du motif TPR (pour tetratricopeptide repeat). Les protéines TPR sont impliquées dans des interactions de type protéine-protéine (Blatch and Lassle 1999; Small and Peeters 2000) et le motif TPR forme une paire d'hélices alpha antiparallèles, qui en répétitions, conduisent à la formation d'une structure super hélicoïdale englobant un sillon central correspondant au site de fixation à un ligand protéique (Das, Cohen et al. 1998).

L'analyse bioinformatique prédictive des motifs PPR a également permis de proposer une structure super hélicoïdale (Small and Peeters 2000). Cependant, à la différence des protéines TPR, les chaînes exposées dans le sillon central des protéines PPR seraient dans la majorité des cas hydrophiles et basiques ce qui suggère que ces protéines se lieraient préférentiellement à des cibles chargées négativement comme des acides aminés acides ou des acides nucléiques (Small and Peeters 2000).

Des données permettent aujourd'hui de confirmer ce type d'interactions pour un certain nombre de protéines PPR et plus précisément avec des molécules d'ARN. En effet des expériences de RIP-CHIP, de retard sur gel ou « d'UV cross linking », ont montré cette interaction entre une protéine PPR et un ARN cible. C'est le cas par exemple de PPR4 chez le maïs qui a été montrée par RIP-CHIP comme interagissant avec l'intron 1 de *rps12* dans les chloroplastes (Schmitz-Linneweber, Williams-Carrier et al. 2006). Cela a été aussi montré

pour HCF152 dont la protéine recombinante se lie avec une haute affinité à la jonction exon-intron de *petB*, permettant sa stabilité et facilitant l'épissage de l'intron (Nakamura, Meierhoff et al. 2003), mais dont l'affinité est diminuée lorsque la protéine est tronquée et ne contient que deux motifs PPR au lieu de 12. C'est le cas également de CRP1 chez le maïs, se liant à une région d'ARN simple brin de 69 nucléotides dans la région 5' UTR des transcrits *psaC* et *petA* (Schmitz-Linneweber, Williams-Carrier et al. 2005), de PPR5, se fixant elle aussi préférentiellement sur de l'ARN simple brin dans une région de 45 nucléotides dans l'intron du *trnG-UCC* (Beick, Schmitz-Linneweber et al. 2008; Williams-Carrier, Kroeger et al. 2008), ou encore plus récemment de PPR10 chez le maïs qui en interagissant en 5' UTR du transcrit d'*atpH* et en 3' UTR de *psaJ* protège ces transcrits face à l'action d'exonucléases (Pfalz, Bayraktar et al. 2009). Il en va également de même pour des protéines PPR impliquées dans l'édition. Ainsi CCR4 et CCR21 se lient sur une région de 25 nucléotides en amont et 10 nucléotides en aval des sites d'édition sur le transcrit de *ndhD* (Okuda, Nakamura et al. 2006). L'interaction entre des molécules d'ARN et des protéines PPR a également été montrée chez d'autres espèces. Ainsi chez l'homme, LRP130 se lie *in vivo* à de l'ARN mitochondrial mais également à de l'ARN nucléaire (Mili and Pinol-Roma 2003).

Toutes les protéines PPR caractérisées à ce jour n'ont pas été montrées se fixant spécifiquement à leur ARN cible, cependant les nombreux exemples décrits permettent de proposer que ce soit une propriété commune à l'ensemble des protéines PPR (Delannoy, Stanley et al. 2007). Cette fixation spécifique à une séquence d'ARN leur permettrait ainsi de jouer un rôle majeur dans la régulation post-transcriptionnelle des ARN mitochondriaux et chloroplastiques. En revanche, aucune donnée ne permet aujourd'hui de comprendre le mécanisme moléculaire permettant cette interaction spécifique mis à part des études *in silico* de comparaisons de structure avec les protéines TPR. Un enjeu clef de la compréhension de la famille PPR sera donc de connaître la structure des protéines par cristallographie ainsi que la structure de l'interaction afin de comprendre comment cette interaction s'établit et pourquoi elle est spécifique de la séquence de l'ARN.

I.2.1.2 Composition de la famille

Les protéines PPR, selon les types de motifs qu'elles portent, peuvent être classées en deux sous-familles : les PPR de type « Pure » comportant seulement des motifs classiques et les PPR de type « PLS » comportant en plus des motifs classiques, d'autres motifs PPR plus courts (S pour Short) ou plus longs (L pour Long) (Figure 1).

La sous-famille « Pure » est composée de 251 membres chez *Arabidopsis*. La sous-famille « PLS » est assez particulière et est elle-même divisée en 4 sous-groupes selon les motifs retrouvés en partie C-terminale. Le premier sous-groupe comprend des protéines PPR de type PLS sans aucun motif C-terminal additionnel (6 membres chez *Arabidopsis*), le deuxième, des protéines PPR PLS ayant en partie C-terminale un motif E pour « Elongated » (47 membres chez *Arabidopsis*), le troisième, des protéines PPR PLS avec un motif E plus un motif E+ (60 membres chez *Arabidopsis*) et enfin le quatrième sous-groupe comprend des protéines PPR PLS avec un motif E, E+ et un domaine dénommé DYW (87 membres chez *Arabidopsis*). Les motifs E et E+ sont dans l'ensemble des motifs plus ou moins dégénérés dont l'identification reste difficile, en revanche le domaine DYW est un domaine d'environ 100 acides aminés très conservé entre les différents membres de ce sous-groupe. Ce motif est composé entre autre de résidus histidines, cystéines et d'un triplet final aspartate (D)-tyrosine (Y) tryptophane (W) quasiment invariants. Ce domaine très conservé pourrait avoir une activité catalytique (Aubourg, Boudet et al. 2000; Lurin, Andres et al. 2004), activité qui a été proposée par homologie de séquence comme étant une cytidine désaminase intervenant dans l'édition des ARNm (Salone, Rudinger et al. 2007).

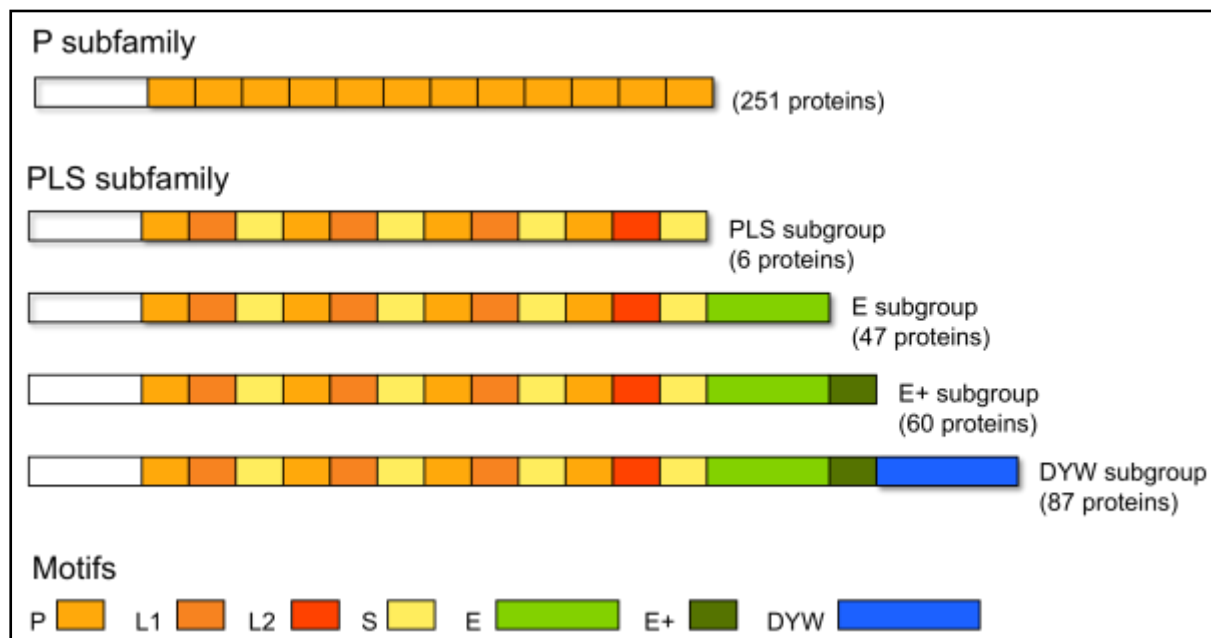


Figure 1 : Composition de la famille des protéines PPR chez *Arabidopsis thaliana*. La famille est divisée en deux sous-familles, la sous-famille « Pure » (P subfamily) comportant des motifs classiques représentés en orange et la sous-famille PLS (PLS subfamily) caractérisée par des motifs courts (S pour Short) en jaune, des motifs longs (L1 et L2 pour Long) en orange foncé et l'addition pour certains cas de motif E en vert clair, de motif E+ en vert foncé, et d'un motif DYW en bleu.

La présence de motifs additionnels non PPR a également été décrite dans des PPR de type « Pure ». Ainsi un domaine RRM (pour RNA Recognition Motif) en N-terminal est présent sur PPR4 (Schmitz-Linneweber, Williams-Carrier et al. 2006) permettant un deuxième type de reconnaissance de l'ARN ; deux domaines LAGLIDADG sont décrits en C-terminal de OTP51 (de Longevialle, Hendrickson et al. 2008). Une dizaine de protéines PPR ont en partie C-terminale un domaine SMR apparenté au domaine terminal des protéines bactériennes MutS2 dont la fonction est d'inhiber la recombinaison homologue. La fonction de ce domaine chez les PPR n'est pas connue, mais il pourrait intervenir dans des interactions de type protéine-protéine ou porter une fonction d'endonucléase avec une reconnaissance de l'ADN simple ou double brin (Moreira and Philippe 1999; Malik and Henikoff 2000; Fukui, Kosaka et al. 2007; Diercks, Ab et al. 2008). Enfin, une autre PPR de type « Pure », AT5G10690, comprend un domaine CBS (pour Cystathionine beta-synthase) en partie C-terminale.

Contrairement à la présence des motifs E, E+ et DYW très répandus dans la famille, la présence de ces motifs additionnels reste sporadique suggérant une fonction particulière pour ces protéines PPR portant de tels motifs supplémentaires.

I.2.1.3 Distribution de la famille

Les protéines PPR sont présentes chez tous les eucaryotes mais presque totalement absentes du monde des procaryotes sauf dans quelques cas où des protéines PPR sont retrouvées dans certaines bactéries.

C'est le cas chez *Erwinia amylovora* où la protéine CFBP1430 YP_003530178.1 présente 7 motifs PPR identifiés sous TPRpred (<http://toolkit.tuebingen.mpg.de/tpred>) et chez *Rhodobacter sphaeroides* où la protéine ATCC 17029 YP_001044140.1 présente 11 motifs PPR identifiés également sous TPRpred. *Rhodobacter* est une bactérie photosynthétique qui présente l'originalité de posséder deux chromosomes ce qui est rare chez les procaryotes. *Erwinia amylovora* est un agent pathogène de la famille des Enterobacteriaceae responsable

du feu bactérien qui infecte majoritairement les poiriers, les pommiers, les cognassiers et les néfliers par le biais des fleurs ou de blessures. Dans ce cas particulier, il est possible que cette PPR soit le résultat d'un transfert horizontal.

La distribution de ces protéines dans le monde des eucaryotes n'est cependant pas très homogène. En effet, il est très intéressant de voir que la famille des protéines PPR a connu une très grande expansion dans le monde végétal (Figure 2).

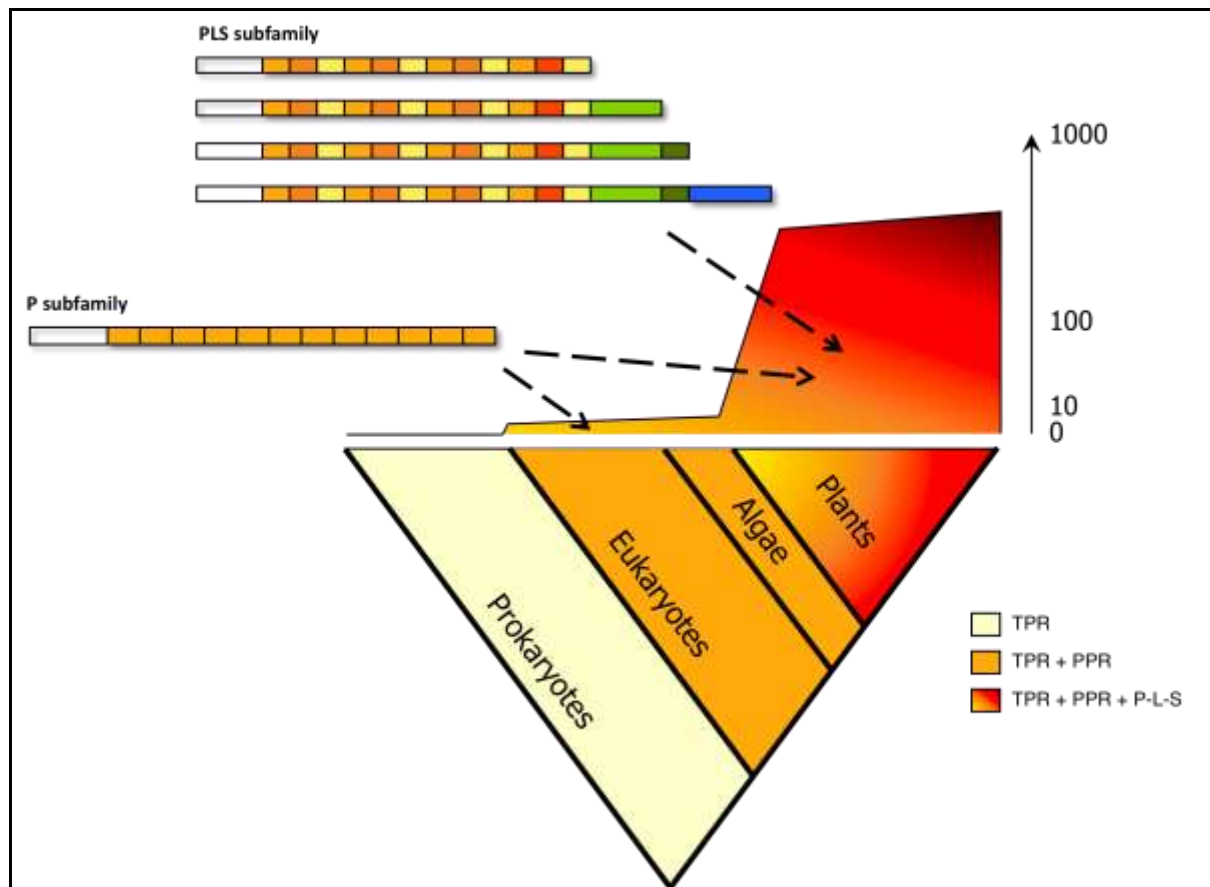


Figure 2 : Distribution des protéines PPR dans différentes espèces.

Alors que l'on comptabilise seulement 7 protéines PPR chez l'homme (Lightowlers and Chrzanowska-Lightowlers 2008), 3 protéines PPR chez la levure *Saccharomyces cerevisiae* et 11 protéines PPR chez l'algue *Chlamydomonas reinhardtii* (Johnson, Wostrickoff et al. 2010), on en compte plus de 450 chez *Arabidopsis thaliana* et chez le riz *Oryza sativa* (O'Toole, Hattori et al. 2008; Schmitz-Linneweber and Small 2008). Il est néanmoins possible que ce nombre soit sous estimé chez les espèces autres que les plantes. En effet, les motifs PPR sont assez dégénérés, donc difficiles à détecter, et les matrices de détection ont été générées essentiellement à partir de motifs « végétaux ». Cependant, même si leur détection était améliorée, leur nombre resterait certainement largement inférieur à celui des plantes. Comme nous le verrons plus tard dans le chapitre concernant la fonction des protéines PPR, cette expansion chez les plantes pourrait résulter du nombre de gènes restants dans les organites et d'une grande complexité de la maturation des ARN dans les organites. Un autre point assez intéressant à noter est que cette expansion des PPR chez les plantes s'est également faite avec l'apparition de la sous-famille PLS (Figure 2). En effet cette sous-famille n'est trouvée que chez les plantes supérieures, tous les autres eucaryotes n'ayant que des protéines PPR de type « Pure ». Plusieurs hypothèses peuvent être faites sur l'apparition de la sous-famille de type

PLS, mais la plus probable serait que cette expansion serait liée à la présence d'un événement particulier et très répandu chez les plantes terrestres qu'est l'édition des ARN. En effet, Véronique Salone, au cours de sa thèse a montré, par une analyse phylogénétique, une corrélation stricte entre la présence de séquences codants pour le motif DYW dans les génomes et l'existence du phénomène d'édition dans les organites (Salone, Rudinger et al. 2007). Par ailleurs, une comparaison de la famille des protéines PPR d'*Arabidopsis thaliana*, d'*Oryza sativa* et de *Physcomitrella patens* a permis de montrer qu'il existe une corrélation entre l'expansion de la sous-famille des protéines PPR de type PLS et celle du processus d'édition chez les plantes terrestres (O'Toole, Hattori et al. 2008). Enfin, l'homologie du motif DYW avec une cytidine désaminase permet de proposer que ces protéines soient impliquées dans l'édition des ARNm (Salone, Rudinger et al. 2007).

I.2.1.4 Données de localisation subcellulaire

Les logiciels de prédiction de localisation subcellulaire proposent une localisation pour la plupart des protéines PPR. Ainsi une grande majorité des PPR est prédite dans les mitochondries et certaines autres dans les chloroplastes. Des données de protéomique permettent de confirmer une partie de ces données de prédiction, mais, du fait de leur faible expression, d'une demi-vie courte des ARN (Narsai, Howell et al. 2007) et d'une faible abondance des protéines PPR, cette confirmation ne peut pas être faite pour une grande majorité d'entre elles.

Néanmoins, parmi les 450 protéines PPR, près d'un quart n'ont pas de claire prédiction de localisation dans les organites. Ainsi, selon les cas, des localisations en dehors des organites sont proposées ou bien il existe une trop grande contradiction entre les différents logiciels de prédiction. Pour évaluer l'existence de protéines PPR localisées en dehors des organites, une étude systématique de la localisation subcellulaire de 167 protéines PPR sans prédiction précise a été entreprise par le laboratoire de Claire Lurin. Ces expériences de localisation sont sur le point d'être terminées par Laure Heurtevin et permettent dès maintenant de montrer que la grande majorité des protéines PPR sans prédiction claire sont malgré tout localisées dans les organites. Cependant, parmi les protéines testées, de façon très intéressante, 17 protéines PPR se trouvent doublement adressées dans les chloroplastes et les mitochondries et 8 autres ont une localisation cytoplasmique ou nucléaire. Cette analyse étant faite en utilisant des protéines de fusion entre les 100 premiers acides aminés des PPR et la DsRedII, la localisation reste à être confirmée en utilisant des fusions avec les protéines entières.

I.2.2 Rôle des protéines PPR

Depuis la découverte de la famille il y a dix ans, de nombreuses protéines PPR ont été étudiées et caractérisées, permettant de démontrer leur implication majeure dans divers phénomènes de régulation transcriptionnelle et post-transcriptionnelle dans les organites. Les tableaux 1, 2 et 3 listent les différentes PPR caractérisées à ce jour, et décrivent leur fonction lorsque celle-ci a été identifiée.

INTRODUCTION

| Gène | Organisme | code AGI | orthologue Arabidopsis | Orthologue riz | PPR | Ref | Processus physiologique | role | cible |
|----------------------------|-------------------------------------|-----------|------------------------|----------------|-----|---|-------------------------|------------------------|--|
| PPR4 | maïs | - | At5g04810 | Os04g58780 | P | Schmitz-Linneweber <i>et al.</i> 2006 | Biogénèse des ribosomes | trans-épissage | <i>rps12</i> intron1 |
| HCF152 | Arabidopsis | At3g09660 | - | Os12g01210 | P | Nakamura <i>et al.</i> 2003; Meierhoff <i>et al.</i> 2003; Nakamura <i>et al.</i> 2004 | Photosynthèse | maturation et épissage | <i>petB</i> |
| OTPS1 | Arabidopsis | At2g15820 | - | Os02G47360 | P | de Longevialle <i>et al.</i> 2008 | Photosynthèse | Cis-épissage | <i>ycf3</i> intron2 |
| PPR5 | maïs | - | AT4G39620 | Os02G51480 | P | Beick <i>et al.</i> 2008; Williams-Carrier <i>et al.</i> 2008 | Traduction | Cis-épissage | <i>trnG</i> |
| CRP1 | maïs | - | AT5G42310 | Os07G36390 | P | Barkan <i>et al.</i> 1994; Fisk <i>et al.</i> 1999; Schmitz-Linneweber <i>et al.</i> 2005 | Photosynthèse | traduction | <i>pet4</i> , <i>psaC</i> |
| PPR2 | maïs | - | At3g06430 | Os03g19650 | P | Williams and Barkan 2003 | Biogénèse des ribosomes | - | - |
| PTAC2 | Arabidopsis | At1g74850 | - | Os03g60910 | P | Pfalz <i>et al.</i> 2006 | Transcription | - | - |
| DG1 | Arabidopsis | AT5G67570 | - | Os05G25060 | P | Chi <i>et al.</i> 2008 | Transcription | - | - |
| PPR10 | maïs | - | At2g18940 | Os05g19380 | P | Pfalz <i>et al.</i> 2009 | Photosynthèse | stabilisation | <i>atp1-atpH</i> et <i>psa2-rpl33</i> |
| PGR3 | Arabidopsis | At4g31850 | - | Os10g28600 | P | Yamazaki <i>et al.</i> 2004 | Photosynthèse | maturation | operon <i>petL</i> |
| MRL1 | <i>Chlamydomonas</i> et Arabidopsis | AT4g34830 | - | Os10G10170 | P | Johnson <i>et al.</i> 2010 | Photosynthèse | Stabilisation | <i>rbcl</i> |
| CRR4 | Arabidopsis | At2g45350 | - | OsPPR_07g15640 | E+ | Kotera <i>et al.</i> 2005; Okuda <i>et al.</i> 2006; Okuda <i>et al.</i> 2008 | Photosynthèse | édition | <i>ndhD-1</i> |
| CLB19/Pde 247/ Yel1 | Arabidopsis | AT1G05750 | - | Os02G53050 | E+ | Chateigner-Boutin <i>et al.</i> 2008 | Transcription | édition | <i>rpoA</i> et <i>dcpP</i> |
| CRR21 | Arabidopsis | AT5G55740 | - | Os01G42280 | E+ | Okuda <i>et al.</i> 2007 | Photosynthèse | édition | <i>ndhD-2</i> |
| OTP80 | Arabidopsis | At5g59200 | - | Os07G02280 | E | Hammani <i>et al.</i> 2009 | Traduction | édition | <i>rpl23</i> |
| OsPPR1 | riz | - | AT1G74600 | Os09g24680 | E | Gothandam <i>et al.</i> 2005 | Biogénèse des plastes | - | - |
| YS1 | Arabidopsis | AT3G2690 | - | Os01G25270 | DYW | Zhou <i>et al.</i> 2008 | Transcription | édition | <i>rpoB</i> |
| CRR22 | Arabidopsis | At1g11290 | - | Os06G02200 | DYW | Okuda <i>et al.</i> 2009 | Photosynthèse | édition | <i>ndhB-7</i> , <i>ndhD-5</i> et <i>rpoB-3</i> |
| CRR28 | Arabidopsis | At1g59720 | - | Os07G09370 | DYW | Okuda <i>et al.</i> 2009 | Photosynthèse | édition | <i>ndhB-2</i> et <i>ndhD-3</i> |
| RARE1 | Arabidopsis | At5g13270 | - | Os01G25270 | DYW | John <i>et al.</i> 2009 | - | édition | <i>accD-1</i> |
| LPA66 | Arabidopsis | At5g48910 | - | Os03G60690 | DYW | Wenhe <i>et al.</i> 2009 | Photosynthèse | édition | <i>psbF</i> |
| ATECB2 | Arabidopsis | AT1G15510 | - | Os05G49920 | DYW | Yu <i>et al.</i> 2009 | Biogénèse des plastes | édition | <i>accD-1</i> |
| OTP82 | Arabidopsis | At1g08070 | - | Os04G14130 | DYW | Okuda <i>et al.</i> 2009 | Photosynthèse | édition | <i>ndhB</i> et <i>ndhG</i> |
| OTP81 | Arabidopsis | At2g29760 | - | Os01G08120 | DYW | Hammani <i>et al.</i> 2009 | - | édition | <i>rps12</i> intron |
| OTP84 | Arabidopsis | At3g57430 | - | Os04G02850 | DYW | Hammani <i>et al.</i> 2009 | Photosynthèse | édition | <i>psbZ</i> , <i>ndhB</i> , <i>ndhF</i> |
| OTP85 | Arabidopsis | At2g02980 | - | Os03G58100 | DYW | Hammani <i>et al.</i> 2009 | Photosynthèse | édition | <i>ndhD</i> |
| OTP86 | Arabidopsis | At3g63370 | - | Os04G39410 | DYW | Hammani <i>et al.</i> 2009 | Biogénèse des ribosomes | édition | <i>rps14</i> |
| VAC1 | Arabidopsis | At1g15510 | - | Os05G49920 | DYW | Tseng <i>et al.</i> 2010 | Photosynthèse | édition | <i>accD</i> et <i>ndhF</i> |
| Emb175 | Arabidopsis | At5g03800 | - | Os07g07620 | DYW | Cushing <i>et al.</i> 2005 | Embryogénèse | - | - |
| CRR2 | Arabidopsis | At3g46790 | - | Os01g62910 | DYW | Hashimoto <i>et al.</i> 2003 | Photosynthèse | maturation | <i>ndhB</i> |
| DOT4 | Arabidopsis | AT4G18750 | - | Os01G10800 | DYW | Petrica <i>et al.</i> 2008 | - | - | - |
| GUN1 | Arabidopsis | At2g31400 | - | Os12g37550 | P | Koussevitzk <i>et al.</i> 2007 | Signaux rétrograde | - | - |
| P67 | Arabidopsis et radis | AT4G16390 | - | Os03g11670 | P | Lahmy <i>et al.</i> 2000 | - | Fixation à l'ARN | - |

Tableau 1 : Protéines PPR chloroplastiques chez les plantes supérieures décrites dans la littérature à la date du 25 juin 2010.

INTRODUCTION

| Gène | Organisme | code AGI | orthologue Arabidopsis | Orthologue riz | PPR | Ref | Processus physiologique | role | cible |
|-----------------------------------|-------------|-----------|------------------------|----------------|-----|--|---------------------------|------------------------------|--|
| CB_1265 | Arabidopsis | At1g53330 | - | OS05G11700 | P | Kocabek et al. 2006 | Respiration | - | - |
| OTP43 | Arabidopsis | At1g74900 | - | Os02g45590 | P | de Longevialle et al. 2007 | Respiration | trans-épissage | nad1 intron1 |
| PPR40 | Arabidopsis | At3g16890 | - | OS12G37100 | P | Zsigmond et al. 2008 | Respiration | - | - |
| PPR596 | Arabidopsis | AT1G80270 | - | OS12G07260 | P | Doniwa et al. 2010 | Respiration | effet indirect sur l'édition | rps3 |
| GRP23 | Arabidopsis | At1g10270 | - | OS05G30240 | P | Ding et al. 2006 | Embryogénèse | Transcription | - |
| LO3 | Arabidopsis | At2g39230 | - | OS01g42620 | P | Prasad et al. 2005 | - | - | - |
| EMP4 | maïs | - | At3g49730, At5g65820 | OS_03g51840 | P | Gutierrez-Marcos et al. 2007 | Développement des graines | - | - |
| RPF2 (Rf like) | Arabidopsis | At1g62670 | - | OS10G35436 | P | Jonietz et al. 2010 | Respiration | Clivage | cox3 et nad9 |
| OGR1 | Riz | - | AT1G34160 | OS12g17080 | DYW | Sung-Ryul et al. 2009 | Respiration | Edition | nad4-C401, nad4-C416, nad4-C433, nad2-C1457, ccmC-C458, cox2-C167, cox3-C572 |
| MEF11 or LO11 | Arabidopsis | At4g14850 | - | OS11G37330 | DYW | Verbitskiy et al. 2010; Jianwei et al. 2009; Kobayashi et al. 2007 | Respiration | Edition | cox3-C422, nad4-C124 et ccb203C-344 |
| MEF1 | Arabidopsis | AT5G52630 | - | OS01G46230 | DYW | Zehrmann et al. 2009 | Respiration | Edition | rps4-C956, nad7-C963 et nad2C-1160 |
| MEF18 | Arabidopsis | At5g19020 | - | OS01G29430 | E | Takenaka et al. 2010 | - | Edition | nad4-C1355 |
| MEF19 | Arabidopsis | At3g05240 | - | OS08G40020 | E | Takenaka et al. 2010 | - | Edition | ccb206-C566 |
| MEF20 | Arabidopsis | At3g18970 | - | OS03G58100 | E | Takenaka et al. 2010 | - | Edition | rps4-C226 |
| MEF21 | Arabidopsis | At2g20540 | - | OS05G36250 | E+ | Takenaka et al. 2010 | - | Edition | cox3-C257 |
| MEF22 | Arabidopsis | At3g12770 | - | OS3G27880 | DYW | Takenaka et al. 2010 | - | Edition | nad3-C149 |
| PPR13 | sorghum | - | - | - | E | Klein et al. 2005 | CMS | - | - |
| At2g02960 | Arabidopsis | At2g02960 | - | OS03G58100 | DYW | Nakamura et Sugita 2008 | Respiration | - | - |
| RF1A, RF1B, PPR791, PPR8-1 | Riz | - | At1g62930 ? | - | P | Kazama et Toriyama 2003; Komori et al. 2004; Akagi et al. 2004; Wang et al. 2006 | CMS | Clivage | ATP6/ORF79 |
| PPR336 | Arabidopsis | AT1G61870 | - | OS01G58080 | P | Uyttewaal et al. 2008 | Traduction | Traduction | associée aux polysomes |
| SLO1 | Arabidopsis | - | - | - | E+ | Sung et al. 2010 | Respiration | Edition | nad4-C449 et nad9-C328 |

Tableau 2 : Protéines PPR mitochondriales chez les plantes supérieures décrites dans la littérature à la date du 25 juin 2010.

| Gène | Organisme | PPR | Ref | Loc. | Processus physiologique | role | cible |
|----------------------------|---------------------------|-----|---|------|-------------------------|-----------------------------|------------------|
| TbPPR1 | Trypanosoma brucei | P | Mingler et al. 2006 | Mt | - | Transcription | COII, ND4 et ND5 |
| LRPPRC or LRP130 | Homme | P | Mili et Pinol-Roma 2003; Xu et al. 2004 | Mt | Respiration | Stabilisation et traduction | coxIII |
| PTCD2 | Souris | P | Xu et al. 2008 | Mt | Respiration | Maturation | ND5-CYTB |
| Pet309 | Yeast | P | Tavares-Carreón et al. 2008 | Mt | Respiration | Traduction | COX1 |
| DMR1 | Saccharomyces | P | Puchta et al. 2010 | Mt | - | Biogénèse des ribosomes | - |
| PpPPR_71 | Physcomitrella patens | DYW | Tasaki et al. 2010 | Mt | - | Edition | ccmF-1 |
| PPR531.11/ PpPPR_38 | Physcomitrella patens | P | Hattori et al. 2007; Hattori et Sugita 2009 | Cp | Biogénèse des plastides | Stabilisation et épissage | clpP |
| MCA1 | Chlamydomonas reinhardtii | P | Loiselay et al. 2008 | Cp | Photosynthèse | Stabilisation et traduction | petA |
| PTCD3 | homme | P | Davies et al. 2009 | Mt | Respiration | Traduction | - |
| PTCD1 | homme | P | Rackham et al. 2009 | Mt | Respiration | Traduction | - |
| AEP3p | levure | P | Ellis et al. 2004; Lee et al. 2009 | Mt | Respiration | Maturation et traduction | atp8-atp6 |
| BSF | Drosophile | P | Mancebo et al. 2001 | Mt | - | Stabilisation | 3' de BICOID |

Tableau 3 : Protéines PPR décrites dans la littérature à la date du 25 juin 2010 chez les organismes autres que les plantes supérieures.

I.2.2.1 Les protéines PPR dans l'initiation de la transcription

La première étape de la régulation de l'expression des gènes s'effectue au niveau de la transcription des gènes. Dans les plastes et les mitochondries, plusieurs protéines sont impliquées dans ce mécanisme, que ce soit des ARN polymérase ou des protéines permettant la reconnaissance de l'initiation de la transcription par ces polymérase comme par exemple les facteurs sigmas reconnaissant des sites d'initiation spécifiques dans le génome chloroplastique (Kanamaru and Tanaka 2004). Dans cette première étape, certaines protéines PPR ont également un rôle primordial.

La première PPR à avoir été montrée comme étant active dans la transcription est P63. En effet cette protéine PPR de type « Pure », identifiée lors d'une extraction de protéines se liant spécifiquement à l'ADN mitochondrial, possède une forte affinité pour l'ADN et plus particulièrement pour la région promotrice du gène *cox2* (Ikeda and Gray 1999).

Dans les chloroplastes, la transcription s'effectue via deux types de complexes de polymérase. Le premier, la NEP, est constitué de protéines codées par le génome nucléaire et le deuxième, la PEP, contient des protéines codées par le génome chloroplastique. Ces deux types de polymérase n'interviennent pas toujours dans la transcription des mêmes gènes et il existe donc des gènes transcrits spécifiquement par la NEP : les gènes *accD*, *rpoB* et *rpoC* et des gènes transcrits spécifiquement par la PEP comme une grande partie des gènes codant des protéines impliquées dans la photosynthèse, ainsi que des gènes transcrits par les deux mais avec des sites d'initiation de la transcription différents. Ainsi, l'activité de la polymérase PEP est essentielle dans la transcription des gènes de la photosynthèse et celle de la NEP dans les gènes dits « housekeeping » pour les fonctions basales (Hajdukiewicz, Allison et al. 1997; Liere and Maliga 1999).

Plusieurs facteurs influencent et régulent étroitement ces deux activités, ainsi certaines protéines PPR ont un rôle majeur dans la transcription et font partie intégrante de ces complexes. C'est le cas de PTAC2 (pour Platidial Transcriptional Active Complex), une protéine PPR de type « Pure » qui fait partie également des protéines PPR possédant des motifs supplémentaires car elle comporte un motif SMR dont la fonction reste à déterminer. Cette PPR a été montrée associée à la PEP et le mutant *ptac2*, dont le phénotype macroscopique est très affecté, présente un défaut majeur dans la transcription des gènes dépendant de la polymérase de type PEP (Pfalz, Liere et al. 2006). On ne connaît malheureusement pas comment cette protéine régule cette activité transcriptionnelle, et quel est son rôle exact dans le complexe actif transcriptionnel. La protéine PPR DG1 (pour Delayed Greening 1) est aussi impliquée dans l'activité transcriptionnelle chloroplastique (Chi, Ma et al. 2008). Cette protéine contrairement à PTAC2 n'est pas associée au complexe actif de transcription. En revanche l'activité de la PEP est fortement affectée chez le mutant.

La protéine PPR GRP23 impliquée dans le développement précoce de l'embryon, interagit avec la sous-unité III de l'ARN polymérase de type II, mais non pas dans les organites comme la majorité des PPR mais dans le noyau (Ding, Liu et al. 2006). Il est possible que cette interaction permette une régulation de l'expression de gènes nucléaires via la PPR. Cependant un autre rôle au niveau des organites n'est pas à exclure car cette PPR a également été localisée dans les mitochondries au sein du laboratoire de Claire Lurin.

Chez les mammifères, il a été montré un lien direct entre les ARN polymérase mitochondriales et les protéines PPR, car des motifs PPR sont retrouvés directement dans celles-ci (Shadel 2004). Chez la levure, des protéines PPR associées à l'ARN polymérase mitochondriale se lient aux chaînes d'ARN néosynthétisées afin de les stabiliser (Shadel 2004). Ces différents exemples montrent qu'il existe un lien direct entre l'ARN polymérase, la transcription et les protéines PPR. Néanmoins, le type d'interaction moléculaire avec les

acides nucléiques (PPR-ARN ou PPR-ADN) n'est pas toujours très clair. Les PPR pourraient jouer un rôle directement en aval de la transcription en stabilisant les ARN néosynthétisés et/ou en permettant leur maturation de façon couplée à la transcription.

I.2.2.2 Les protéines PPR dans la maturation des ARNm

Une fois le transcrite immature synthétisé, un certain nombre d'étapes de maturation sont nécessaires avant la phase traductionnelle ou une phase de régulation conduisant à la dégradation des ARN synthétisés. Dans les organites ces étapes sont hautement régulées de façon spécifique à chaque transcrite; ceci implique un grand nombre de protéines parmi lesquelles de nombreuses protéines PPR.

I.2.2.2.1 Maturation en 5' et 3' des ARNm

Par leurs origines, la majorité des gènes des organites sont transcrits sous des formes polycistroniques. De ce fait, l'expression de chaque gène nécessite différentes étapes de clivages sur chaque transcrite. Ces clivages en 5' et 3' de chaque transcrite nécessitent l'intervention d'enzymes de type RNase ayant des activités exoribonucléase ou endoribonucléase. L'action spécifique de ces RNase dans la reconnaissance ou dans leur arrêt nécessite soit des protéines de type « RNA binding » soit une structure secondaire spécifique de l'ARN. Parmi les protéines reconnaissantes des séquences spécifiques de l'ARN, les protéines PPR sont de très bon candidats et, à ce jour, un certain nombre de protéines PPR ont été identifiées dans ce type de maturation.

CRR2 (pour chlororespiratory reduction) est la première PPR identifiée dans les chloroplastes ayant un rôle dans la maturation des ARN (Hashimoto, Endo et al. 2003). Cette PPR de type DYW est impliquée dans le clivage du transcrite polycistronique *rps7-ndhB*. Une autre PPR du même type, AT2G02980, mais localisée dans les mitochondries a également été montrée comme ayant une activité endoribonucléase *in vitro* (Nakamura and Sugita 2008) mais aucune cible n'a été déterminée. Récemment une autre PPR, RPF2 appartenant à une sous-famille particulière des PPR qui sera décrite ultérieurement, les Rf-like, a été montrée impliquée dans le clivage en 5' des transcrits de *nad9* et *cox3* (Jonietz, Forner et al. 2010). Enfin, toujours chez *Arabidopsis*, deux autres PPR de type « Pure », PGR3 (pour Proton Gradient Regulation) et HCF152 (pour High Chlorophyll Fluorescence), sont impliquées dans la maturation des transcrits polycistroniques *petL-petG-psaJ* et *psbT-psbH-petB-petD*, respectivement (Meierhoff, Felder et al. 2003; Yamazaki, Tasaka et al. 2004), avec également un rôle dans l'épissage de l'intron de *petB* pour HCF152.

Dans d'autres espèces, des protéines PPR ont également été caractérisées comme ayant un rôle dans la maturation des transcrits. Chez la souris, PTC2, est impliquée dans le clivage du transcrite polycistronique *nd5-cytB* (Xu, Ackerley et al. 2008). Chez *Physcomitrella patens* PPR531.11, renommée PpPPR_38 est impliquée dans la maturation de *clpP* (Hattori, Miyake et al. 2007) et permet le clivage entre *clpP* et *rps12* en se liant directement sur cette région (Hattori and Sugita 2009). De façon intéressante cette protéine se lie également sur une région chevauchant l'exon1 et l'intron1 de *clpP* facilitant ainsi l'épissage de celui-ci, donnant à cette protéine une double fonction dans la maturation de ce transcrite. Chez la levure *Saccharomyces cerevisiae*, le mutant *Aep3p* présente un défaut de clivage de l'ARN polycistronique *atp8-atp6* (Ellis, Helfenbein et al. 2004), mais il vient également d'être montré que cette protéine aurait un rôle dans l'initiation de la traduction en se liant avec mIF2, un initiateur de la traduction dans les mitochondries (Lee, Tibbetts et al. 2009).

1.2.2.2.2 L'édition des ARNm

Une autre étape de la maturation des ARNm dans les organites est l'édition. L'édition est une modification de la séquence nucléotidique mature de l'ARNm ; dans les organites des plantes supérieures, elle consiste généralement en une substitution d'un nucléotide Cytidine en Uridine. Ce changement, en modifiant la séquence de l'ARNm, peut avoir plusieurs impacts sur la synthèse de la protéine finale. Par exemple, l'édition peut conduire à la création d'un codon d'initiation de la traduction comme dans le cas du site *ndhD-1* par le changement de ACG en AUG. L'édition a aussi un impact sur la fonctionnalité de la protéine en changeant un acide aminé par un autre via l'édition de l'ARN. Mais il existe également certains sites d'édition qui n'ont aucun impact sur la séquence en acides aminés, mais dont on pense qu'ils auraient un rôle dans la structure secondaire de l'ARN et permettraient ou non la maturation correcte de celui-ci. Ainsi, l'édition d'ARNm pourrait influencer l'efficacité de l'épissage lorsque le site d'édition se situe à proximité de la jonction exon-intron ou au sein même de l'intron.

34 sites d'édition ont été décrits dans les ARN chloroplastiques (Tillich, Funk et al. 2005; Chateigner-Boutin and Small 2007) et 526 dans les ARN mitochondriaux chez *Arabidopsis* (Giege and Brennicke 1999; Bentolila, Chateigner-Boutin et al. 2005; de Longevialle, Meyer et al. 2007; Bentolila, Elliott et al. 2008; Zehrmann, van der Merwe et al. 2008). Le mécanisme moléculaire de l'édition dans les organites des plantes supérieures n'est toujours pas réellement connu, en revanche, de plus en plus de protéines sont identifiées comme étant strictement nécessaires à l'édition de certains sites d'édition. De façon très intéressante, ces protéines sont exclusivement des protéines PPR, et de façon encore plus remarquable, toutes celles décrites jusqu'à ce jour sont des protéines PPR de la sous-famille PLS et plus particulièrement des protéines PPR de type E, E+ ou DYW.

Dans les chloroplastes, à ce jour, 16 protéines PPR, représentées dans la figure 3, ont été caractérisées comme permettant l'édition de 21 sites d'édition sur les 34 totaux.

CRR4 est la première protéine PPR de type E identifiée dans l'édition, elle est nécessaire à l'édition du site *ndhD-1*, créant ainsi un site d'initiation de la traduction (Kotera, Tasaka et al. 2005). CRR21, une protéine PPR de type E+ est impliquée dans l'édition du site *ndhD-2* (Okuda, Myouga et al. 2007). OTP80, une protéine de type E est nécessaire à l'édition du site *rpl23* (Hammani, Okuda et al. 2009). Enfin CLB19, PPR de type E+, est impliquée dans l'édition des deux sites *rpoA* et *clpP*, dont les défauts d'édition observés chez le mutant créent un problème dans la transcription par la PEP polymérase (Chateigner-Boutin, Ramos-Vega et al. 2008).

Douze protéines de type DYW ont également été décrites nécessaires dans l'édition d'un ou plusieurs sites chloroplastiques. C'est le cas de YS1 impliquée dans l'édition de *rpoB-1* (Zhou, Cheng et al. 2008), et de FLV impliquée dans l'édition du site *rpoC1* (Anne-Laure Chateigner-Boutin et al., résultats non publiés). LPA66 est impliquée dans l'édition du site *psbF* (Cai, Ji et al. 2009). OTP85 et OTP86 sont nécessaires pour l'édition respective des sites *ndhD* et *rps14* (Hammani, Okuda et al. 2009). Et OTP81 permet l'édition du site dans l'intron de *rps12*, ce dernier n'ayant visiblement aucun impact sur l'efficacité de l'épissage (Hammani, Okuda et al. 2009).

Dans le cas de l'édition du site *accD-1* (*accD C794*), deux protéines PPR de type DYW ont été décrites nécessaires chez *Arabidopsis*, RARE1 (Robbins, Heller et al. 2009) et ECB2 (Yu, Jiang et al. 2009). Cependant, une caractérisation récente de la protéine nommée VAC1 codée par le même gène que *ECB2* a mis en évidence un résultat différent de celui publié précédemment (Tseng, Sung et al. 2010). Ainsi, il a été montré que le mutant *vac1* présente

une diminution de l'édition du site *accD*-1, et non une absence de l'édition, mais en revanche également une forte diminution du site en 3'UTR d'*accD* et du site sur *ndhF*.

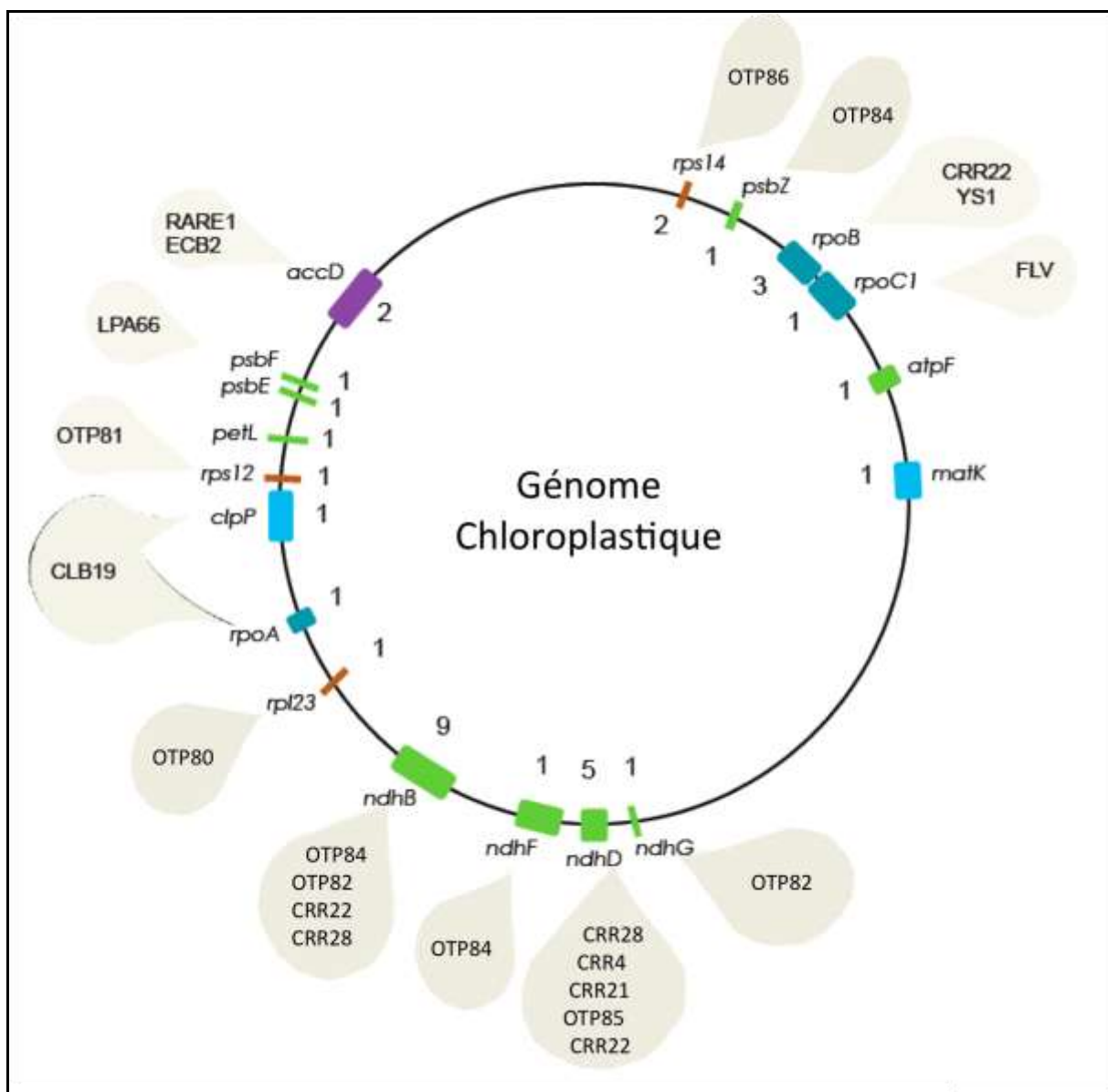


Figure 3 : Schématisation des protéines PPR impliquées dans l'édition des différents sites dans les chloroplastes chez *Arabidopsis*.

Les quatre dernières PPR-DYW décrites sont impliquées dans l'édition de plusieurs sites. Ainsi, CRR22 permet l'édition des sites *ndhB*-7, *ndhD*-5 et *rpoB*-3 (Okuda, Chateigner-Boutin et al. 2009). L'édition des sites *ndhB*-2 et *ndhD*-3 est sous le contrôle de CRR28 (Okuda, Chateigner-Boutin et al. 2009). OTP82 est impliquée dans l'édition des sites *ndhB*-7 (C-95644) et *ndhG* (Okuda, Hammani et al. 2010). Et les sites *psbZ*, *ndhB*-10 et *ndhF* nécessitent OTP84 (Hammani, Okuda et al. 2009).

Ainsi, au total, cinq protéines PPR (de type E+ ou DYW) sont nécessaires à l'édition de plusieurs sites dans les chloroplastes. Une analyse de la séquence aux alentours de chaque site a permis d'identifier des séquences consensus potentielles de reconnaissance pour 4 de ces protéines PPR. En revanche, ce type d'analyse n'a pas permis de proposer une séquence consensus de fixation pour CRR22 (Hammani, Okuda et al. 2009).

Dans les mitochondries, seulement 4 protéines PPR de type DYW et 4 autres de type E et E+ ont été identifiées dans l'édition, et donc la quasi totalité des sites d'édition mitochondriaux n'a aucun facteur d'édition connu. Cependant, il est fort probable que d'ici peu, la majorité d'entre eux sera caractérisée. Dans le riz, OGR1 est essentielle pour l'édition de 7 sites d'édition répartis sur 5 transcrits différents nad4-C401, nad4-C416, nad4-C433, nad2-C1457, ccmC-C458, cox2-C167, cox3-C572 (Kim, Yang et al. 2009). MEF1 (pour mitochondrial RNA Editing Factor) est impliquée dans l'édition de rps4-C956, nad7-C963, nad2-C1160 (Zehrman, Verbitskiy et al. 2009). MEF11 ou LOI1 (pour LOvastatin Insensitive) est nécessaire pour l'édition de 3 sites, cox3-C422, nad4-C124 and ccb203-C344 (Verbitskiy, Zehrman et al. 2010). Et MEF22, la dernière protéine DYW, est impliquée dans l'édition du site nad3-C149 (Takenaka, Verbitskiy et al. 2010). En ce qui concerne les protéines PPR de type E et E+, SLO1 (pour SLOW GROWTH1), une protéine PPR de type E+, est impliquée dans l'édition des deux sites nad4-449 and nad9-328 pour lesquels une séquence consensus est détectable en 5' du site (Sung, Tseng et al. 2010). Et les protéines MEF18, MEF19, MEF20, et MEF21 sont impliquées dans les sites respectifs : nad4-C1355 ; ccb206-C566 ; rps4-C226 et cox3-C257 (Takenaka, Verbitskiy et al. 2010).

Chez la mousse *Physcomitrella patens*, la protéine PpPPR_71, une protéine PPR de type DYW, est impliquée dans l'édition de ccmF-1 (Tasaki, Hattori et al. 2010).

I.2.2.2.3 L'épissage des ARNm

Comme il a déjà été discuté dans la première partie de l'introduction, l'épissage des ARNm dans les organites fait partie des étapes de maturation indispensables à la maturation correcte de l'ARN avant la traduction. Certaines PPR sont nécessaires à ce mécanisme. Les protéines PPR impliquées dans l'épissage dans les organites chez *Arabidopsis* sont décrites dans l'article 1 présenté dans l'introduction. Certaines protéines PPR non décrites dans la revue, car identifiées chez d'autres organismes que les plantes supérieures, sont également impliquées dans l'épissage des introns dans les organites.

Ainsi la protéine Pet309 est nécessaire pour l'accumulation de la forme non épissée de *cox1* chez la levure, mais son rôle direct dans l'épissage reste cependant à être prouvé (Manthey and McEwen 1995) et CCM1 (pour COB and COX1 mRNA Maturation) est impliquée dans l'épissage de l'intron 4 de *cob* et de l'intron 4 de *cox1*, tous deux décrits comme des introns de group I (Moreno, Buie et al. 2009). Chez la mousse *Physcomitrella patens*, la protéine PPR531-11 ou PpPPR_38 est impliquée dans l'épissage de l'ARNm de *clpP* dans les chloroplastes en se fixant sur la région exon1-intron1 de l'ARN pré-messager de *clpP* (Hattori, Miyake et al. 2007; Hattori and Sugita 2009), cependant cette protéine n'est pas essentielle pour l'épissage car des formes épissées de *clpP* sont toujours détectables chez le mutant.

I.2.2.3 Les PPRs dans la stabilisation des ARNm et la traduction

Certaines protéines PPR permettent le clivage en 5' ou en 3' des ARNm, mais il existe aussi des protéines PPR dont la fonction essentielle est de protéger les transcrits face à des dégradations de type exonucléase et des clivages en jouant un rôle protecteur et de stabilisation de l'ARNm mature.

Chez le maïs, PPR10 stabilise deux transcrits chloroplastiques en se fixant sur deux séquences similaires (Pfalz, Bayraktar et al. 2009). Cette interaction en 5' UTR du transcrit d'*atpH* et en 3' UTR de *psaJ* permet ainsi la protection des transcrits face à l'action des 5'-3' ou 3'-5' exoribonucléases et augmente donc ainsi la durée de vie de ces transcrits et permet également de montrer qu'une protection protéique est nécessaire pour les transcrits n'ayant pas de structure de type boucle en 3'UTR. Plus récemment, chez *Chlamydomonas* et

Arabidopsis, il a été montré que MRL1, une protéine PPR de type « Pure » se trouvant dans un complexe de haut poids moléculaire, est nécessaire à la stabilisation et la maturation du transcrite *rbcL* en se fixant certainement dans la région en 5'UTR (Johnson, Wostrikoff et al. 2010). De la même manière, MCA1 permet la stabilisation du transcrite *petA* en se fixant sur les premiers 25 nucléotides de la région 5'UTR et permet sa protection contre des 5'-3' RNases (Loiselay, Gumpel et al. 2008).

Chez la drosophile, la protéine PPR BSF (pour Bicoïd mRNA Stability Factor) stabiliserait le transcrite *bicoïd* codant pour une protéine importante dans le développement embryonnaire en se fixant sur la région 3'UTR du transcrite (Mancebo, Zhou et al. 2001). Cependant, l'homologue humain de cette protéine, LRP130 est localisé dans les mitochondries, alors que le transcrite *bicoïd* se situe dans le cytosol, cette différence de compartimentation remet partiellement en cause cette interaction. La protéine PPR LRP130 est également importante dans la stabilité des ARNm. Des mutations dans le gène codant pour cette protéine conduisent à une maladie neurodégénérative : la LSFC (pour Leigh Syndrome French Canadian), une déficience de la cytochrome c oxydase chez l'homme (Mootha, Lepage et al. 2003). Chez les patients atteints de cette maladie, les transcrits *cox1* et *cox3* sont très réduits suggérant un rôle de LRP130 dans la stabilisation des transcrits *cox1* et *cox3* (Xu, Morin et al. 2004).

La stabilisation des ARN matures permet d'augmenter leur durée de vie et ainsi influence leur taux de traduction. Cette dernière étape implique également d'autres facteurs, notamment dans la reconnaissance des sites d'initiation de la traduction, mais aussi dans la biogénèse des sous-unités ribosomales, et, là encore, des protéines PPR ont été montrées essentielles dans ces procédés dans les organites.

La protéine CRP1 chez le maïs est nécessaire pour la traduction des transcrits *petA* et *petD* en se liant dans la région 5'UTR de chaque transcrite (Barkan, Walker et al. 1994; Schmitz-Linneweber, Williams-Carrier et al. 2005) et agit ainsi comme régulateur de la traduction de ces deux transcrits. Le mutant *ppr2* chez le maïs présente une absence de ribosomes chloroplastiques suggérant que la protéine PPR2 est nécessaire dans l'assemblage de la machinerie traductionnelle (Williams and Barkan 2003). Une autre PPR, PPR336, identifiée chez Arabidopsis, serait certainement impliquée dans la traduction au niveau mitochondrial du fait de son association avec les polysomes (Uyttewaal, Mireau et al. 2008).

Chez l'homme, PTCD3 (pour pentatricopeptide repeat domain protein) s'associe directement avec la petite sous-unité ribosomale mitochondriale (ARNr 12S) et régule ainsi directement l'activité traductionnelle de façon positive (Davies, Rackham et al. 2009). En revanche, PTCD1, une autre protéine PPR chez l'homme, a un rôle de régulation négative sur l'abondance de l'ARN de transfert Leucine et par ce fait a donc un rôle indirect sur la traduction, avec un effet négatif sur son activité (Rackham, Davies et al. 2009).

Dans ce même type de régulation de la traduction, il a été montré récemment chez la levure *Saccharomyces cerevisiae* que la protéine PPR DMR1 s'associe avec la sous-unité ribosomale ARNr 15S et serait impliquée dans la biogénèse des ribosomes mitochondriaux et ainsi dans la traduction (Puchta, Lubas et al. 2010). Chez la levure, la protéine AEP3p est nécessaire dans la maturation du transcrite polycistronique *atp8-atp6* (Ellis, Helfenbein et al. 2004), mais pourrait également avoir un rôle dans l'initiation de la traduction en se liant avec mIF2, un initiateur de la traduction dans les mitochondries (Lee, Tibbetts et al. 2009). Enfin, la protéine PPR, PET309, en plus de son rôle dans la stabilité de l'ARN *cox1*, est également impliquée dans la traduction de ce transcrite en se fixant dans la région non traduite 5'UTR (Manthey and McEwen 1995) et a été montrée par double hybride comme interagissant avec Nam1p, une protéine importante dans le métabolisme des ARN mitochondriaux qui permet le transfert des ARN vers leur site de traduction au niveau de la membrane interne (Naithani, Saracco et al.

2003). Plus récemment, des délétions des différents motifs PPR composant cette protéine ont permis de montrer que tous les motifs PPR sont importants pour la traduction du transcrit *cox1*, en revanche, la délétion de ces motifs n'a pas d'impact sur la stabilité du transcrit (Tavares-Carreon, Camacho-Villasana et al. 2008), montrant ainsi l'importance de cette PPR dans la traduction et visiblement moins dans la stabilisation.

I.2.2.4 Les protéines PPR impliquées dans la restauration de la fertilité mâle cytoplasmique (CMS)

La stérilité mâle cytoplasmique (CMS) est un phénomène par lequel une plante hermaphrodite n'est plus capable de produire des gamètes mâles fertiles empêchant ainsi son autofécondation. Le terme cytoplasmique évoque le fait que le phénotype est provoqué par un facteur porté par le génome mitochondrial. Ce phénomène de grand intérêt agronomique est dû à l'expression d'un gène mitochondrial « toxique », nommé généralement *ORF-CMS*. Dans la nature, ce processus favorise les croisements entre individus afin d'augmenter la diversité génétique. En agronomie, ce processus a deux intérêts majeurs : faciliter la production de descendance hybride chez des plantes naturellement autogames et bénéficier ainsi du phénomène de vigueur hybride ou hétérosis, et, grâce aux gènes restaurateurs, produire des hétérozygotes fertiles. En effet, sans gène de restauration, la stérilité mâle cytoplasmique est transmise à la descendance. Très souvent, c'est une production de graines qui est recherchée pour la plante hybride et il est nécessaire de restaurer la fertilité dans l'hybride. C'est le parent mâle de l'hybride qui aura cette fonction en apportant dans son génome nucléaire le gène de restauration de fertilité.

De façon intéressante, hormis le cas d'une aldéhyde déshydrogénase codée par le gene *Rf2* du maïs (Cui, Wise et al. 1996), tous les gènes restaurateurs de fertilité découverts à ce jour codent pour des protéines PPR. Ces protéines PPR appartiennent en très grande majorité à la sous-famille « Pure » et, étrangement, font partie d'une petite sous-famille de PPR qui a évolué très rapidement. Par exemple, ces deux sous-familles sont très différentes chez le riz et chez *Arabidopsis*, les gènes d'*Arabidopsis* n'ayant pas d'orthologues chez le riz (O'Toole, Hattori et al. 2008).

Rf-PPR592, restaurateur de fertilité de type PPR, a été isolé chez le pétunia (Bentolila, Alfonso et al. 2002). Chez cette espèce, la stérilité male cytoplasmique est conférée par l'expression d'un gène *PCF* qui est présent au sein d'un transcrit polycistronique comprenant *atp9*, *cox2* et cette *orf* de fonction inconnue. En présence de la protéine PPR592 fonctionnelle, l'expression du transcrit *pcf* est fortement affectée ainsi que l'expression de la protéine résultante apportant ainsi une restauration de la fertilité. Ainsi la protéine PPR592, faisant partie d'un large complexe protéique, aurait une action sur la maturation post-transcriptionnelle de ce transcrit avec un effet de régulation négative sur l'expression du gène *PCF* en s'associant directement au transcrit (Gillman, Bentolila et al. 2007).

Chez le riz, la CMS est induite par l'expression de l'ORF79 codant pour un peptide cytotoxique et dont le gène se situe en aval du gène *atp6* dans les mitochondries. Lors de la CMS, un seul transcrit de 2 kb comprenant *atp6* et l'*orf79* est observé alors que deux transcrits sont normalement formés en présence du restaurateur de fertilité (Kazama and Toriyama 2003). En réalité deux gènes PPR sont capables de restaurer la fertilité chez le riz, *Rf1a* and *Rf1b* (Wang, Zou et al. 2006). Le premier, *Rf1a* code pour une protéine PPR impliquée dans le clivage du transcrit de 2 kb mais aussi dans l'édition d'*atp6* et le deuxième, *Rf1b* ou *Ifr1* code pour une PPR impliquée dans la dégradation du transcrit dicistronique *atp6-orf79*. En effet il a été montré récemment que la protéine codée par le gène *Ifr1* permettrait de réduire la quantité de transcrit *atp6-orf79* mais n'affecterait en aucun cas le clivage de ce

transcrit, montrant ainsi que les deux restaurateurs de fertilité chez le riz induisent deux effets différents sur les transcrits selon leur site de fixation (Ohta, Ogino et al. 2010).

Chez le radis, le locus *Rfo* comprenant 3 gènes *PPR* (*PPR-A*, *PPR-B/Rfkl* et *PPR-C*) réduisent l'expression des *ORF* mitochondriales 138 et 125 responsables de la stérilité mâle cytoplasmique (Brown, Formanova et al. 2003; Desloire, Gherbi et al. 2003; Koizuka, Imai et al. 2003). Parmi les 3 gènes du locus *Rfo*, il a été montré que seul le gène *PPR-B* était le restaurateur de fertilité en s'associant *in vivo* à l'*orf138* (Uyttewaal, Arnal et al. 2008). Ainsi la protéine *PPR-B* agirait directement sur la régulation de la traduction du transcrit de l'*orf138*.

Le restaurateur de fertilité chez le sorgho (*Sorghum bicolor*) est également une protéine *PPR*, cependant le mécanisme d'action de cette restauration n'est pas claire pour le moment car le gène responsable de la stérilité n'a toujours pas été identifié (Klein, Klein et al. 2005).

Les protéines *PPR* impliquées dans la restauration de la fertilité n'ont pas un rôle réellement différent des autres *PPR* car le mécanisme moléculaire réside dans la maturation des transcrits et/ou dans la régulation de la traduction. Cependant, la présence de la protéine *PPR* restauratrice de fertilité a un effet négatif sur l'expression de l'*Orf-cms* (au niveau transcriptionnel, post-transcriptionnel ou traductionnel) au contraire de l'effet plus classiquement positif des protéines *PPR* sur l'expression des transcrits chloroplastiques et mitochondriaux. Il est également intéressant de noter qu'il est difficile de trouver des orthologues pour ces protéines *PPR* entre différentes espèces. Ceci peut s'expliquer par la mise en place de différents mécanismes pour la stérilité mâle cytoplasmique. Cette diversité des mécanismes a ainsi forcé l'évolution spécifique de ces protéines *PPR*, créant ainsi une plus ou moins grande divergence. Ces protéines devenant donc uniques.

PRESENTATION DU SUJET DE THESE :

Depuis la découverte de la famille des protéines PPR, il a été montré que ces protéines ont un rôle majeur dans la régulation transcriptionnelle, post-transcriptionnelle et traductionnelle du génome des organites. Parmi les différents processus, l'épissage des ARNm comprenant des introns est un événement essentiel de maturation. Le génome chloroplastique d'*Arabidopsis* est composé de 20 introns de groupe II et d'un intron de groupe I et le génome mitochondrial est composé de 23 introns de groupe II. Ces introns ne peuvent pas s'auto-épisser et nécessitent donc un certain nombre de facteurs facilitant la mise en place d'une structure secondaire adéquate pour les deux étapes de trans-estérification caractérisant ce processus. Avant de commencer ma thèse, peu de facteurs intervenant dans cette maturation des ARNm avaient été identifiés, mais parmi les quelques caractérisés, deux protéines PPR avaient été identifiées dans ce phénomène. Ainsi PPR4, strictement nécessaire pour l'épissage en trans de l'intron 1 de *rps12*, fut la première protéine PPR décrite dans l'épissage en *trans* (Schmitz-Linneweber, Williams-Carrier et al. 2006) et HCF152, la première PPR identifiée dans l'épissage en stabilisant et en permettant l'épissage de l'intron de *petB* (Meierhoff, Felder et al. 2003). Cette découverte, le nombre important d'introns décrits dans les génomes des organites, ainsi que la capacité des protéines PPR à interagir avec les ARN, suggéraient que ces deux protéines PPR n'étaient certainement pas des cas isolés et que d'autres protéines PPR pouvaient avoir un rôle majeur dans le processus de l'épissage. Dans le but de connaître et de comprendre le fonctionnement des protéines PPR dans les différentes étapes de la maturation des ARNm, il était donc important de trouver d'autres protéines PPR impliquées dans le processus d'épissage. C'est dans ce cadre que mon sujet de thèse a été élaboré.

C'est au cours du projet AGRİKOLA, dans lequel j'étais responsable des étapes de clonage, que j'ai eu l'opportunité de commencer la caractérisation de certains mutants pour des protéines PPR d'intérêt au sein du laboratoire de Ian Small et de Claire Lurin à l'URGV. Ce projet Européen consistait à utiliser le système « ARNi » afin de créer une collection de mutants 'knock-down' pour tous les gènes d'*Arabidopsis* (Hilson, Small et al. 2003). La famille PPR était une famille cible intéressante pour ce projet car, d'une part, un certain nombre de gènes codant pour des PPR ne présentent pas de mutant d'insertion ADN-T dans les différentes collections disponibles et, d'autre part, s'il existe des mutants d'insertion, une grande partie d'entre-eux est décrite comme embryo-letaux empêchant toute caractérisation fonctionnelle de ces protéines par une approche classique de génétique inverse. Le projet AGRİKOLA pouvait donc permettre d'obtenir par la méthode de « ARNi » des mutants dans lesquels l'expression des gènes PPR serait dérégulée mais pas totalement afin d'obtenir des plantes suffisamment viables pour leur étude. C'est dans ce cadre que j'ai commencé une analyse systématique des mutants « ARNi » pour les 48 PPRs sélectionnées dans le laboratoire. Mon analyse s'est alors plus particulièrement focalisée sur une protéine PPR, OTP43, dont les mutants « ARNi » présentaient des retards importants de développement. Afin de valider la corrélation entre l'extinction du gène *OTP43* et le phénotype, j'ai étudié le mutant d'insertion ADN-T associé et commencé la caractérisation de celui-ci. Par la suite, l'élaboration de différents outils permettant de cribler des mutants pour des défauts d'épissage m'a permis de caractériser un certain nombre de protéines PPR plus ou moins essentielles dans le processus d'épissage, créant ainsi une carte plus complète des facteurs d'épissage chloroplastiques et mitochondriaux. Plus récemment, j'ai mené deux nouvelles approches dans le but de mieux comprendre la régulation de la fonction des protéines PPR. Ainsi en 3^{ème} partie de mon manuscrit, je présente une étude de l'épissage alternatif des transcrits PPR et de l'impact éventuel des différentes isoformes sur la fonction des protéines. Enfin, des données très récentes d'interactions protéine-protéine pour les protéines PPR sont présentées et permettent de proposer de nouvelles hypothèses sur le fonctionnement des protéines PPR impliquées notamment dans l'édition des ARN.

CHAPITRE II : RESULTATS

II.1 Identification de protéines PPR impliquées dans l'épissage des ARNm dans les chloroplastes.

Le génome chloroplastique d'*Arabidopsis* est composé de 20 introns de groupe II et d'un intron de groupe I (cf Introduction). Ces introns, comme il a été décrit dans la revue présentée dans l'introduction, nécessitent l'intervention de différents facteurs pour leur épissage. Au début de ma thèse, une grande partie de ces facteurs avait déjà été identifiée dans le laboratoire d'Alice Barkan, mais, néanmoins, les facteurs nécessaires à l'épissage de certains introns n'avaient pas encore été identifiés. Parmi les protéines impliquées dans l'épissage, une majorité fait partie d'une famille contenant des motifs CRM, mais on retrouve également d'autres protéines provenant d'autres familles de protéines comme des protéines PPR. Seulement deux protéines PPR avaient été identifiées comme jouant un rôle dans l'épissage des introns dans le chloroplaste au début de ma thèse. PPR4, impliquée dans l'épissage en *trans* de l'intron 1 de *rps12* et HCF152, impliquée dans la stabilisation et l'épissage de l'intron de *petB*. Ces deux protéines PPR permettent de démontrer que les protéines PPR, parmi les nombreuses fonctions dans la modification post-transcriptionnelle, sont impliquées dans le phénomène d'épissage. Environ un quart des PPRs (soit environ une centaine de protéines) sont adressées dans les chloroplastes. En dehors des fonctions dans l'édition ou le processing en 5' et 3', il est fort probable que certaines d'entre-elles soient impliquées dans l'épissage des introns dans les organites.

II.1.1 OTP51

Un des candidats les plus probables était OTP51. Cette « Pure » PPR a une caractéristique bien particulière. En effet, en plus de ses 7 motifs PPR de type « Pure », deux autres motifs de type « LAGLIDADG » sont décrits par Pfam (PF00961) en C-terminal. Cette PPR n'est pas la seule à présenter des motifs supplémentaires aux motifs PPRs. En revanche, OTP51 est la seule protéine PPR, mais aussi la seule protéine décrite chez *Arabidopsis*, à avoir des motifs de type « LAGLIDADG ». Ces motifs sont majoritairement trouvés chez d'autres espèces dans des endonucléases codées par des introns de group I. Grace à leur activité endonucléase, ces protéines coupent une séquence d'ADN double brin spécifique dans les gènes ne contenant pas ce type d'intron, permettant l'invasion de cet intron codant dans le génome. Mais certaines protéines LAGLIDADG ont aussi acquis une fonction de maturase, permettant l'épissage de l'intron, d'autres encore ont perdu l'activité endonucléase mais conservé l'activité de maturase. C'est le cas certainement d'OTP51, comme nous le verrons dans l'article suivant, qui a perdu les acides aminés essentiels pour l'activité endonucléase, mais a en revanche conservé les acides aminés essentiels pour la fonction de maturase. Le fait qu'OTP51 soit prédite dans les chloroplastes et que la fonction de maturase soit conservée laissent à penser que cette PPR pourrait être fortement impliquée dans l'épissage d'introns chloroplastiques. L'article suivant montre la caractérisation de cette protéine et de son rôle primordial dans l'épissage de l'intron 2 de *ycf3*, un des introns chloroplastiques pour lesquels aucun facteur d'épissage n'avait été identifié jusqu'à présent.

Article 2 : de Longevialle A.F., Hendrickson L., Taylor N.L., Delannoy E., Lurin C., Badger M., Millar A.H. and Small I.D. (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. The Plant Journal, 56: 157 – 168.

RESULTATS

Supplemental Table 1. qPCR primer sequences

| transcript | forward | reverse |
|--------------|----------------------------|-----------------------------|
| <i>accD</i> | TGTGGATTCAATGCGACAAT | TTTTGCGCAGAGTCAATACG |
| <i>atpA</i> | CGGAAATCTTACCTCGACCA | ATGGGTGACGGTTTGTATGAT |
| <i>atpB</i> | CCGTTTCGTACAAGCAGGAT | CGGGGTCAAGTCAAATCATCT |
| <i>atpE</i> | TCCACAAGAAGCTCAGCAAA | GTGTCCGAGCTCGTCTGAG |
| <i>atpF</i> | GTCCTTCACGCAGTTCTTC | TACTTGGGTCACTGGCCATC |
| <i>atpH</i> | ATCCACTGGTTTCTGCTGCT | TTCTTCTGCCTCAGGTTGT |
| <i>atpI</i> | ATTGGCAAATAGGGGGTTTC | GCCGTCAAGTTGGAATTGTTT |
| <i>ccsA</i> | CACAATAACTGCGCCAAGTG | AACAGAGCGCCATAGCCTAA |
| <i>cemA</i> | TTTGCCCTGGTTGATCTCTC | TTGGATCGTTTCTTTGTGGA |
| <i>ClpP1</i> | GTCGGAGGAGCAATTACCAA | GTGATGGTTTCGCGAAGTTT |
| <i>matK</i> | ATCCTTTGTTGCCAGAATGC | TTTTTCTACGCAAGCGGTCT |
| <i>ndhA</i> | TTGACGCCACAAATTCAT | TTAGGTGGTCTGCGAGCTG |
| <i>ndhB</i> | CCAGAAGAAGATGCCATTCA | TCATCAATGGACTCCTGACG |
| <i>ndhC</i> | TATAGAACCGATCGGGGATG | AACTCATTGCCACGGATAC |
| <i>ndhD</i> | TGGAGAATGGGAATAGATGGAC | TCCCGAGAAGAAAATGATCCTA |
| <i>ndhE</i> | TGGATTGATCACAAAGTCGAAA | AGCGGCTGCAATTGCTATAA |
| <i>ndhF</i> | CGGCGGGTATTTTCTTGTA | GGCTAAACCCCGCTTAATGT |
| <i>ndhG</i> | TTGCCTGGACCAATACATGA | ACATTTATGGCCCCCATA |
| <i>ndhH</i> | ATGGGAAATTCATGGCAAA | TCAAAGCCCCTGCTTTCTAA |
| <i>ndhI</i> | TTTGCTGTTGTTGATTGGA | ATTGGTAAACGACCCAAAGC |
| <i>ndhJ</i> | CGTTTTCTGGGTTTGGAAAA | AGGCCACCCTATCCAACCTCT |
| <i>ndhK</i> | GCAGTCCGCATATTGGAAAT | CGTGGGACGATACTGGACTT |
| <i>petA</i> | CAGAGGGCGAATCCATTA | GCCAAAACAACCGATCCTAA |
| <i>petB</i> | ATTGGGCGGTCAAAATTGTA | AGACGGCCGTAAAGAAGAGGT |
| <i>petD</i> | TCCTTTTGCAACTCCTTTGG | CCGCTGGTACTGAAACCATT |
| <i>petG</i> | TCTAATTCCTATTACTTTGGCTGGA | CCAAGTATCACCACGTCTG |
| <i>petI</i> | TTTCGGTTTTCTACTAGCAGCTT | TGCTTAGACCAATAAACAGAACTGA |
| <i>petN</i> | CGCATGGGCTGCTTTAAT | GAGTCCACTTCTTCCCACA |
| <i>psaA</i> | GCCAAGAAATCCTGAATGGA | CATCTTGGAAACCAAGCCAAT |
| <i>psaB</i> | GGACCCCACTACTCGTCGTA | ATTGCTAATTGCCGAAATG |
| <i>psaC</i> | GAGCATGCCCTACAGCGTA | CAGGCGGATTACATCTCTT |
| <i>psaI</i> | ACTTACCCTCTATTTTGTGCCTTT | TGAATATGAAGAAATAAAGAAGCCATT |
| <i>psaJ</i> | ATGGTTCGGTTCGTTAGCAG | GGGAAATGTTAATGCATCTGG |
| <i>psbA</i> | GAGCAGCAATGAATGCGATA | CCTATGGGGTTCGCTTCTGTA |
| <i>psbB</i> | CGTGCAGCTTTGAAATCTGA | TAGCACCATGCCAAATGTGT |
| <i>psbC</i> | ACTTCCCACCTAGCCACTT | AGCCAAAACCTGCAGAAGAA |
| <i>psbD</i> | CACAAATCTTTGGGGTTGCT | CCATCCAAGCACGAATACCT |
| <i>psbE</i> | TGTCTGGAAGCACAGGAGAA | AACCGGTGCTGACGAATAAC |
| <i>psbF</i> | GGACCTATCCAATTTTACAGTGC | GTTGGATGAACTGCATTGCT |
| <i>psbH</i> | TCTAGATCTGGTCCAAGAAGCA | CATTGCAACACCCATCAAAG |
| <i>psbI</i> | TTTCTCTTTCATATTTGGATTCT | TTCTTACGTCCCAGGATTAC |
| <i>psbJ</i> | CTGGAAGGATTCTCTTTGG | CAGGGATGAACCTAATCCTGA |
| <i>psbK</i> | AGGCCTACGCCTTTTGAAT | CGAAAACCTACAGCGGCTTG |
| <i>psbL</i> | CAATCAAATCCGAACGAACA | GAAATAATTCGAAAATAAACAGCAA |
| <i>psbM</i> | TGCACTTTCATTCTCGTTCC | TCATTTTGACTAACGGTTTACG |
| <i>psbN</i> | GGAAACAGCAACCCTAGTCG | CGTGTTCCTCGAATGGATCT |
| <i>psbT</i> | GGAAGCATTGGTTTATACATTTCTCT | AAATTTTAGGTGGTTCCCAGAA |
| <i>psbZ</i> | TGCTTTCCAATTGGCAGTTT | GTTACTCGACCAACCATCAGG |
| <i>rbcl</i> | GTGTTGGGTTCAAAGCTGGT | CATCGGTCCACACAGTTGTC |
| <i>rpL14</i> | AGCGGGGCTAGAGAATTGAT | ACTGCGGCATTGTCATCATA |
| <i>rpL16</i> | TGTACGACGTGGTGGAAAA | GCATTTTGTATGCCGCTATT |
| <i>rpL2</i> | CGGACCTCTCCAGAAGGTAAT | AAATGGGAAATGCCCTACCT |
| <i>rpL20</i> | TCGGAGGCGTAGAACAAAAC | CGATGAGCCGAAACTAAAGC |
| <i>rpL22</i> | AAAGCTGAGGTGAACCAAGG | TGTCCCATAGGCCTCCACTA |
| <i>rpL23</i> | CGGTTATTGGGAAAAATCA | TTTTAACCTTTCCGGGGAGT |
| <i>rpL32</i> | CTCGAAAAAGCGTATTCGTA | TGAAAAAGCTTTCAACGATGTC |
| <i>rpL33</i> | GCCAAGGGTAAAGATGTTTCG | TTGATTTCCCCTGGAATTGT |
| <i>rpL36</i> | AAATAAGGGCTTCCGTTTCGT | CCTCGGGTTGGAACAAATTA |
| <i>rpoA</i> | GCGATGCGAAGAGCTTACT | CCAGGACCTTGACACAAAT |

RESULTATS

| | | |
|--------------------------------|---------------------------|------------------------------|
| <i>rpoB</i> | AAAAAGCACGGATACGGATG | CTTCTTGAATGCCCGATTA |
| <i>rpoC1</i> | TCGGATACGAAGATATCAAATGG | TTAGTTATGGGCCTAGCAAAGA |
| <i>rpoC2</i> | ATGGAGCCCGTAAAGGAGTT | CGTCTGCTAAGACACGACCA |
| <i>rpS11</i> | TACTTGTGGATTCCGGGGTA | CAGCTCGTTGCATACCTTGA |
| <i>rpS12A</i> | TCTCACACCGGGTAAATCCT | ATCCGAAACGTCACGAAATC |
| <i>rps14</i> | AATCCCCACCGCGTAATAGT | AACATGCCTGAACCATTTC |
| <i>rpS15</i> | CAGGGGATCCGTTGAATTT | CGTTGACGTTTTCCAGAAT |
| <i>rpS18</i> | CAAGCGATCTTTTCGTAGGC | AAAGTCACTCTATTACCCGTCT |
| <i>rpS19</i> | CACAATGATTGGCCATACGA | TTTGGCATGTCCTCGAAAA |
| <i>rpS2</i> | GGGCTCGGTGTCATTATGTT | TCTTCAACACAGCTGCATCC |
| <i>rpS3</i> | CAATCCGTATGGGGATCCTA | GATCCATTCAACACGTGCAA |
| <i>rps4</i> | CGATTGGGTATGGCTTTGAC | ATGGTTTGGCAATTCCTCAG |
| <i>rpS7</i> | AAACTGCAAAATCCGATCCA | ATGAGTTGACCCGCTACAC |
| <i>rpS8</i> | CGACCGGGTCTACGAATCTA | ATTTCTCCGCCGATTCTTTT |
| <i>ycf1</i> | TTTCGGAAGAAGGGGAAGAT | TTCGAACGTGGAATTCATCA |
| <i>ycf15</i> | GCGAACAACCGGAGCTATTA | CCGACATGCGTATTTTTGATT |
| <i>ycf2.1</i> | TAGCCCTCGGTCTATTGGTG | GGATCCACTTTTTGGGGAAT |
| <i>ycf3</i> | TCCAATACTCAGCGGCTTG | TTCGGGCATTAGAACGAAAC |
| <i>ycf4</i> | TTTCTATGGGATCGCAGGTC | GGAAATCCCCAACGAAAAAT |
| <i>rRNA 23S</i> | GGGCGACTGTTTACCAAAAA | TTACCCGACAAGGAATTCG |
| <i>rRNA 16S</i> | CGGTATCTGGGGAATAAGCA | GATTTGACGGCGGACTTAA |
| | | |
| | | |
| transcript | forward | reverse |
| spliced <i>atpF</i> | GTCCTTCACGCAGTTCTTC | TACTTGGGTCACTGGCCATC |
| unspliced <i>atpF</i> | GTCCTTCACGCAGTTCTTC | TTCGGGAAGGGATCATAGAA |
| spliced <i>rpoC1</i> | ATTCAAATGAACCTCGTAATCGTAA | TATGGTATTTGAAACGTCTTCTAGTTAT |
| unspliced <i>rpoC1</i> | TTAGTTATGGGCCTAGCAAAGA | AACTACTTGAGCCGGATGAGA |
| spliced <i>ycf3</i> intron2 | TCCAATACTCAGCGGCTTG | TTCGGGCATTAGAACGAAAC |
| unspliced <i>ycf3</i> intron2 | TCCAATACTCAGCGGCTTG | AGTTGGTTGTCGAGCCGTAT |
| spliced <i>ycf3</i> intron1 | CGAGTCATTCCGACAACCTTC | TGTGGTAAAAAGGGGTTTCG |
| unspliced <i>ycf3</i> intron2 | TGTGGTAAAAAGGGGTTTCG | TTCGTTAAAGCGAATGACGA |
| spliced <i>rps12</i> intron1 | TCTCACACCGGGTAAATCCT | ATCCGAAACGTCACGAAATC |
| unspliced <i>rps12</i> intron1 | TCTCACACCGGGTAAATCCT | GGAGCCGTATGAGGTGAAAA |
| spliced <i>clpP</i> intron2 | GTCGGAGGAGCAATTACCAA | GTGATGGTTTTCGCGAAGTTT |
| unspliced <i>clpP</i> intron2 | GTGATGGTTTTCGCGAAGTTT | TCATTCTGCGAAATAGAAAAACC |
| spliced <i>clpP</i> intron1 | AATAAGTTGATTTCGAGATTTGCGT | TTCGAAGTCTGGAGAAGGA |
| unspliced <i>clpP</i> intron1 | AATAAGTTGATTTCGAGATTTGCGT | GAACCGTATGCACCAAAGG |
| spliced <i>petB</i> | CATTGTATATTTCCGGAATATGAG | TATGTTGACATGCGGAGGAA |
| unspliced <i>petB</i> | TCTTGGAGGGGAGTAACCT | TATGTTGACATGCGGAGGAA |
| spliced <i>petD</i> | GAAGAGATAATGGATTATGGGAGTG | GGGTTCCCCGTAATAATTGTG |
| unspliced <i>petD</i> | AAAAATTATCATGTCCGGTTCC | GGGTTCCCCGTAATAATTGTG |
| spliced <i>rpl16</i> | CCAAATTTTCCACCACGTC | TCCTTTGATATAATTGCTATGCTTAGT |
| unspliced <i>rpl16</i> | CCAAATTTTCCACCACGTC | GAAACTCTCACGTTTCAGTTCTGT |
| spliced <i>rpl2</i> | CGGACCTCTCCAGAAGGTAAT | AAATGGGAAATGCCCTACCT |
| unspliced <i>rpl2</i> | CGGACCTCTCCAGAAGGTAAT | GCCGTATGCTTTGGAAGAAG |
| spliced <i>ndhB</i> | CCAGAAGAAGATGCCATTCA | TCATCAATGGAATCCTGACG |
| unspliced <i>ndhB</i> | CCAGAAGAAGATGCCATTCA | AGTCTCATGCACGGTTTTGA |
| spliced <i>rps12</i> intron2 | TTTGGCTTTTTGACCCATA | CGTAAAGTTGCCAGAGTACGA |
| unspliced <i>rps12</i> intron2 | TTTGGCTTTTTGACCCATA | TGTGGAAGCCGTATTTCGAT |
| spliced <i>ndhA</i> | TTGACGCCACAAATTCAT | TTAGGTGGTCTGCGAGCTG |
| unspliced <i>ndhA</i> | TTGACGCCACAAATTCAT | AGGCCAAGACCTCATGTACG |

Supplemental Table 2. Spots identified by mass spectrometry.

| ColO | | | | | | |
|------|--------------|---|-------|--------------|----|--------------|
| Spot | Accession | Protein description | Score | protein mass | MW | prot_matches |
| 1 | AtCg00680.1 | PSBB encodes for CP47, subunit of the photosystem II reaction center. photosystem II 47 kDa protein | 50 | 56001 | 2 | |
| 2 | AtCg00280.1 | PSBC chloroplast gene encoding a CP43 subunit of the photosystem II reaction center. promoter contains a | 45 | 51835 | 4 | |
| 3 | AtSg01530.1 | LHC84.2 LHC84.2 (LIGHT HARVESTING COMPLEX PSII); chlorophyll binding similar to chlorophyll a/b-bindin | 48 | 31120 | 3 | |
| 4 | AtMg10340.1 | LHC85 LHC85 (LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5); chlorophyll binding chlorophyll a/b-t | 39 | 24986 | 2 | |
| 5 | AtCg00350.1 | PSAA Encodes psaa protein comprising the reaction center for photosystem I along with psbB protein; hydr | 40 | 30138 | 4 | |
| 6 | AtCg00340.1 | PSAB Encodes the D1 subunit of photosystem I and II reaction centers. photosystem I P700 apoprotein A2 | 39 | 83178 | 2 | |
| 7 | AtCg00350.1 | PSAA Encodes psaa protein comprising the reaction center for photosystem I along with psbB protein; hydr | 64 | 82423 | 3 | |
| 8 | AtCg00680.1 | PSBB encodes for CP47, subunit of the photosystem II reaction center. photosystem II 47 kDa protein | 58 | 83178 | 4 | |
| 9 | AtMg12800.1 | PSAL PSAL (photosystem I subunit L) photosystem I chain XI precursor - spinach, PIR2:S35151;supported l | 72 | 56001 | 1 | |
| 10 | AtMg12800.1 | PSAL PSAL (photosystem I subunit L) photosystem I chain XI precursor - spinach, PIR2:S35151;supported l | 47 | 23037 | 2 | |
| 11 | AtMg12800.1 | PSAF PSAF (photosystem I subunit F) similar to photosystem I subunit III precursor GB-U08135 GI:854707 | 49 | 24158 | 3 | |
| 12 | AtMg12800.1 | PSAF PSAF (photosystem I subunit F) similar to photosystem I subunit III precursor GB-U08135 GI:854707 | 88 | 24158 | 4 | |
| 13 | AtMg12800.1 | PSAL PSAL (photosystem I subunit L) photosystem I chain XI precursor - spinach, PIR2:S35151;supported l | 49 | 23037 | 3 | |
| 14 | AtMg12800.1 | PSAF PSAF (photosystem I subunit F) similar to photosystem I subunit III precursor GB-U08135 GI:854707 | 47 | 24158 | 1 | |
| 15 | AtMg12800.1 | PSAF PSAF (photosystem I subunit F) similar to photosystem I subunit III precursor GB-U08135 GI:854707 | 146 | 29163 | 5 | |
| 16 | AtCg00490.1 | RBCL large subunit of RUBISCO. | 206 | 52922 | 9 | |
| 17 | AtCg00480.1 | ATPB chloroplast-encoded gene for beta subunit of ATP synthase ATP synthase CF1 beta chain | 111 | 53900 | 7 | |
| 18 | AtCg00350.1 | PSAA Encodes psaa protein comprising the reaction center for photosystem I along with psbB protein; hydr | 58 | 83178 | 3 | |
| 19 | AtCg00120.1 | ATPA Encodes the ATPase alpha subunit, which is a subunit of ATP synthase and part of the CF1 portion whi | 250 | 55294 | 17 | |
| 20 | AtCg00480.1 | ATPB chloroplast-encoded gene for beta subunit of ATP synthase ATP synthase CF1 beta chain | 627 | 53900 | 21 | |
| 21 | AtSg08670.1 | ATP2a ATP synthase beta chain 1, mitochondrial supported by cDNA: gi_15809908_gb_AY054222.1_ | 74 | 59634 | 3 | |
| 22 | AtCg00680.1 | PSBB encodes for CP47, subunit of the photosystem II reaction center. photosystem II 47 kDa protein | 72 | 56001 | 3 | |
| 23 | AtCg00280.1 | PSBC chloroplast gene encoding a CP43 subunit of the photosystem II reaction center. promoter contains a | 51 | 51835 | 4 | |
| 24 | AtCg00540.1 | PETA Encodes cytochrome f apoprotein; involved in photosynthetic electron transport chain; encoded by the | 123 | 35335 | 9 | |
| 25 | AtCg00270.1 | PSBD PSII D2 protein photosystem II protein D2 | 68 | 39522 | 2 | |
| 26 | AtCg00270.1 | PSBD PSII D2 protein photosystem II protein D2 | 63 | 39522 | 3 | |
| 27 | AtSg01530.1 | chlorophyll A-B binding protein CP29 (LHC84) supported by cDNA: gi_15081738_gb_AY048262.1_ | 93 | 31120 | 5 | |
| 28 | AtSg08940.2 | LHC84.2 LHC84.2 (LIGHT HARVESTING COMPLEX PSII) | 55 | 31174 | 4 | |
| 29 | AtMg15820.1 | CP24, LHC86 LHC86 (LIGHT HARVESTING COMPLEX PSII); chlorophyll binding similar to chlorophyll binding | 53 | 27505 | 1 | |
| 30 | AtMg15820.1 | PSAD-2 PSAD-2 (photosystem I subunit D-2) Strong similarity to gb X14017 from Spinacia oleracea. ESTs g | 44 | 22293 | 3 | |
| 31 | AtMg15820.1 | PSAD-1 PSAD-1 (photosystem I subunit D-1) similar to spinach PSI-D, GenBank accession number P12353; | 103 | 22584 | 7 | |
| 32 | AtMg15820.1 | CAB4, LHCA4 LHCA4 (Photosystem I light harvesting complex gene 4); chlorophyll binding CHLOROPHYLL A | 122 | 27716 | 4 | |
| 33 | AtMg15820.1 | PSBE PSII cytochrome b559. There have been many speculations about the function of Cyt b559, but the m | 62 | 9391 | 2 | |
| 34 | AtCg00540.1 | PETA Encodes cytochrome f apoprotein; involved in photosynthetic electron transport chain; encoded by the | 163 | 35335 | 8 | |
| 35 | AtCg00820.1 | PSB02, PSB0-2 PSB0-2/PSB02 (PHOTOSYSTEM II SUBUNIT O-2); oxygen evolving identical to OEC33, SP: | 64 | 34998 | 4 | |
| 36 | AtMg129910.1 | CAB3, AB180, LHC81.2 CAB3 (CHLOROPHYLL A/B BINDING PROTEIN 3); chlorophyll binding similar to phot | 207 | 28209 | 8 | |
| 37 | AtMg129910.1 | CAB3, AB180, LHC81.2 CAB3 (CHLOROPHYLL A/B BINDING PROTEIN 3); chlorophyll binding similar to phot | 83 | 28209 | 6 | |
| 38 | AtMg129910.1 | LHC83*1, LHC83 LHC83 (LIGHT-HARVESTING CHLOROPHYLL BINDING PROTEIN 3) supported by cDNA: gi_ | 46 | 28688 | 2 | |
| 39 | AtMg129910.1 | PSBS, NPQ4 NPQ4 (NONPHOTOCHEMICAL QUENCHING) | 42 | 21967 | 2 | |
| 40 | AtMg129910.1 | CAB4, LHCA4 LHCA4 (Photosystem I light harvesting complex gene 4); chlorophyll binding CHLOROPHYLL A | 42 | 27716 | 1 | |
| 41 | AtCg00280.1 | PSBC chloroplast gene encoding a CP43 subunit of the photosystem II reaction center. promoter contains a | 81 | 51835 | 3 | |
| 42 | AtCg00820.1 | PSB02, PSB0-2 PSB0-2/PSB02 (PHOTOSYSTEM II SUBUNIT O-2); oxygen evolving identical to OEC33, SP: | 62 | 34998 | 2 | |
| 43 | AtSg01530.1 | chlorophyll A-B binding protein CP29 (LHC84) supported by cDNA: gi_15081738_gb_AY048262.1_ | 119 | 31120 | 6 | |
| 44 | AtSg08940.2 | LHC84.2 LHC84.2 (LIGHT HARVESTING COMPLEX PSII) | 116 | 31174 | 8 | |
| 45 | AtMg10340.1 | LHC85 LHC85 (LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5); chlorophyll binding chlorophyll a/b-t | 249 | 30138 | 6 | |
| 46 | AtMg129910.1 | CAB3, AB180, LHC81.2 CAB3 (CHLOROPHYLL A/B BINDING PROTEIN 3); chlorophyll binding similar to phot | 61 | 28209 | 3 | |
| 47 | AtSg54270.1 | LHC83*1, LHC83 LHC83 (LIGHT-HARVESTING CHLOROPHYLL BINDING PROTEIN 3) supported by cDNA: gi_ | 46 | 28688 | 2 | |
| 48 | AtMg15820.1 | CP24, LHC86 LHC86 (LIGHT HARVESTING COMPLEX PSII); chlorophyll binding similar to chlorophyll binding | 96 | 27505 | 5 | |
| 49 | AtCg00680.1 | PSBB encodes for CP47, subunit of the photosystem II reaction center. photosystem II 47 kDa protein | 54 | 56001 | 2 | |

| otp51 | | | | | | |
|-------|--------------|---|-------|--------------|----|--------------|
| Spot | Accession | Protein description | Score | protein mass | MW | prot_matches |
| 1 | AtCg00490.1 | RBCL large subunit of RUBISCO. | 59 | 52922 | 5 | |
| 2 | AtMg01190.2 | ATP1 RNA edited; added by Plant Energy, novel AGI | 114 | 55011 | 4 | |
| 3 | AtMg07698.1 | ATP1 ATP synthase alpha chain, mitochondrial, putative | 95 | 85879 | 4 | |
| 4 | AtCg00120.1 | ATPA Encodes the ATPase alpha subunit, which is a subunit of ATP synthase and part of the CF1 portion whi | 46 | 55294 | 3 | |
| 5 | AtSg08670.1 | ATP2a ATP synthase beta chain 1, mitochondrial supported by cDNA: gi_15809908_gb_AY054222.1_ | 140 | 59634 | 10 | |
| 6 | AtCg00480.1 | ATPB chloroplast-encoded gene for beta subunit of ATP synthase ATP synthase CF1 beta chain | 36 | 53900 | 3 | |
| 7 | AtCg00480.1 | ATPB chloroplast-encoded gene for beta subunit of ATP synthase ATP synthase CF1 beta chain | 100 | 53900 | 9 | |
| 8 | AtCg00120.1 | ATPA Encodes the ATPase alpha subunit, which is a subunit of ATP synthase and part of the CF1 portion whi | 95 | 55294 | 10 | |
| 9 | AtSg01530.1 | chlorophyll A-B binding protein CP29 (LHC84) supported by cDNA: gi_15081738_gb_AY048262.1_ | 93 | 31120 | 3 | |
| 10 | AtMg10340.1 | LHC85 LHC85 (LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5); chlorophyll binding chlorophyll a/b-t | 50 | 30138 | 3 | |
| 11 | AtMg129910.1 | CAB3, AB180, LHC81.2 CAB3 (CHLOROPHYLL A/B BINDING PROTEIN 3); chlorophyll binding similar to phot | 41 | 28209 | 2 | |

II.1.2 Identification de défauts d'épissage en utilisant la PCR quantitative.

Plusieurs méthodes ont été mises au point pour identifier et quantifier des défauts d'épissage d'ARNm. La plus communément utilisée est le *northern blot*. Celle-ci permet non seulement de vérifier le taux d'introns épissés, mais aussi de regarder en parallèle la maturation des transcrits. Une autre technique, qui a été fortement utilisée lors de l'identification de facteurs d'épissage chloroplastiques, est la PPE (Poisoned Primer Extension). Cette méthode, décrite dans la Figure 4, permet grâce à sa grande sensibilité de quantifier très précisément un défaut d'épissage.

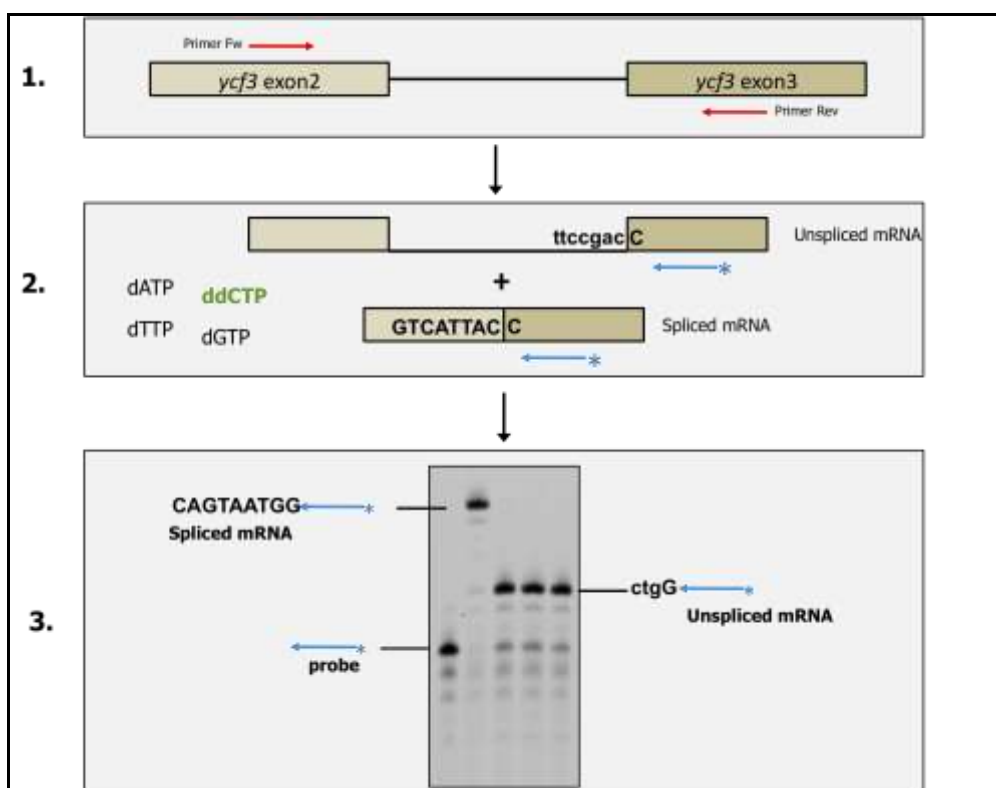


Figure 4 : Principe de la PPE (Poisonned Primer Extension), exemple de détection de l'épissage de l'intron 2 d'*ycf3*.

1. Une PCR ayant pour matrice de l'ADNc dilué est réalisée en présence d'oligonucléotides spécifiques de l'exon 2 et l'exon 3 de *ycf3*. Cette PCR permet d'amplifier à la fois la forme non épissée et la forme épissée de l'ARNm. Les produits PCR sont alors purifiés afin d'éliminer les oligonucléotides et les dnTPs restants.

2. Une réaction de type « séquençage » sur les produits PCR est réalisée en présence d'un oligonucléotide marqué en 5' par de la Fluorescéine (FAM) et spécifique de la partie en 5' de l'exon 3, de dnTPs (dATP, dGTP, dTTP) et d'un dideoxynucléotide (ddCTP). Lors de l'inclusion du ddCTP au cours de la synthèse, la réaction d'élongation est alors bloquée.

3. La séquence en 3' de l'intron et en 3' de l'exon 2 n'étant pas les mêmes, deux types de produits PPE sont alors obtenus. Un premier produit où l'inclusion du ddCTP se fait 4 nucléotides après l'oligonucléotide, alors spécifique de la forme non épissée de l'intron 2 d'*ycf3*, et un autre produit plus grand car l'inclusion du ddCTP s'effectue 9 nucléotides plus loin. Ce dernier produit est spécifique de la forme épissée d'*ycf3*. Ces produits sont dénaturés et déposés sur un gel d'acrylamide afin de permettre une séparation à un nucléotide près des produits de PPE et de comparer les ratios des formes épissées et non épissées obtenues.

Lors de crible de mutants, ces deux méthodes ne sont malheureusement pas envisageables et dans l'optique d'identifier des protéines PPR impliquées dans l'épissage, il fallait mettre en

place une technique adéquate. La méthode la plus simple était alors d'amplifier par PCR la forme épissée en utilisant des oligonucléotides sur chaque exon et de vérifier la présence de chaque transcrit épissé pour chaque événement d'épissage chloroplastique mais aussi mitochondrial. Même si cette méthode, très rapide, permet de cribler un défaut total d'épissage chez les mutants étudiés, elle n'est pas quantitative et permet difficilement d'identifier des défauts partiels. De plus, un certain nombre d'introns se trouvant dans les ARNt ne peuvent être analysés par cette méthode étant donné l'impossibilité de dessiner des oligonucléotides dans ces zones très courtes.

Afin de pallier au problème de la quantitativité, nous avons choisi d'utiliser une approche de PCR quantitative. Pour chaque événement d'épissage des oligonucléotides ont été dessinés sur les exons et sur l'intron associé, permettant ainsi d'amplifier la forme épissée et la forme non épissée. Le ratio « forme épissée/ forme non épissée » pour chaque événement d'épissage est alors comparé entre le mutant étudié et le sauvage (Figure 5).

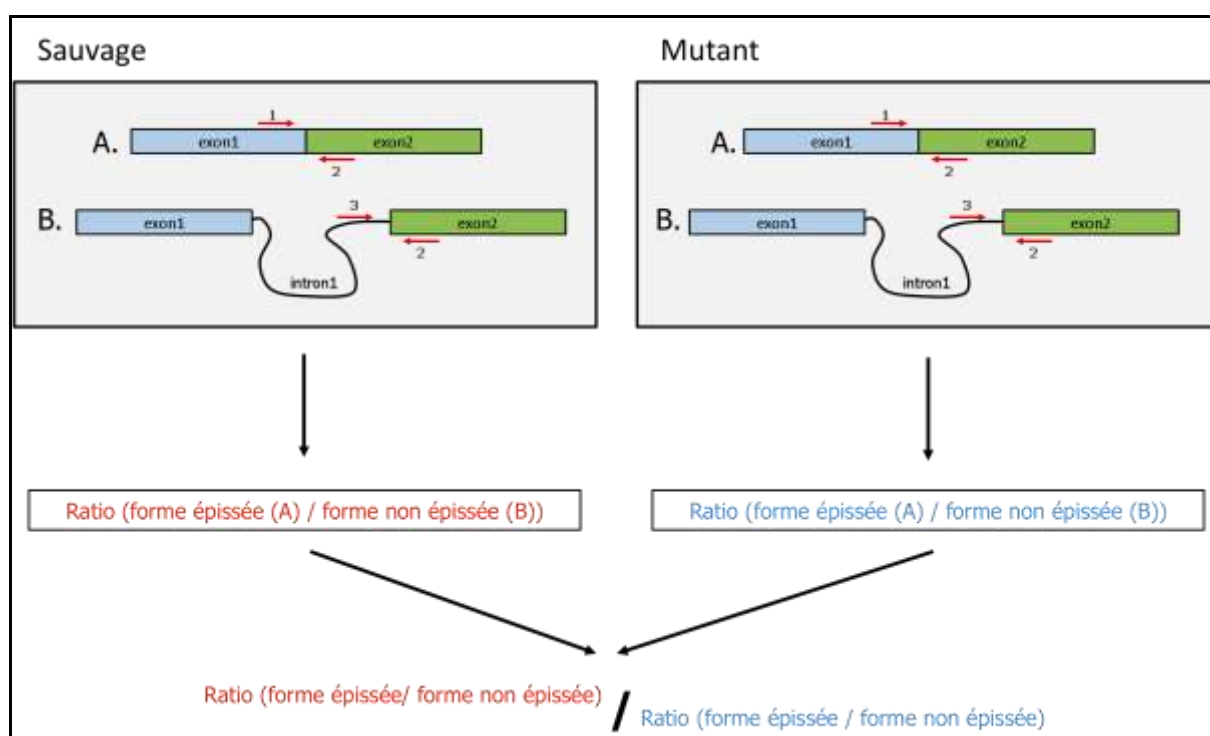


Figure 5 : Schéma représentant la technique de quantification de l'épissage par PCR quantitative.

Les conditions et le dessin des oligonucléotides sont décrits dans l'article 2 (supplemental table 1) pour les événements d'épissage chloroplastiques et dans l'article 3 (supplemental table 1) pour les événements d'épissage mitochondriaux.

Comme il a été montré pour OTP51, cette approche permet de quantifier très précisément des défauts d'épissage au même titre que de la PPE. Cependant le problème de dessin d'oligonucléotides pour de la PCR quantitative sur les introns des ARNt réside, et seul le *northern blot* permet de regarder ces événements.

Grâce à ce nouvel outil, il m'a été possible de quantifier les événements d'épissage pour un certain nombre de mutants *ppr* et ainsi identifier de nouvelles protéines PPR impliquées dans cette étape importante de modification post-transcriptionnelle.

II.1.2.1 OTP52 (At3g06430, emb2750)

OTP52 est une PPR de type « Pure » prédite dans les chloroplastes sous Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>) et TargetP (<http://www.cbs.dtu.dk/services/TargetP/>). Le mutant d'insertion ADN-T dans ce gène, *emb2750*, est décrit comme « embryo-letal » chez *Arabidopsis*. Chez les dicotylédones, contrairement aux monocotylédones, un certain nombre de gènes essentiels pour le développement des chloroplastes sont codés dans le génome chloroplastique (*accD*, *ycf1*, *ycf2*), alors qu'ils sont absents du génome chloroplastique chez les monocotylédones (Konishi, Shinohara et al. 1996; Wakasugi, Tsudzuki et al. 2001). Un défaut de traduction ou de certaines modifications post-transcriptionnelles au niveau de ces gènes peuvent donc aboutir à un phénotype très marqué, l'arrêt de développement des plastes conduisant à un arrêt précoce du développement embryonnaire. Ainsi, il n'est pas possible d'obtenir des mutants homozygotes pour des gènes essentiels dans la traduction plastidiale. Le gène *accD* code pour une β -carboxyl transférase, sous-unité de l'acetyl-CoA carboxylase. Cette enzyme a un rôle essentiel dans la synthèse des acides gras (Kode, Mudd et al. 2005) et se trouve essentielle lors de la germination de la graine. Etienne Delannoy (chercheur post-doctoral dans l'équipe de Ian Small à Perth) a envisagé une approche permettant de sauver des mutants homozygotes en les complétant avec un transgène *accD* nucléaire codant pour une protéine ACC-D portant un peptide d'adressage chloroplastique. En complétant avec cette construction chimérique, il serait éventuellement possible de « passer » certains stades de développement embryonnaire, et ainsi sauver en partie certains mutants décrits comme embryo-letaux.

Parmi les mutants testés, se trouve le mutant « embryo-letal » *otp52*. La complémentation par le gène *accD* n'a pas très bien fonctionné, mais il a été possible de récupérer des mutants d'insertion ADN-T homozygotes pour ce gène. Ces mutants présentent un très fort phénotype : très faible taux de germination mais supérieur chez les mutants complétés par le gène *accD* que chez le mutant non complété ; des plantules blanches et très petites en condition *in vitro*.

Des analyses réalisées par Etienne sur les ARNt chloroplastiques par *northern blot* ont montré que le mutant présente un défaut d'épissage de l'ARNt Val UAC. En effet, aucun ARNt Val mature n'est observé alors que l'on observe une accumulation de la forme non épissée de cet ARNt. En addition de cette absence quasi totale de l'épissage de cet intron, le mutant présente également une diminution de l'épissage des introns ARNt Ile et Ala, celle-ci étant probablement due à un défaut de maturation de l'ARN 23S. Afin de voir si d'autres événements d'épissage sont affectés chez ce mutant, j'ai réalisé une analyse de l'épissage des ARN chloroplastiques par PCR quantitative. Comme il est montré sur la Figure 6, en plus du défaut d'épissage de l'ARNt Val, le mutant présente un très fort défaut d'épissage de l'intron d'*atpF*, pratiquement aucun ARNm d'*atpF* mature ne pouvant être détecté et également un défaut partiel d'épissage de l'intron 2 de *rps12*.

Un défaut d'épissage d'*atpF* a déjà été observé pour un grand nombre de mutants, tels que *crs1* et *why1* (Jenkins, Kulhanek et al. 1997; Till, Schmitz-Linneweber et al. 2001; Asakura and Barkan 2006; Prikryl, Watkins et al. 2008). Et ce défaut chez ces mutants conduit à un phénotype assez similaire à celui d'*otp52*. L'absence d'épissage de l'ARNt Val devrait conduire à un problème majeur dans la traduction chloroplastique. Et on peut se demander si les défauts d'épissage observés sur les autres introns sont des effets indirects ou des conséquences du défaut de traduction engendré par le défaut d'épissage de l'ARNt Val.

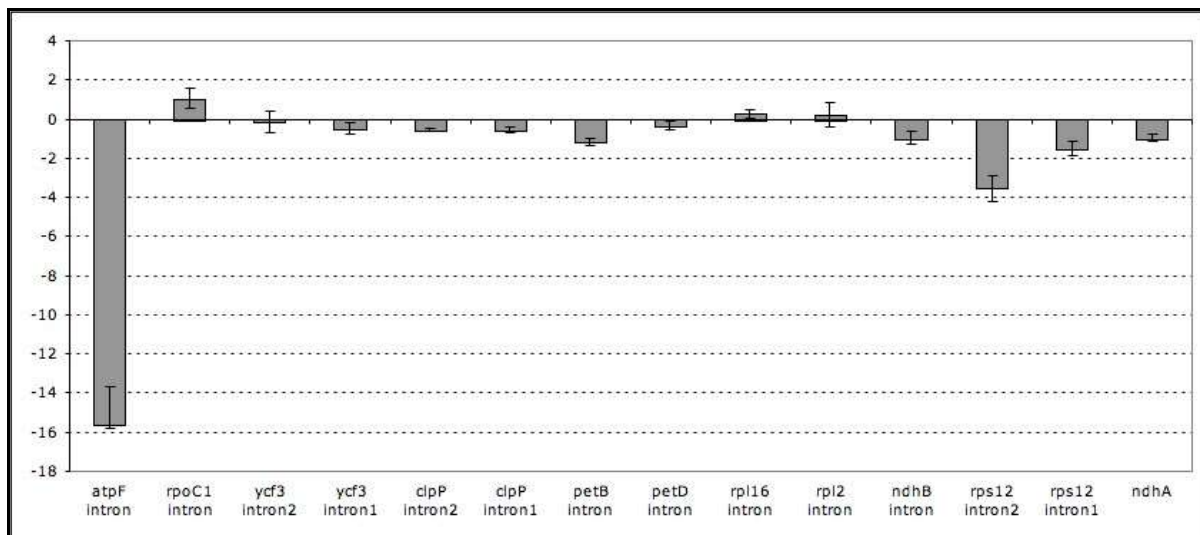


Figure 6 : Détection des événements d'épissage par qPCR chez le mutant *otp52*. Pour chaque intron, le ratio entre les formes épissées et non épissées chez le mutant est comparé au sauvage et est représenté en abscisse sous la forme d'un $\text{Log}_2 : \text{Log}_2 [(forme\ épissée / forme\ non\ épissée)\ mutant / ((forme\ épissée / forme\ non\ épissée)\ sauvage)]$. L'épissage des ARNt n'est pas représenté sur ce graphique car la détection de celui-ci a été quantifiée par *northern blot* (données non montrées).

Deux études différentes permettent en partie de répondre à cette question. La première, qui a été réalisée au sein du laboratoire de Ian Small à Perth et qui est expliquée ultérieurement dans ce chapitre, consiste à regarder le profil d'épissage chez des mutants affectés dans la transcription de type PEP. En effet, un des effets de l'arrêt de la traduction conduit à un défaut de la transcription impliquant les ARN polymérase codées par le génome chloroplastique (PEP). Et ce défaut de transcription sur des transcrits ayant des introns pourrait provoquer des défauts apparents d'épissage.

La deuxième étude consiste à connaître quels sont les introns dépendants de la maturase chloroplastique, MatK. Un des effets de l'arrêt de la traduction est l'absence de synthèse de la maturase chloroplastique, MatK, codée dans l'intron *trnK*. Cette protéine qui a longtemps été décrite comme jouant certainement un rôle essentiel dans l'épissage d'introns chloroplastiques, vient récemment d'être mieux caractérisée. MatK pourrait être impliquée dans l'épissage de l'intron d'*atpF*, de *rpl2*, de l'intron2 de *rps12*, de l'intron de *trnV*, *trnK*, *trnA* et de *trnI* puisque son association avec ces différents introns a été révélée par RIP-CHIP (Zoschke, Nakamura et al. 2010).

Chez le mutant *otp52*, cinq introns sur les sept présentant une interaction avec MatK sont affectés, néanmoins aucun défaut d'épissage n'est observé pour l'intron de *rpl2* et de *trnK*. Il faut remarquer que l'interaction de MatK avec ces 7 introns ne signifie pas nécessairement que cette maturase ait un rôle essentiel dans leur épissage. Et il est donc envisageable que les effets observés chez le mutant *otp52* ne correspondent qu'à des effets indirects dus à l'absence de traduction. Cependant l'absence de défaut sur l'intron de *trnK*, dont l'épissage serait certainement très lié à la maturase codée par son intron, suggère qu'il existe un certain taux de traduction de cette maturase et donc que la protéine OTP52 a un effet plutôt direct sur l'épissage des introns *trnV* et *atpF* et qu'une partie du défaut de l'ARNt^{Val} mature serait compensée par un autre mécanisme permettant un taux « minimal » de la traduction. En effet, des données de *western blot* effectuées par Etienne Delannoy montrent une quantité significative des protéines chloroplastiques AtpB, RbcL, PsaD et PsbA et de nouvelles données obtenues par Catherine Colas-des-Francis ne montrent aucune différence

dans l'accumulation de la protéine AccD entre les mutants *otp52* et *crs1* et, en revanche, une sur-accumulation comparée au sauvage. En conclusion, ceci suggère que le mutant *otp52* est plus apparenté à un mutant de l'ATP synthase, au même titre que *crs1*, et que le défaut de traduction lié aux problèmes d'épissage des ARNt est relativement mineur.

II.1.2.2 Profil d'épissage des mutants affectés dans la transcription de type PEP

Un certain nombre de phénomènes de régulation transcriptionnelle et post-transcriptionnelle peuvent affecter l'efficacité de l'épissage et il est important, mais difficile, de dissocier les effets indirects d'une mutation aux effets directs de celle-ci sur l'épissage. L'outil de PCR quantitative n'étant pas brin spécifique, il est fort possible que, dans un cas où il existe un défaut de transcription, on observe moins d'épissage pour un intron mais que la forme non épissée observée reste au même niveau que dans le sauvage alors que celle-ci devrait être réduite en raison du défaut transcriptionnel pour ce transcrit. Cela peut s'expliquer par l'amplification de transcrits chevauchants, en général en sens opposé, qui ne sont pas forcément affectés et qui peuvent au contraire s'accumuler et ainsi biaiser le résultat observé. Seul un *northern blot* permettrait de discriminer parmi ces deux événements. Cependant, il est possible d'avoir une idée du profil d'épissage des transcrits chloroplastiques dans le cas de mutants affectés dans la transcription et ainsi de définir quels sont les événements d'épissage affectés dans ces mutants.

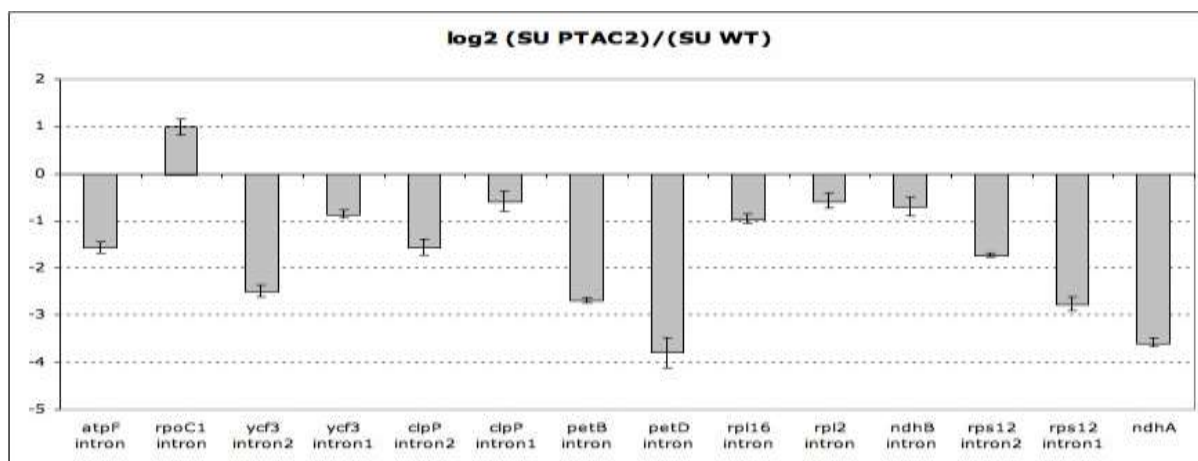
La transcription chloroplastique est effectuée par deux types de polymérases. Deux polymérases codées par le génome nucléaire, RpoTp et RpoTmp, nommées NEP pour « Nuclear Encoded Polymerase », et une autre codée par le génome chloroplastique, la PEP (Plastid-Encoded Polymerase), comprenant quatre sous-unités, RpoA, RpoB, RpoC1, RpoC2. Ces deux types de polymérases ont un rôle bien défini dans la transcription chloroplastique et, alors que ces trois polymérases vont transcrire un ensemble de gènes en commun, d'autres gènes sont transcrits préférentiellement par la PEP ou les polymérases de type NEP. Deux mutants *ppr* (*ptac2* et *clb19*) affectés dans la transcription par la PEP ont été identifiés et présentent le même phénotype. PTAC2 est une PPR de type « Pure » comportant un domaine endonucléase de type SMR en C-terminal. Plusieurs autres protéines PPR comportant ce motif sont codées par le génome nucléaire d'*Arabidopsis* et sont soit localisées dans les chloroplastes comme GUN1 intervenant dans le signal rétrograde (Koussevitzky, Nott et al. 2007) ou encore P67 identifiée comme se fixant à l'ARN (Lahmy, Barneche et al. 2000), soit dans les mitochondries.

Le mutant *ptac2* présente un défaut de transcription de la PEP (Pfalz, Liere et al. 2006). Cette protéine PPR trouvée dans des complexes de haut poids moléculaire associés à des protéines impliquées dans la transcription, serait directement liée à l'activité transcriptionnelle de type PEP, cependant sa fonction moléculaire exacte n'est pas encore élucidée.

CLB19 est une PPR de type E+ impliquée dans l'édition des deux sites chloroplastiques *clpP* et *rpoA* (Chateigner-Boutin, Ramos-Vega et al. 2008). ClpP est une protéase et RpoA est une sous-unité de la PEP. Le défaut d'édition de ce transcrit codant pour cette sous-unité conduit à un défaut de la transcription par la PEP équivalent à celui observé dans le mutant *ptac2* et de ce fait le profil transcriptionnel du génome chloroplastique des deux mutants est identique.

Afin de déterminer le profil d'épissage chez un mutant affecté dans la PEP, ces deux mutants ont été étudiés. Les résultats des profils d'épissage pour chacun sont représentés dans la figure 7.

A.



B.

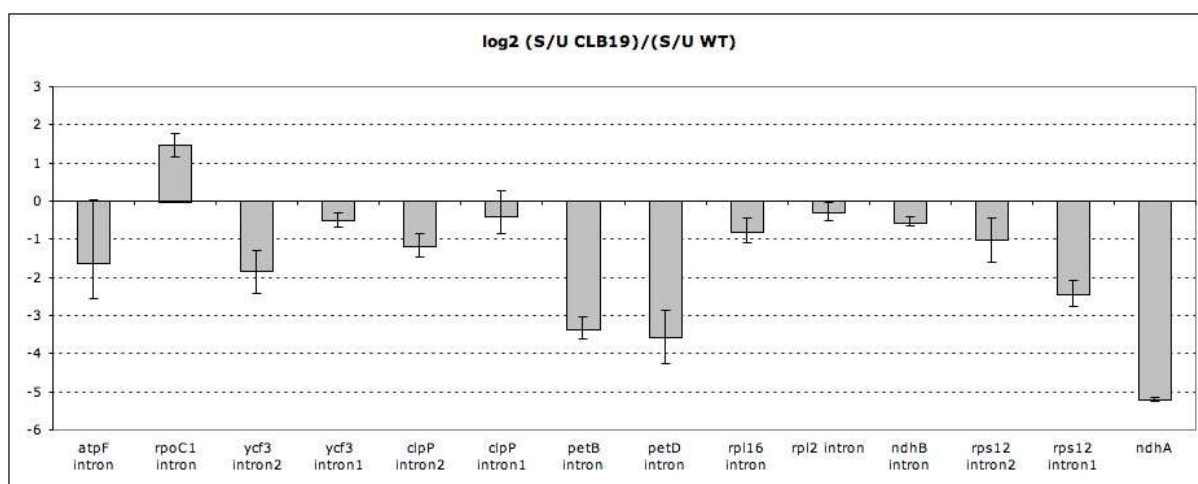


Figure 7 : Profil de l'épissage des introns chloroplastiques chez les mutants *ptac2* (A.) et *clb19* (B.) comparés à des plantes de type sauvage. Les ratios des formes épissées sur les formes non épissées entre le mutant et le sauvage sont représentés en Log2 : $\text{Log}_2 \left[\frac{\text{forme épissée}}{\text{forme non épissée}} \right]_{\text{mutant}} / \left[\frac{\text{forme épissée}}{\text{forme non épissée}} \right]_{\text{sauvage}}$.

Seuls les introns des gènes chloroplastiques codants pour des protéines ont été analysés dans cette expérience, et aucune donnée n'a été obtenue concernant l'épissage des ARNt. Lorsque l'on regarde les deux profils d'épissage chez les deux mutants, on observe les mêmes défauts d'épissage. Huit introns présentent un ratio formes épissées sur formes non épissées au moins deux fois inférieur à celui du sauvage. Ces deux mutants présentent des défauts de transcription et donc le ratio pourrait être biaisé du fait de l'existence de zones chevauchantes. C'est donc pour cette raison qu'il est aussi important de regarder le taux des formes épissées avant de comparer les ratios. En établissant cette comparaison représentée dans la figure 8, les deux mutants présentent moins de formes épissées pour *petB*, *petD* et *ndhA* et étrangement plus de formes épissées pour *rpoC1*. Le transcriptome chloroplastique de ces deux mutants montre une diminution des transcrits de *petB* et *petD* qui sont des gènes majoritairement sous le contrôle de la PEP. La diminution des formes épissées pour ces transcrits peut donc s'expliquer par une diminution de l'abondance des transcrits primaires.

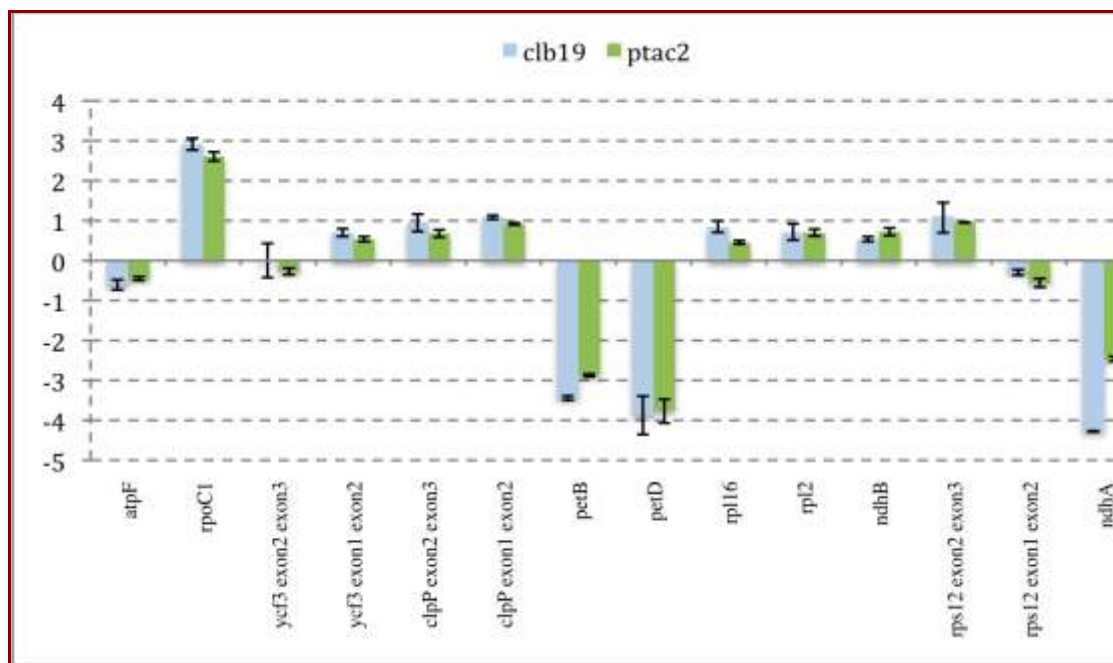


Figure 8 : Profil des transcrits chloroplastiques épissés des mutants *clb19* et *ptac2*. Log2 des ratios des formes épissées des ARNm chloroplastiques chez les mutants *clb19* et *ptac2* sur les formes épissées chez les sauvages Col0 : Log2 (forme épissée mutant/ forme épissée sauvage).

En parallèle, chez ces deux mutants, un certain nombre de transcrits spécifiques des polymérases de type NEP s'accumulent. Comme si en absence de l'activité de la PEP, les transcrits NEP ont une durée de vie plus importante et de ce fait s'accumulent. C'est le cas de *rpoC1* dont le transcrit s'accumule, mais aussi la forme épissée, suggérant une durée de vie plus importante de ces transcrits ou bien que ce défaut de la PEP s'accompagne d'une surexpression des facteurs d'épissage liés à l'intron de *rpoC1*. En ce qui concerne *ndhA*, il n'y a pas de réduction de l'expression des transcrits polycistroniques *ndhH-ndhA-ndhI-ndhG-ndhE-psaC-ndhD*, en revanche une forte diminution de l'épissage de l'intron de *ndhA* est observée chez les deux mutants. Pour ce transcrit, 5 précurseurs possibles sont décrits et il est possible que le précurseur adéquat pour l'épissage soit sous représenté chez les mutants. Les deux polymérases sont impliquées dans la transcription de ces transcrits mais les précurseurs ne sont pas les mêmes selon la polymérase. La reconnaissance par les facteurs d'épissage pourrait donc s'effectuer via un seul type de précurseur, qui serait préférentiellement transcrit via la PEP.

En conclusion, ce profil d'épissage est spécifique des mutants ayant un défaut de la polymérase PEP, et permet désormais d'avoir une certaine signature de ce type de mutant. Ces défauts ne sont visiblement pas liés à des défauts d'épissage mais plus à une diminution des transcrits ou des formes correctes pour l'épissage.

Concernant OTP52, le profil d'épissage du mutant correspondant n'est pas similaire au profil des mutants PEP. Des défauts partiels de l'épissage de *ndhA*, *petB*, *petD* et une augmentation de l'épissage de *rpoC1* sont observés chez *otp52* lorsque l'on compare les taux de formes épissées plutôt que les ratios. Cependant ces défauts sont beaucoup moins importants que ceux observés chez *clb19* et *ptac2*, suggérant encore qu'une activité de traduction existe chez le mutant *otp52* malgré le défaut d'épissage de l'ARNt Val.

II.2. Identification de protéines PPR impliquées dans l'épissage des ARNm dans les mitochondries.

Il existe une réelle disproportion entre les facteurs d'épissage chloroplastiques identifiés et les facteurs mitochondriaux. En effet, alors que tous les introns chloroplastiques ont au moins un facteur d'épissage identifié, au niveau mitochondrial, la quasi-totalité des facteurs d'épissage reste à être caractérisée.

Au début de ma thèse, seulement un mutant dans un probable facteur d'épissage avait été identifié (*nms1*) (Brangeon, Sabar et al. 2000), mais pour lequel aucun clonage positionnel n'a été entrepris. Et de ce fait, au niveau mitochondrial, tous les facteurs d'épissage restaient donc à découvrir.

23 introns de group II sont décrits dans le génome mitochondrial chez *Arabidopsis*. Un certain nombre d'homologues de protéines CRM identifiées comme des facteurs d'épissage dans les chloroplastes sont prédits être localisés dans les mitochondries et pourraient avoir un rôle dans l'épissage des introns mitochondriaux (cf article 1) mais jusqu'à présent aucune preuve n'a été apportée. Comme il a été montré précédemment, les protéines PPR sont impliquées dans l'épissage d'un certain nombre d'introns chloroplastiques et il est fort possible que ces protéines PPR aient aussi un rôle dans l'épissage dans les mitochondries au même titre que les protéines CRM.

En utilisant une approche de génétique inverse, j'ai pu identifier quatre protéines PPR ayant une fonction plus ou moins majeure dans l'épissage de certains introns mitochondriaux.

II.2.1 OTP43

OTP43 est la première protéine PPR que j'ai étudiée et caractérisée. Cette protéine PPR de type « Pure » fait partie des 48 protéines PPR pour lesquelles j'avais mis en place une étude systématique par « "ARNi" » dans le cadre du projet AGRIKOLA. Le « mutant "ARNi" » d'OTP43 présente un phénotype de retard de croissance assez fort comparé au sauvage. Dans le but de comparer un « mutant "ARNi" » à un mutant « knockout » de type insertion ADN-T pour le même gène, j'ai entrepris l'étude du mutant ADN-T Sail_867F07 dont l'insertion se situe en 3' du gène *OTP43*. Cette étude a montré, dans un premier temps, que le phénotype du mutant d'insertion était plus marqué. En plus du retard de croissance observé dans les lignées « "ARNi" », des problèmes de germination, des retards de croissance plus forts et une stérilité des graines issues des homozygotes sont observés chez le mutant Sail_867F07. Afin de comprendre le rôle de cette protéine PPR, des études plus poussées ont été réalisées et ont pu montrer que cette protéine est impliquée dans l'épissage en *trans* de l'intron 1 de *nad1*. Cette protéine dont la caractérisation est décrite dans le papier suivant est la première protéine PPR identifiée au niveau mitochondrial ayant un rôle dans l'épissage.

Article 3 : de Longevialle A.F., Meyer E.H., Andrés C., Taylor N.L., Lurin C., Millar A.H. and Small I.D. (2007) *The Pentatricopeptide Repeat Gene OTP43 Is Required for trans-Splicing of the Mitochondrial nad1 Intron 1 in Arabidopsis thaliana. The Plant Cell*, 19:3256-3265

RESULTATS

Supplemental Table 1

| | Forward primer | genome position | Reverse primer | genome position |
|-----------------------------|-----------------------------|------------------------|----------------------------|----------------------------|
| <i>rpl2</i> | CCGAAGACGGATCAAGGTAA | 155542..155561 | CGCAATTCATCACCATTITG | complement(157249..157268) |
| <i>rpl2</i> intron1 exon2 | TTAGGAAGAGCCGTACGAGG | 157142..157161 | CGCAATTCATCACCATTITG | complement(157249..157268) |
| <i>rps3</i> | AGCCGAAGGTGAGTCTCGTA | 26990..27009 | CCGATTCGGTAAGACTTGG | complement(28696..28715) |
| <i>rps3</i> intron1 exon2 | AGCCGAAGGTGAGTCTCGTA | 26990..27009 | TCTACGGCGGGTCACTAT | complement(27088..27106) |
| <i>cox2</i> | TGGGGGATTAATTGATTGGA | 40518..40537 | TGATGCTGTACTGGTCTGTT | complement(42016..42035) |
| <i>cox2</i> intron1 exon2 | TGGGGGATTAATTGATTGGA | 40518..40537 | AGCAGTACGAGCTGAAAGGC | complement(40640..40659) |
| <i>ccb452</i> | GTGGGTCCATGTAATGATCG | 51765..51785 | CACATGGAGGAGTGTGCATC | complement(52856..52875) |
| <i>ccb452</i> intron1 exon1 | CCCAGGATCGAATCAGAGTT | 52747..52765 | CACATGGAGGAGTGTGCATC | complement(52856..52875) |
| <i>nad1</i> exon1-2 | GACCAATAGATACTTCATAAGAGACCA | 289014..289040 | TTGCCATATCTTCGCTAGGTG | complement(318039..318059) |
| <i>nad1</i> intron1 exon2 | GACCAATAGATACTTCATAAGAGACCA | 289014..289040 | CGTGTCTGACGGTTCATAG | complement(289133..289152) |
| <i>nad1</i> exon2-3 | ATTCAGCTCCCGTCTCTGG | 287943..287961 | TCTGCAGCTCAATGGTCTC | complement(289034..289053) |
| <i>nad1</i> intron2 exon2 | GGTTGGGTAGGGGAACATC | 288936..288955 | TCTGCAGCTCAATGGTCTC | complement(289034..289053) |
| <i>nad1</i> exon3-4 | AAAAGAGCAGACCCCATTTGA | 147026..147045 | TCCGTTTGTATCTCCAGAAG | complement(287955..287974) |
| <i>nad1</i> intron3 exon4 | AAAAGAGCAGACCCCATTTGA | 147026..147045 | GGGAGCTGTATGAGCGGTAA | complement(147113..147132) |
| <i>nad1</i> exon4-5 | AGCCCGGATCTTCTTGA | 143401..143418 | TCTCAATGGGGTCTGCTC | complement(147030..147048) |
| <i>nad1</i> intron4 exon5 | AGCCCGGATCTTCTTGA | 143401..143418 | ACGGAGCTGCATCCTACT | complement(143482..143500) |
| <i>nad2</i> exon1-2 | GCGAGCAGAAGCAAGGTTAT | 80109..80128 | GGATCTCCACACATGTTT | complement(81259..81278) |
| <i>nad2</i> intron1 exon2 | GCGAGCAGAAGCAAGGTTAT | 80109..80128 | CCCATTCTCAACAGTGGAG | complement(80272..80291) |
| <i>nad2</i> exon2-3 | AAAGGAAGCTGAGTGTATTTGA | 332947..332966 | AATATTTGATCTTAGGTGCATTTT | complement(79761..79785) |
| <i>nad2</i> intron2 exon2 | CCCAGTCCGATAGTTACAA | 79641..79660 | AATATTTGATCTTAGGTGCATTTT | complement(79761..79785) |
| <i>nad2</i> exon3-4 | GCGCAATAGAAAGGAATGCT | 330240..330259 | CTATGGGTCTACTGGAGTACCC | complement(333071..333093) |
| <i>nad2</i> intron3 exon4 | GCGCAATAGAAAGGAATGCT | 330240..330259 | GGCGAATTTCAAACCTTGTGG | complement(330377..330396) |
| <i>nad2</i> exon4-5 | CAAAGGAGAGGGGTATAGCAA | 327932..327952 | TATTTGTTCTTCGCCGCTTT | complement(329793..329812) |
| <i>nad2</i> intron4 exon4 | CTTATTCGTGGCAACCTTCC | 329705..329724 | TATTTGTTCTTCGCCGCTTT | complement(329793..329812) |
| <i>nad4</i> exon1-2 | ATTCTATGTTTTCCCGAAAGC | 162115..162136 | GAAAACTGATATGCTGCCTTG | complement(163664..163685) |
| <i>nad4</i> intron1 exon2 | CCGTATGATGCGGAAGTCTC | 163512..163531 | GAAAACTGATATGCTGCCTTG | complement(163664..163685) |
| <i>nad4</i> exon2-3 | AATACCATGTTTCCCGAAG | 163967..163986 | TGCTACCTCCAATCCCTGT | complement(167228..167247) |
| <i>nad4</i> intron2 exon3 | GCGGAAGCAGCCAGAAAATA | 167110..167129 | TGCTACCTCCAATCCCTGT | complement(167228..167247) |
| <i>nad4</i> exon3-4 | TTCTCCATAAATCTCCGATT | 167577..167598 | TGAAATTTGCCATGTTGCAC | complement(169651..169670) |
| <i>nad4</i> intron3 exon4 | TCTAGCTTGGTTCGGAGAGC | 169498..169517 | TGAAATTTGCCATGTTGCAC | complement(169651..169670) |
| <i>nad5</i> exon1-2 | TGGACCAAGCTACTTATGGATG | 141880..141901 | CCATGGATCTCATCGGAAAT | complement(142793..142812) |
| <i>nad5</i> intron1 exon2 | TGGACCAAGCTACTTATGGATG | 141880..141901 | TTCGCAATAGGTCGGACT | complement(141962..141980) |
| <i>nad5</i> exon2-4 | AACTCGGATTCGGCAAGAA | 22023..22041 | CTGGCTCTCGGGAGTCTCTT | complement(140743..140762) |
| <i>nad5</i> intron2 exon2 | GTACGATCGTGTCCGGTGA | 140656..140674 | CTGGCTCTCGGGAGTCTCTT | complement(140743..140762) |
| <i>nad5</i> exon2-4 | AACTCGGATTCGGCAAGAA | 22023..22041 | CTGGCTCTCGGGAGTCTCTT | complement(140743..140762) |
| <i>nad5</i> intron3 exon4 | AACTCGGATTCGGCAAGAA | 22023..22041 | GCCGTGTAATAGGCGACCA | complement(22150..22168) |
| <i>nad5</i> exon4-5 | AACATTGCAAAGGCATAATGA | 20695..20715 | GTTCTCGGTTTCGGATATG | complement(21808..21827) |
| <i>nad5</i> intron4 exon5 | AACATTGCAAAGGCATAATGA | 20695..20715 | CCTGTAAACCCCATGATGT | complement(20827..20846) |
| <i>nad7</i> exon1-2 | ACCTCAACATCTGCTGCTC | 132118..132137 | AAGGTAAAGCTTGAAGATAAGTTTGT | complement(133203..133229) |
| <i>nad7</i> intron1 exon2 | ACGGTTTTTAGGGGGATCTG | 133128..133147 | AAGGTAAAGCTTGAAGATAAGTTTGT | complement(133203..133229) |
| <i>nad7</i> exon2-3 | GAGGGACTGAGAAATTAAGAGTACA | 133179..133205 | TGGTACCTCGCAATTCAAAA | complement(134362..134381) |
| <i>nad7</i> intron2 exon3 | AGTGGGAGAGCCGTGTTATG | 134182..134201 | TGGTACCTCGCAATTCAAAA | complement(134362..134381) |
| <i>nad7</i> exon3-4 | ACTGTCACTGCACAGCAAGC | 134718..134737 | CATTGCACAATGATCCGAAG | complement(135963..135982) |
| <i>nad7</i> intron3 exon4 | TAAAGTGAAGTGGTGGCCT | 135797..135816 | CATTGCACAATGATCCGAAG | complement(135963..135982) |
| <i>nad7</i> exon4-5 | GATCAAAGCCGATGATCGTAA | 136007..136027 | AGGTGCTTCAACTGCGGTAT | complement(137951..137970) |
| <i>nad7</i> intron4 exon5 | CGGCCAAATGACTACAGGAT | 137856..137875 | AGGTGCTTCAACTGCGGTAT | complement(137951..137970) |

RESULTATS

Supplemental Table 2

| spot | ID | Accession | calc MW | Score | Nb of peptides | |
|------|---|-------------------------------|---------|-------|----------------|-------------|
| 1 | NDUFS1 : 75kDa subunit | At5g37510 | 81,13 | 652 | 18 | Complex I |
| 2 | NDUFV1 : 51kDa subunit | At5g08530 | 53,415 | 125 | 5 | |
| | ND5 : subunit 5 | AtMg00060 | 74,267 | 76 | 2 | |
| 3 | ND7 : subunit 7 | AtMg00510 | 44,933 | 266 | 9 | |
| 4 | NDUFA9 : 39kDa subunit | At2g20360 | 43,908 | 391 | 12 | |
| | ND2 : subunit 2 | AtMg00285 | 55,413 | 38 | 2 | |
| 5 | CAL1 : carbonic anhydrase like 1 | At5g63510 | 27,552 | 105 | 4 | |
| | NDUFS8-1 : 23kDa subunit | At1g79010 | 25,487 | 94 | 3 | |
| | ND1 : subunit 1 | AtMg00516/AtMg01120/AtMg01275 | 36,115 | 61 | 2 | |
| | MADS-box protein (AGL42) MADS-box protein | At5g62165 | | 38 | 1 | |
| 6 | ND9 : subunit 9 | AtMg00070 | 22,865 | 293 | 12 | |
| 7 | NDU16 : 16kDa subunit | At2g27730 | 11,94 | 130 | 3 | |
| | NDUFB10-2 : PDSW subunit | At3g18410 | 12,13 | 108 | 4 | |
| 8 | NDUFB10-1 : PDSW subunit | At1g49140 | 12,522 | 102 | 4 | |
| | NDUFA6 : B14 subunit | At3g12260 | 15,073 | 137 | 5 | |
| 9 | NDUFA6 : B14 subunit | At3g12260 | 15,073 | 74 | 4 | |
| | NDUFB7 : B18 subunit | At2g02050 | 11,732 | 69 | 2 | |
| 10 | ATP2 | At5g08670.1 | 59634 | 1027 | 24 | complex V |
| | ATP1 | AtMg01190.2 | 55011 | 541 | 16 | |
| 11 | ATP3 | At2g33040.1 | 35426 | 364 | 10 | |
| 12 | ATP FAD | At2g21870.1 | 27579 | 213 | 6 | |
| 13 | ATP8 | AtMg00480 | 18199 | 39 | 2 | complex III |
| 14 | MPP b | At3g02090.1 | 59123 | 586 | 15 | |
| 15 | MPP a | At1g51980.1 | 54368 | 659 | 16 | |
| 16 | Cyt c1-1 | At3g27240.1 | 33629 | 238 | 5 | |
| | Cyt c1-2 | At5g40810.1 | 33669 | 193 | 5 | |
| | COB | AtMg00220.1 | 44118 | 182 | 4 | |
| 17 | URC1 | At5g13430.1 | 29589 | 61 | 3 | complex III |
| 18 | QCR7 | At4g32470.1 | 14518 | 323 | 9 | |
| 19 | ATP2 | At5g08670.1 | 59634 | 1087 | 23 | complex V |
| | ATP1 | AtMg01190.2 | 55011 | 780 | 21 | |
| 20 | ATP3 | At2g33040.1 | 35426 | 417 | 10 | |
| 21 | ATP FAD | At2g21870.1 | 27579 | 248 | 8 | |
| 22 | ATP8 | AtMg00480 | 18199 | 42 | 2 | complex III |
| 23 | MPP b | At3g02090.1 | 59123 | 923 | 24 | |
| 24 | MPP a | At1g51980.1 | 54368 | 651 | 16 | |
| 25 | COB | AtMg00220.1 | 33629 | 194 | 5 | |
| | Cyt c1-1 | At3g27240.1 | 33669 | 193 | 5 | |
| | Cyt c1-2 | At5g40810.1 | 44118 | 169 | 4 | |
| 26 | UCR1 | At5g13430.1 | 29589 | 343 | 10 | |
| 27 | QCR7 | At4g32470.1 | 14518 | 307 | 13 | |

II.2.2 TANG2

TANG2 est une PPR de type « Pure » qui a été identifiée par Yunhai Li dans l'équipe de Mike Bevan au cours d'un cribble de mutants sensibles au glucose. Le mutant présente un retard de développement, et des graines plus foncées. Ce phénotype très similaire au mutant *otp43* ainsi qu'à d'autres mutants des sous-unités du complexe I étudiés par Etienne Meyer comme *ndufs4* (Meyer, Tomaz et al. 2009) suggère un défaut éventuel de l'activité du complexe I. En collaboration avec l'équipe de Mike Bevan, j'ai donc eu l'opportunité d'étudier plus en détails le mutant *tang2* et ainsi d'identifier le défaut moléculaire associé. Cette étude décrite dans le papier suivant montre que cette protéine PPR est impliquée dans l'épissage en *trans* de l'intron 3 de *nad5*, ce qui fait de cette protéine la deuxième protéine PPR décrite comme ayant un rôle dans l'épissage d'introns mitochondriaux. Mon rôle dans la caractérisation de ce mutant a été d'identifier le défaut moléculaire en utilisant les outils de transcriptomique mis en place dans le laboratoire de Ian Small. Ainsi, des expériences de RT-PCR, de PCR quantitative sur l'expression des gènes mitochondriaux, mais aussi sur les événements d'épissage, et des expériences de *northern blot* sur certains transcrits mitochondriaux ont permis de mettre en évidence en grande partie la fonction de TANG2.

Article 4 : Li Y., Smith C., Colas des Francs-Small C., de Longevialle A.F., Corke F., Zheng L., Tanz S., Small I.D. and Bevan M.W. (2010) TANG2, a Glucose-Responsive Gene, Encodes a Pentatricopeptide Repeat Protein Required for Splicing of Mitochondrial nad5 Transcripts, in submission.

***TANG2*, a Glucose-Responsive Gene, Encodes a
Pentatricopeptide Repeat
Protein Required for Splicing of Mitochondrial *nad5*
Transcripts**

Yunhai Li ^{a,b,1}, Caroline Smith ^a, Catherine Colas des Francs-Small ^c, Andéol Falcon de Longevialle ^c, Fiona Corke ^a, Leiying Zheng ^a, Sandra K. Tanz ^c, Ian Small ^c, and Michael W. Bevan ^{a,1}

^a Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK.

^b State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

^c Australian Research Council of Excellence in Plant Energy Biology, University of Western Australia, Crawley 6009 WA, Australia.

¹ Corresponding authors

E-mail: yhli@genetics.ac.cn; Telephone: 86 (10) 64807856; Fax: 86 (10) 64807856

E-mail: michael.bevan@bbsrc.ac.uk Telephone: 44 (1603) 450520; Fax: 44 (1603) 450025.

Running title: The *TANG2* PPR protein splices mitochondrial *nad5* transcripts

Key words: *tang2*, sugar responses, PPR, Complex I, splicing of mitochondrial genes

ABSTRACT

Pentatricopeptide repeat (PPR) proteins are RNA binding proteins required for RNA editing, splicing, cleavage and translation of mitochondrial and chloroplast transcripts. Mutations in PPR genes lead to diverse organelle defects, cytoplasmic male sterility and a wide variety and severity of other phenotypes. We have identified an Arabidopsis nuclear PPR gene called *TANG2* that is specifically required for splicing of intron 3 of the mitochondrial *nad5* gene. This defect leads to substantially reduced, but not absent, Complex I activity. *tang2* mutant plants have substantially elevated glucose, sucrose and anthocyanin levels, increased expression of sugar-responsive genes, and greatly enhanced sensitivity of seedling growth and development to high glucose levels. Expression of the *TANG2* gene is rapidly induced by glucose and is expressed in tissues with a high energy demand, suggesting *nad5* splicing by *TANG2* responds to available energy sources such as sugars.

INTRODUCTION

Metabolism of sugars in the glycolytic and TCA pathways produces reducing equivalents that generate ATP in the oxidative phosphorylation pathway in mitochondria. The oxidative phosphorylation system is located on the inner mitochondrial membrane and is composed of five enzyme complexes. Complex I (NADH:ubiquinone oxidoreductase) is the main entry point for electrons and is composed of more than 40 subunits in *Arabidopsis* (Heazlewood, Howell et al. 2003), of which nine are encoded by the mitochondrial genome. The centrally important role of Complex I in cellular energetics is reflected in the multiple effects of mutations and chemicals that reduce Complex I function. In plants the best-known effect of altered Complex I mitochondrial loci is cytoplasmic male sterility (CMS) (Chase 2007). This maternally-inherited trait is characterised by mitochondrial dysfunctions that result in non-viable pollen. The *Nicotiana sylvestris* CMSII mutant, which lacks functional Complex I, respire using alternative NADH dehydrogenases and alternative oxidase, and manages a high photosynthetic capacity (Dutilleul, Driscoll et al. 2003). However, Complex I activity was needed for optimal photosynthetic performance when both respiratory and photorespiratory substrates needed to be oxidised simultaneously. Inhibition of Complex I activity by rotenone (Garmier, Carroll et al. 2008) caused a transient decrease in respiration, a subsequent acclimation by the induction of alternative oxidase, and reduced growth. This study also identified a physical and functional dissociation of glycolytic enzymes from mitochondria in response to reduced Complex I activity and associated changes in carbon metabolism reflecting the activation of compensatory pathways.

Rotenone treatment also led to increased expression of nuclear genes encoding mitochondrial proteins. This was thought to be due to retrograde signalling from the mitochondria to the nucleus mediated by cellular redox status (Garmier, Carroll et al. 2008). Another link between reduced Complex I function and nuclear gene expression was shown by the reduced cold-induced gene expression in the *fro1* mutant, which is defective in the 18-kD Fe-S subunit of complex I (Lee, Lee et al. 2002). *fro1* mutant lines exhibited reduced cold acclimation and accumulate reactive oxygen species (ROS) that are involved in cold signalling.

Five of the nine mitochondrial *nad* genes encoding Complex I subunits are composed of exons that need to be spliced to create functional complex I subunits. The *nad1*, *nad2* and *nad5* transcripts are *trans*-spliced *via* type II introns (Malek and Knoop 1998) in a process that probably requires RNA splicing factors. Several factors involved in splicing analogous chloroplast introns have been identified, including a pentatricopeptide repeat (PPR) protein that is required for *trans*-splicing of *rps12* transcripts (Schmitz-Linneweber, Williams-Carrier et al. 2006). PPR proteins are also necessary for RNA editing, processing and stabilisation

(Schmitz-Linneweber and Small 2008; Zhou, Cheng et al. 2008; Okuda, Chateigner-Boutin et al. 2009; Zehrmann, Verbitskiy et al. 2009). PPR proteins are encoded by a relatively large family of 450 genes in *Arabidopsis* (Lurin, Andres et al. 2004) and other land plants (O'Toole, Hattori et al. 2008) that are predicted to be predominantly located to mitochondria or plastids. Biochemical studies have shown the PPR repeat regions, comprising a canonical 35 amino acid motif repeated in tandem up to 30 times, bind different types of RNA. An understanding of PPR function is emerging in which PPR repeats function as RNA-binding specificity determinants and other domains and interacting proteins provide functions for splicing and editing (Schmitz-Linneweber and Small 2008).

A reverse genetic screen of members of the *Arabidopsis* PPR gene family using RNAi (Hilson, Allemeersch et al. 2004) identified the PPR gene *OTP43* as a mitochondrial *trans*-splicing factor that is specifically required for splicing of intron 1 of the *nad1* transcript in *Arabidopsis* (de Longevialle, Meyer et al. 2007). Complex 1 activity was absent in the mutant, which exhibited severely impaired germination and reduced growth and development. The *otp43* mutant also led to increased expression of alternative NAD(P)H dehydrogenases and alternative oxidases to bypass the Complex 1 defect, similar to that seen in rotenone- treated cells (Garmier, Carroll et al. 2008) and the *Nicotiana* CMSII mutant (Dutilleul, Driscoll et al. 2003).

Key questions in the biology of PPR proteins concern the specific role of individual family members in organelle RNA metabolism and whether PPR proteins may have a role, as nuclear- encoded genes, in coordinating the functions of organelles in a cellular context. Here we characterise the *tang2* mutant, exhibiting extreme sugar-sensitive growth, as being defective in a sugar-inducible PPR gene. Complex I activity is reduced in *tang2* due to decreased splicing of mitochondrial *nad5* transcripts, demonstrating a specific role of TANG2 in mitochondrial RNA processing.

RESULTS

The Sugar-Inducible Gene *TANG2* Encodes a Pentatricopeptide Repeat Protein

Microarray data has revealed that the expression of thousands of genes was rapidly altered in response to glucose (Li, Lee et al. 2006). To further understand the biological functions of these sugar-responsive genes, we are characterizing their functions, and here we describe the function of the glucose-inducible gene At1g19290. We first confirmed the glucose induction of this gene using quantitative RT-PCR (Fig.1A) and called the gene *TANG2* (TANG means "sugar" in Chinese). This gene has a single exon and is predicted to

encode a pentatricopeptide repeat (PPR) protein comprising 17 conserved PPR motifs (Fig. 1B). In different plant species, genome sequencing revealed a large number of PPR domain proteins (O'Toole, Hattori et al. 2008). *TANG2* homologs sharing highly conserved domain structure were identified in *Vitis vinifera* and rice (Supplementary Fig.1).

Expression pattern of *TANG2*

The spatial expression patterns of *TANG2* were revealed by histochemical assays of β -glucuronidase (GUS) activity of transgenic plants grown on Murashige and Skoog (MS) medium or in the greenhouse expressing *TANG2* promoter:GUS fusions (*pTANG2 GUS*). Histochemical staining showed the relatively high levels of GUS activity in the shoot and root apices, lateral root primordia, young flower buds, and stamens (Fig.1C-G). GUS activity was essentially undetectable in mature tissues of stems, leaves and flowers.

The *tang2* Mutant has Altered Growth and Development

To determine the biological functions of *TANG2* in plant growth and development, we obtained a T-DNA insertion allele of *TANG2*. The position of the T-DNA insert in a canonical PPR repeat was confirmed by PCR and sequencing. The line Salk_003139 (hereafter called *tang2*) contained the T-DNA insert within the coding region, 1436 bp upstream of the stop codon (Fig.2A). *tang2* was transformed with a genomic DNA fragment carrying the wild-type At1g19290 gene, and two lines expressing *TANG2* were identified. One of these lines, *TANG2COM2*, was used in further studies; it fully complemented the morphological and growth phenotypes of *tang2* (Fig. 2B-E). In addition, transgenic plants expressing an antisense transcript to *TANG2* (*35S:antiTANG2*) showed similar phenotypes to *tang2* mutants (data not shown).

On Murashige and Skoog (MS) agar medium, *tang2* seeds germinated about 1 day later than wild-type seeds. After germination, *tang2* plants also grow more slowly than the wild type (Fig.2B), and plate-grown *tang2* mutants also exhibited smaller stature and shorter roots compared with wild type (Fig.2B). Soil-grown *tang2* plants have smaller stature, with narrow, dark green leaves, and delayed growth compared with the wild type (Fig.2C and D). Seeds of the *tang2* mutant are larger than those of the wild type (Fig.2E) but seed set was reduced. *tang2* flowers also have shorter stamen filaments than the wild type (data not shown) such that *tang2* pollen is not able to reach stigmatic papillae directly; this may explain the decreased fertility. These results indicated that *TANG2* is required for normal growth and development.

***tang2* Seedlings Are Sensitive to Glucose.**

We investigated whether the *tang2* mutant has altered sugar responses since the expression of *TANG2* is induced by glucose. Seedling development on different concentrations of glucose is a widely used assay for assessing glucose sensitivity (Sheen, Zhou et al. 1999). Seedlings of the wild-type grown on media containing 3% (w/v) glucose for 10 days developed two pairs of true leaves, while *tang2* seedlings were arrested at an early growth stage and did not green or develop true leaves and the cotyledons accumulated anthocyanins. This demonstrated that that *tang2* is hypersensitive to glucose (Fig.3A). This hypersensitivity of *tang2* in response to glucose is not a result of an osmotic effect, because *tang2* seedlings were fully established on medium supplemented with 3% mannitol (data not shown). Quantitative analysis of seedling establishment on medium containing 3% glucose showed that *tang2* was extremely sensitive to 3% glucose, with less than 5% of the seedlings (n=100) greening or developing true leaves (Fig.3B). In these conditions all Col-0 seedlings rapidly developed true leaves.

We assessed the expression levels of several sugar-responsive genes known to have altered expression patterns in sugar-response mutants (Moore, Zhou et al. 2003; Li, Smith et al. 2007). For example, genes involved in starch and flavonoid biosynthesis are sugar-inducible (Rook, Corke et al. 2001), while in glucose-hypersensitive lines the expression of these genes is induced at lower glucose concentrations than in wild-type (Baier, Hemmann et al. 2004; Li, Smith et al. 2007). Expression of the sugar-inducible *ApL3* (Rook, Corke et al. 2001) and *A β -Amylase* (Mita, Suzuki-Fujii et al. 1995) genes was significantly increased in *tang2* in response to 1% (w/v) glucose compared with wild type (Fig. 3C). Furthermore the expression level of both the *TT4* and *DRF* genes involved in flavonoid biosynthesis (Weisshaar and Jenkins 1998) was also higher in *tang2* than in wild type (Fig. 3C).

***tang2* Plants Have Altered Starch, Anthocyanin and Sugar Levels**

The *ApL3* gene encodes one of four large subunits of ADPG-pyrophosphorylase (AGPase), which catalyses the first committed step in starch synthesis, and *TT4* and *DRF* are required for anthocyanin biosynthesis. Consistent with the elevated expression of these genes, starch levels were slightly but significantly increased in the *tang2* mutant compared with the wild-type (Fig. 4A), and anthocyanins also accumulated to higher levels in *tang2* compared with the wild type (Fig.4B). These results indicated that the *tang2* mutation enhances several glucose-mediated metabolic responses.

Increased cellular levels of sugars could contribute to enhanced responses of *tang2* plants to exogenous glucose. Therefore glucose, fructose and sucrose levels were measured in 7-

day-old seedlings of the *tang2* mutant and the wild type. The *tang2* seedlings have significantly increased levels of glucose, fructose, and sucrose compared to wild type seedlings of the same age (Fig. 4C), suggesting that altered intracellular sugar levels may account for the enhanced glucose responses seen in the *tang2* mutant.

TANG2 Is Involved in the biogenesis of Complex I.

The phenotypes of the *tang2* mutant, such as slow growth and darker, narrower leaves, are reminiscent of *Arabidopsis* mitochondrial Complex I mutants (de Longevialle, Meyer et al. 2007). We first demonstrated that a GFP fusion of TANG2 is co-localised with a mitochondrial-specific protein fused to the red fluorescent protein (Figs. 5A-C). Having determined that TANG2 is mitochondrial, we used BN-PAGE (Blue-Native polyacrylamide gel electrophoresis) and NADH oxidase staining to investigate the activity of Complex I in *tang2* mitochondria. As shown in Fig.5D, NADH oxidase activity at the position expected for Complex I was greatly reduced in the *tang2* mutant, compared with mitochondria from Col-0 and complemented *tang2COM2* plants. A similar BN-PAGE gel was blotted and probed with a NAD9 antibody (Fig.5E), proving that the high molecular weight band of NADH oxidase activity is indeed Complex I, and that the *tang2* mitochondria not only have very little complex I activity, but also very little assembled complex I protein. Two-dimensional blue-native/SDS-PAGE of mitochondrial proteins from the *tang2* mutant further confirmed the specific reduction of Complex I (arrowed in Figs. 5F and 5G), whereas none of the other complexes appear to be affected, indicating that TANG2 is specifically required for the biogenesis of mitochondrial Complex I.

The Splicing of Mitochondrial Transcripts is altered in the *tang2* Mutant

As several PPR proteins have been shown to be involved in RNA processing in plant organelles (Schmitz-Linneweber and Small 2008), we performed RT-PCR on *tang2* RNA to analyse the mitochondrial transcripts encoding subunits of Complex I. Fig.6A shows that the mature, fully-processed *nad5* transcript was hardly detectable in *tang2* RNA, whereas other *nad* transcripts accumulated normally, suggesting that transcription, processing or stability of *nad5* RNA is defective in *tang2*. As several other PPR proteins are required for transcript-specific splicing (Meierhoff, Felder et al. 2003; Schmitz-Linneweber, Williams-Carrier et al. 2006; de Longevialle, Meyer et al. 2007; Hattori, Miyake et al. 2007; Beick, Schmitz-Linneweber et al. 2008; de Longevialle, Hendrickson et al. 2008; Williams-Carrier, Kroeger et al. 2008), we performed an RT-PCR splicing assay designed to quantify the extent of splicing of each mitochondrial intron in *Arabidopsis*. The results showed clear defects in splicing of

nad2 intron 1 and, much more dramatically, in splicing of *nad5* intron 3 (Fig.6,B and C). Nevertheless, low levels of correctly spliced *nad5* transcripts were recovered in *tang2* and verified by sequencing (data not shown). The splicing of other transcripts is not significantly altered in *tang2* mutants. The defect in splicing of *nad5* intron 3 is sufficient to explain the dramatic reduction of fully-processed *nad5* transcript, and in turn, that of Complex I activity in *tang2* mutants.

DISCUSSION

Plant organellar genomes contain introns that are excised with the aid of nuclear-encoded and/or organelle-encoded factors. Several PPR proteins have been implicated directly or indirectly in organellar RNA splicing. HCF152 increases splicing of the *petB-petD* precursor RNA coding for subunits of the cytochrome *b₆f* complex (Meierhoff, Felder et al. 2003). The maize chloroplast proteins PPR4 and PPR5 are involved in splicing introns in *rps12* and *trnG(UCC)* transcripts, respectively, and have been shown to associate with these introns *in vivo* (Schmitz-Linneweber, Williams-Carrier et al. 2006; Beick, Schmitz-Linneweber et al. 2008; Williams-Carrier, Kroeger et al. 2008). OTP43 is specifically required for *trans*-splicing of the mitochondrial *nad1* intron 1 in *Arabidopsis* (de Longevialle, Meyer et al. 2007). *Arabidopsis* *OTP51* is needed for the splicing of chloroplast *ycf3* intron 2, and also influences the splicing of several other group-IIa introns in plastids (de Longevialle, Hendrickson et al. 2008). *TANG2* encodes a pentatricopeptide repeat protein similar to OTP43 and OTP51, and loss of *TANG2* gene function causes a 100x reduction in splicing of *nad5* intron 3 and a much smaller reduction in *nad2* intron 1 splicing (Fig.6,B and C). As *nad2* intron1 splicing is also affected to the same degree in *opt43* (de Longevialle, Meyer et al. 2007), this may be a secondary effect. The splicing of other mitochondrial transcripts was not significantly altered in *tang2* mutants, indicating a high degree of specificity.

PPR motifs bind RNA directly (Williams-Carrier, Kroeger et al. 2008; Okuda, Chateigner-Boutin et al. 2009) and it is thought that diversity in PPR motif organisation provides this binding specificity reviewed in (Delannoy, Stanley et al. 2007), although direct biochemical evidence is sparse. The characterisation of *TANG2* provides further genetic evidence for a high degree of specificity of mitochondrial intron recognition by PPR proteins.

Mitochondrial Complex I is the main entry point for the electron transport chain in plant and animal mitochondria and is of central importance in cellular energy homeostasis (Sterky and Larsson 2008). A T-DNA insertion in the *TANG2* gene causes a specific major reduction in the levels of mitochondrial NADH:ubiquinone oxidoreductase (Complex I) as determined by western blotting and two-dimensional blue-native/SDS-PAGE of mitochondrial proteins

isolated from the *tang2* mutant (Figs. 5F and G). None of the other complexes appear to be significantly affected, indicating that *TANG2* is specifically involved in the biogenesis of mitochondrial Complex I. The link between electron transport and altered sugar levels and responses we observed in *tang2* have also been noticed previously in another mutant that may affect Complex I activity. The Arabidopsis *css1* mutation in a type II intron maturase gene that reduces *nad4* splicing also exhibits sugar hypersensitive seedling establishment phenotypes, increased amino acid and starch levels, and reduced cellulose synthesis (Nakagawa and Sakurai 2006). However the effects of this mutation on Complex I levels was not determined.

The *tang2* T-DNA insertion mutant enhances a wide range of responses to physiologically relevant sugar levels, including increased starch and anthocyanin levels, glucose-sensitive seedling growth, and enhanced expression of sugar-inducible genes compared with the wild type (Fig. 3C). These phenotypes are similar to those observed in several mutations that lead to hypersensitive responses to glucose (Baier, Hemmann et al. 2004; Li, Lee et al. 2006). *tang2* mutant plants also exhibited a range of other defects such as growth retardation, small stature, narrow and dark green leaves, short roots, reduced fertility and large seeds, showing that *TANG2* is required for normal plant growth and development. Although other mutations that reduce Complex I levels such as *fro1* and *otp43* (Lee, Lee et al. 2002; de Longevialle, Meyer et al. 2007) exhibit some similar phenotypes to *tang2* such as retarded growth and small stature, these mutants exhibit very distinct phenotypes. No difference in root growth between *fro1* and the wild type was observed, in contrast to the shorter primary roots of *tang2*. The *otp43* mutant has smaller and darker seeds than wild type which have severely reduced germination, whereas *tang2* has larger seeds compared to wild type with only slightly reduced germination. Mutations in *fro1* cause reduced cold induction of stress-responsive genes such as *RD29A*, *KIN1*, *COR15A*, and *COR47*, in contrast to the elevated starch and anthocyanin biosynthesis gene expression seen in *tang2*.

We showed that *tang2* mutants accumulate substantially higher levels of glucose, fructose and sucrose (Fig. 4C). This could be explained by reduced Complex 1 activity, as Complex I inhibition in cell cultures by rotenone affects carbon metabolism by slowing the TCA cycle and reducing the association of glycolytic enzymes with mitochondria (Garmier, Carroll et al. 2008). The increased glucose, fructose and sucrose levels seen in *tang2* plants is consistent with reduced TCA cycle activity. Reduced activity of the electron transport chain also leads to compensatory increases in the activity of alternative oxidase pathways (Dutilleul, Driscoll et al. 2003; Dutilleul, Lelarge et al. 2005; Vidal, Ribas-Carbo et al. 2007). These contribute to altered balances between ATP/ADP, NAD/NADH and ROS, which are known to regulate metabolism and multiple cellular processes, and may therefore also account for altered

sugar- responsive gene expression and growth. However, a recent analysis of an Arabidopsis mutant specifically lacking Complex I activity (Meyer, Tomaz et al. 2009) did not reveal increased glucose or fructose levels, nor was the expression of sugar-responsive genes altered despite similar growth conditions (1% glucose in this work and 1% sucrose in (Meyer, Tomaz et al. 2009)), and there was no reduction in germination in *tang2*. This may be due to differences in Complex I activity, although it is unlikely that a partial reduction in Complex I activity leads to strong sugar-response phenotypes that are not seen when Complex I activity is abolished. This suggests that *tang2* may specifically alter sugar levels and sugar-responsive gene expression and growth.

The rapid induction of *TANG2* expression in response to sugars (Fig. 1A) and its preponderant expression in tissues with high energy demand such as meristem and floral tissues, and its virtual absence from mature tissues Fig. 1,B-G), suggest that *TANG2* may mediate *nad5* splicing levels and Complex I activity in response to sugar levels. As sugars are the primary mobile form of photosynthate and the main source of energy in heterotrophic tissues, it is possible that *TANG2* expression could be a link between energy substrates and energy generation in the mitochondria.

METHODS

Plant Materials and Growth Conditions

All lines were in the Col-0 background. Seeds were surface-sterilized with 100% isopropanol and 20% (v/v) household bleach, washed at least five times with sterile water, stratified at 4°C for 4d in the dark, dispersed on Murashige and Skoog (MS) medium (Duchefa Biochemie BV, Haarlem, The Netherlands) supplemented with 0.9% agar and 1% and 3% glucose or mannitol, and then grown at 22°C.

Identification of *tang2* Mutant and Complementation Tests

The *tang2* mutant (SALK_003139) was obtained from the Nottingham Arabidopsis Stock Centre. SALK_003139 T-DNA insertion in the *TANG2* gene was confirmed by PCR and sequencing, by using primers SALK_003139LP (CTGCAATAGCAAACAAGCCTC), SALK_003139RP (TGTACGGTGTGTTGATGGATG) and LBa1.

A genomic DNA fragment containing the entire *TANG2* coding region, the 1.5kb upstream sequence, and the 1.0kb downstream sequence was inserted into the binary vector pCAMBIA1300 to generate the transformation plasmid COM for complementation. The

RESULTATS

plasmid COM was introduced into the *tang2* mutant using *Agrobacterium tumefaciens* GV3101 and transformants were selected on hygromycin-containing medium.

GUS Staining

The *TANG2* promoter-GUS construct (*pTANG2:GUS*) was made using a PCR-based Gateway system. The promoter specific primers for the *TANG2* gene were TOPOTANG2PROM-F (CACC AGC TAT ACT TTG TAG CCA CGG) and TOPOTANG2-R (GTC AAA GAC TGT AGG AGA GAA G). PCR products were subcloned into the pENTR/D-TOPO vector (Invitrogen) using TOPO enzyme and sequenced. The *TANG2* promoter was then subcloned into Gateway Binary Vector (pGWB3) containing the GUS reporter gene. Seedlings were grown in the soil or on MS medium supplemented with 1% glucose.

TANG2 Localisation

The first 300 bp of the coding sequence of *TANG2* were amplified including adapters containing the *attB* Gateway cloning sequences with the primers 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGCTCCGACGG TCTCCG and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCCACCTCCGGATCCGAACCTTTTGTGCT TTGAAGCCAA using the Expand High Fidelity PCR system (Roche Diagnostics, Germany). The PCR product was cloned into Gateway® vector pDONR207 (Invitrogen, USA) and sequenced to check PCR accuracy. The entry vector and a Gateway® cloning cassette were recombined to clone the targeting sequence of *TANG2* in frame with the coding region of the green fluorescent protein (GFP). For co-localization studies, the mitochondrial targeting sequence of yeast *ScCox4* fused to mCherry in pBIN20 (Nelson BK, Cai X, Nebenfuehr A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J 51: 1126-1136) was used as a mitochondrial control. The constructs were biolistically transformed into Arabidopsis cultured cells as previously described. In brief, the GFP and mCherry plasmids (5 µg each) were co-precipitated onto 1µm gold particles and transformed using the biolistic PDS-1000/He system (Bio-Rad, USA). Particles were bombarded onto 2 ml of Arabidopsis cell suspension resting on filter paper on osmoticum plates. After bombardment, the Arabidopsis cell suspension was placed in the dark at 22°C. Fluorescence images were obtained 24 hrs after transformation using an Olympus BX61 epifluorescence microscope with HQ GFP (U-MGFPHQ) and RFP (U-MRFPHQ) filters and the Cell® imaging software.

RNA Isolation, Reverse Transcription (RT)-PCR and Quantitative Real-Time RT-PCR Analysis

Total RNA was extracted from Arabidopsis seedlings grown on MS medium supplemented with 1% glucose using an RNeasy Plant Mini kit (Qiagen). Reverse transcription (RT)-PCR analysis was performed as described (Li, Lee et al. 2006). Quantitative real-time RT-PCR analysis was performed with Biorad system using the SYBR Green JumpStart Taq readyMix (Sigma). *ACTIN* mRNA was used as an internal control, and relative amounts of mRNA were calculated using the comparative threshold cycle method. The primers used for Quantitative real-time RT-PCR are as follows: AT5G09810-*ACTIN* (AtactF -5'-GAGAAGATGACTCAGATC-3' and AtactR -5'-ATCCTTCCTGATATCGAC-3'); AT4G39210-*ApL3* (*ApL3F*-5'-CGTCTGAATCATGCAACC-3' and *ApL3R*-5'-GCATTCCTGATCTTTGTATCCTCG-3'); AT4G15210 - *-Amylase* (-AmyF-5'-CGGAGAAGGGGAAGTTTTTC-3' and -AmyR-5'-AATCTCATGCCCGTACTTCG-3'); AT5G13930-*TT4* (*TT4F*-5'-GGT TCA TGG CTA GTG ATG CG-3' and *TT4R*-5'-A ACG CTG TGC AAG ACG ACT G-3'); *AT5G42800-DRF* (*DRFF*-5'-GGT TCA TGG CTA GTG ATG CG-3' and *DRFR*-5'-AGA CGT TGT GAT GAA TGG ACC-3') and *At1g19290-TANG2* (*TANG2QRTE*-5'-ATA ACA CTC TGA TAT CTG GTG C-3' and *TANG2QRTR*-5'-TAA TGT GTG CTT ACC TGA GGC-3').

Carbohydrate and Anthocyanin Analysis

Carbohydrate and anthocyanin analysis of 12-15 day old seedlings grown on MS medium supplemented with 1% glucose were performed as described (Baier, Hemmann et al. 2004).

Glucose, fructose, and sucrose were measured sequentially in cleared supernatants of K₂CO₃-neutralized HClO₄ extracts of ground, frozen plant material (Galtier and Gouy 1995).

Mitochondrial Isolation, BN PAGE and Immunoblot Analysis

Mitochondria were isolated from 8-week-old rosette leaves of greenhouse grown plants and BN PAGE was performed as previously described (de Longevialle, Meyer et al. 2007). Proteins separated by BN PAGE were transferred onto PVDF membrane (GE Healthcare) in cathode buffer for 15 h at 40 mA, using a BioRad Mini transblot cell. The membrane was then destained with ethanol and restained with Coomassie blue to visualise the complexes, and subsequently destained before any further treatment. The western blot was performed using an anti NAD9 antibody at a 1/50,000 dilution, and a goat anti-rabbit conjugated to

horseradish peroxidase secondary antibody at a 1/10,000 dilution, and subsequently revealed using ECL reagents (GE Healthcare).

Isolation of Mitochondrial RNA and Transcript Analysis

Total leaf RNA was isolated from 6-week-old rosettes of greenhouse grown plants using the plant RNeasy extraction kit (Qiagen) according to the manufacturers instructions, and genomic DNA was removed using DNafree DNase (Ambion). Complete removal of both mitochondrial and nuclear DNA was checked by PCR on diluted RNA prior to reverse transcription. The reverse transcription step was performed on 3 µg of DNase-treated RNA using Superscript III (Invitrogen) and random hexamer primers as described previously (de Longevialle, Meyer et al. 2007). RT-PCR was performed with 1 µl of 1/10 dilution of the reverse transcriptase reaction (except for *nad5* for which 3.5 µl of 1/10 dilution were used and *nad3* and *nad6* for which 3 µl of 1/10 dilution were used). The primers used for detecting the spliced *nad* transcripts are given in Supplementary Table 1.

To quantify the splicing of mitochondrial mRNAs, primers were designed to amplify 100-200 bp regions spanning intron-exon junctions (unspliced forms) or spanning splice junctions (spliced forms) of each gene (see example given for *nad5* in Fig.5C). The complete list of these primers is given in Supplementary Table 2. Quantitative RT-PCR was performed using LightCycler 480 SYBR Green I Master Mix (Roche Diagnostics) and a Roche LightCycler 480 real-time PCR system with the following thermal cycling program: 95°C for 10 min, followed by 45 cycles of 95°C for 10 s, 60°C for 10 s, and 72°C for 10 s. The data were analysed using the LightCycler 480 software version 1.5 (Roche Diagnostics). Three independent experiments were performed, and each sample at a 1/200 dilution was run in triplicate. Nuclear and mitochondrial 18S rRNA genes were used as internal controls. The primers used for quantitative RT-PCR assays are given in Supplementary Table 2.

The same RNA preparations and data analysis methods were used to quantify the expression of *TANG2* in the complemented lines in Fig. 2, Two independent experiments were performed, and each sample at a 1/200 dilution was run in triplicate. The primers used were tang2 F; 5' TTGTTCAAGCTTGGCGATTT 3' and R: 5' TGGCTTCATTGACCTTCTCC 3'.

ACKNOWLEDGMENTS

This work was supported by BBSRC Grant BB/C515620, the John Innes Centre Core Strategic Grant, the National Basic Research Program of China (2009CB941503), Australian Research Council Centre of Excellence grant CE0561495, and DEST International Science Linkages grant CG120098. IS is supported as a WA State Premier's Fellow.

REFERENCES

- Baier, M., et al. (2004). Characterization of mutants in Arabidopsis showing increased sugar-specific gene expression, growth, and developmental responses. *Plant Physiol* 134:81-91.
- Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. (2008). The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol Cell Biol* 28:5337-5347.
- Chase, C.D. (2007). Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet* 23:81-90.
- de Longevialle, A.F., et al. (2008). The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. *Plant J* 56:157-168.
- de Longevialle, A.F., et al. (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana. *Plant Cell* 19:3256-3265.
- Delannoy, E., Stanley, W.A., Bond, C.S., and Small, I.D. (2007). Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem Soc Trans* 35:1643-1647.
- Dutilleul, C., Driscoll, S., Cornic, G., De Paepe, R., Foyer, C.H., and Noctor, G. (2003). Functional mitochondrial complex I is required by tobacco leaves for optimal photosynthetic performance in photorespiratory conditions and during transients. *Plant Physiol* 131:264-275.
- Dutilleul, C., Lelarge, C., Prioul, J.L., De Paepe, R., Foyer, C.H., and Noctor, G. (2005). Mitochondria-driven changes in leaf NAD status exert a crucial influence on the control of nitrate assimilation and the integration of carbon and nitrogen metabolism. *Plant Physiol* 139:64-78.
- Galtier, N., and Gouy, M. (1995). Inferring phylogenies from DNA sequences of unequal base compositions. *Proc Natl Acad Sci U S A* 92:11317-11321.
- Garmier, M., et al. (2008). Complex I dysfunction redirects cellular and mitochondrial metabolism in Arabidopsis. *Plant Physiol* 148:1324-1341.
- Hattori, M., Miyake, H., and 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.
- Heazlewood, J.L., Howell, K.A., and Millar, A.H. (2003). Mitochondrial complex I from Arabidopsis and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits. *Biochim Biophys Acta* 1604:159-169.
- Hilson, P., et al. (2004). Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. *Genome Res* 14:2176-2189.
- Lee, B.H., Lee, H., Xiong, L., and Zhu, J.K. (2002). A mitochondrial complex I defect impairs cold-regulated nuclear gene expression. *Plant Cell* 14:1235-1251.
- Li, Y., et al. (2006). Establishing glucose- and ABA-regulated transcription networks in Arabidopsis by microarray analysis and promoter classification using a Relevance Vector Machine. *Genome Res* 16:414-427.
- Li, Y., et al. (2007). Signaling from an altered cell wall to the nucleus mediates sugar-responsive growth and development in Arabidopsis thaliana. *Plant Cell* 19:2500-2515.

- Lurin, C., et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16:2089-2103.
- Malek, O., and Knoop, V. (1998). Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *Rna* 4:1599-1609.
- Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003). HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* 15:1480-1495.
- Meyer, E.H., et al. (2009). Remodeled respiration in *ndufs4* with low phosphorylation efficiency suppresses Arabidopsis germination and growth and alters control of metabolism at night. *Plant Physiol* 151:603-619.
- Mita, S., Suzuki-Fujii, K., and Nakamura, K. (1995). Sugar-inducible expression of a gene for beta-amylase in Arabidopsis thaliana. *Plant Physiol* 107:895-904.
- Moore, B., et al. (2003). Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* 300:332-336.
- Nakagawa, N., and 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.
- O'Toole, N., et al. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* 25:1120-1128.
- Okuda, K., et al. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* 21:146-156.
- Rook, F., Corke, F., Card, R., Munz, G., Smith, C., and Bevan, M.W. (2001). Impaired sucrose-induction mutants reveal the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling. *Plant J* 26:421-433.
- Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13:663-670.
- Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas, A., and Barkan, A. (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* 18:2650-2663.
- Sheen, J., Zhou, L., and Jang, J.C. (1999). Sugars as signaling molecules. *Curr Opin Plant Biol* 2:410-418.
- Sterky, F.H., and Larsson, N.G. (2008). Complex I: a complex gateway to the powerhouse. *Cell Metab* 7:278-279.
- Vidal, G., et al. (2007). Lack of respiratory chain complex I impairs alternative oxidase engagement and modulates redox signaling during elicitor-induced cell death in tobacco. *Plant Cell* 19:640-655.
- Weisshaar, B., and Jenkins, G.I. (1998). Phenylpropanoid biosynthesis and its regulation. *Curr Opin Plant Biol* 1:251-257.
- Williams-Carrier, R., Kroeger, T., and Barkan, A. (2008). Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *Rna* 14:1930-1941.
- Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA

RESULTATS

editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21:558-567.

Zhou, W., et al. (2008). The *Arabidopsis* gene *YS1* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J.*

Supplementary table 1: Primers Used for Amplifying the Whole nad Transcripts

| | Forward primer | Reverse primer |
|--------------|-------------------------|-----------------------|
| nad1 exon1-5 | AGCCCGGGATCTTCTTGA | TTGCCATATCTTCGCTAGGTG |
| nad2 exon1-5 | CAAAGGAGAGGGGTATAGCAA | GGATCCTCCCACACATGTTT |
| nad3 | GCACCCCTTTTCCATTCCATA | GATGTCAGAAATTTGCACCAA |
| nad4 exon1-4 | ATTCTATGTTTTTCCGAAAGC | TGAAATTTGCCATGTTGCAC |
| nad5 exon1-5 | AACATTGCAAAGGCATAATGA | ACCATGGATCTCATCGGAAAT |
| nad6 | CGTCCTCCTCATTATAGTCCTCC | CTCTGGTTTGATGGTTGCAC |
| nad7 exon1-5 | ACCTCAACATCCTGCTGCTC | AGGTGCTTCAACTGCGGTAT |
| nad9 | TTGGGTCATCTCAATGGGTT | CGATCGATATTTGCGGAGTT |

Supplemental table 2: Primers Used for the Splicing qRT-PCR Experiment.

For nad5, the letters refer to the primers shown on the nad5 diagram in Materials and Methods.

| | Forward primer | Reverse primer |
|----------------------|-----------------------------|--------------------------|
| rpl2 | CCGAAGACGGATCAAGGTAA | CGCAATTCATCACCAATTTTG |
| rpl2 intron exon2 | TTAGGAAGAGCCGTACGAGG | CGCAATTCATCACCAATTTTG |
| rps3 | AGCCGAAGGTGAGTCTCGTA | CCGATTTTCGGTAAGACTTGG |
| rps3 intron1 exon2 | AGCCGAAGGTGAGTCTCGTA | TCTACGGCGGGGTCACTAT |
| cox2 | TGGGGGATTAATTGATTGGA | TGATGCTGTACCTGGTCGTT |
| cox2 intron1 exon2 | TGGGGGATTAATTGATTGGA | AGCAGTACGAGCTGAAAGGC |
| ccb452 | GTGGGTCCATGTAATGATCG | CACATGGAGGAGTGTGCATC |
| ccb452 intron1 exon1 | CCCGGATCGAATCAGAGTT | CACATGGAGGAGTGTGCATC |
| nad1 exon1-2 | GACCAATAGATACTTCATAAGAGACCA | TTGCCATATCTTCGCTAGGTG |
| nad1 intron1 exon2 | GACCAATAGATACTTCATAAGAGACCA | CGTGCTCGTACGGTTCATAG |
| nad1 exon2-3 | ATTCAGCTTCCGCTTCTGG | TCTGCAGCTCAAATGGTCTC |
| nad1 intron2 exon2 | GGTTGGGTTAGGGGAACATC | TCTGCAGCTCAAATGGTCTC |
| nad1 exon3-4 | AAAAGAGCAGACCCCATGGA | TCCGTTTGATCTCCGAGAAG |
| nad1 intron3 exon4 | AAAAGAGCAGACCCCATGGA | GGGAGCTGTATGAGCGGTAA |
| nad1 exon4-5 | AGCCCGGGATCTTCTTGA | TCTTCAATGGGGTCTGCTC |
| nad1 intron4 exon5 | AGCCCGGGATCTTCTTGA | ACGGAGCTGCATCCCTACT |
| nad2 exon1-2 | GCGAGCAGAAGCAAGGTTAT | GGATCCTCCCACACATGTTT |
| nad2 intron1 exon2 | GCGAGCAGAAGCAAGGTTAT | CCCATTCTTAACCAGTGGAG |
| nad2 exon2-3 | AAAGGAAGTGCAGTGATCTTGA | AATATTTGATCTTAGGTGCATTTT |
| nad2 intron2 exon2 | CCCGATCCGATAGTTTACAA | AATATTTGATCTTAGGTGCATTTT |
| nad2 exon3-4 | GCGCAATAGAAAGGAATGCT | CTATGGGTCTACTGGAGCTACCC |
| nad2 intron3 exon4 | GCGCAATAGAAAGGAATGCT | GGCGAATTTCAAACCTGTGG |
| nad2 exon4-5 | CAAAGGAGAGGGGTATAGCAA | TATTTGTTCTTCGCCGCTTT |
| nad2 intron4 exon4 | CTTATTCGTGGCAACCTTCC | TATTTGTTCTTCGCCGCTTT |
| nad4 exon1-2 | ATTCTATGTTTTTCCGAAAGC | GAAAAACTGATATGCTGCCTTG |
| nad4 intron1 exon2 | CCGTATGATGCGGAAGTCTC | GAAAAACTGATATGCTGCCTTG |
| nad5 intronE-I | AACTCGGATTCGGCAAGAA | GCCGTGTAATAGGCGACCA |
| nad7 intron4 exon5 | CGGCCAAATGACTACAGGAT | AGGTGCTTCAACTGCGGTAT |
| nad5 exon K-J | AACATTGCAAAGGCATAATGA | GTTCTGCGTTTCGGATATG |
| nad5 intronK-L | AACATTGCAAAGGCATAATGA | CCTGTAAACCCCATGATGT |
| nad7 intron1 exon2 | ACGGTTTTTAGGGGGATCTG | AAGGTAAAGCTTGAAGATAAGTTT |
| nad7 intron2 exon3 | AGTGGGAGAGCCGTGTTATG | TGGTACCTCGCAATTCAAAA |
| nad7 intron3 exon4 | TAAAGTGAAGTGGTGGGCCT | CATTGCACAATGATCCGAAG |
| nad4 exon2-3 | AATACCCATGTTTCCCGAAG | TGCTACCTCCAATTCCTGT |
| nad4 intron2 exon3 | GCGGAACGACCAGAAAAATA | TGCTACCTCCAATTCCTGT |
| nad4 exon3-4 | TTCCTCCATAAATTCTCCGATT | TGAAATTTGCCATGTTGCAC |
| nad4 intron3 exon4 | TCTAGCTTGGTTCGGAGAGC | TGAAATTTGCCATGTTGCAC |
| nad5 exon B-A | TGGACCAAGCTACTTATGGATG | CCATGGATCTCATCGGAAAT |
| nad5 intron B-C | TGGACCAAGCTACTTATGGATG | TTCGCAAATAGGTCCGACT |
| nad5 exon E-D | AACTCGGATTCGGCAAGAA | CTGGCTCTCGGGAGTCTCTT |
| nad5 exon F-D | GTGAGTCCCATGCGTCAGT | CTGGCTCTCGGGAGTCTCTT |
| nad5 intro G-D | GTACGATCGTGTCCGGTGA | CTGGCTCTCGGGAGTCTCTT |
| nad5 exon E-H | AACTCGGATTCGGCAAGAA | GTTTTGAAGCCGAGCCTTT |
| nad7 exon1-2 | ACCTCAACATCCTGCTGCTC | AAGGTAAAGCTTGAAGATAAGTT |
| nad7 exon2-3 | GAGGGACTGAGAAATTAATAGAGTACA | TGGTACCTCGCAATTCAAAA |
| nad7 exon3-4 | ACTGTCACTGCACAGCAAGC | CATTGCACAATGATCCGAAG |
| nad7 exon4-5 | GATCAAAGCCGATGATCGTAA | AGGTGCTTCAACTGCGGTAT |

RESULTATS

La protéine TANG2 n'est pas totalement essentielle pour l'épissage en *trans* de l'intron 3 de *nad5*. D'autres défauts mineurs sont présents chez le mutant *tang2*, et il est intéressant de voir que ces défauts se retrouvent également chez le mutant *otp43*. Par exemple, ces deux mutants présentent aussi une réduction de l'épissage de l'intron 1 de *nad2* et une accumulation de la forme non épissée (Figure 10). Ces deux protéines ont-elles une action plus ou moins directe sur l'efficacité de l'épissage de cet intron ou s'agit-il d'un effet de régulation dû à l'absence de l'assemblage des sous-unités du complexe I ? A ce stade, il est difficile de répondre à cette question, en revanche il serait intéressant de vérifier ce phénomène chez d'autres mutants affectés dans l'épissage des introns des gènes *nad* ou chez d'autres mutants affectés dans les sous-unités de ce complexe. En collaboration avec Etienne Meyer, j'ai eu l'occasion de regarder le profil chez un mutant affecté dans une des sous-unités du complexe I. Le défaut d'épissage de l'intron 1 de *nad2* n'est pas retrouvé alors que le mutant *ndufs4* n'a plus de complexe I (Figure 10). Ce défaut n'est donc pas typique des mutants de complexe I mais il est plus probable que OTP43 et TANG2 aient une fonction mineure (directe ou indirecte) dans l'épissage de l'intron 1 de *nad2*.

Un autre point intéressant à noter est la surexpression du transcrit *nad6* chez les trois mutants du complexe I. Comme il est montré sur la figure 9, le transcrit *nad6* semble être plus abondant chez les mutants *otp43*, *tang2* et *ndufs4* comparés au sauvage et aux mutants complémentés. Ce résultat suggère qu'une régulation de la transcription ou une régulation post-transcriptionnelle pourrait se mettre en place lorsqu'une des sous-unités du complexe I n'est plus présente.

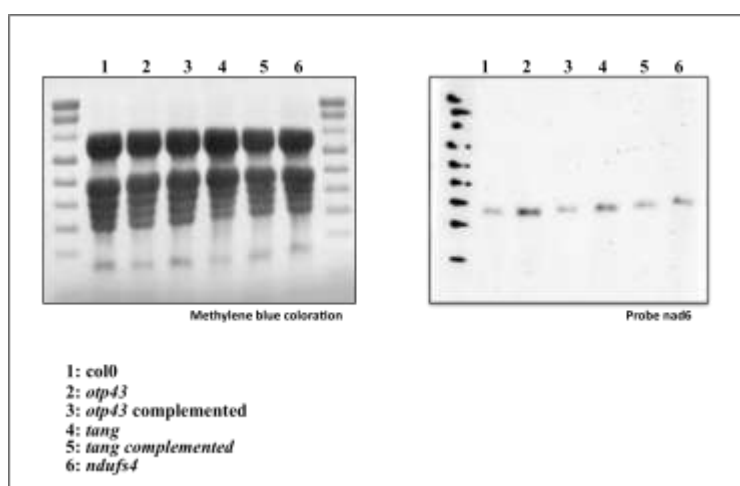


Figure 9 : Analyse par *northern blot* des ARN issus des mutants de complexe I et des mutants complémentés avec la sonde *nad6*.

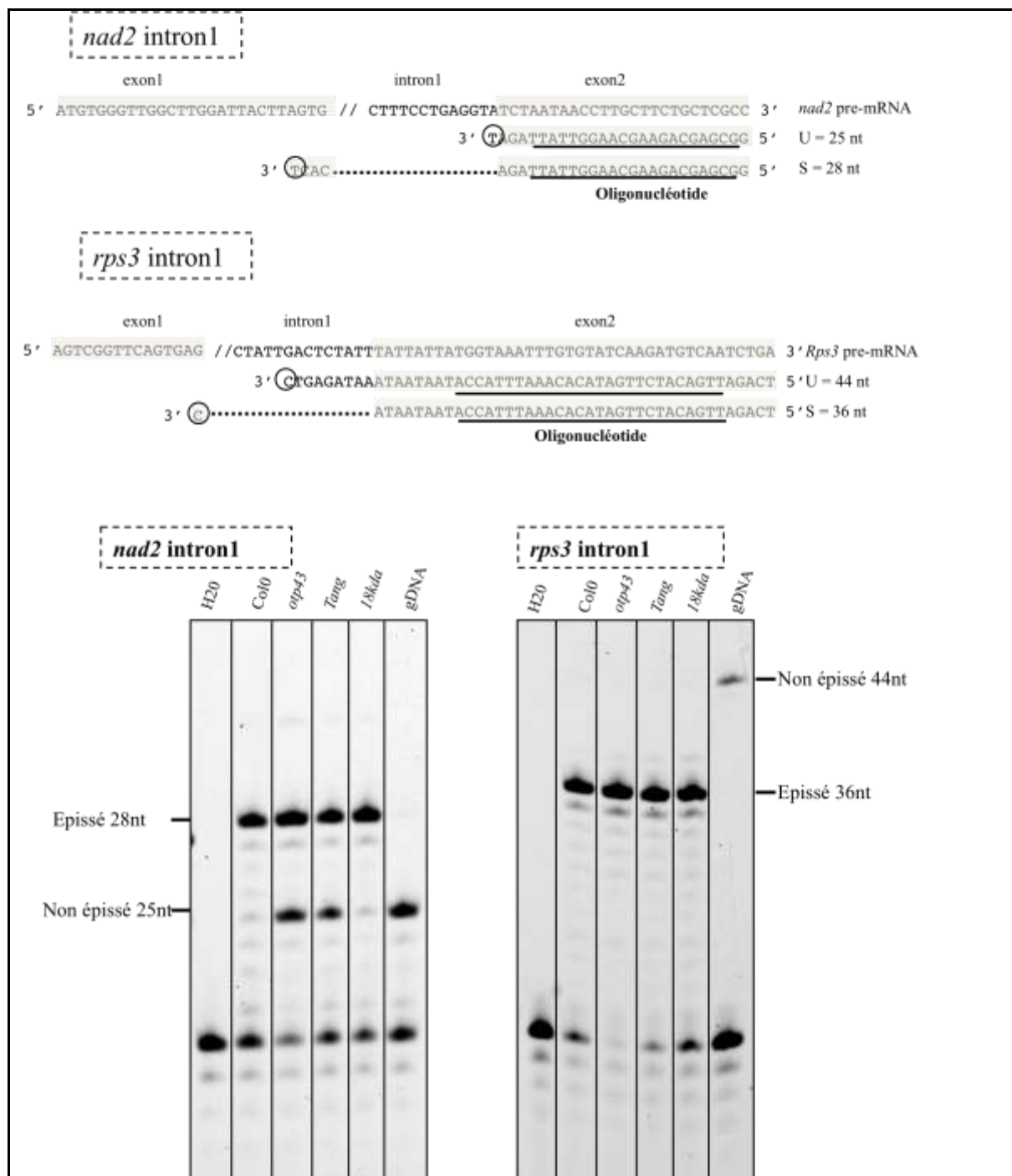


Figure 10 : Quantification par PPE des évènements d'épissage de l'intron1 de *nad2* et de l'intron de *rps3* chez les mutants *otp43*, *tang2*, *ndusf4* comparés au sauvage.

II.2.3PPR40

PPR40 est également une PPR de type « Pure » faisant partie des 48 protéines PPR que nous avons sélectionnées lors de l'étude systématique des protéines PPR. Hormis des clonages et des données de localisation subcellulaire montrant que cette PPR est localisée dans les mitochondries, aucune étude n'avait été réalisée sur les mutants d'insertion pour cette PPR dans l'équipe de Ian Small et de Claire Lurin. En revanche, une analyse de deux mutants d'insertion faite par l'équipe de L. Szabados a permis d'élucider en partie le rôle de cette protéine PPR (Zsigmond, Rigo et al. 2008).

RESULTATS

Cette protéine PPR fusionnée à un TAG HA en C-terminal a été montrée comme étant apparemment associée au complexe III, cependant des données de spectrométrie de masse sur le complexe III provenant d'Etienne Meyer n'ont jamais permis d'identifier cette protéine PPR. Néanmoins, la majorité des protéines PPR étant faiblement exprimées, il est possible qu'aucune identification par cette technique ne puisse être faite. Dans les résultats obtenus, Zsigmond et ses collaborateurs montrent que le mutant présente une diminution du complexe III. Or le Blue-Native présenté dans le papier semble montrer également une diminution du complexe I. Les événements d'épissage pour les introns de *nad1*, *nad4*, et *nad7* sont présentés dans l'article mais aucune mention n'est faite des introns des gènes *nad2* et *nad5* alors que l'on en dénombre 8 pour ces gènes.

Le phénotype des deux allèles de *ppr40* est caractérisé par des plantes plus petites avec un retard de croissance, phénotype assez similaire à celui des mutants de complexe I, même si cependant d'autres défauts pourraient conduire au même phénotype. Ces différentes observations, ou manque d'informations, concernant les introns de *nad2* et *nad5* me laissaient penser que ce mutant pouvait éventuellement présenter un défaut au niveau du complexe I. Nous avons au laboratoire un des deux allèles décrits dans le papier, *ppr40-2* (SALK_071712), mais pour lequel le phénotype est moins marqué que *ppr40-1*. J'ai donc décidé de regarder les événements d'épissage chez ce mutant afin de voir si le phénotype observé pouvait s'expliquer par un défaut dans ce processus au niveau des introns de *nad2* et *nad5*. La PCR quantitative sur les formes épissées et non épissées chez le mutant montre des défauts partiels d'épissage pour certains introns (figure 11). En effet, le mutant présente huit fois moins de forme épissée des introns 1 et 4 de *nad5*. Cette diminution d'épissage pourrait expliquer la diminution de complexe I observée dans le BN-PAGE décrit dans le papier de Zsigmond et al.

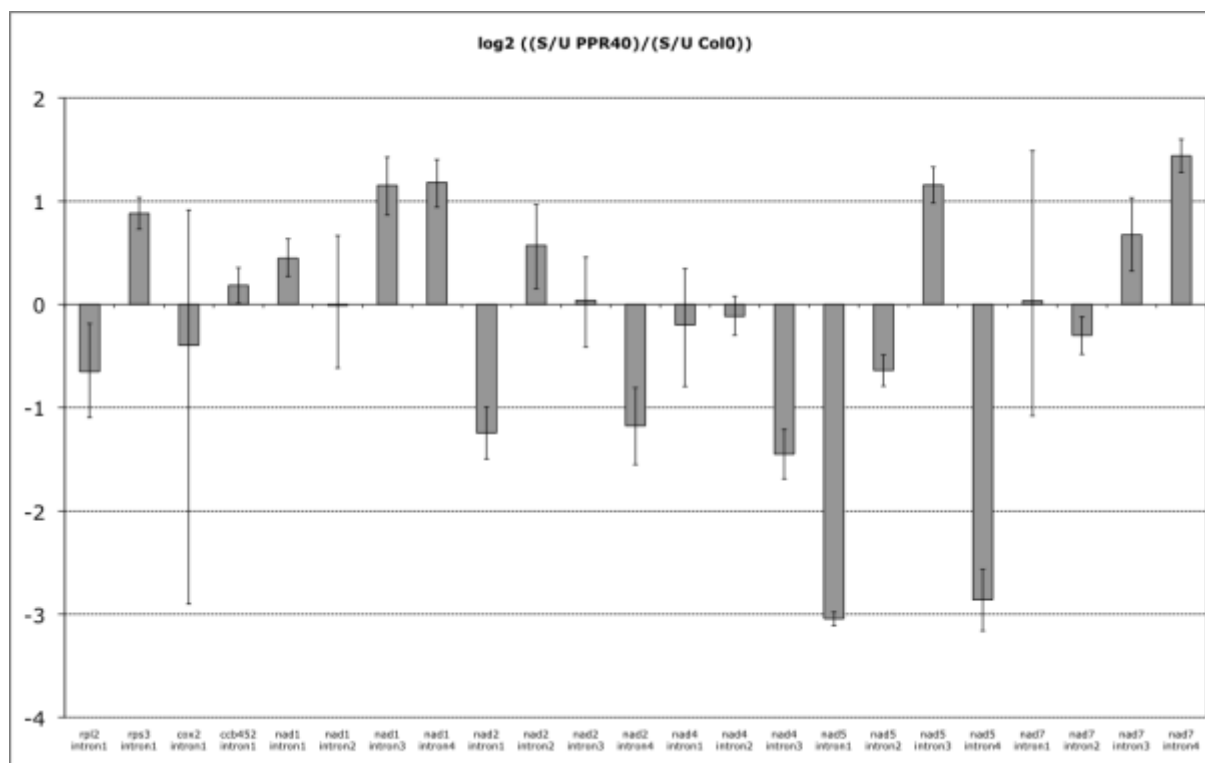


Figure 11 : Profil de l'épissage mitochondrial du mutant *ppr40-2*. Les ratios des formes épissées sur les formes non épissées sont représentés en Log2 en abscisse.

Le défaut observé dans l'épissage de certains introns chez le mutant *ppr40* ne signifie pas que la protéine PPR40 est un facteur d'épissage. En effet, ce défaut est une diminution et non une absence d'épissage et suffisamment de formes épissées sont présentes pour la synthèse et la formation de complexe I. Cependant, l'analyse faite ici se porte sur le deuxième allèle dont le phénotype est moins marqué que celui de *ppr40-1*. Une nouvelle analyse vient d'être faite sur cet allèle par Catherine Colas-des-Francis dans le laboratoire de Ian Small et montre qu'il n'y a pas plus de défaut d'épissage chez le mutant *ppr40-1* que pour *ppr40-2*. Le phénotype étant plus marqué il est possible qu'il existe une autre mutation dans le mutant *ppr40-1*. Quoiqu'il en soit, cette mutation influence clairement l'efficacité de l'épissage.

Comme il a été décrit dans l'introduction, les protéines PPR sont impliquées dans de nombreuses modifications post-transcriptionnelles. Il est possible que PPR40 ne soit pas impliquée directement dans l'épissage car cette protéine ne se trouve pas être indispensable pour les deux introns de *nad5* pour lesquels un défaut d'épissage a été observé, mais elle pourrait avoir un rôle dans la stabilité de certaines formes d'ARNm pré-épissées ou dans des maturations en 5' ou en 3' des transcrits primaires, influençant ainsi l'épissage. Il reste donc de nombreuses analyses à effectuer afin de comprendre exactement le rôle de cette protéine PPR.

II.2.4 Comment identifier de nouvelles PPRs impliquées dans l'épissage des introns *nad* ?

La majorité des introns dans le génome mitochondrial se situe dans les gènes *nad*. Ces gènes codent pour des sous-unités du complexe NADH dehydrogenase. Ces sous-unités faisant partie du même complexe, il est possible qu'une co-régulation de l'expression des gènes codant ces sous-unités ait lieu. Cette régulation pourrait s'effectuer à différents niveaux, en commençant par la transcription, la maturation des ARNm et enfin par l'activité traductionnelle. La régulation de l'épissage pouvant être un mode de régulation de l'expression, il est possible que les introns des gènes *nad* soient épissés et régulés de façon conjointe. Si c'est le cas, les facteurs impliqués dans l'épissage de ces introns doivent être exprimés de façon conjointe. S'il existe une régulation quelconque de l'épissage, il pourrait être possible d'identifier d'autres protéines PPR dont la régulation de l'expression des gènes serait similaire à celle des deux gènes codant des protéines PPR décrites dans l'épissage des introns *nad*, et ainsi identifier d'autres protéines PPR impliquées dans l'épissage.

II.2.4.1 Co-régulation des PPRs

Comme le montre le graphique d'expression réalisé sous Genecat (<http://genecat.mpg.de/>) dans la figure 12, le profil d'expression des gènes *OTP43* et *TANG2* au cours du développement est assez similaire.

Les données de ce même logiciel donnent un coefficient de corrélation de Pearson de 0,67 entre ces deux gènes, montrant ainsi une corrélation moyenne entre l'expression des deux gènes. En utilisant ce logiciel, de nombreux gènes codant pour des protéines PPR sont identifiés comme étant co-exprimés avec *OTP43* et *TANG2*, mais il est hasardeux de supposer que ces protéines soient impliquées dans un mécanisme similaire car les coefficients restent assez faibles.

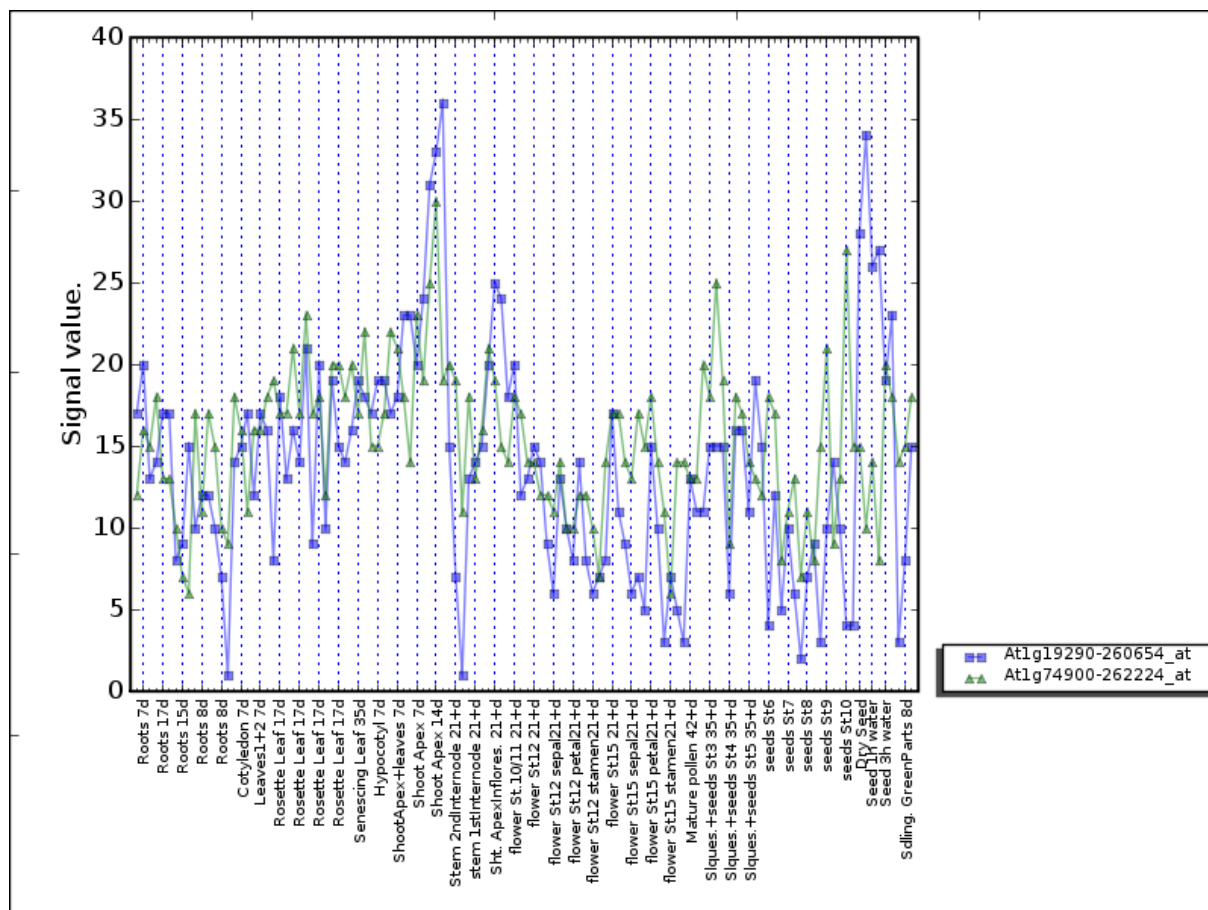


Figure 12 : Expression des gènes *OTP43* (At1g74900) et *TANG2* (At1g19290) au cours du développement. Analyse réalisée sous Genecat (<http://genecat.mpg.de/>) en utilisant l’option « profil d’expression » et sans normalisation du signal d’expression.

Le mutant *tang2* a été identifié au cours de tests de sensibilité au glucose. En présence de glucose ce gène est très fortement exprimé. Il est difficile d’établir une relation directe entre le glucose et le fonctionnement de la protéine TANG2 dans l’épissage, mais cette information pourrait être utile dans la recherche de nouveaux facteurs intervenant dans ce processus.

Sur Genevestigator (<https://www.genevestigator.com>), logiciel regroupant un grand nombre d’expériences de « microarrays », il existe peu de conditions pour lesquelles les deux gènes réagissent de la même manière dans leur expression (ce qui confirme l’analyse effectuée sous Genecat). Seule une condition montre une réelle co-expression : les 2 gènes *OTP43* et *TANG2* sont surexprimés en présence de fortes concentrations en saccharose. Une étude a donc été entreprise à l’aide de Genevestigator afin d’identifier les autres gènes PPR surexprimés dans les mêmes conditions. L’idée était de tester par la suite en génétique inverse si les protéines codées par ces gènes ont un rôle dans l’épissage des introns mitochondriaux. Il est clair que cette approche reste très critiquable, mais, n’ayant au début que peu d’éléments concernant les facteurs d’épissage de type PPR dans les mitochondries, cette approche me semblait plus appropriée qu’une approche de criblage complètement aléatoire de mutants *ppr*.

II.2.4.2 Les nouveaux facteurs d’épissage identifiés

Parmi les 450 gènes codant pour des protéines PPR, seulement 378 possèdent une sonde de type Affymetrix (description dans le tableau 1 en annexe). Une analyse de l’expression sur saccharose grâce à Genevestigator a donc été faite pour ces 378 gènes. En plus des 2 gènes

OTP43 et *TANG2*, 19 autres gènes codant pour des protéines PPR sont surexprimés d'un facteur au moins deux dans des expériences de haute concentration en saccharose (Figure 13).

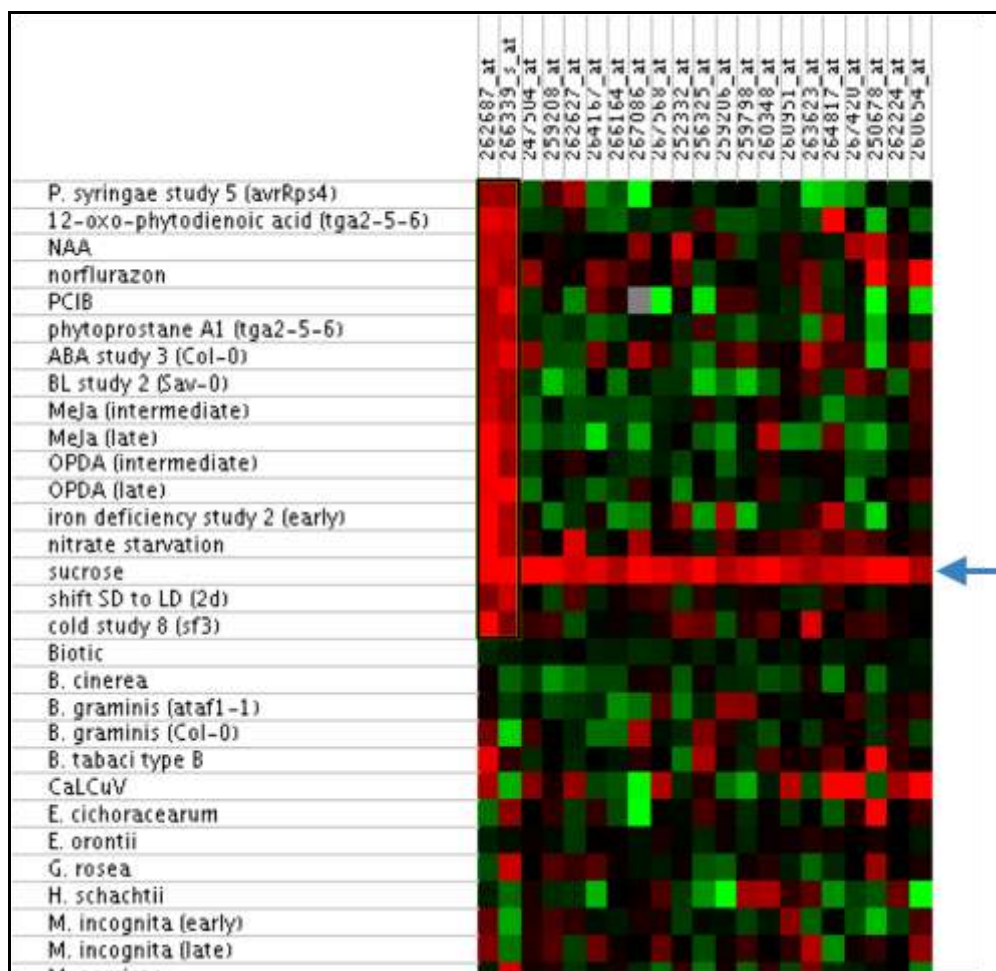


Figure 13 : Capture d'image de l'analyse de l'expression des 19 gènes codant pour des protéines PPR sous GENEVESTIGATOR. La flèche bleue montre la surexpression de tous ces gènes en présence d'une forte concentration en saccharose.

A l'exception d'une protéine montrée expérimentalement par fusion GFP localisée dans les mitochondries, aucune donnée de localisation n'a été réalisée pour les 18 autres. Cependant, les données de prédiction de localisation obtenues sous SUBA (<http://suba.plantenergy.uwa.edu.au/>) permettent de prédire une éventuelle fonction dans les mitochondries pour la majorité d'entre-elles (tableau 4).

Parmi ces 19 protéines PPR, 10 sont des PPR de type P, 3 de type E, 5 de type E+ et une de type DYW. Les protéines PPR de la sous classe PLS sont généralement impliquées dans l'édition ou le clivage des ARNm. De ce fait, si ces gènes surexprimés sur saccharose ont un rôle quelconque dans l'épissage, il est fort possible que l'effet sur ce type de maturation chez ces mutants soit indirect, en raison d'un défaut soit dans l'édition de sites proches des sites d'épissage (intron ou exon), soit d'un défaut de clivage des ARN pré-messagers. Pour cette raison, je me suis focalisé dans un premier temps sur l'étude des lignées d'insertion ADN-T dans les gènes codant pour les 10 protéines PPR-P, car jusqu'à présent seules des protéines PPR-P ont été décrites comme jouant un rôle dans l'épissage.

RESULTATS

| locus | Type | location_targetp | location_mitoprot2 | location_ipsort | location_predotar | location_mitopred |
|-------------|------|------------------|--------------------|-----------------|-----------------------|-------------------|
| AT1G02060.1 | Pure | plastid | | mitochondrion | mitochondrion | |
| AT1G03540.1 | E | mitochondrion | | mitochondrion | mitochondrion | mitochondrion |
| AT1G06150.1 | E+ | | | plastid | | mitochondrion |
| AT1G06580.1 | Pure | mitochondrion | mitochondrion | mitochondrion | endoplasmic reticulum | mitochondrion |
| AT1G62670.1 | Pure | mitochondrion | mitochondrion | mitochondrion | mitochondrion | |
| AT1G64310.1 | E | mitochondrion | mitochondrion | mitochondrion | plastid | mitochondrion |
| AT1G69350.1 | E+ | mitochondrion | mitochondrion | mitochondrion | mitochondrion | mitochondrion |
| AT2G01390.1 | Pure | plastid | mitochondrion | mitochondrion | mitochondrion | |
| AT2G04860.1 | E | | | | | |
| AT2G28050.1 | Pure | | | mitochondrion | mitochondrion | mitochondrion |
| AT2G30780.1 | Pure | mitochondrion | mitochondrion | mitochondrion | mitochondrion | mitochondrion |
| AT2G32630.1 | Pure | mitochondrion | mitochondrion | mitochondrion | | mitochondrion |
| AT2G35030.1 | E+ | mitochondrion | mitochondrion | mitochondrion | mitochondrion | mitochondrion |
| AT3G02330.1 | E+ | | | | | mitochondrion |
| AT3G09040.1 | E+ | mitochondrion | | mitochondrion | mitochondrion | mitochondrion |
| AT3G09060.1 | Pure | | | plastid | mitochondrion | mitochondrion |
| AT3G48810.1 | Pure | extracellular | | mitochondrion | | |
| AT5G06540.1 | DYW | extracellular | | mitochondrion | mitochondrion | mitochondrion |
| AT5G61990.1 | Pure | mitochondrion | mitochondrion | mitochondrion | mitochondrion | mitochondrion |

Tableau 4 : Prédiction de la localisation subcellulaire des 19 protéines PPR surexprimées en présence de saccharose. Prédictions extraites de la base de données SUBA (<http://suba.plantenergy.uwa.edu.au/>).

II.2.4.2.1 « Pure » PPRs (OTP439)

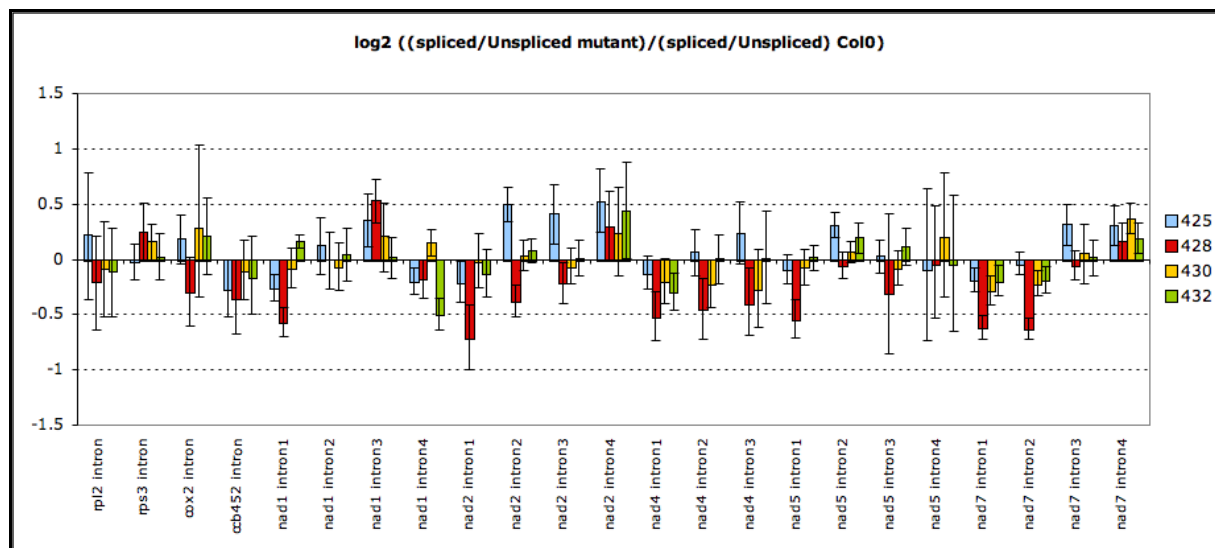
Parmi les 10 gènes codant pour des PPR-P, il a été possible d'identifier des mutants d'insertion ADN-T disponibles dans les bases de données pour 9 d'entre-elles. Ces lignées ont été commandées, semées, et un génotypage ainsi qu'une étude du phénotype ont été réalisés pour chaque lignée. Les résultats sont présentés dans le tableau 5. Pour deux gènes PPR, At5g61990 et At3g09060, aucune plante homozygote pour l'insertion n'a pu être obtenue. Pour deux autres gènes, At2g32630 et At2g30780, il n'a pas été possible d'amplifier le transgène pour toutes les lignées décrites, l'insertion n'a donc pas été confirmée. Finalement, seuls 5 gènes sur les 10 de départ ont permis d'identifier des mutants homozygotes pour l'insertion de l'ADN-T. Une lignée homozygote pour chacun des 5 gènes a été sélectionnée et une détection des événements d'épissage mitochondriaux a été entreprise. Les 4 lignées homozygotes pour les 4 gènes, At1g06580, At1g62670, At1g02060, At2g01390 ne présentent pas de phénotype particulier et sans surprise aucun défaut d'épissage n'est détecté par PCR quantitative comme le montre la figure 14.

En revanche, la dernière lignée d'insertion homozygote, correspondant au gène At3g48810 codant pour une protéine PPR nommée OTP439 par la suite, présente un défaut de retard de développement. La détection des événements d'épissage représentée sur la figure 15 montre plusieurs défauts partiels dans l'épissage de l'intron 1 de *nad2*, de l'intron 1 de *nad5* et de l'intron 2 de *nad5* (dans cette expérience les oligonucléotides dessinés pour la PCR quantitative ne permettaient pas de distinguer entre l'intron 2 et 3 de *nad5*. C'est au cours d'une nouvelle analyse non montrée ici, en redessinant des oligonucléotides, que l'on a pu distinguer le défaut entre ces deux introns). A ce stade il est intéressant de noter que l'on retrouve un défaut partiel de l'épissage de l'intron 1 de *nad2* identique aux défauts observés chez les autres mutants, *tang2*, *otp43* et *ppr40*, affectés dans l'épissage des introns mitochondriaux.

RESULTATS

| AGI | lignées T-DNA | génotypage | phenotype |
|-----------|---------------|--------------------------|-------------------------|
| At5g61990 | SALK_143317 | pas de plantes HO | |
| At3g09060 | SALK_144391 | pas de plantes HO | |
| At1g06580 | SAIL_633_D09 | plantes Homozygotes | sauvage |
| | SALK_022879 | plantes Homozygotes | sauvage |
| | SALK_032203 | plantes Homozygotes | sauvage |
| | SALK_032204 | mauvaise insertion | |
| At1g62670 | SALK_146198 | plantes Homozygotes | sauvage |
| | SALK_146196 | mauvaise insertion | |
| At1g02060 | SAIL_423_B07 | plantes Homozygotes | sauvage |
| At2g28050 | SALK_058434 | pas de lignée disponible | |
| At2g01390 | SALK_038623 | mauvaise insertion | |
| | SALK_025648 | plantes Homozygotes | sauvage |
| | SALK_049291 | mauvaise insertion | |
| | SALK_049289 | mauvaise insertion | |
| At2g32630 | SALK_043094 | mauvaise insertion | |
| | SALK_050020 | mauvaise insertion | |
| | SALK_043509 | mauvaise insertion | |
| | SALK_082869 | mauvaise insertion | |
| At2g30780 | SALK_082869 | mauvaise insertion | |
| At3g48810 | SALK_089911 | plantes Homozygotes | retard de développement |
| | SALK_009125 | pas de lignée disponible | |

Tableau 5 : Résumé des génotypages et de l'analyse des phénotypes associés des mutants présentant des insertions dans des gènes PPR de type « Pure » sur-exprimés sur saccharose.



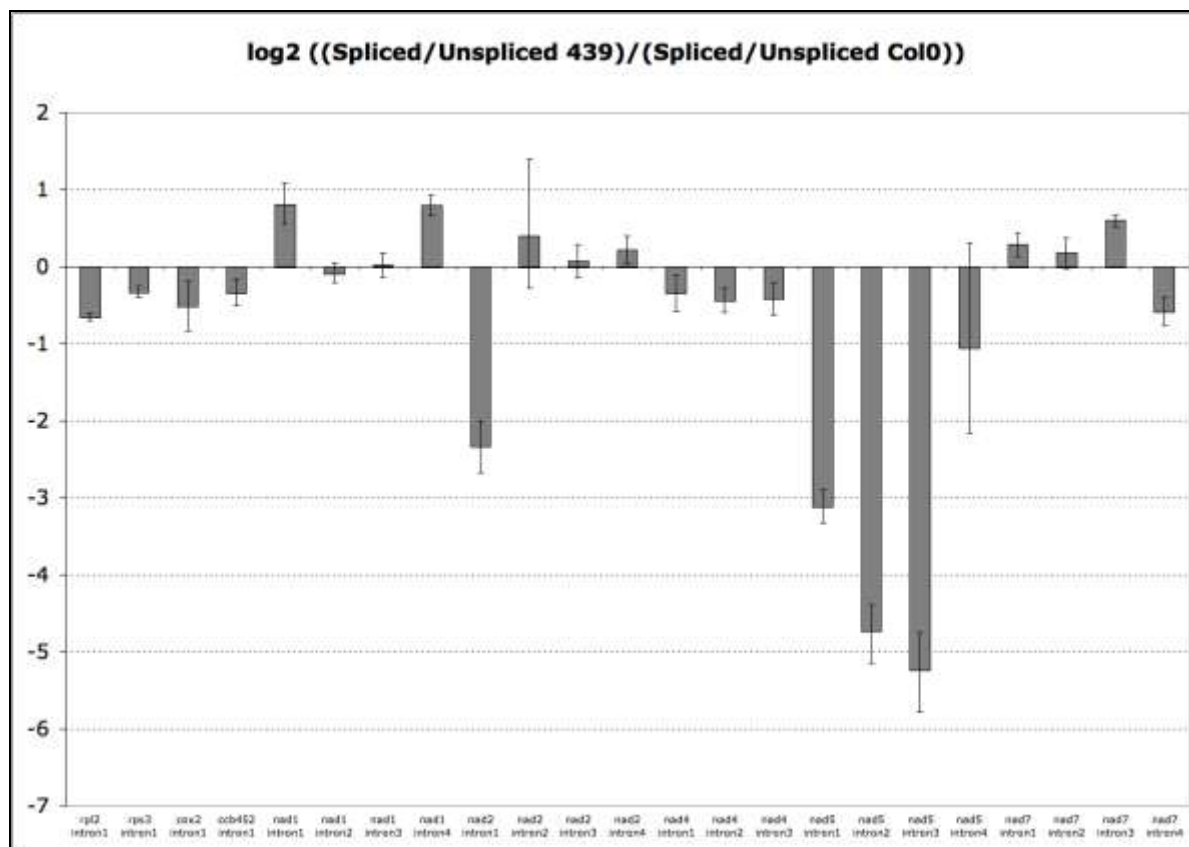


Figure 15 : Profil de l'épissage mitochondrial chez le mutant *otp439*.

Ce retard de développement, observé chez ce mutant est typique des mutants de complexe I et peut s'expliquer par une réduction des formes épissées des transcrits *nad2* et *nad5*. Cette réduction s'accompagne d'une diminution mais non une absence de complexe I, ainsi que de la présence d'un sous-complexe partiel de taille plus petite comprenant la sous-unité NAD6 (résultats obtenus par Catherine Colas-des-Francis à Perth). Cette présence de ce sous-complexe est accompagnée d'une sur-expression de la protéine NAD6 et est spécifique du mutant *otp439*. Le mutant *tang2* est également affecté dans l'épissage des transcrits *nad5* mais, en revanche, il ne présente pas d'accumulation de la protéine NAD6 et de sous-complexe comme pour *otp439*. Il est donc possible qu'OTP439, en plus d'influencer l'épissage, ait une autre action. Même si le défaut d'épissage est assez marqué chez *otp439*, ce défaut n'est pas total et on ne peut donc pas conclure sans expérience additionnelle que la protéine OTP439 est un facteur d'épissage mitochondrial, en revanche cette protéine est importante pour l'efficacité d'épissage de certains introns et est ainsi la quatrième protéine PPR identifiée influençant l'épissage mitochondrial. Il sera intéressant d'effectuer de plus amples études afin de comprendre comment cette protéine influence l'épissage, si c'est en stabilisant l'ARNm ou alors en agissant sur sa maturation.

II.2.4.2.2 Influence de l'édition dans l'épissage

Neuf gènes codant pour des protéines de type PLS sont également surexprimés sur saccharose. Aucune donnée ne permet de montrer la fonction de ces neuf protéines au niveau mitochondrial, cependant le fait qu'elles appartiennent à la sous-famille PLS suggère un éventuel rôle dans l'édition ou dans le clivage. Certains sites d'édition dans le génome mitochondrial sont décrits à proximité de jonctions exon-intron ou au sein même de l'intron. Ces sites pourraient avoir une influence sur l'efficacité de l'épissage en agissant sur la structure secondaire des ARN pré-messagers comme il a déjà été montré pour l'épissage de

l'intron de *cox2* chez le blé où l'édition d'un site dans la zone IBS (pour Intron Binding Site) de l'exon1 est essentielle pour l'efficacité de l'épissage (Kurihara-Yonemoto and Handa 2001).

Afin de savoir si ces neuf gènes ont un rôle quelconque dans l'épissage, une approche de génétique inverse a donc été mise en place. Deux gènes, At1g06145 et At2g04860 n'ont pas de lignées d'insertion ADN-T associée dans les bases de données. Pour les 7 autres gènes, une étude des lignées d'insertion a permis d'obtenir des mutants homozygotes pour 3 gènes uniquement. Cependant, aucune de ces 3 lignées ne présente de phénotype macroscopique. Les mutants affectés dans le complexe I présentent des délais de germination plus ou moins sévères selon la mutation. Ainsi pour *otp43*, des plantes homozygotes ne peuvent être obtenues qu'en semant des graines issues d'hétérozygotes sur un milieu *in vitro* standard et après un long délai avant germination. Ainsi, contrairement à des graines issues de plantes sauvages ou à trois quart des graines issues d'hétérozygotes, il faut attendre entre 2 et 3 semaines avant que des graines homozygotes ne germent. Dans le cas des lignées pour lesquelles aucun mutant homozygote n'avait été obtenu dans une première étude, j'ai prolongé les temps de germination et transféré en terre les plantules ayant germé très tardivement. En utilisant cette méthode, des plantes homozygotes pour l'insertion ADN-T SALK_043866 dans le gène At1g03540 ont été obtenues très récemment. Ce mutant présente un délai de germination très important et montre également un fort délai de développement ainsi qu'un phénotype nain (Figure 16). Alors que les graines des plantes hétérozygotes pour l'insertion sont déjà récoltées, la majorité des plantes homozygotes commencent seulement à avoir une hampe florale comme le montre la photo dans la figure 16. Seule une lignée ADN-T pour ce gène montre ce phénotype. Pour les deux autres lignées, il n'a jamais été possible de détecter la présence du transgène par PCR sur ADNg.

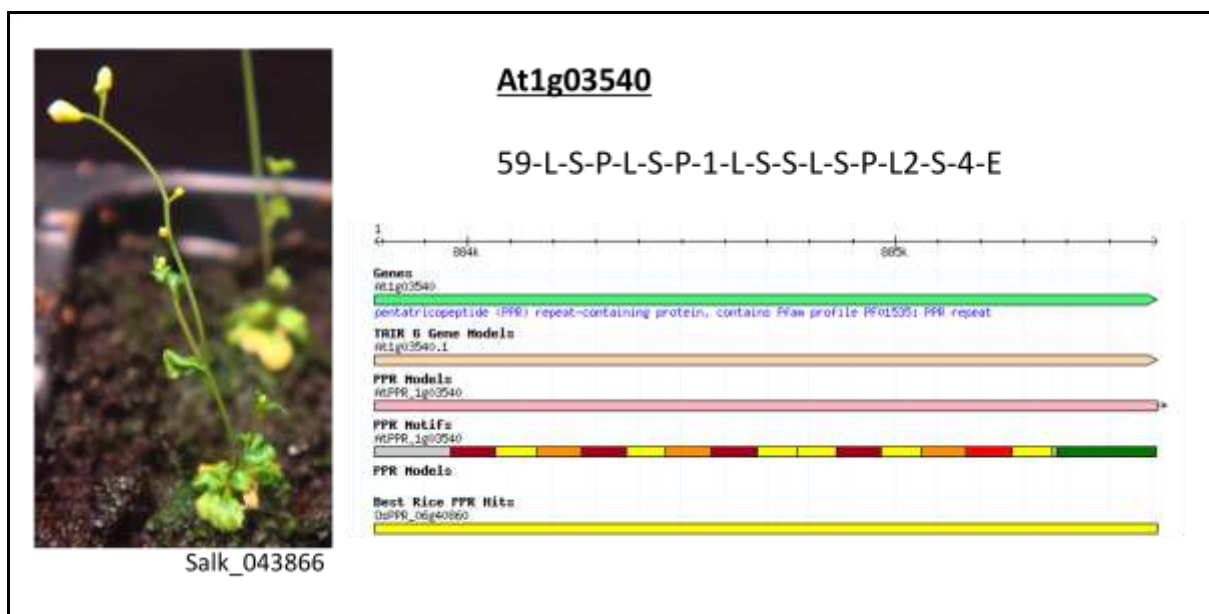


Figure 16 : Photos du mutant salk_043866 et description des motifs PPR qui composent la protéine At1g03540.

Ce phénotype très marqué est probablement le résultat d'un dysfonctionnement mitochondrial. Afin d'identifier le défaut moléculaire, j'ai tout d'abord caractérisé les événements d'épissage chez ce mutant par PCR quantitative dans l'espoir qu'un défaut d'édition ou de clivage pourrait avoir un impact sur l'épissage. Lorsque la comparaison entre le mutant et le sauvage est effectuée entre les taux de formes épissées, on observe une

RESULTATS

diminution de l'épissage de l'intron1 de *nad2*, diminution déjà observée chez les autres mutants étudiés précédemment (figure 17A). En revanche, lorsque cette comparaison est effectuée entre les ratios de formes épissées sur les formes non épissées (Figure 17B), le mutant montre une diminution relativement importante pour une majorité des introns. Ainsi, le taux d'épissage ne semble pas être affecté à l'exception de l'intron 1 de *nad2*, mais le taux de transcrits non épissés est fortement augmenté chez le mutant. Ce mutant est actuellement en cours d'étude afin de comprendre quel est le défaut moléculaire primaire. Dans le cadre d'un projet de mise au point des nouvelles approches de séquençage « haut-débit » sur appareil de type « Solexa », ce mutant va être utilisé pour évaluer notre capacité à détecter des défauts moléculaires chez des mutants PPR par ce type d'approche.

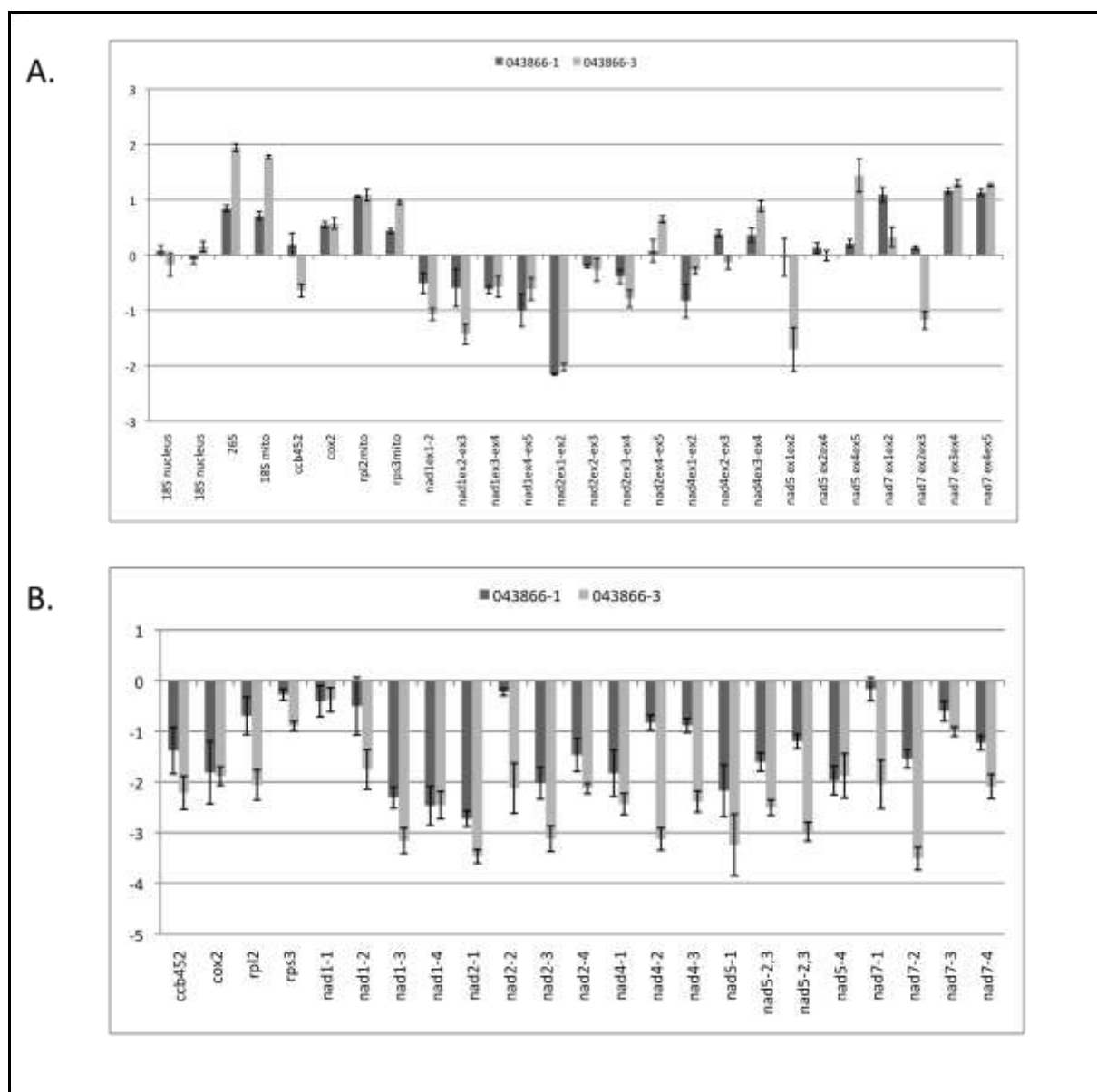


Figure 17 : Profil de l'épissage des transcrits mitochondriaux du mutant salk_043866. 043866-1 et 043866-3 sont deux répétitions biologiques. A. Ratios des formes épissées entre le mutant et le sauvage représentés en Log2 en abscisse. B. Ratios des formes épissées sur les formes non épissées entre le mutant et le sauvage représentés en Log2 en abscisse.

II.2.4.2.3 *Effet du sucre sur l'épissage*

Le saccharose ou le glucose ont visiblement une influence sur la transcription de certains gènes codant pour des protéines PPR impliquées dans l'épissage des introns mitochondriaux, mais il reste à vérifier si cette surexpression a un impact sur l'efficacité de l'épissage. En d'autres termes : Y a-t-il une variation de l'épissage lorsque des graines ou des plantules sont cultivées sur des milieux plus ou moins riches en sucre ? En vue de mieux comprendre le phénomène observé et de tenter d'expliquer cette surexpression des gènes codant ces protéines PPR, j'ai donc entrepris différentes expériences de culture d'*Arabidopsis* de type sauvage (Col-0) sur des milieux plus ou moins riches en sucre et analysé les différents événements d'épissage afin d'observer d'éventuels changements de ratios entre les formes épissées et non épissées pour chaque intron.

Deux types d'expériences ont été réalisées, la première a été de caractériser l'épissage chez des plantules cultivées sur des milieux ½ MS avec des concentrations de saccharose variant de 0 à 3% et d'effectuer des comparaisons de ratios avec des plantules poussées sur 1% de saccharose. La deuxième expérience a consisté à effectuer la même analyse pour des plantules cultivées sur des concentrations de glucose variant de 0 à 3% comparées à des plantules cultivées sur des concentrations en mannitol similaires. Comme il est montré dans les deux figures suivantes (figures 18 et 19), aucun changement significatif n'a été détecté dans ces deux types d'expériences. Alors qu'une forte concentration en sucres influence l'expression de gènes codant pour des PPR impliquées dans l'épissage, il n'existe pas de lien entre cet effet et le taux de transcrits mitochondriaux épissés.

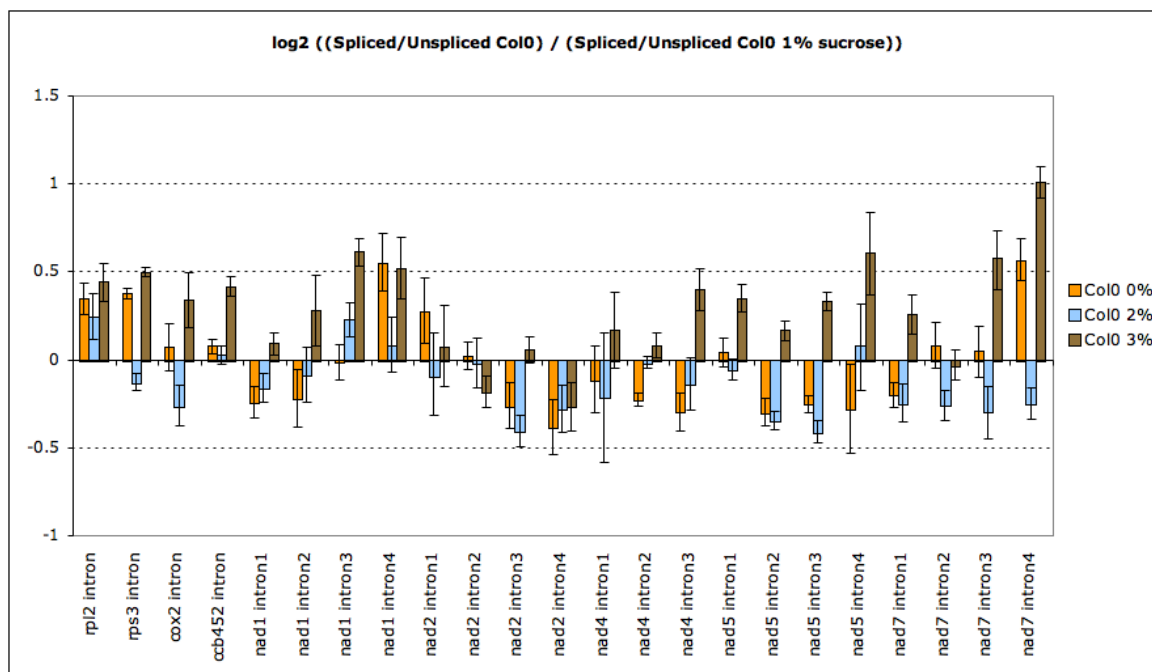


Figure 18 : Profil de l'épissage mitochondrial chez des plantules cultivées en présence de différentes concentrations de saccharose. La comparaison est faite avec des plantules cultivées sur 1% de saccharose.

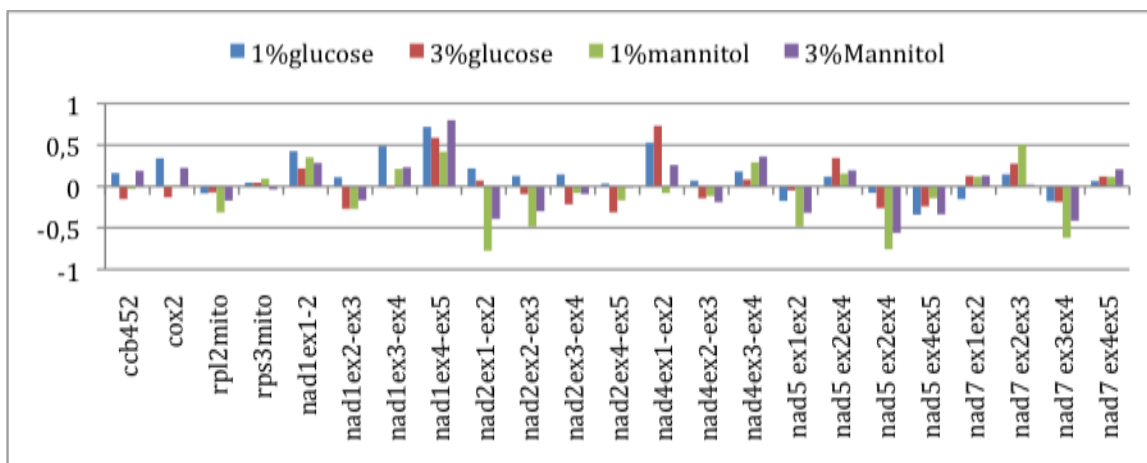


Figure 19 : Profil de l'épissage mitochondrial chez des plantules cultivées en présence de différentes concentrations de glucose et de mannitol.

Il n'a pas été possible dans cette expérience de regarder en détails le taux de transcription mitochondrial en présence de sucre, mais il est possible qu'en condition de forte concentration de sucre, l'activité transcriptionnelle mitochondriale soit augmentée et que les facteurs d'épissage soient donc aussi sur-exprimés.

II.3 L'épissage alternatif des PPR

Préambule : Au cours de clonages, j'ai été amené à découvrir par hasard des événements d'épissage alternatif dans des transcrits codant pour des protéines PPR et j'ai donc voulu évaluer ce phénomène au sein de la famille.

L'épissage alternatif est un phénomène permettant à un gène de coder pour plusieurs protéines. Ce phénomène s'établit soit par l'épissage de façon alternative de un ou plusieurs exons, soit par la rétention d'un ou plusieurs introns ou par l'utilisation de sites donneurs ou accepteurs d'épissage alternatifs. Ce phénomène chez l'homme joue un rôle très important dans le développement des cellules, l'organisation des tissus et même dans le développement d'un individu. Le dogme de « un gène => une protéine » est donc révolu et il est maintenant montré par séquençage haut-débit que plus de 90 % des gènes chez l'homme subissent de l'épissage alternatif (Wang, Sandberg et al. 2008) créant ainsi une grande variété de protéines largement supérieure à celle estimée. Ainsi 24 000 gènes chez l'homme vont permettre de coder pour environ 100 000 protéines (Modrek and Lee 2002).

Chez les plantes, et notamment chez Arabidopsis, l'épissage alternatif a longtemps été largement sous-estimé. Alors que les données d'EST dénombraient moins de 20% des gènes chez Arabidopsis possédant des formes alternatives d'épissage (Kim, Magen et al. 2007), les nouvelles données provenant de séquençage haut-débit montrent que 42% des gènes comprenant des introns ont des formes d'épissage alternatives (Filichkin, Priest et al. 2010). Contrairement au monde animal, chez les plantes peu d'études ont été réalisées concernant la fonction des différents isoformes au niveau protéique cependant il est certain que ces isoformes permettent d'augmenter la diversité protéique.

L'épissage alternatif est régulé de façon temporelle et spatiale et est souvent induit lors de stress environnementaux (Reddy 2007), permettant ainsi la régulation de l'expression des gènes. Dans ce type de régulation, deux mécanismes majeurs sont impliqués. De nombreux isoformes découverts par séquençage haut-débit aboutissent à des formes pour lesquelles l'épissage alternatif conduit à un changement de phase et à l'apparition d'un codon stop prématuré (PTC). Lorsque ce codon apparaît 50 nucléotides avant le codon stop originel, un processus de dégradation de ces transcrits est alors mis en place par la voie NMD (pour Nonsense mediated mRNA Decay) qui va reconnaître les transcrits aberrants grâce à un complexe protéique et ainsi provoquer la dégradation de ceux-ci (Chang, Imam et al. 2007). Un deuxième mécanisme, qui induit également la voie de dégradation NMD, est l'apparition de formes aberrantes d'épissage de type « RUST » (pour Regulated of unproductive splicing and translation). Cette voie permet ainsi la régulation de l'expression de certains gènes et est notamment décrite chez les gènes codants les protéines SR impliquées dans l'épissage alternatif (Lareau, Inada et al. 2007).

Il existe plusieurs types d'épissage alternatif décrits dans la figure 20. Quatre vont permettre de changer la séquence protéique et deux autres permettent de contrôler l'expression et la stabilité des ARNm. Selon les règnes, ces événements ne sont pas présents de la même manière et, alors qu'un type d'événement est majoritairement présent dans un cas, il peut se trouver très rare chez un autre. C'est le cas du phénomène de rétention d'introns très répandu chez les plantes alors qu'il est rare chez les vertébrés et les invertébrés (Keren, Lev-Maor et al. 2010). Ainsi, chez les plantes supérieures dont le génome a été séquencé et pour lesquelles suffisamment de séquences d'ESTs et d'ADNc sont disponibles, la proportion des événements d'épissage alternatif est différente des vertébrés et des invertébrés. Comme le montre la figure 21, les deux phénomènes majoritaires trouvés chez Arabidopsis, le riz et le maïs sont des événements d'épissage alternatif de type « rétention d'intron » et « site

d'épissage alternatif accepteur », représentant à eux deux plus des trois quart des formes trouvées (Barbazuk, Fu et al. 2008).

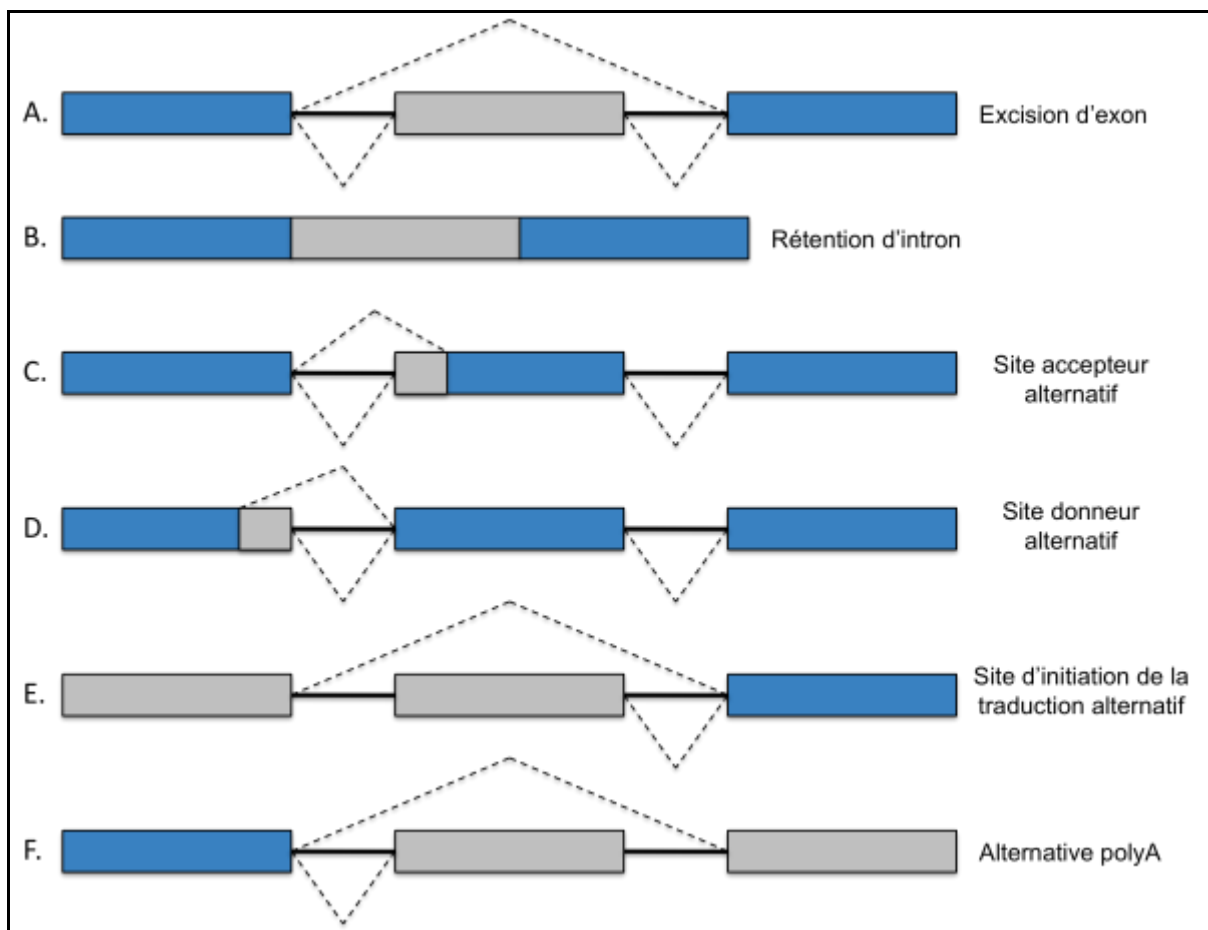


Figure 20 : Description des différents événements possibles d'épissage alternatif. L'épissage alternatif peut être divisé en quatre sous-groupes : A. L'excision d'exon est la forme d'épissage alternatif la plus répandue chez les eucaryotes supérieurs et représente 40 % des événements d'épissage, en revanche ce phénomène est moins répandu voir même très rare chez les eucaryotes inférieurs. B. La rétention d'intron est un mécanisme dans lequel l'intron n'étant pas excisé, est présent dans l'ARNm mature et va donc être traduit. Ce mécanisme peu répandu chez les vertébrés et les invertébrés (5% des événements d'épissage alternatif), est assez répandu chez les plantes, les mousses et les protozoaires. C. Le site accepteur alternatif et D. le site donneur alternatif sont deux mécanismes qui ont lieu lorsque plusieurs sites d'épissage accepteurs ou donneurs sont présents au sein d'un même exon. Ces deux événements représentent respectivement 18,4% et 7,9% des événements d'épissage alternatif chez les eucaryotes supérieurs. De façon assez rare, l'épissage alternatif peut avoir lieu sur des séquences non codantes, dans les régions 5' ou 3' UTR, conduisant ainsi à la possibilité d'obtenir différents sites d'initiation de la traduction (E) ou des sites alternatifs de poly Adénylation (F) (Keren, Lev-Maor et al. 2010).

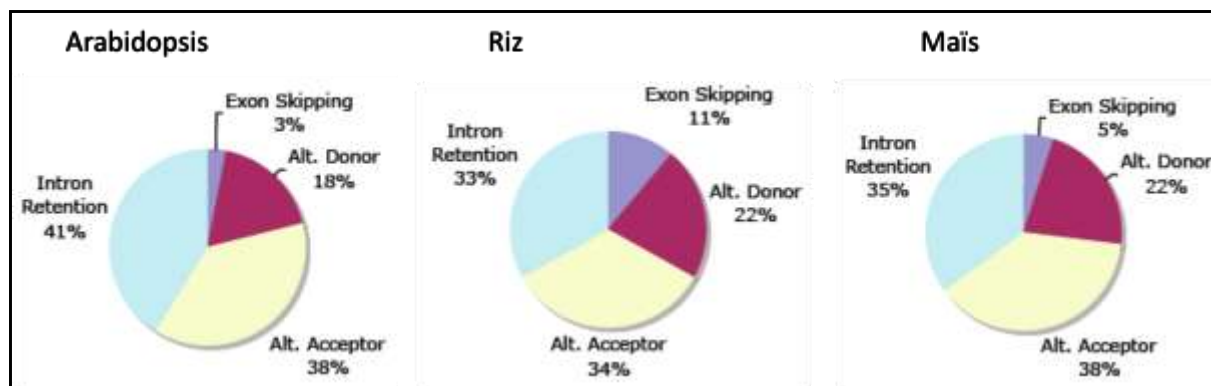


Figure 21 : Proportion des événements d'épissage alternatif chez Arabidopsis, le riz et le maïs. Figure extraite de Barbazuk et al. (2008).

Qu'en est-il des transcrits codant des protéines PPR ? Il faut tout d'abord noter que les gènes *PPR* sont une classe de gènes à part. En effet, la majorité de ces gènes ne contient pas d'introns. Alors qu'une moyenne de 5 introns par gène est observée chez Arabidopsis, un ratio bien inférieur est observé pour la famille PPR. Environ 80 % des gènes *PPR* n'a pas d'introns, 12 % ont un intron et seulement 8 % ont plus d'un intron. Du fait de leur architecture, si de l'épissage existe pour ces gènes sans intron il est donc fort probable que l'épissage alternatif au sein de cette famille soit différent de ce qui est observé globalement chez Arabidopsis.

II.3.1 Epissage alternatif identifié dans les banques d'EST

Cette première étude a été réalisée en utilisant FLAGdb++ (<http://urgv.evry.inra.fr/FLAGdb>) (Samson, Brunaud et al. 2004). Pour chaque gène *PPR* d'Arabidopsis, les ESTs et les ADNc décrits dans cette base de données ont été analysés afin de voir si des formes alternatives d'épissage étaient décrites. La majorité des gènes PPR sont très faiblement exprimés et de ce fait aucun ADNc ou EST n'existe pour une grande partie d'entre eux. Ainsi, sur les 450 gènes codant pour des protéines PPR, seulement 201 ont au moins un ADNc décrit et de fait mon étude s'est donc portée sur ces derniers.

Sur ces 201 gènes *PPR*, 6 gènes (At1g13040, At2g02980, At5g60960, At1g09410, At5g06540, At1g74900) ont des transcrits présentant au moins une isoforme d'épissage identifiée par des ADNc ou des ESTs. 4 autres gènes (At1g61870, At1g62350, At1g74900, At2g02750) présentent également des formes épissées alternatives, cependant, pour ces derniers, il est difficile de conclure car l'EST montrant ce phénomène ne couvre qu'une faible partie du gène.

Cette première observation suggère que 5% des gènes *PPR* caractérisés ont des transcrits alternatifs d'épissage. Cette proportion est certainement sous estimée car parmi les 201 gènes *PPR* étudiés, la majorité n'a seulement qu'un ou deux ADNc décrits ce qui ne permet pas d'avoir une couverture statistiquement satisfaisante de la structure des transcrits. Néanmoins, cette première analyse a permis de confirmer l'existence d'épissage alternatif au sein de plusieurs gènes *PPR* sans intron.

Sur les six gènes présentant différentes isoformes, se trouve le gène *OTP43* qui, par ailleurs, est le premier cas pour lequel des formes alternatives d'épissage ont été mises en évidence au laboratoire. Une seule isoforme est décrite dans les différentes banques, alors que j'ai pu mettre en évidence quatre autres formes au cours de ma thèse. Le cas d'*OTP43* sera présenté dans le chapitre « cas d'*OTP43* ».

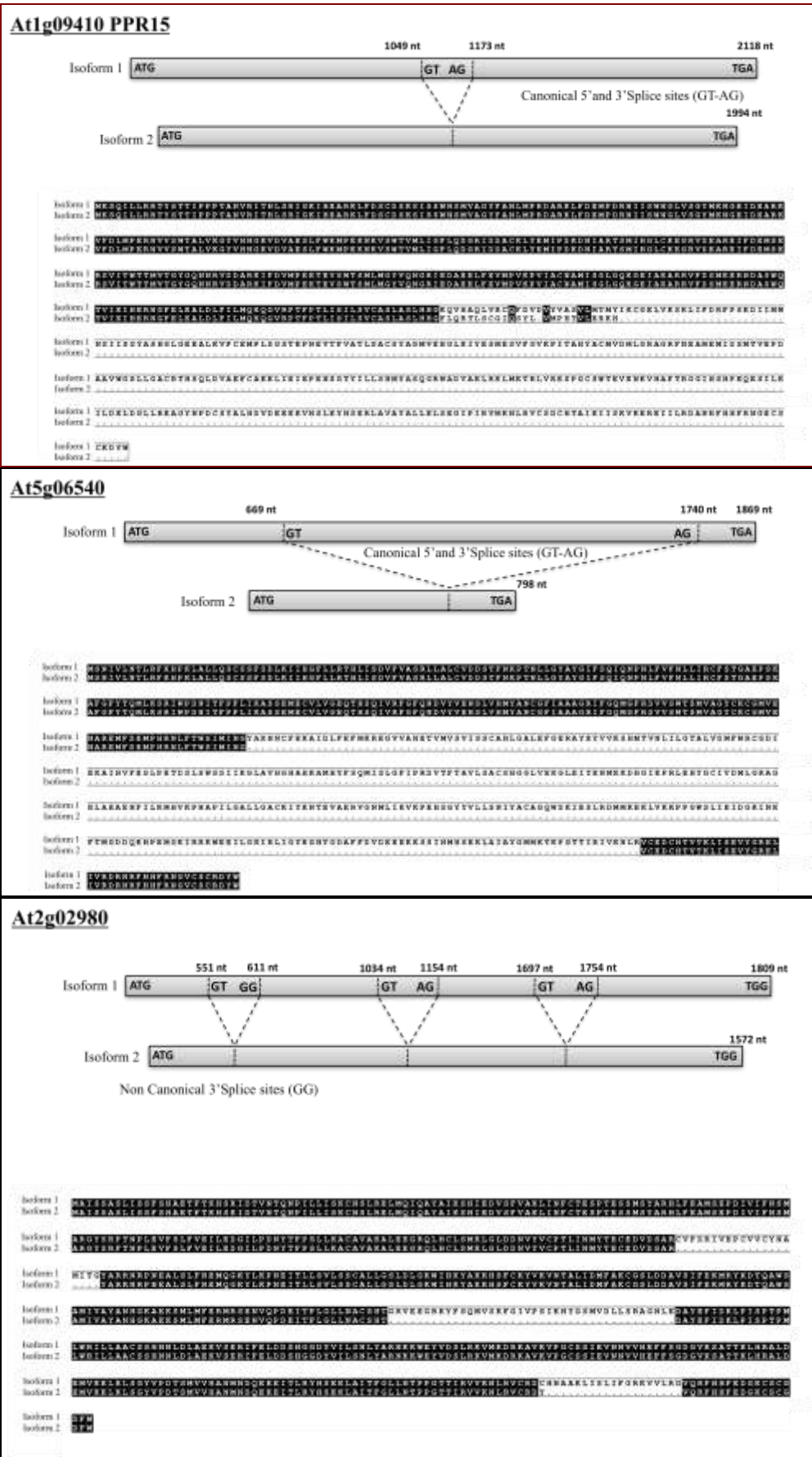
Les différentes isoformes décrites dans les banques de d'ADNc sont présentées dans la figure 22, chaque isoforme après traduction est montrée sous la forme d'un alignement de séquence avec la protéine issue d'un ARNm non épissé. L'épissage alternatif sur ces 6 gènes conduit à 3 différentes classes d'isoformes selon le type d'impact sur l'ARNm final et la synthèse protéique.

La première classe d'isoformes se caractérise par l'apparition d'un codon stop prématuré comme le montre l'exemple AT1g09410, conduisant certainement à une dégradation de cet ARN via la voie NMD. Ce mécanisme, comme cela a été décrit précédemment, permettrait la régulation de l'expression de cette PPR.

La deuxième classe, visiblement la plus répandue, est caractérisée par des isoformes dans lesquelles l'excision d'une partie codante ne change pas la phase de lecture. Si l'isoforme est traduite, les codons d'initiation et d'arrêt de la traduction seront identiques, mais la protéine présentera un certain nombre de délétions de peptides en comparaison avec la forme non épissée. Ces délétions de peptides pourraient créer ainsi des protéines différentes pouvant avoir d'autres fonctions. De tels événements ont été identifiés pour les gènes At5g06540, At2g02980, At5g60960.

La troisième classe est définie par l'apparition potentielle d'un nouveau codon « start » en aval du codon d'initiation de la traduction originel. Deux cas sont présentés : At5g60960 et At1g13040. La majorité des protéines PPR possèdent une séquence d'adressage à leur extrémité N-terminale. Ce changement de codon d'initiation de la traduction pourrait ainsi avoir un impact sur la localisation subcellulaire de ces nouvelles isoformes créées, même s'il n'est pas prouvé que ces différentes isoformes soient effectivement traduites.

Ces différentes formes observées sont, dans la majorité des cas, le résultat d'une excision par l'intermédiaire de sites donneurs et accepteurs d'épissage conventionnels, GT-AG. Cependant trois autres cas sont décrits ici. La présence d'un site accepteur GG pour At2g02980, l'intervention d'un site donneur GC pour la synthèse de l'isoforme 3 de At5g60960, et la présence d'un site donneur CT et accepteur AT non conventionnels pour l'élaboration de l'isoforme 2 de At5g60960.



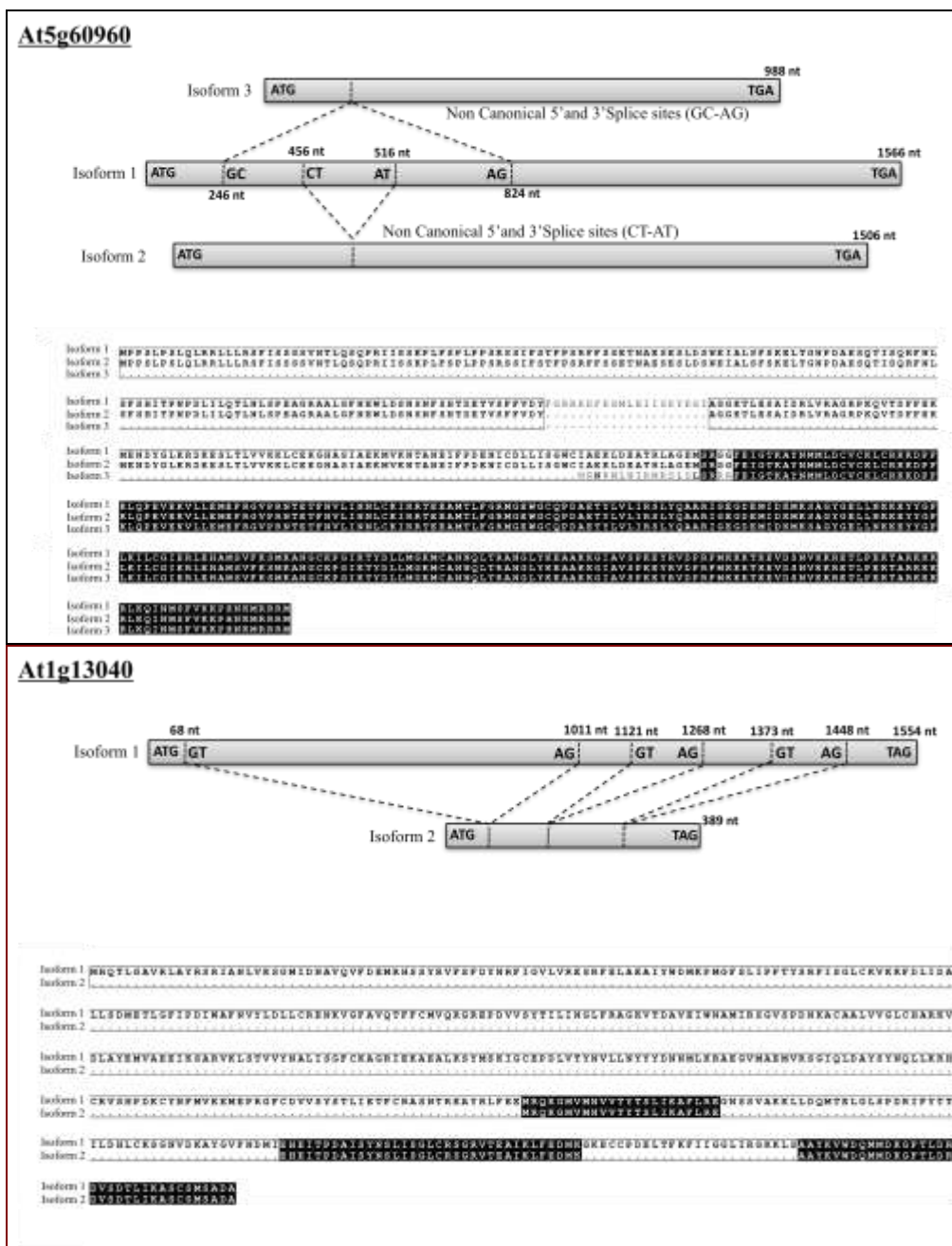


Figure 22 : Description des isoformes de At1g13040, At2g02980, At5g60960, At1g09410, At5g06540. Les isoformes d'ARNm sont représentées dans un premier temps sous la forme d'un schéma sur lequel les sites donneurs et accepteurs d'épissage sont annotés. Les protéines résultantes de ces isoformes sont ensuite représentées dans un alignement avec la forme non épissée traduite sous Multalin.

Depuis cette première analyse, d'autres cas ont été identifiés. C'est le cas de At5g04810 qui, par séquençage haut-débit, a montré une isoforme provenant d'un site accepteur alternatif dans l'exon 4, 8 nucléotides après le site conventionnel (Filichkin, Priest et al. 2010). Ce changement conduit à un codon de terminaison de la transcription prématuré pour lequel le transcrit va certainement être pris en charge par la voie de dégradation NMD. D'autres exemples sont également répertoriés dans la base de données plantgdb (<http://www.plantgdb.org/>). La rétention d'intron pour At4g26800 conduit à un probable changement de codon d'initiation alors que la rétention d'intron dans le cas de At5g02830 et de At2g17670 conduit à un codon stop prématuré.

II.3.2 Analyse préliminaire des 48 PPRs « du laboratoire »

La majorité des gènes *PPR* étant faiblement exprimée, il existe peu d'ADNc décrits. Afin d'avoir une meilleure idée de la proportion de gènes *PPR* présentant des formes alternatives d'épissage, une RT-PCR à partir d'ADNc, suivie d'un clonage, a été effectuée pour un échantillon de gènes *PPR* représentatif. Quarante huit gènes *PPR* ont été sélectionnés par le laboratoire il y a une dizaine d'années afin de mener de nombreuses études fonctionnelles de la famille des protéines PPR. En vue de réaliser ces études, un clonage systématique avait été réalisé en utilisant le système de clonage par recombinaison GATEWAY. Ces gènes *PPR* n'ayant pas d'introns, tous les clonages avaient donc été faits à partir d'ADNg. Nous avons toujours au laboratoire les oligonucléotides permettant d'amplifier ces gènes *PPRs*. Une PCR sur ADNg et ADNc après synthèse à partir d'ARN total de feuilles d'*Arabidopsis* a donc été faite (figure 23).

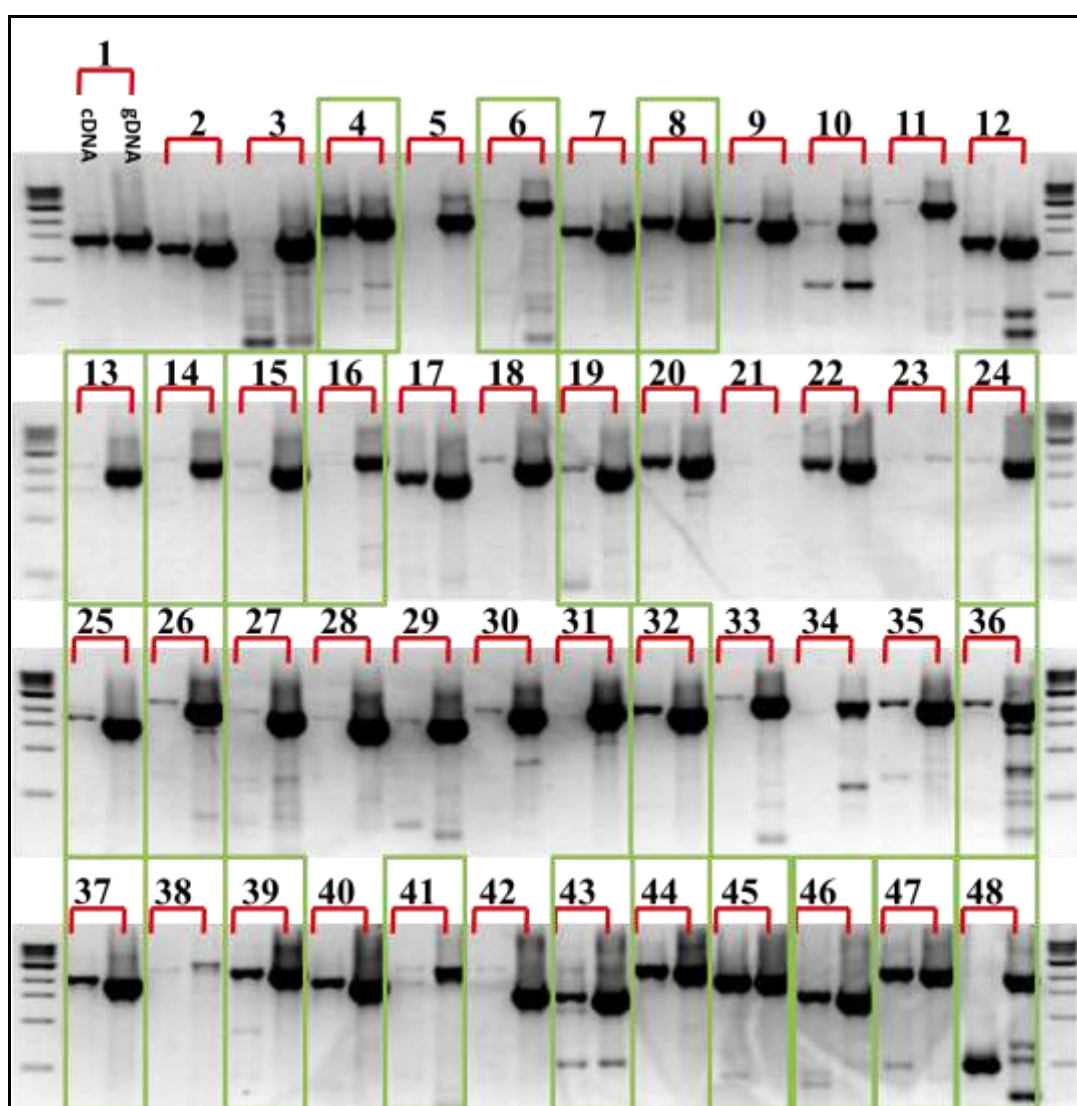


Figure 23 : Analyse par PCR comparative sur ADNc et ADNg de la présence d'épissage alternatif au sein des 48 transcrits PPR du laboratoire. La présence de profils d'amplification est encadrée en vert. Dépôt sur gel 1,2% TAE des produits PCR obtenus à partir d'ADNc (puits de gauche) et d'ADNg (puits de droite). Le marqueur de taille utilisé est le « 1kb DNA ladder » de chez Biolabs (ref : N3232L).

RESULTATS

Cette comparaison montre des profils d'amplification différents selon la matrice utilisée. Ainsi, des bandes supplémentaires ont été observées lorsque la matrice utilisée est de l'ADNc, alors qu'elles sont absentes lorsque la matrice est de l'ADNg. Sur les 48 amplifications, 21 présentent des bandes supplémentaires plus ou moins intenses mais qui ne sont pas retrouvées lorsque l'amplification est réalisée sur de l'ADNg (Figure 23). Ces bandes peuvent correspondre à des bandes parasites dues à une amplification aspécifique ou à des formes d'épissage alternatif. Afin de conclure, les produits PCR ont été clonés dans le vecteur pGEMTeasy, des bactéries DH5 α ont été transformées et sélectionnées sur carbenicilline/X-gal/IPTG. Les colonies blanches positives ont été testées par PCR et les plasmides extraits ont été séquencés en utilisant les amorces M13Fw et M13rev.

Tous les produits n'ont pu être caractérisés entièrement mais l'étude a été menée jusqu'au bout pour un certain nombre. Nous avons pu montrer que les bandes supplémentaires obtenues dans certains puits étaient des amplifications aspécifiques, comme par exemple l'amplification d'une palmitoyl thioesterase (At3g60340) avec les oligonucléotides spécifiques de PPR28, ou d'une Purine transmembranaire AtPUP14 avec les oligonucléotides spécifiques de PPR16. Pour d'autres, comme PPR36, PPR4, PPR13, PPR19, il n'a pas été possible de cloner les produits intermédiaires. En revanche, pour deux autres gènes, PPR24 et PPR47, il a été possible de montrer l'existence d'épissage alternatif.

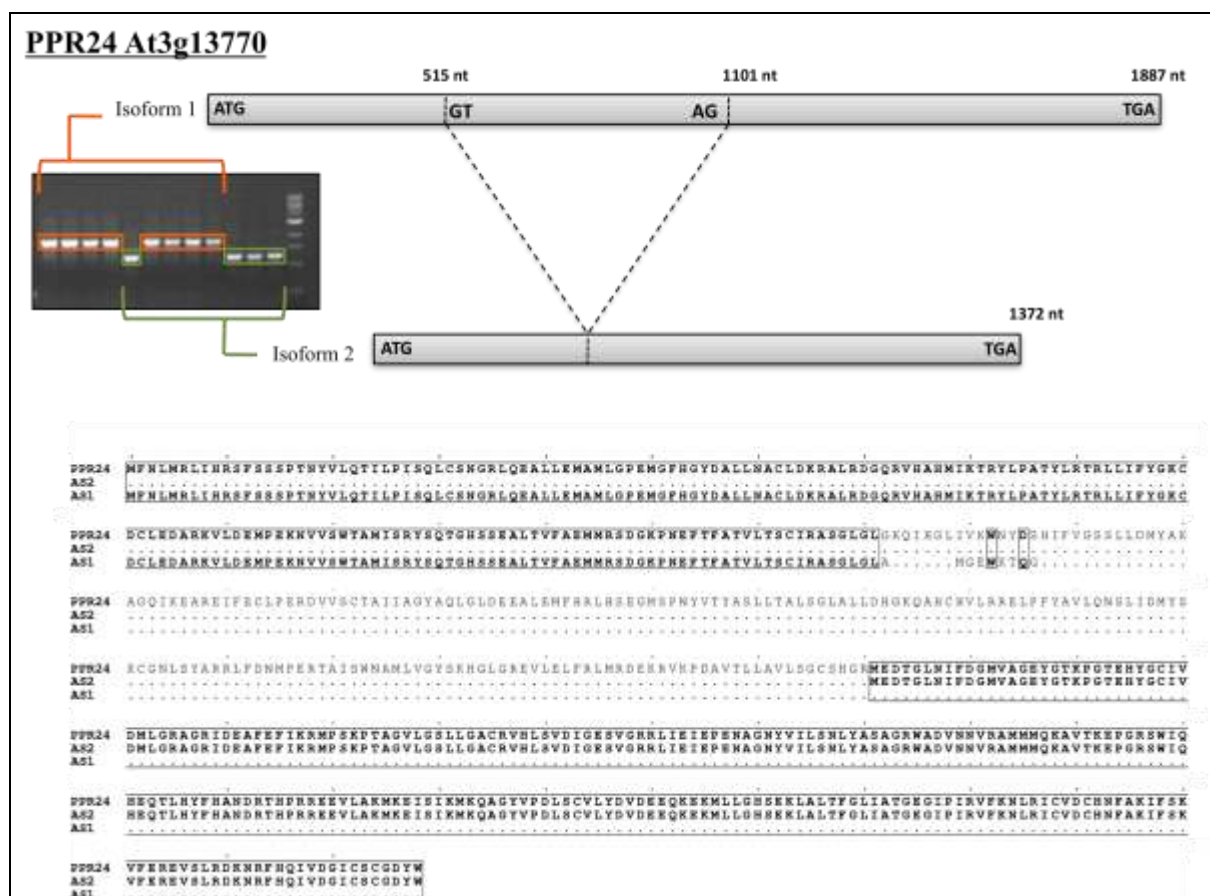


Figure 24 : Isoformes d'épissage alternatif de PPR24 détectées par RT-PCR suivie d'un clonage. Les deux isoformes amplifiées et déposées sur un gel contenant 1% d'agarose dans du TAE (marqueur de taille : « 1kb DNA ladder » de chez Biolabs (ref : N3232L)), sont représentées sous la forme d'un schéma sur lequel les sites donneurs et accepteurs d'épissage sont annotés. Le transcrit isoforme 2 peut coder théoriquement pour 2 protéines comme montré dans l'alignement effectué sous Multalin.

RESULTATS

PPR24 est une protéine PPR de type DYW. Par RT-PCR et dépôt sur gel, deux produits d'amplification sont détectés en utilisant les oligonucléotides « start » et « stop » de PPR24 (figure 24). Ces deux produits correspondent à deux transcrits *ppr24*. Le premier est la forme pleine longueur sans épissage et le deuxième est le résultat d'une excision de 586 nucléotides entre la position 515 et 1101 (Figure XX). Cet épissage crée un changement de phase de lecture avec la création d'un codon de terminaison de la traduction prématuré (PTC) (AS1 sur la figure XX) et le transcrit résultant devrait donc être dégradé via la voie NMD. Cependant un autre codon d'initiation de la traduction pourrait être utilisé aboutissant à une nouvelle protéine identique à PPR24 mais tronquée en N-terminal (AS2 sur la Figure 24). Ceci n'est qu'une hypothèse très théorique qui demande à être confirmée.

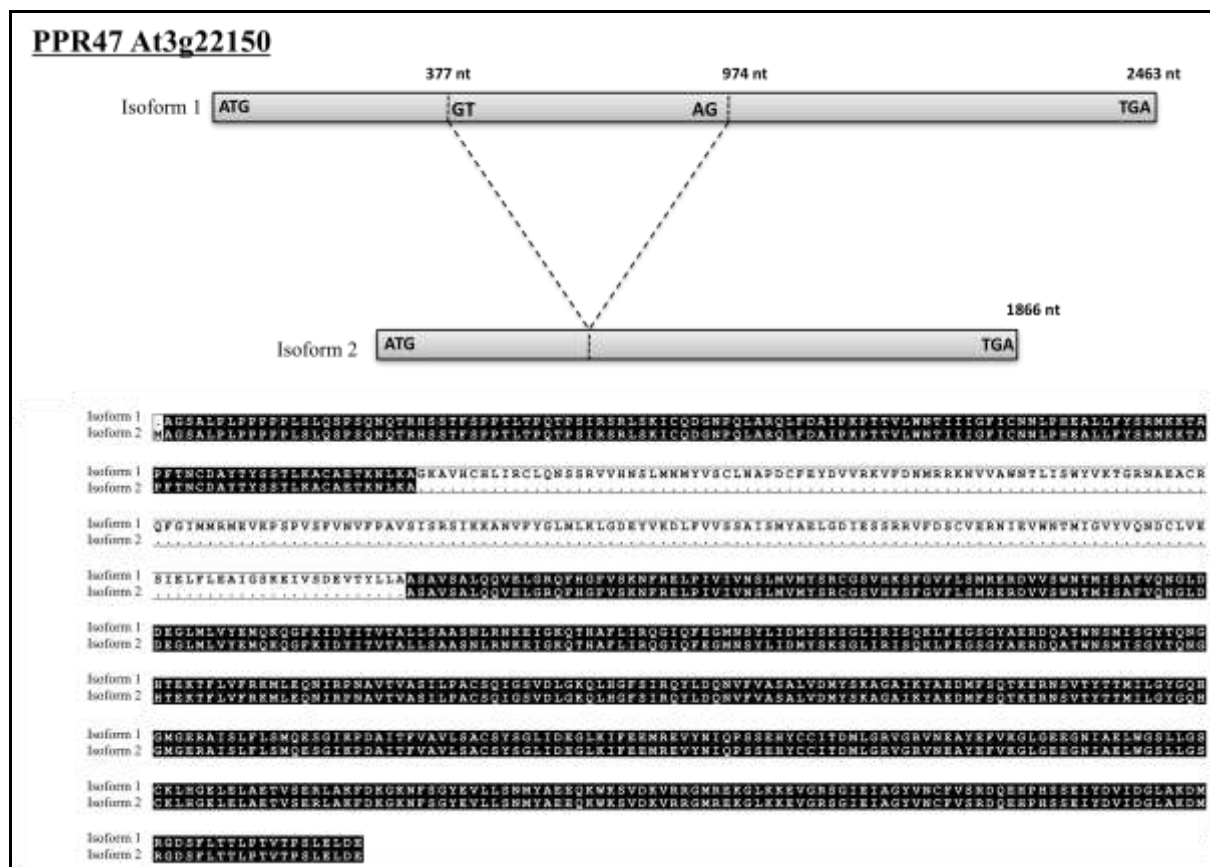


Figure 25 : Isoformes d'épissage alternatif de PPR47 détectées par RT-PCR suivie d'un clonage. Les deux isoformes amplifiées et déposées sur un gel contenant 1% d'agarose dans du TAE (marqueur de taille : « 1kb DNA ladder » de chez Biolabs (ref : N3232L)), sont représentées sous la forme d'un schéma sur lequel les sites donneurs et accepteurs d'épissage sont annotés. Les deux transcrits peuvent coder pour 2 protéines différentes comme montré dans l'alignement effectué sous Multalin.

PPR47 est une protéine PPR de type E+ localisée dans les chloroplastes, dont le mutant d'insertion SALK_040629 ne présente pas de défaut d'édition (Hammani, Okuda et al. 2009). Deux isoformes ont pu être mises en évidence par séquençage. La première correspond à la forme pleine longueur du transcrit *PPR47* et correspond au transcrit majoritaire détecté par RT-PCR en utilisant les oligonucléotides « start » et « stop » de PPR47 (Figure 25). La deuxième isoforme (isoforme 2 sur la Figure 25) correspond à un transcrit épissé, dans lequel 597 nucléotides sont excisés. L'épissage de cette séquence codante ne change pas la phase de lecture et, comme le montre l'alignement de séquence protéique entre les deux isoformes, l'isoforme 2 conduit à une protéine similaire ayant une séquence interne tronquée. Il est possible, que cette isoforme, même si elle est visiblement très peu présente (indétectable par RT-PCR et dépôt sur gel) ait une fonction alternative.

Des isoformes d'épissage alternatif ont également été détectées pour PPR14, cependant comme le montre la Figure 26, il est fort possible que celles-ci n'existent pas réellement et ne soient finalement que le résultat d'un artefact lors de la synthèse du premier brin d'ADNc ou de la PCR. En effet au niveau de chacune des jonctions des sites d'épissage putatifs, on trouve des séquences identiques de 7 et 9 nucléotides (marquées en rouge sur la Figure 26). Lors de la synthèse d'ADNc, il est possible que ces séquences s'hybrident, éliminant ainsi la séquence interne. Il serait nécessaire d'effectuer une synthèse d'ADNc avec de la Superscript III pour laquelle la température de synthèse est de 50°C (au lieu de 42°C pour la Superscript II) afin de vérifier si ces formes sont conservées à cette température plus haute et s'il s'agit, ou non, de formes probables d'épissage alternatif. Faute de temps, cette expérience n'a pas été réalisée. Quoi qu'il en soit ces formes pourraient également résulter d'événements moléculaires liés à la PCR qu'il sera plus difficile d'éliminer simplement.



Figure 26 : Isoformes détectées pour PPR14. En majuscule sont décrites les zones non épissées. En rouge, les nucléotides dont la séquence identique pourrait créer des artefacts lors de la réaction de transcription inverse de synthèse d'ADNc ou au cours de la PCR sur ADNc.

En conclusion, cette analyse n'est pas encore terminée et d'autres clonages doivent être entrepris, cependant, sur les 48 gènes PPR sélectionnés, il a été possible de montrer qu'au moins 4 gènes, *OTP43*, *PPR15*, *PPR24* et *PPR47* présentent des formes alternatives d'épissage. Ces données préliminaires montrent à nouveau que cet événement n'est pas isolé, et doit donc jouer un rôle soit dans la régulation de l'expression, soit dans l'élaboration de nouvelles protéines fonctionnelles.

Faute de temps, malheureusement cette expérience n'a jamais été complètement terminée. Mais, il a tout de même été possible de caractériser deux nouveaux cas et aussi de montrer que des artefacts pouvaient avoir lieu au cours de la synthèse du premier brin d'ADNc.

II.3.3 Cas d'OTP43 et de ses différentes isoformes

II.3.2.1 Identification des isoformes par RT-PCR

OTP43, comme il a été décrit auparavant, est une protéine PPR de type « Pure », dont la fonction dans l'épissage en *trans* de l'intron 1 de *nad1* a été montrée (de Longevialle, Meyer et al. 2007). Cette protéine PPR fait partie des 48 PPR sélectionnées lors de l'étude préliminaire de la détection des formes alternatives d'épissage par RT-PCR. Un intérêt majeur étant porté à celle-ci, une étude plus approfondie a été entreprise. L'annotation TAIR 9 propose l'existence d'au moins une isoforme. En effet, alors qu'une annotation manuelle de ce gène décrit un gène sans intron, l'annotation basée sur le séquençage d'un ADNc propose un gène présentant un intron dans la région 3' du gène. Plusieurs logiciels de prédiction de sites alternatifs d'épissage sont disponibles en ligne. En utilisant le logiciel NNSPLICE version 9.0 (http://www.fruitfly.org/seq_tools/splice.html), deux sites donneurs et deux sites accepteurs sont détectés et permettent de prédire 5 isoformes possibles pour l'ARNm *otp43*. Une analyse similaire en utilisant le logiciel ASSP (<http://www.es.embnet.org/~mwang/assp.html>), permet de prédire deux sites donneurs constitutifs, un site accepteur constitutif et un autre site accepteur alternatif. Ces sites correspondent aux sites décrits en utilisant le premier logiciel. Ces données *in silico* montrent que le transcrit d'*OTP43* pourrait avoir au moins 5 isoformes possibles. Afin de vérifier l'existence de la forme pleine longueur sans intron et de trouver les autres formes prédites par les logiciels de prédiction, une RT-PCR, suivie d'un clonage en pGEMT easy et de séquençage ont été effectués sur les transcrits d'*OTP43*.

Après le séquençage de nombreux plasmides, j'ai pu identifier 4 isoformes correspondants à *OTP43* (figure 27), mais étrangement je n'ai pas réussi à cloner la forme décrite dans la collection d'ADNc sous TAIR9, représentée dans la Figure 27 par le transcrit noté *otp43-2*. Comme le montre le gel (figure 27A), ces différentes formes ne sont pas présentes de manière équivalente. La forme prédominante est la forme pleine longueur sans épissage et les autres formes sont moins représentées.

Parmi les 5 isoformes, quatre ont la même phase de lecture que le transcrit non épissé (*OTP43-1*, *OTP43-2*, *OTP43-3*, *OTP43-4*). Ces isoformes après traduction codent pour des protéines ayant des extrémités N- et C-terminales identiques mais présentant des délétions de motifs plus ou moins importantes selon l'épissage. La dernière isoforme, *OTP43-5*, est l'isoforme la plus courte et est issue de l'épissage entre le premier site donneur et le dernier site accepteur. Ce processus crée un changement de phase de lecture et conduit à l'apparition d'un codon de terminaison de la traduction prématuré. Ce codon se situe à moins de 50 nucléotides de la fin du transcrit, suggérant que ce transcrit n'est éventuellement pas dégradé via la voie NMD et pourrait être traduit.

Il existe cinq isoformes différentes pour *OTP43*, en revanche seule la forme non épissée est « fortement » exprimée. Toutes ces expériences ont été réalisées à partir d'ARN extrait de feuilles. L'épissage alternatif est un événement qui peut se produire selon le stade de développement, selon les organes ou lors de conditions de stress. J'ai voulu rechercher quelle est la proportion des différentes isoformes d'*OTP43* dans différents organes et si l'on pouvait observer une accumulation ou disparition de certaines formes selon les organes examinés.

Afin de commencer à apporter un début de réponse à cette question, j'ai réalisé une RT-PCR en utilisant des ARN provenant de différents organes et les oligonucléotides « start » et « stop » d'*OTP43*. Comme il est montré sur la Figure 28, dans cette expérience, il semble que la proportion des différents isoformes diffère selon l'organe.

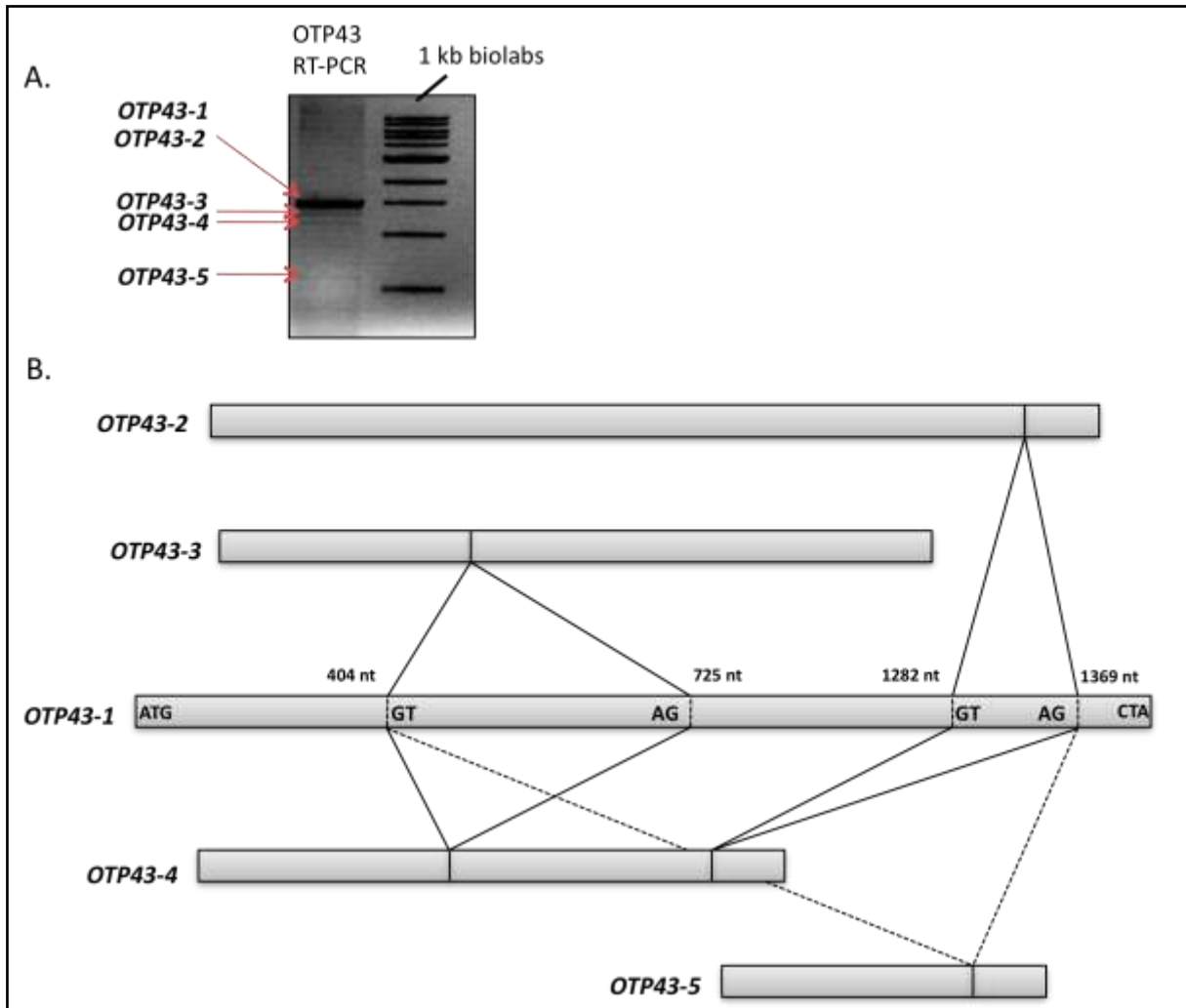


Figure 27 : Isoformes issues d'épissage alternatif du transcrite OTP43. A. PCR sur ADNc avec les amorces « Start » et « Stop » correspondants à OTP43. B. Représentation schématique des 5 isoformes avec la position des sites donneurs et accepteurs d'épissage.

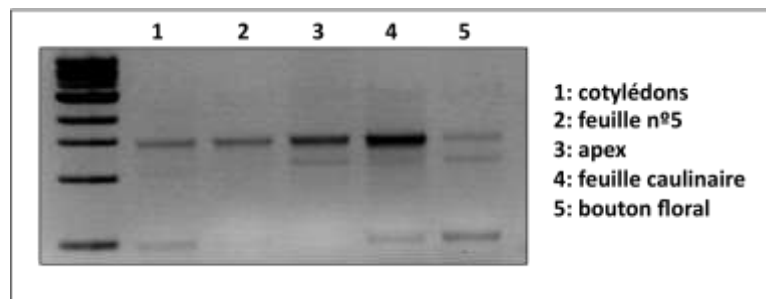


Figure 28 : Caractérisation des différents transcrits *otp43* dans différents organes d'Arabidopsis. RT-PCR à partir d'ARN extraits de cotylédons, de feuilles, d'apex, de feuilles caulinaires ou de boutons floraux. La PCR sur ADNc est faite en présence des amorces « start » et « stop » d'OTP43. 10 μ l sont déposés sur un gel TAE contenant 1,2% d'agarose. Le marqueur de taille est du « 1kb DNA ladder » de chez Biolabs (ref : N3232L).

L'isoforme *OTP43-5* n'est présente que dans les cotylédons, les feuilles caulinaires et les boutons floraux. Alors qu'elle est fortement diminuée dans les feuilles, elle n'est pas détectable au niveau de l'apex. L'isoforme *OTP43-4* ne semble être détectable que dans les

cotylédons et la forme *OTP43-3* est présente dans l'apex, les feuilles caulinaires et les boutons floraux. En parallèle, *OTP43-1* n'est pas exprimée de façon équivalente dans tous les organes, alors qu'elle est fortement exprimée au niveau de l'apex et des feuilles caulinaires, son expression est diminuée dans les boutons floraux. Ainsi la forme *OTP43-5* est la plus abondante dans les boutons floraux alors que dans tous les autres organes testés, c'est la forme pleine longueur sans épissage qui domine. Ces premières données suggèrent qu'il existe une régulation de l'épissage alternatif dans les organes. Malheureusement, une répétition de cette expérience n'a pas été très concluante et m'a conduit à envisager une approche plus quantitative. Afin de démontrer l'existence d'une régulation de l'épissage du transcrit *OTP43*, une autre approche « semi-quantitative » par *northern blot* ou par PCR quantitative serait en effet plus judicieuse. En utilisant des oligonucléotides chevauchant les différentes régions de jonction exon-exon après épissage, j'ai essayé de quantifier les différentes isoformes, mais, malheureusement, certains oligonucléotides n'étant pas compatibles avec la PCR quantitative et la majorité des produits PCR étant amplifiés trop tardivement, je n'ai pas été capable d'obtenir des conditions permettant une quantification. Une approche par *northern blot* devra être envisagée dans le futur pour confirmer ces résultats.

II.3.2.2 Preuves de traduction des différents transcrits :

La détection de ces différents transcrits ne signifie pas que tous vont être traduits et coder ainsi pour différentes protéines. Néanmoins, il faut remarquer que la majorité d'entre eux conservent les mêmes codons d'initiation et de terminaison de la traduction ce qui rend plus probable leur traduction. Il est possible que ces isoformes d'ARNm aient juste un rôle dans la régulation de l'expression d'*OTP43*. Et il est donc nécessaire de vérifier leur existence en tant que « protéine ». Dans les différentes bases de données de protéomiques, aucune donnée concernant *OT43* n'est disponible, ainsi aucun peptide spécifique de cette protéine n'a été séquencé par spectrométrie de masse (ce résultat illustre à nouveau la faible expression des protéines PPR). Cependant, *OTP43* a fait l'objet, comme nous en discuterons ultérieurement, d'une expérience de TAP-TAG dans le but d'identifier ses partenaires protéiques éventuels. Une surexpression dans des cellules d'*Arabidopsis* de la protéine étiquetée en C-terminal, nous a permis de la purifier sur colonne d'affinité et de séquencer par spectrométrie de masse les protéines co-purifiées avec notre protéine d'intérêt. Au cours de ce séquençage, un certain nombre de peptides spécifiques d'*OTP43* ont été observés, mais également des peptides spécifiques de certaines isoformes. La figure 29 montre un alignement protéique entre les différentes isoformes possibles d'*OTP43* ainsi que les zones spécifiques permettant de différencier les différentes isoformes. Aucun peptide permettant d'identifier un épissage de la zone A n'a pu être détecté, en revanche un peptide spécifique de cette zone A permettrait de prouver sa rétention dans certaines isoformes (*OTP43*isoform1 et *OTP43*isoform2). La zone B est présente pour les isoformes 1 et 3 (*OTP43*isoform1 et *OTP43*isoform3) et est absente pour les isoformes 2 et 4 (*OTP43*isoform2 et *OTP43*isoform4). Deux peptides indiqués en rouge au niveau de cette zone dans la figure 29 permettent de distinguer entre ces deux types d'isoformes et la conjugaison entre la présence et l'absence de la zone A et la zone B selon les zones où le séquençage a eu lieu sur le gel, permettent de préciser la présence ou l'absence de telle ou telle isoforme. Ainsi, dans la zone des 35kDa sur le gel SDS-PAGE, le peptide spécifique de la zone B est détecté alors qu'il n'y a aucune présence de la zone A. Seuls les peptides après cette zone sont détectés. Il est possible que cette protéine corresponde à l'isoforme *OTP43*-isoform3 mais avec une dégradation en N-terminal. Dans la zone des 42 kDa, aucun peptide spécifique des zones A et B n'est détecté, en revanche le peptide spécifique de l'absence de la zone B est trouvé. Ceci correspond donc à l'isoforme *OTP43*isoform4 avec un poids moléculaire théorique de 42,1 kDa. Dans la zone des 55 kDa,

nous avons une preuve de la présence de la zone A et de l'absence de la zone B, correspondant donc à l'isoforme OT43isoform2 avec un poids moléculaire théorique de 54,4 kDa. Enfin dans la zone supérieure, nous retrouvons l'isoforme OTP43isoform1 provenant de la traduction du transcrit non épissé. Dans cette expérience, il n'a pas été possible d'identifier l'isoforme la plus courte. En effet, cette isoforme présentant un codon de terminaison de la traduction prématuré, l'étiquette en C-terminal n'est pas traduite. Cette isoforme n'est donc pas étiquetée et ne peut donc être purifiée. En revanche, les 4 autres isoformes, en phase avec l'étiquette, ont pu être mises en évidence au niveau protéique démontrant ainsi une traduction de chacune d'entre elles dans ce système d'expression constitutif. Nous n'avons malheureusement jamais réussi à produire des anticorps dirigés spécifiquement contre OTP43, mais ceux-ci seraient très utiles pour tenter de détecter les différentes isoformes chez une plante sauvage ayant un niveau d'expression normal du transcrit *otp43*. Cependant une alternative à ce système pourrait être de compléter le mutant *otp43* avec une forme étiquetée d'OTP43 sous le contrôle de son propre promoteur et de détecter les éventuelles isoformes par *western blot* en utilisant un anticorps spécifique de l'étiquette.

RESULTATS



Figure 29 : Alignement des différentes isoformes d'OTP43 et définition des peptides permettant de différencier les différentes protéines.

II.3.2.3 Complémentation fonctionnelle avec les isoformes

OTP43 présente donc 5 isoformes, dont 4 ont pu être détectées au niveau protéique, laissant suggérer que chacune a un éventuel rôle fonctionnel. Cette fonction est elle importante pour la régulation de l'expression d'OTP43 ? Les isoformes ont-elles des fonctions additionnelles et complémentaires dans l'épissage en *trans* de l'intron 1 de *nad1* ? Ou bien des fonctions complètement nouvelles ? Finalement, quelle est l'isoforme nécessaire pour l'épissage en *trans* de l'intron 1 de *nad1* ?

Le mutant *otp43* présente un phénotype beaucoup plus marqué que n'importe quel autre mutant de complexe I. Il se peut que l'absence de la sous-unité NAD1, sous-unité importante dans l'ancrage du complexe dans la membrane du complexe ait un impact plus important que n'importe quelle autre sous-unité, mais il est aussi possible que l'absence éventuelle des isoformes d'OTP43 explique une partie du phénotype.

Afin de répondre à certaines de ces questions, et d'essayer d'élucider la fonction de chaque isoforme, j'ai entrepris une complémentation fonctionnelle du mutant *otp43* avec les différentes isoformes d'OTP43, ainsi qu'avec une forme pleine longueur dont les sites accepteurs d'épissage ont été mutés par mutagenèse dirigée de telle sorte qu'aucun événement d'épissage ne puisse avoir lieu.

Quatre constructions ont été réalisées et ont été introduites dans un vecteur binaire pGWB2 dont l'expression est sous le contrôle du promoteur 35S (Figure 30). La première construction contient la forme pleine longueur ayant les sites accepteurs d'épissage mutés (OTP43 AS sites mutés), dans laquelle les nucléotides AG sont modifiés en AA, cette mutation ne change pas l'acide aminé codé par le transcrit. La deuxième contient la forme sauvage d'OTP43 (OTP43-1). La troisième contient l'isoforme dont l'exon 2 est excisé (OTP43-3). La quatrième contient la plus petite isoforme dans laquelle les exons 2, 3 et 4 sont excisés (OTP43-5). Toutes ces constructions, après transfert dans une souche d'*Agrobacterium* C58C1 pMP90, ont été utilisées pour la transformation de plantes hétérozygotes pour la mutation *otp43* (en effet, les plantes homozygotes étant stériles, elles ne peuvent pas être transformées directement). Une sélection des transformants sur kanamycine et un génotypage ont permis d'identifier différentes plantes homozygotes pour la mutation *otp43* et portant les différentes constructions.

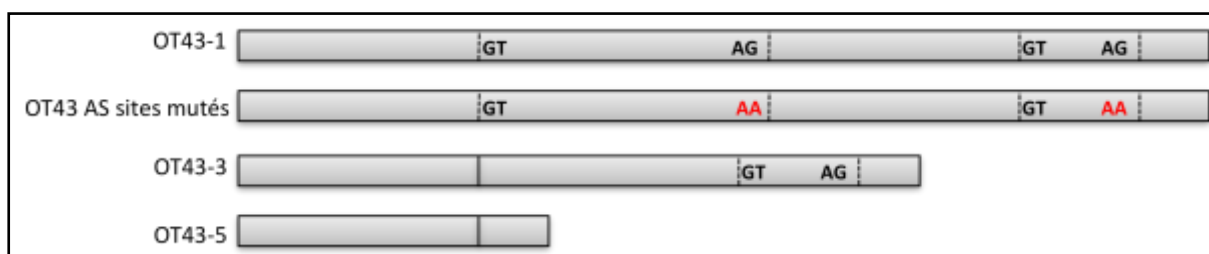


Figure 30 : Schéma décrivant les différentes isoformes utilisées pour la complémentation du mutant *otp43*.

Seule la construction pleine longueur d'OTP43-1 et la forme dans laquelle les sites accepteurs d'épissage ont été mutés (OTP43 AS sites mutés) ont permis d'obtenir des plantes homozygotes pour la mutation *otp43* et présentant un phénotype « sauvage ». Ce résultat montre que ces constructions sont capables de compléter fonctionnellement la mutation *otp43*. Cette première observation permet de montrer que la forme pleine longueur sans épissage permet la complémentation fonctionnelle du mutant. Cependant, une analyse par RT-PCR de l'ARN extrait à partir de deux plantes homozygotes pour l'insertion Sail867F07 complémentées avec la forme mutée a mis en évidence des produits d'amplification très

similaires aux différentes isoformes (gel non présenté). Une identification de ces bandes sera nécessaire pour conclure (ou non) sur la capacité de la forme pleine longueur seule à compléter le phénotype.

En parallèle, aucune plante homozygote pour l'insertion Sail_867F07 n'a été identifiée lors de la sélection des plantes transformées avec les deux autres isoformes (OTP43-3 et OTP43-5). La descendance d'une plante hétérozygote pour l'insertion Sail867F07 contenant le transgène a été semée sur milieu ½ MS contenant 1% de saccharose. Sur les 216 graines semées pour la complémentation par OTP43-5, 149 graines germent normalement, et 56 ont un phénotype de germination retardée similaire au mutant *otp43* (germination deux semaines après le « sauvage »). 48 plantules parmi les 149 germant normalement ont été transférées en terre et génotypées. Aucun homozygote pour l'insertion Sail867F07 n'a été identifié. En revanche il a été possible de sélectionner un certain nombre de mutants portant la construction codant pour l'isoforme OTP43-5 parmi les 56 graines ayant germé tardivement. Ainsi, parmi les 35 plantules génotypées, 19 portent la construction et 3 autres sont seulement homozygotes pour l'insertion Sail867F07. Cependant aucune complémentation du phénotype macroscopique n'a été observée, suggérant que cette isoforme n'a pas de rôle dans l'épissage.



Figure 31 : Photographies de plantes homozygotes pour la mutation *otp43* (Sail867F07) portant la construction OTP43-5. Le phénotype de plantes homozygotes et hétérozygotes pour la mutation est également présenté.

En ce qui concerne la complémentation par l'isoforme OTP43-3, parmi les 161 graines semées issues d'un hétérozygote, 125 germent normalement et 22 seulement ont germé tardivement. 48 plantules sur les 125 ont été génotypées et aucun homozygote n'a été détecté, en revanche, 38 plantes étaient hétérozygotes pour la mutation et 5 de génotype « sauvage » ; elles contenaient toutes le transgène. Sur les 22 plantules ayant germé tardivement, 16 ont été transférées en terre et seulement une plantule présentant le phénotype du mutant *otp43* était homozygote et portait le transgène, 13 étaient des mutants *otp43* et 2 plantes étaient de type sauvage. Il est assez étrange que la proportion d'homozygotes portant le transgène soit si faible alors que dans les 48 plantes de type sauvage génotypées toutes montrent la présence du transgène. Ce résultat très étonnant suggère fortement qu'il y ait eu un problème au cours du génotypage et ces expériences sont donc à refaire.

Il reste beaucoup d'analyses avant de comprendre quel est le rôle de chaque isoforme. Grace aux différentes expériences de complémentation, on peut penser que l'isoforme OT43-5 ne complémente pas le phénotype macroscopique d'*otp43*, suggérant donc que cette protéine n'a pas de fonction dans l'épissage de l'intron1 de *nad1*. Néanmoins, le fait que la descendance d'une seule plante transgénique ait été analysée ne permet pas de conclure définitivement. Pour les isoformes OTP43-3 et OTP43-4 (de façon indirecte), une seule plante homozygote pour l'insertion de l'ADN-T ayant le transgène a été obtenue et le phénotype n'est pas complétement. L'étude d'autres plantes est nécessaire, mais il semblerait que ces deux isoformes ne sont pas non plus impliquées dans l'épissage. Des confirmations sont nécessaires afin de déterminer quelle isoforme est responsable de l'épissage de l'intron 1 de *nad1*. La complémentation avec la forme pleine longueur mutée sur les sites accepteurs d'épissage complémente le phénotype, mais il semblerait que l'épissage alternatif ait encore lieu. Rien ne permet pour le moment de savoir laquelle des ces deux isoformes OTP43-1 et OTP43-2 est impliquée dans ce processus. Et il reste donc à découvrir quelle est la fonction de chaque isoforme et si elles en ont une.

Mise à part OTP43-5, pour lequel il n'a pas été possible de vérifier sa présence au niveau protéique, toutes les autres isoformes sont traduites. Ces différentes isoformes ne sont donc pas prises en charge majoritairement par une voie de dégradation de l'ARN et leur rôle ne semble pas être dans la régulation de l'expression d'OTP43. De plus, il faut noter qu'au moins pour les formes OTP43-3, -4 et -5, la surexpression n'a aucun impact phénotypique lorsque ces isoformes sont exprimées dans un fond génétique sauvage ou hétérozygote.

Il n'est pas encore possible de définir un rôle pour chaque isoforme d'OTP43, mais il est cependant possible d'émettre quelques hypothèses. En vue du phénotype plus marqué du mutant *otp43* comparé à d'autres mutants de complexe I, il est possible que certaines isoformes aient un rôle important autre que dans l'épissage de l'intron 1 de *nad1*, comme par exemple dans la maturation des graines ou dans la germination. Cependant il n'a pas encore été montré qu'une seule forme est nécessaire pour l'épissage et il est également possible que l'isoforme OTP43-2 ait un rôle complémentaire dans la reconnaissance des deux parties de l'intron1.

Tous ces résultats de complémentation sont préliminaires et une dernière vérification qui n'a malheureusement pas été encore faite permettra de pouvoir conclure sur le rôle de certaines isoformes. En effet l'insertion de l'ADN-T du mutant *otp43* se situe 130 pb avant le codon de terminaison de la traduction, et est donc dans la dernière séquence de 87 nucléotides qui peut être excisée. Si cette insertion n'a pas d'impact sur l'épissage, alors 3 isoformes peuvent exister dans le mutant, OTP43-2, OTP43-4 et OTP43-5. Cette hypothèse pourrait expliquer le fait que l'on observe des bandes chez les plantes complémentées avec la forme mutée.

II.4 Identification de partenaires des protéines PPR

A ce jour, environ 10% des protéines PPR d'Arabidopsis ont été plus ou moins caractérisées. Leurs fonctions dans l'expression, la maturation des ARN ou la traduction sont très spécifiques et ne nécessitent visiblement, dans la majorité des cas, aucun autre facteur. Cependant, la plupart des caractérisations a été faite par une approche de génétique inverse et très peu de données de type « interaction protéique » ont été obtenues. Certains résultats permettent néanmoins de proposer que, probablement en fonction de leur rôle, certaines protéines PPR sont associées à d'autres facteurs ou font partie de complexes protéiques. C'est le cas de PTAC2 qui est associé au complexe actif transcritptionnel dans les chloroplastes (Pfalz, Liere et al. 2006), de GRP23 qui interagit avec la sous-unité III de l'ARN polymérase II (Ding, Liu et al. 2006), de PPR40 visiblement associée au complexe III mitochondrial (Zsigmond, Rigo et al. 2008) et encore de PPR5 associée à un complexe ribonucléoprotéique (Beick, Schmitz-Linneweber et al. 2008). Ces résultats d'interactions montrent que les protéines PPR peuvent en plus de leur reconnaissance à l'ARN, établir des interactions protéiques ou être associées à des complexes. Pour ces protéines, mise à part PPR5, la fonction moléculaire n'est pas encore élucidée. Qu'en est-il des autres protéines PPR impliquées dans l'épissage ou dans l'édition des ARNm ? Il a été montré par des expériences de RIP-CHIP que certaines protéines PPR, telles que PPR4 et PPR5 sont liées directement à des séquences spécifiques au sein de l'intron, permettant certainement la mise en place de la structure secondaire optimale de l'intron pour les deux étapes de trans-esterification. Contrairement à PPR5, il semblerait que PPR4 ne soit pas lié au complexe d'épissage mais aucune donnée ne permet de dire qu'aucune interaction n'existe.

Dans l'optique de mieux comprendre comment les protéines PPR fonctionnent, il est important de connaître les interactions de type protéine-protéine pour chacune. Ces informations permettront de mieux comprendre les mécanismes de régulation de l'expression du génome des organites et de découvrir de nouveaux facteurs impliqués.

Ce type d'information est malheureusement difficile à obtenir pour le moment, mais des projets d'interaction protéine-protéine comme le test matriciel systématique de plus de 10 000 protéines d'Arabidopsis en système double hybride chez la levure effectué actuellement dans le laboratoire de Marc Vidal à Boston vont permettre de révéler un grand nombre d'interactions. Ces données ne sont pas encore rendues publiques, mais par l'intermédiaire d'une collaboration du laboratoire de Claire Lurin avec ce grand projet NSF, nous avons eu accès aux premiers résultats. Parmi les protéines PPR faisant partie de la première matrice d'étude, certaines interagissent avec une ou plusieurs protéines autres que des protéines PPR ouvrant ainsi un nouveau champ d'étude. Malheureusement toutes ces interactions observées ne concernent pas nos candidats ou des protéines PPR déjà identifiées, en revanche ce seront d'excellents candidats pour des études de caractérisation à venir.

Pour nos protéines candidates à la fois impliquées dans l'épissage ou dans l'édition, il était donc important de trouver une méthode nous permettant d'identifier potentiellement des candidats d'interaction. Que ces protéines fassent parties d'un important complexe protéique ou juste présente une simple interaction binaire avec une unique autre protéine, une des méthodes les plus appropriées est une approche de purification des protéines d'intérêt marquées par TAP-TAG. C'est donc dans ce cadre que nous avons établie une collaboration avec Geert De Jaeger au VIB à Gent (Belgique) qui a mis en place une plateforme de purification de protéines marquées après expression dans des cultures cellulaires d'Arabidopsis.

II.4.1 Méthode utilisée : TAP-TAG

La technique de « Tandem Affinity Purification » de protéines étiquetées (TAP-TAG) est une technique qui est de plus en plus utilisée dans la recherche de partenaires mais surtout dans l'identification de complexes protéiques car les interactions observées ne sont pas seulement binaires comme dans la plupart des autres techniques d'interactions protéine-protéine utilisées. Grace aux adaptations qui ont été apportées dans l'équipe de Geert De Jaeger, cette méthode a été optimisée dans des cultures cellulaires d'*Arabidopsis* et permet d'obtenir une purification avec un minimum de protéines non-spécifiques grâce à l'utilisation de deux purifications successives différentes avec des étiquettes adaptées (Van Leene, Witters et al. 2008). Le vecteur GATEWAY d'expression utilisé, décrit dans le papier de Van Leene *et al.*, permet une fusion en C-terminal de la proie avec la construction GS-TAG qui comprend deux domaines « IgG-binding » de la protéine G et un peptide de liaison à la streptavidine SBP séparés par un site de clivage à la protéase de type TEV. Une première purification sur colonne en présence de streptavidine permet d'éliminer un grand nombre de protéines n'interagissant pas avec la proie, cette première étiquette est ensuite éliminée par la protéase TEV et une seconde purification sur colonne en présence d'un anticorps IgG permet ainsi une meilleure co-purification des partenaires.

Quatre protéines PPR, deux impliquées dans l'épissage, OTP43 et OTP51, et deux autres dans l'édition, CLB19 et DYW1, ont été sélectionnées et clonées dans un vecteur d'entrée pDONR221 sans codon stop afin de permettre une fusion en C-terminal (Figure 32). Parmi les quatre, seul le gène codant pour OTP51 n'a pas pu être cloné malgré trois tentatives. Les expériences de TAP-TAG se sont donc limitées aux trois autres. La réaction de recombinaison LR a été faite à Gent ainsi que les transformations de cellules d'*Arabidopsis* et le reste des purifications et analyses.

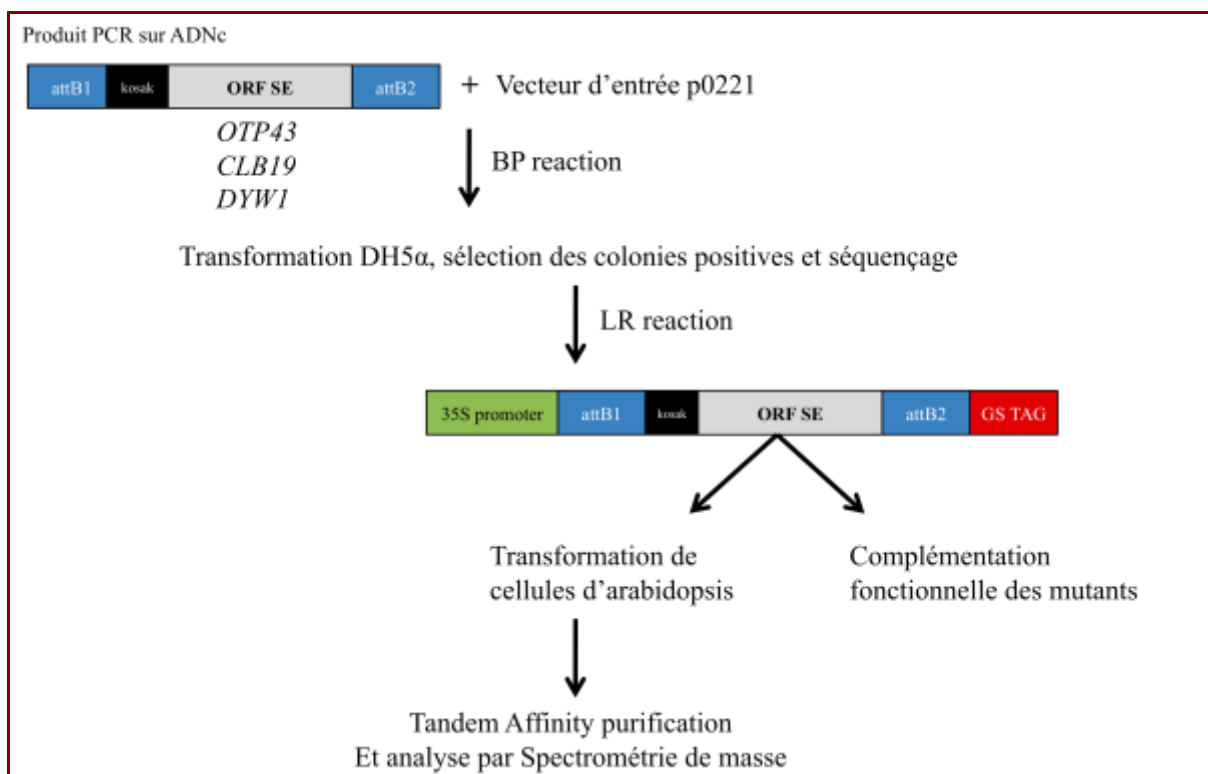


Figure 32 : Résumé du protocole adopté pour les expériences de TAP-TAG réalisées pour OTP43, CLB19 et DYW1.

Comme décrit dans l'article de Van Leene, Witters *et al*, après transformation, les cellules transformées et sélectionnées sur antibiotiques sont mises en culture et une purification protéique est alors faite sur l'extrait cellulaire total. Après deux purifications successives, une élution permet de récupérer l'éluât comprenant la proie avec ses éventuels partenaires associés. Cet éluât est alors déposé sur un gel SDS-PAGE afin de dissocier toutes les protéines présentes. Puis le gel est fractionné en 48 bandes qui sont envoyées pour le séquençage par spectrométrie de masse. Deux extractions indépendantes et analyses ont été faites pour chaque construction.

En parallèle, afin d'évaluer si les constructions chimériques étaient encore fonctionnelles, j'ai effectué une complémentation des mutants *otp43*, *clb19* et *dyw1*. Ainsi, j'ai pu montrer que la fusion des protéines OTP43 et CLB19 en C-terminal avec les deux étiquettes ne modifie pas leur fonction car ces deux constructions complètent le phénotype macroscopique des deux mutants. En revanche, il semblerait qu'aucune complémentation fonctionnelle n'a lieu dans le cas de DYW1 car tous les homozygotes contenant la construction chimérique présentent toujours le même défaut moléculaire.

II.4.2 Partenaires d'OTP43

Cette expérience de TAP-TAG réalisée sur OTP43 nous a permis dans un premier temps de vérifier la présence des isoformes d'épissage alternatif décrits précédemment. Sur les cinq isoformes décrites, les données de séquençage par spectrométrie de masse permettent d'en identifier trois de façon certaine et une par déduction. Seule la forme la plus courte OTP43-5 n'a pas pu être détectée (cf Chapitre II.3.2.2).

Après la double purification sur colonne décrite dans le paragraphe sur la méthode du TAP-TAG, notre protéine étiquetée a été purifiée par élution avec ses partenaires et le tout est déposé sur un gel SDS-PAGE (figure 33). Les résultats de séquençage de chaque bande découpée obtenus sont alors filtrés en fonction du nombre de peptides trouvés pour chaque protéine, du pourcentage de recouvrement, un score de confiance est attribué et toutes les protéines ne l'atteignant pas sont éliminées de la liste.

Enfin, les protéines contaminantes trouvées très régulièrement dans ce type d'extraction sont éliminées de la liste. Toutefois, si certaines d'entre elles sont très abondantes dans une expérience, elles sont conservées. Enfin, une comparaison est faite entre les résultats obtenus entre les deux extraits.

Pour OTP43, une première analyse a révélé un résultat très surprenant. Alors que dans tous les gels de TAP-TAG analysés, la bande majoritaire observée correspond toujours à la protéine étiquetée surexprimée, dans notre expérience, cette bande ne correspond pas à OTP43 mais à plusieurs protéines de type chaperone dites de « choc thermique » ou « Heat Shock protein » (Figure 33). Le plus grand nombre de peptides spécifiques d'OTP43 a été repéré dans les fragments des bandes F3/F4/F6/F15 ainsi que H3/H4/H6 et H15. D'autres peptides sont identifiés ailleurs dans le gel dont certains sont spécifiques des différents isoformes d'OTP43. En revanche, aucun n'a été détecté dans la zone F8 et H8, correspondant à la bande majoritaire. Cette zone ne comprend que majoritairement des peptides spécifiques des protéines HSP60 (AT3G23990), HSP60-2 (AT2G33210) et HSP60-3A (AT3G13860). Dans la liste finale, ces trois protéines sont trouvées en plus d'une autre protéine, APS1 (AT3G22890). HSP60, HSP60-2 et APS1 sont considérées comme des partenaires faux positifs, cependant leur surabondance dans les deux extraits comparée à leur présence générale dans d'autres extraits suggère une réelle interaction avec OTP43. Les trois protéines HSP ont toutes été détectées dans les mitochondries par spectrométrie de masse (données sous

SUBA), montrant une fois de plus que les résultats ne sont certainement pas des artefacts. Cependant leur lien avec OTP43 n'est pas très clair à première vue.

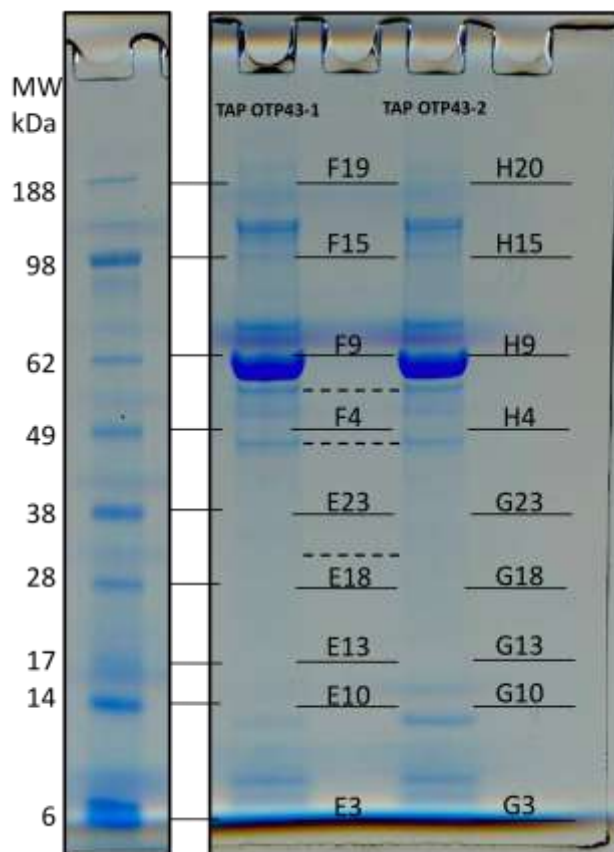


Figure 33 : SDS-PAGE des extraits de TAP-TAG sur OTP43. Les protéines des deux extraits, TAP OTP43-1 et TAP OTP43-2, ont été séparées en fonction de leur masse moléculaire. Chaque piste de gel, découpée en 48 fragments (E1 à F24 pour TAP OTP43-1 et G1 à H24 pour TAP OTP43-2) a été analysée par spectrométrie de masse.

Les protéines de choc thermique sont directement impliquées dans la mise en place de la structure tertiaire et quaternaire des protéines nouvellement synthétisées dans les mitochondries mais elles ont aussi un rôle dans l'import de protéines cytoplasmiques dans les mitochondries et dans leur structure. Par cette fonction primordiale dans le fonctionnement et l'activation de celles-ci, ces protéines sont responsables de la maintenance de la conformation du pool protéique face à des réponses au stress.

Il est fort possible qu'OTP43 et ses différents isoformes nécessitent une conformation bien précise une fois intégrées dans les mitochondries afin d'être pleinement fonctionnelles. Les protéines HSP60 sont généralement très présentes dans les expériences de TAP-TAG du fait de leur fonction dans la conformation protéique et dans la détoxification, mais c'est la première fois qu'elles sont autant représentées. C'est également la première fois qu'une expérience de ce type est réalisée sur des protéines mitochondriales au VIB. Et il est donc possible que cette surreprésentation de ces HSP60 dans les extraits soit tout à fait liée à la surexpression d'une protéine importée dans les mitochondries et non spécifiquement à celle d'OTP43. Il serait intéressant lors de futures expériences de TAP-TAG sur des protéines mitochondriales de voir quelle est la proportion de HSP60 dans les extraits.

L'autre protéine trouvée associée à OTP43 est APS1 (pour ATP sulfurylase 3), une enzyme impliquée dans la voie d'assimilation du sulfate. Cette protéine, contrairement aux

HSP60, est décrite dans les chloroplastes et des données de spectrométrie de masse le confirment sous SUBA. Cette protéine est elle-aussi considérée comme faux positif, et il est fort possible qu'il n'existe pas une réelle interaction avec OTP43.

En conclusion de ces analyses, peu de protéines passant les différents critères d'analyse sont décrites associées à OTP43, et les seules trouvées n'ont probablement pas de lien direct avec la fonction d'épissage dans les mitochondries. OTP43 n'est probablement pas associé à un complexe protéique et notre approche n'a pas permis de montrer l'importance d'un autre facteur dans l'épissage de l'intron 1 de *nad1*.

II.4.3 Partenaires de DYW1

Une des actions des protéines PPR qui a été discutée dans l'introduction est l'édition des ARN des chloroplastes et des mitochondries. Le mécanisme enzymatique d'édition n'a pas encore été élucidé dans les organites, mais à partir des données sur les facteurs impliqués identifiés à ce jour, un modèle peut être proposé.

Seules les protéines PPR de type E, E+ et DYW ont un rôle dans l'édition. Dans ce modèle, les motifs PPR permettent une reconnaissance spécifique de l'ARNm cible et c'est grâce au motif DYW que l'édition a lieu du fait de la présence d'un motif similaire à la cytidine désaminase (Salone, Rudinger et al. 2007), motif qui aurait donc la fonction de désaminer une Cytidine en Uridine. Les protéines PPR E et E+ impliquées dans l'édition n'ont pas ce motif DYW en C-terminal et semblent donc nécessiter un facteur enzymatique additionnel, par exemple un partenaire de type DYW, pour que la réaction catalytique puisse avoir lieu. Une des protéines de ce type que nous avons longtemps supposée comme étant le partenaire de toutes les protéines E et E+ impliquées dans l'édition est DYW1 (At1g47580). Cette protéine PPR, assez spéciale est composée de très peu de motifs PPR et d'un motif DYW très conservé mais sans motifs E et E+.

C'est donc dans l'optique de confirmer notre modèle, que nous avons envisagé des expériences de TAP-TAG sur DYW1. Malheureusement une contamination sur les cultures a retardé la suite des expériences et nous n'avons toujours pas de résultats permettant de montrer quels sont les partenaires de cette protéine.

Cependant, des résultats récents obtenus dans le cadre de la thèse de Clément Boussardon, semblent indiquer que cette protéine serait impliquée uniquement dans l'édition d'un site chloroplastique, *ndhD-1*. Une autre protéine PPR, CRR4 est un facteur essentiel de l'édition de ce site (Okuda, Nakamura et al. 2006). Et il apparaît donc que deux protéines PPR sont essentielles à l'édition de ce site. CRR4 est une protéine PPR de type E, n'ayant pas le motif E+ et de motif DYW décrit comme portant l'activité catalytique. Ainsi, pour le site *ndhD-1*, un modèle peut être dessiné. CRR4 permettrait la reconnaissance spécifique du site *ndhD-1* et DYW1 se liant à CRR4 apporterait l'activité catalytique nécessaire à l'édition. L'interaction est en cours de validation par des techniques de split-luciférase et de split-YFP, mais les futures données de TAP-TAG permettront également de confirmer ce nouveau modèle.

II.4.4 Partenaires de CLB19

Les nouvelles données sur DYW1 modifient en grande partie notre premier modèle, mais le modèle d'association entre une protéine E ou E+ sans motif DYW avec une protéine DYW est toujours d'actualité. Il fallait donc trouver quelles sont les protéines DYW susceptibles de se lier avec les autres PPR E et E+ impliquées dans l'édition afin de redessiner notre modèle sur l'édition. Ainsi une stratégie de type « TAP-TAG » a été privilégiée, et une protéine PPR

RESULTATS

de type E+ a été utilisée comme proie. CLB19, première PPR caractérisée au laboratoire comme ayant un rôle dans l'édition et impliquée dans l'édition de deux sites chloroplastiques, ClpP et RpoA (Chateigner-Boutin, Ramos-Vega et al. 2008), était donc un bon candidat. Les résultats de purification par « TAP-TAG » des partenaires de CLB19 ont permis d'identifier sans aucune ambiguïté deux protéines associées à CLB19. Ces deux protéines sont des protéines PPR. La première est une PPR de type « Pure », AT3G49240 renommée OTP100, pour laquelle des mutants ont été décrits comme embryo-léthaux. La deuxième, de façon très intéressante, est une PPR de type DYW, AT2G15690 renommée DYW2 par la suite. La prédiction de la localisation subcellulaire n'est pas très claire concernant ces deux protéines, en revanche des données de spectrométrie de masse pour OTP100 et des données de localisation en fusion GFP effectuées par Sandra Tanz à Perth pour DYW2 permettent de proposer leur double adressage dans les chloroplastes et les mitochondries. De plus, une forte homologie de séquence en N-terminal de OTP100 avec GRP23 (figure 34) suggère une éventuelle localisation dans le noyau.

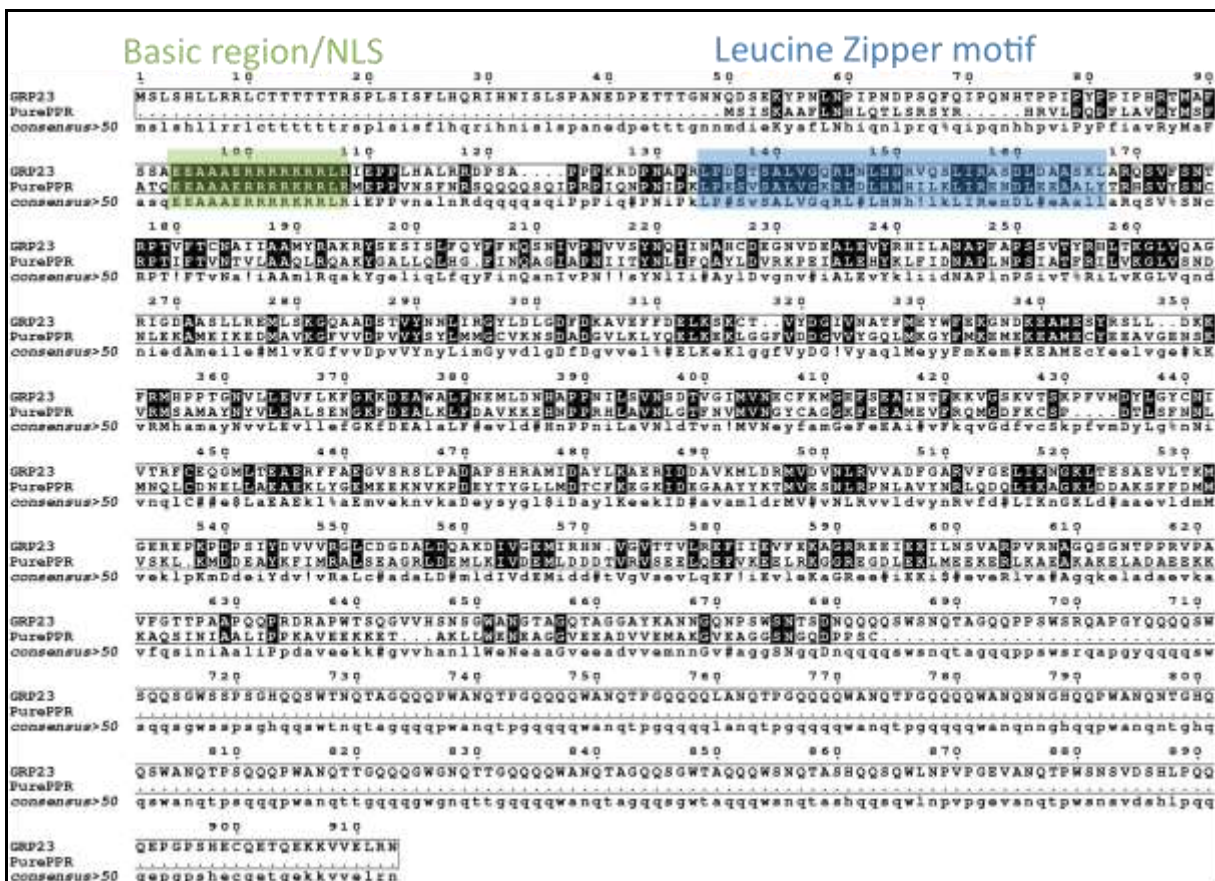


Figure 34 : Alignement grâce au logiciel Multalin des protéines GRP23 et OTP100 (« Pure »PPR). La région NLS est surlignée en vert et la région de liaison à l'ADN (Leucine Zipper motif) en bleu.

Au niveau de l'expression, les deux gènes, *OTP100* et *DYW2* sont relativement co-régulés comme le montre la figure 35, suggérant que les deux protéines codées par ces deux gènes pourraient avoir une fonction similaire au sein des organites.

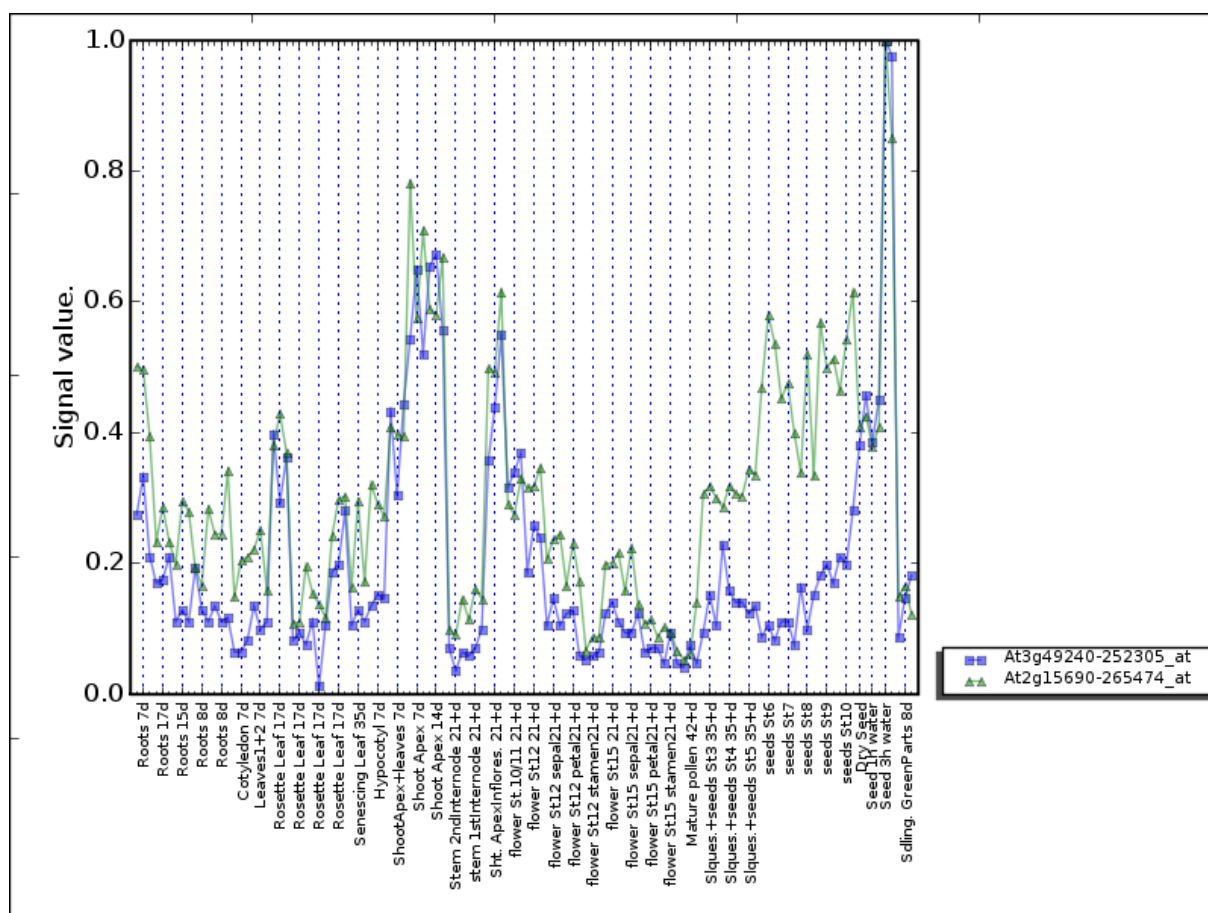


Figure 35 : Analyse de l'expression des gènes *OTP100* et *DYW2* grâce au logiciel « Genecat ». L'expression a été normalisée de 0 à 1.

Afin de tester si ces deux protéines sont associées à CLB19, et potentiellement impliquées dans l'édition des deux sites dans les transcrits *rpoA* et *clpP*, une approche par génétique inverse a été entreprise. Des mutants d'insertion ADN-T ont été commandés, semés et génotypés. Pour le moment, aucun homozygote pour aucune des lignées n'a pu être obtenu. Ces deux protéines semblent donc essentielles pour la survie de l'embryon (ou éventuellement d'un des gamètes) ce qui n'est pas compatible avec une fonction unique dans l'édition des mêmes sites que CLB19 ; en effet le mutant *clb19*, quoique très affecté, est viable. Une première analyse des siliques prélevées sur des plantes hétérozygotes montre deux types de populations de graines : des graines de type « sauvage » et des graines blanches et plus petites. Plusieurs essais de germination en condition *in vitro* n'ont pas permis d'obtenir la germination d'homozygotes et des expériences de « sauvetage d'embryons » sont en cours.

Afin de confirmer les interactions, des expériences de test d'interaction protéine-protéine par le biais de la technique de split-luciférase et de split-YFP sont également en cours. Ces expériences nous permettront ainsi de tester les interactions binaires.

En addition, Clément Boussardon met au point actuellement un essai d'édition *in vitro* en utilisant les extraits purifiés par TAP-TAG sur CLB19 et, en parallèle, il tente de reconstituer un complexe minimal par expression en système bactérien.

Grace à ces nouvelles données, un nouveau modèle a été élaboré (Figure 36). La protéine *DYW2*, est une PPR d'un type *DYW* assez particulier, assez proche de *DYW1*. En effet, contrairement à la plupart des protéines de type *DYW*, mais de façon similaire à *DYW1*, elle

n'a pas de motifs E et E+ et est composée seulement de quelques motifs PPR en plus de son motif DYW très conservé (figure 37). Il est donc proposé que cette protéine doublement adressée se lie à certaines protéines E ou E+ impliquées dans l'édition au niveau chloroplastique et mitochondrial (Figure 36). Ainsi, ces protéines E/E+ reconnaîtraient de façon spécifique les molécules d'ARN à éditer et DYW2 apporterait l'activité catalytique nécessaire pour l'édition dans certains cas. Ce nouveau modèle est assez similaire au précédent mais une spécificité de l'interaction « protéine PPR-E »-« protéine DYW » semblerait exister. Par ailleurs, le rôle de la PPR « Pure » OTP100 dans ce complexe reste à être élucidé.

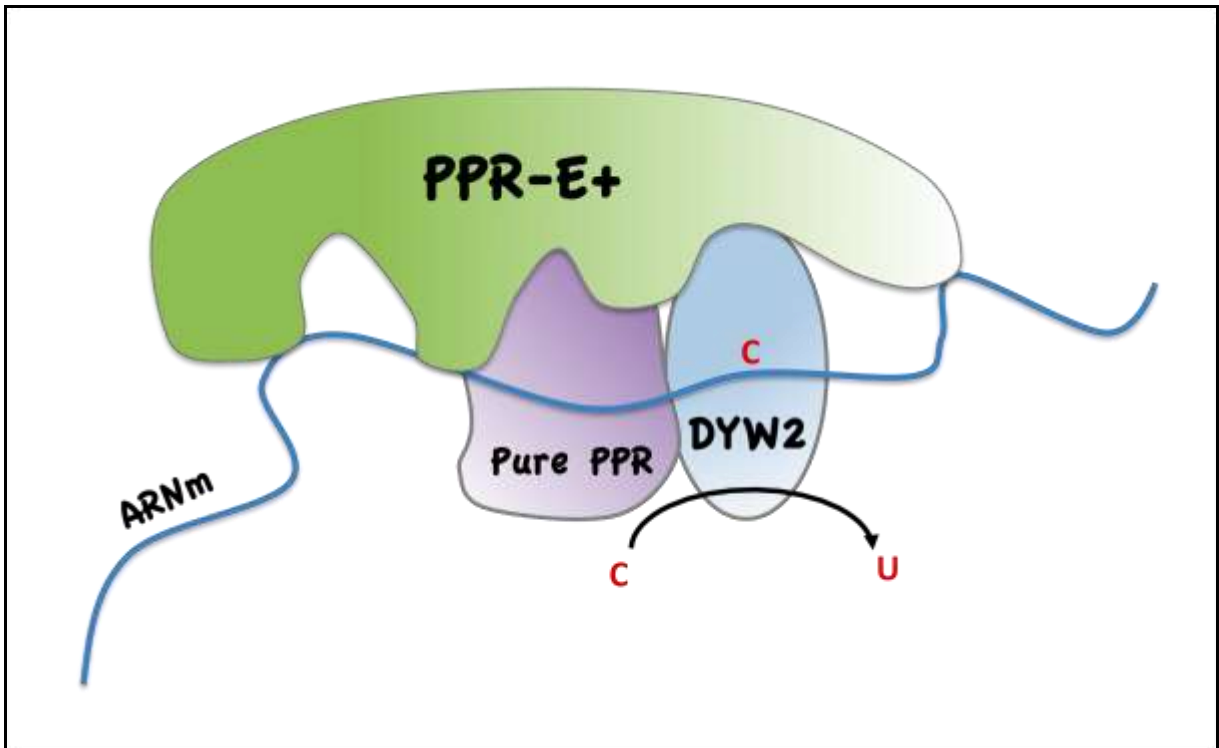


Figure 36 : Modèle d'édition faisant intervenir DYW2, une protéine E+ et la « Pure » PPR OTP100.



Figure 37 : Composition de la protéine DYW2.

CHAPITRE III : CONCLUSION ET DISCUSSION GENERALE

III.1 Les protéines PPR, des facteurs d'épissage intron spécifique :

Depuis la mise en place de l'outil de détection des événements d'épissage chloroplastiques et mitochondriaux par PCR quantitative, d'autres protéines PPR ont pu être caractérisées.

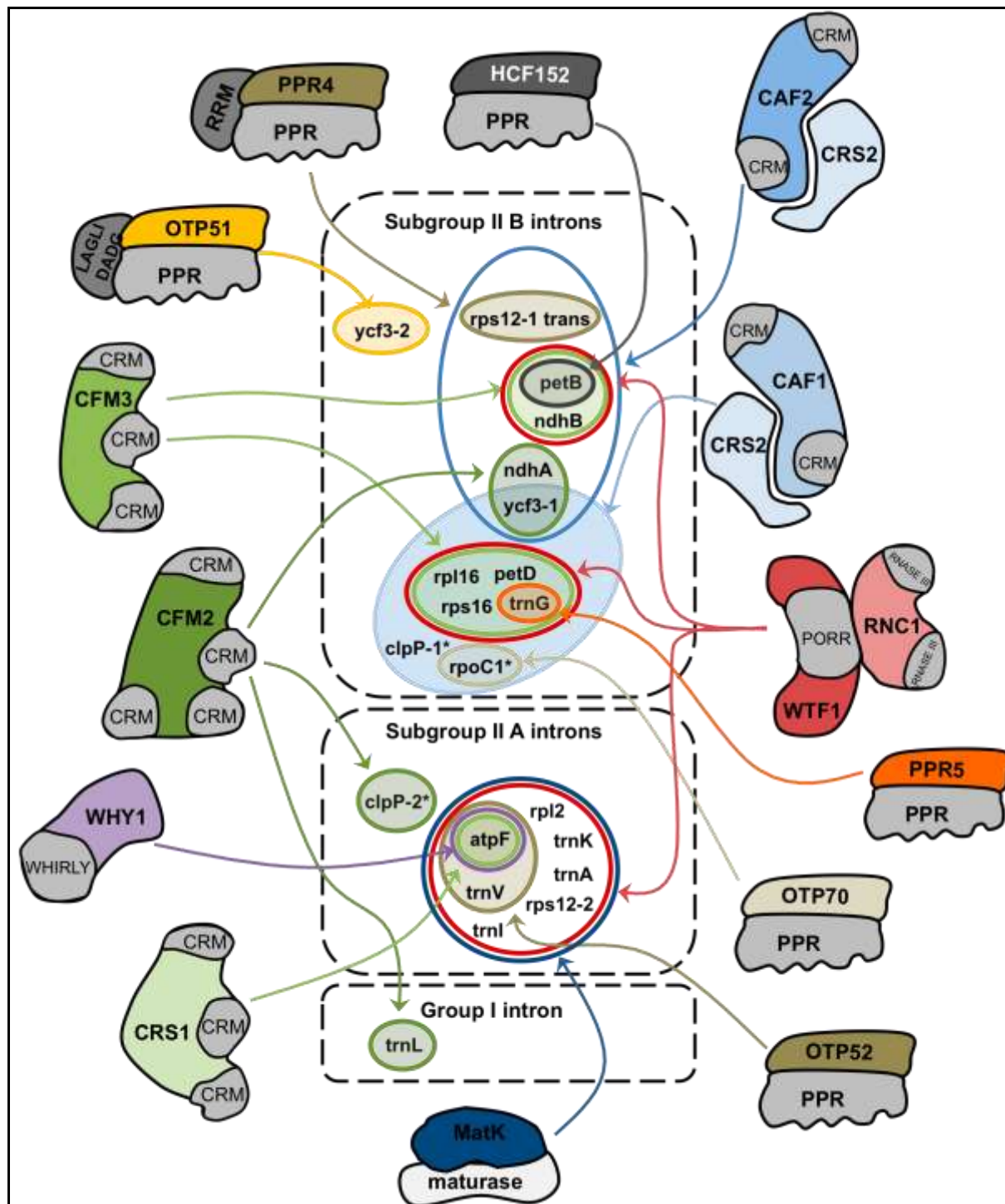


Figure 38 : Facteurs d'épissage chloroplastiques. Les introns marqués par un astérisque sont les introns présents chez Arabidopsis mais absents chez les monocotylédones comme le maïs (Figure contenant des résultats non publiés du laboratoire actualisée à partir de celle de l'article Falcon et al, 2010).

La figure 38 présente les facteurs d'épissage chloroplastiques identifiés à ce jour. Quatre facteurs de type PPR, OTP51, OTP52, OTP70 et PPR5, ont été caractérisés en plus de PPR4 et de HCF152. A l'exception d'OTP52 nécessaire dans l'épissage de deux introns, les cinq autres protéines PPR sont indispensables pour l'épissage d'un seul intron.

Au niveau mitochondrial, les recherches effectuées dans le domaine ont permis d'établir une carte de l'épissage un peu plus détaillée qu'il y a quatre ans. A ce jour, 9 à 10 facteurs ont été caractérisés (figure 39) dont 4 protéines PPR, OTP43, BIR6, OTP439, TANG2, indispensables ou quasiment indispensables pour l'épissage d'introns, et une protéine PPR, PPR40 influençant l'efficacité de l'épissage de certains introns. Comme pour les protéines PPR impliquées dans l'épissage au niveau chloroplastique, la majorité des protéines PPR identifiées dans les mitochondries dans ce processus sont spécifiques d'un intron.

Le mécanisme d'épissage s'effectue par l'intermédiaire d'un complexe ribonucléoprotéique visiblement différent pour chaque intron mais dont les protéines PPR peuvent faire partie. C'est le cas de PPR5 dont l'interaction avec le complexe est plus de l'ordre d'une interaction protéique que via l'ARN (Beick, Schmitz-Linneweber et al. 2008). Pour OTP43, aucune interaction avec une protéine de complexe ribonucléoprotéique n'a pu être identifiée par TAP-TAG, on pourrait donc penser que cette protéine a une action isolée sur l'épissage ; cependant des expériences supplémentaires seraient nécessaires pour affirmer qu'aucune interaction avec d'autres facteurs n'a lieu. OTP43 est impliquée dans l'épissage de l'intron 1 de *nad1*. L'épissage de cet intron s'effectue en *trans* et nécessite donc la jonction de deux transcrits différents. Malheureusement je n'ai pas entrepris d'expérience permettant de mettre en évidence les séquences de liaison avec OTP43 car la production de protéine recombinante a échoué du fait de l'impossibilité d'exprimer celle-ci en système bactérien. Cependant, on peut émettre l'hypothèse que cette protéine ait un rôle dans la reconnaissance des deux parties de l'intron 1 de *nad1*. Cette interaction permettrait ainsi de joindre l'intron 1 en 5' et l'intron 1 en 3' afin que le complexe ribonucléoprotéique se forme. Cette même hypothèse peut également être proposée pour les autres introns en *trans* pour lesquels une protéine PPR est essentielle pour l'épissage comme dans le cas de PPR4 et de l'intron 1 de *rps12* dans les chloroplastes, par exemple.

Ces protéines PPR restent des facteurs d'épissage à part. Leur reconnaissance spécifique d'un intron et d'une séquence spécifique au sein même de l'intron comme il a été montré pour PPR5 (Beick, Schmitz-Linneweber et al. 2008; Williams-Carrier, Kroeger et al. 2008), confèrent à ces protéines un rôle limitant de l'épissage et donc, en quelque sorte, elles pourraient être considérées comme des régulateurs de ce processus de maturation. Cependant aucune donnée ne permet actuellement d'affirmer ce dernier rôle.

Même si de plus en plus de protéines PPR sont identifiées dans l'épissage, leur rôle dans ce processus n'est pas encore totalement élucidé. Comme il a été montré, certaines sont totalement nécessaires dans ce mécanisme alors que d'autres telles qu'OTP439 et TANG2 facilitent uniquement ce processus. Il est possible pour ces dernières que le rôle dans l'épissage soit de type indirect et qu'un défaut dans la maturation des ARN pré-messagers, ou dans leur stabilisation va influencer l'efficacité de l'épissage comme c'est le cas de HCF152 (Meierhoff, Felder et al. 2003).

Pour les autres protéines PPR essentielles dans l'épissage, la reconnaissance spécifique de l'intron pourrait permettre à celui-ci d'adopter une conformation optimale afin d'être reconnu par d'autres facteurs d'épissage plus généraux (hypothèse émise dans l'article 1). Cette mise en place des autres facteurs établirait ainsi la conformation finale divisée en six domaines, typique des introns de groupe II, et permettrait le mécanisme d'épissage caractérisé par deux étapes successives de trans-estérifications.

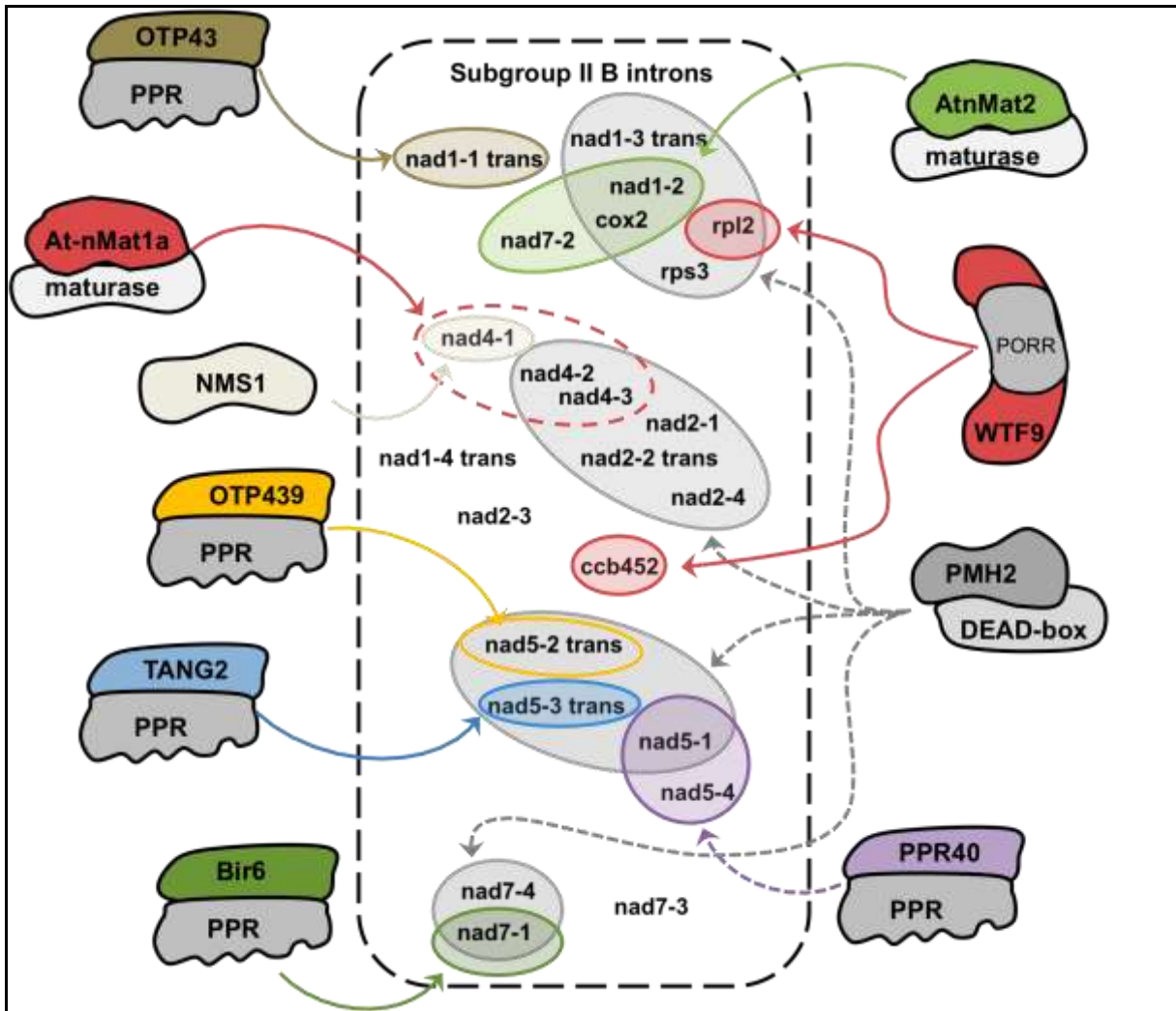


Figure 39 : Facteurs d'épissage mitochondriaux. Les flèches en pointillés indiquent les introns pour lesquels une diminution de l'épissage est observée (Figure contenant des résultats non publiés du laboratoire actualisée à partir de celle de l'article Falcon et al, 2010).

Ce modèle n'est pour le moment qu'une hypothèse et, évidemment, plusieurs expériences sont à mettre en place afin de confirmer celle-ci. Dans ce modèle, la protéine PPR est le premier facteur se liant à l'ARN à épisser et est donc la protéine clef pour la mise en place de ce système. Si cette protéine n'est pas présente, alors le complexe ribonucléoprotéique ne peut se former. Une analyse des introns et des protéines fixées sur ces introns chez les mutants *ppr* permettrait de confirmer ou de réfuter cette hypothèse. Ainsi, dans le cas de PPR5, l'analyse de l'intron *trnG* chez le mutant *ppr5* permettrait de voir si les facteurs CAF1/CRS2, CFM3, ou encore WTF1/RNC1 sont encore liés à l'intron de cet ARNt. La même question pourrait être posée pour OTP52, avec les facteurs MatK, CRS1, WHY1 et WTF1/RNC1.

Une autre étude pourrait être entreprise lorsque les facteurs connus seront cristallisés. La structure de ces protéines et la connaissance des sites de fixation permettront ainsi de modéliser *in silico* l'interaction. Cette interaction modifie probablement la structure secondaire de l'ARN cible et cette « reconfiguration » permettrait l'accès à d'autres facteurs d'épissage. Ainsi, en absence ou présence des PPR, on pourrait prédire l'affinité d'autres facteurs d'épissage. En effet, des études *in silico* permettent aujourd'hui de prédire des interactions avec l'ARN et notamment pour des facteurs d'épissage nucléaires chez l'homme (www.bioquanta.net). Ce type d'étude, mené sur les PPR, permettrait ainsi de montrer

l'affinité des autres facteurs sur l'intron en présence ou en l'absence de la PPR. Et ainsi de montrer leur rôle primordial dans la mise en place des facteurs d'épissage.

Tous les facteurs d'épissage connus dans les organites ont, sauf dans le cas de CRS1, WHY1 et des protéines PPR, au minimum deux introns cibles. Du fait de la grande conservation des séquences nucléotidiques dans les introns de groupe II, il est fort possible que les facteurs plus généraux reconnaissent ces zones et que les PPR reconnaissent des parties non conservées. Ces zones moins conservées provenant de différentes mutations au cours de l'évolution peuvent avoir un impact sur la structure secondaire de l'ARN, et ce serait pour cette raison que des protéines de type PPR sont présentes afin de modifier la structure secondaire pour l'épissage. Ceci irait dans le sens de l'hypothèse émise par Maier *et al.* (Maier, Bozarth et al. 2008; Schmitz-Linneweber and Small 2008), qui proposent que l'expansion des protéines PPR serait liée aux mutations dans le génome et que ces protéines seraient présentes en tant que suppresseur de ces mutations. En effet, le génome des organites mute et évolue beaucoup moins rapidement que le génome nucléaire (Lynch, Koskella et al. 2006). Ainsi, lors de délétions ou de mutations, il serait plus rapide de créer des facteurs de réparation pour lesquels les protéines PPR seraient de très bons candidats.

III.2 L'épissage alternatif, un moyen d'augmenter le nombre de protéines PPR :

Au cours de mon étude sur certains transcrits des gènes *PPR*, j'ai pu identifier des formes alternatives d'épissage. Etrangement, ces formes sont trouvées pour des gènes dans lesquels aucun intron n'avait été identifié. C'est la première fois qu'un tel phénomène est décrit. Lors d'analyses de l'épissage alternatif par des approches de séquençage haut-débit (Filichkin, Priest et al. 2010), les gènes ne contenant pas d'introns n'avaient pas été considérés et il est donc probable que ce phénomène soit en réalité plus général.

Pour les *PPR*, le phénomène observé est plus proche de l'épissage alternatif de type « rétention d'introns » qui se trouve être très fréquent chez les plantes comparées aux mammifères. Mais, contrairement à la majorité des transcrits observés dans ce cas, cette rétention d'introns dans des transcrits *PPR* ne crée pas généralement de codon de terminaison de la traduction prématuré. Ainsi, dans le cas de ce que l'on appelle généralement une « rétention d'intron », le modèle de gène annoté, correspondant au transcrit majoritaire, est celui dans lequel l'intron est épissé. La forme non épissée correspond à une « anormalité » d'épissage. Les pertes de séquences codantes observées pour les transcrits *PPR* correspondent à un phénomène inverse, à ma connaissance jamais observé auparavant. Il est donc difficile de classer ce phénomène puisque l'on pourrait parler plutôt de perte d'exon (« exon skipping »). Afin de préciser ce nouveau phénomène, je propose de le nommer « coding region skipping ».

Le résultat de cet épissage alternatif est de produire un certain nombre de transcrits isoformes. Comme il a été montré pour OTP43, ces ARN isoformes sont traduits, créant ainsi une diversité et une augmentation de la famille des protéines PPR. Dans le cadre de ma thèse, j'ai mis en évidence le phénomène et montré l'existence d'un certain nombre d'isoformes pour plusieurs gènes PPR. Cette analyse étendue à la famille entière, révélera très certainement de nombreuses autres isoformes. En revanche, nous n'avons pas obtenu de données nous permettant de comprendre ce phénomène, et notamment de savoir si ces isoformes ont un rôle fonctionnel ou pas.

Il est néanmoins possible de formuler de nombreuses hypothèses :

1. Les ARN isoformes ne sont le fruit que du hasard de la séquence nucléotidique, et ces formes alternatives ne sont pas prises en charge par des voies de dégradation de l'ARN car ces isoformes gardent la même séquence terminale. Cependant, ce hasard permettrait

d'élaborer de nouvelles protéines intervenant dans la réparation des mutations créées dans le génome des organites. Si cela est vraiment le fruit du hasard, alors 2/3 des événements devraient changer la phase. Nous n'avons malheureusement pas assez d'exemples pour réaliser de véritables tests de probabilité, mais pour l'ensemble des cas décrits dans ma thèse, cette proportion est de 3 cas de changement de phase sur 13 sans changement de phase pour les gènes *PPR* sans intron, donc bien inférieur au fruit du hasard.

2. Les protéines isoformes produites ont une action d'autorégulation de l'expression de l'isoforme majeure qui a seule un réel rôle moléculaire.

3. Un moyen de réguler la fonction d'une protéine à un autre niveau que celui de l'expression du gène est de le faire au niveau protéique. Des isoformes plus courtes pourraient jouer un rôle antagoniste et rentreraient donc en compétition avec la forme fonctionnelle.

4. Les isoformes traduites sont fonctionnelles et ont une fonction propre indépendante de celle codée par l'ARN non épissé. Chaque forme joue alors un rôle particulier dans la maturation des ARNm (les cibles pouvant être différentes).

Par exemple, dans le cas des protéines impliquées dans l'édition des ARNm, il est possible que la formation de certaines isoformes permette la reconnaissance de différents sites. Cette dernière hypothèse pourrait expliquer la reconnaissance de plusieurs sites d'édition sans séquences consensus par une même protéine PPR, qui serait en réalité plusieurs isoformes (cf partie suivante pour le modèle proposé).

La fonction de ce phénomène reste donc un grand mystère mais ouvre de nouvelles voies d'investigations sur la régulation et la fonction des protéines PPR.

Ces isoformes d'épissage alternatif sont-elles régulées en fonction du développement ? sont-elles activées au cours d'un stress ? ou sont-elles exprimées de façon tissu- ou organe-spécifique ? Autant de questions auxquelles malheureusement je n'ai pas réussi à répondre dans le temps de ma thèse mais auxquelles il sera intéressant de répondre dans un avenir proche.

III.3 Rôle de l'épissage alternatif : une nouvelle hypothèse sur la reconnaissance de multi-sites d'édition par une seule protéine PPR ?

Des formes alternatives d'épissage ont été identifiées à la fois chez des PPR de type « Pure » mais aussi sur des protéines PPR de la sous-famille PLS. Cette dernière comprend notamment les sous-familles de PPR de type E, E+ et DYW impliquées pour la plupart dans l'édition. 526 sites d'édition sont décrits au niveau mitochondrial et 34 sites dans les chloroplastes. Seulement 193 protéines PPR de type E, E+ et DYW sont décrites dans le génome d'*Arabidopsis*. Partant de l'hypothèse que seules les protéines PPR sont impliquées dans l'édition, il existe donc un ratio d'un facteur de type PPR pour plusieurs sites d'édition. Ainsi, bien que les protéines PPR aient une reconnaissance spécifique de la séquence de l'ARN, parmi les facteurs d'édition identifiés dans les chloroplastes, cinq protéines PPR (CRR22, CRR28, CLB19, OTP82 et OTP84) sont impliquées dans l'édition de plusieurs sites. Pour ces dernières, il existe potentiellement une séquence de reconnaissance commune entre les sites. En effet, il a été montré récemment qu'une séquence consensus autour des différents sites pouvait être définie pour quatre d'entre elles, même si aucune preuve expérimentale n'est apportée et que ces similitudes restent très faibles (Hammani, Okuda et al. 2009). De façon notable, aucune séquence consensus ne peut être identifiée pour les trois sites d'édition impliquant CRR22.

CONCLUSION ET DISCUSSION GENERALE

Cette difficulté, voire impossibilité, à reconnaître une séquence consensus peut être expliquée par trois hypothèses. La première serait que d'autres facteurs seraient impliqués dans l'édition, apportant pour certains la reconnaissance spécifique de l'ARN. La deuxième serait que la zone de reconnaissance est une très courte séquence composée éventuellement de plusieurs éléments cis ou alors que la reconnaissance s'effectue par l'élaboration d'une structure secondaire spécifique de l'ARN. La troisième, enfin, pourrait être que des formes alternatives d'épissage permettraient de reconnaître ces différents sites. Ainsi, mon travail de thèse me permet de proposer, par exemple, que le gène *CRR22* pourrait coder pour trois protéines différentes, et chaque protéine serait impliquée dans l'édition d'un des trois sites d'édition décrits pour *CRR22*. Cette dernière hypothèse, très attractive, peut se tester en regardant s'il existe des formes alternatives d'épissage pour cette PPR mais également pour d'autres PPRs impliquées dans l'édition de plusieurs sites (même si une séquence consensus a pu être définie). Aucun ADNc ne permet de vérifier cette hypothèse pour *CRR22*, en revanche des prédictions de sites donneurs et accepteurs peuvent être effectuées (figure 40).

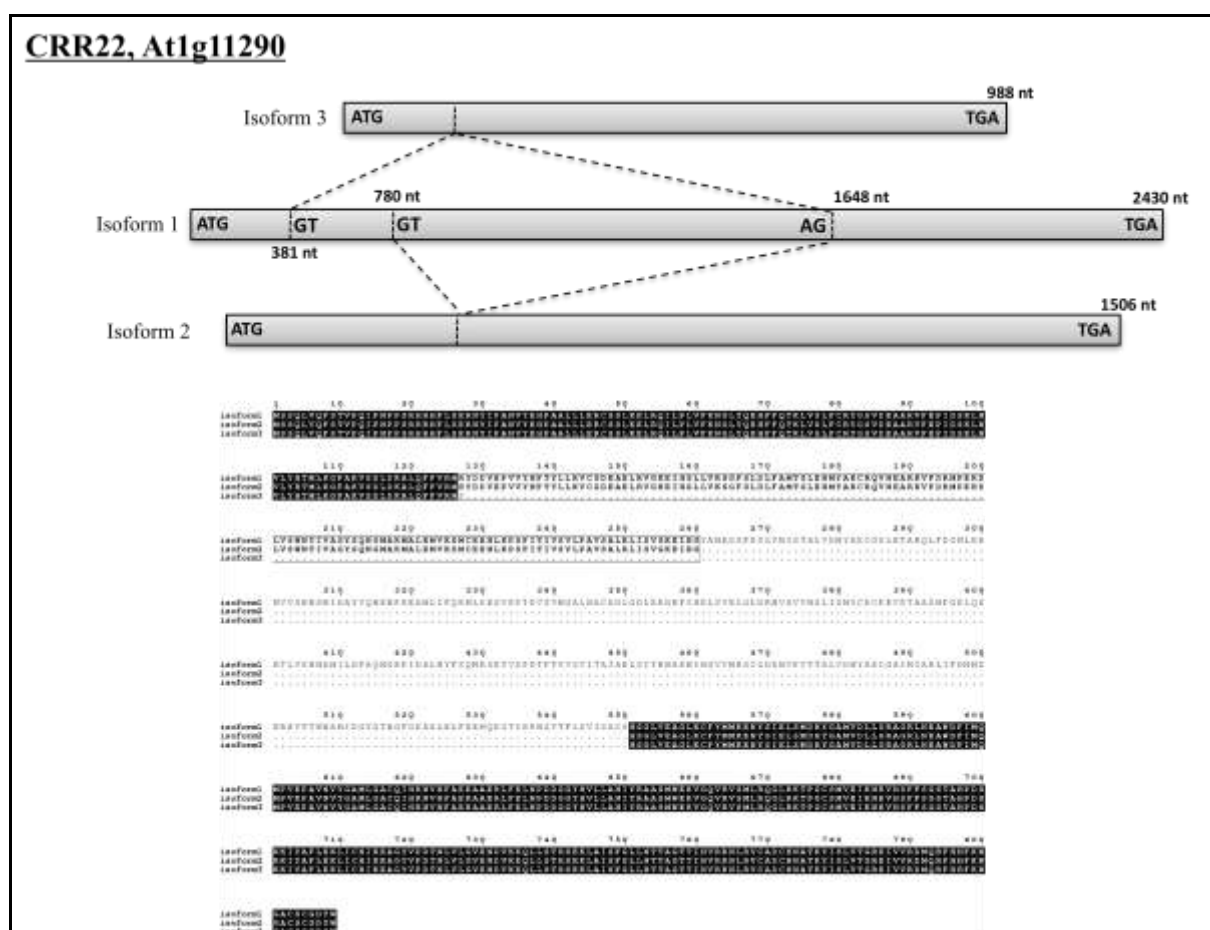


Figure 40 : Prédiction des isoformes d'épissage alternatif pour *CRR22* en utilisant le logiciel http://www.fruitfly.org/seq_tools/splice.html. Un alignement sous « Multalin » des différentes isoformes putatives traduites est représenté.

Ces prédictions permettent de proposer l'existence de trois isoformes putatives. Une sans épissage et deux autres avec une excision d'une séquence codante ne changeant pas la phase de lecture. Ces isoformes ne sont que des prédictions et une analyse par RT-PCR suivie d'un clonage serait nécessaire afin de confirmer l'hypothèse de l'existence de plusieurs isoformes. Des premières données d'amplification sur ADNc faites par Sandra Tanz dans l'équipe de Ian

Small montrent que seule la forme pleine longueur est présente, cependant le signal d'amplification étant très faible, ces résultats ne sont pas définitifs et demandent à être confirmés par une nouvelle expérience d'amplification suivie d'un clonage.

Pour tester un éventuel épissage alternatif d'autres transcrits PPR codant pour une protéine importante dans l'édition, nous avons en notre possession les oligonucléotides utilisés pour le clonage de *CLB19*. Cette PPR de type E+ chloroplastique est impliquée dans l'édition de deux sites, un sur le transcrit *ClpP* et l'autre sur *RpoA* (Chateigner-Boutin, Ramos-Vega et al. 2008). Elle s'avère donc être une autre candidate pour tester si des formes alternatives peuvent expliquer la reconnaissance de deux sites. J'ai effectué quelques expériences dans ce but au cours de ma thèse (résultats non présentés). Malheureusement, elles n'ont pas permis d'identifier une autre forme que celle décrite, suggérant qu'aucune forme alternative d'épissage n'existe dans le cas de *CLB19* et que la reconnaissance des deux sites s'effectuerait par un autre mécanisme.

J'ai également regardé tous les ESTs disponibles pour toutes les PPR E, E+ et DYW décrites dans les différentes espèces présentes dans la base de données FlagDB++, *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa* et *Vitis Vinefera*. Malheureusement trop peu d'ADNc ou d'ESTs couvrent ces gènes PPR et aucune forme alternative n'a été détectée. Des sites alternatifs d'épissage peuvent être prédits pour la majorité d'entre eux mais seule une PCR sur ADNc suivie d'un clonage et d'un séquençage permettrait de confirmer ces prédictions.

III.4 Action des protéines PPR, une action seule ou en partenariat :

Les premières données d'interaction protéine-protéine obtenues par TAP-TAG mais aussi par double hybride montrent que certaines protéines PPR interagissent avec d'autres protéines PPR mais également avec d'autres protéines.

Jusqu'ici, dans le cas des protéines PPR impliquées dans l'épissage des ARNm, seule PPR5 a été montrée associée à un complexe ribonucléoprotéique (Beick, Schmitz-Linneweber et al. 2008). Aucune donnée ne permet de montrer que PPR4 est associée avec d'autres protéines et les données obtenues par TAP-TAG ne montrent pas d'interaction entre OTP43 et des protéines pouvant être impliquées dans ce phénomène post-transcriptionnel. Ces deux dernières protéines sont impliquées dans l'épissage en *trans* de l'intron 2 de *rps12* (Schmitz-Linneweber, Williams-Carrier et al. 2006) et l'intron 1 de *nad1* (de Longevialle, Meyer et al. 2007). Le rôle de ces deux protéines pourrait être de reconnaître de façon spécifique des séquences dans les deux parties de l'intron et de permettre ainsi l'assemblage du reste du complexe ribonucléoprotéique sans forcément nécessiter une interaction physique avec les autres facteurs. Cet assemblage pourrait avoir lieu grâce à la formation d'interactions entre elles-mêmes, formation qui a visiblement été suspectée dans le cas de PPR4.

Dans le cas de PPR5, il a été montré que l'interaction avec le complexe ne se faisait pas via l'ARN et il est donc possible que cette protéine interagisse avec un ou plusieurs autres facteurs d'épissage.

Dans le cas de protéines PPR impliquées dans l'édition, le laboratoire a proposé un modèle depuis plusieurs années dans lequel les protéines de type E et E+ interagissent avec une autre protéine PPR ayant le motif DYW, mais sans le motif E+. En revanche, pour les protéines PPR de type DYW avec la présence des motifs E et E+, il est proposé que ces protéines agissent seules dans l'édition, sans aucun autre facteur. Grâce à l'approche de TAP-TAG sur *CLB19*, nous avons pu montrer une interaction entre deux ou voire trois protéines PPR impliquées dans l'édition. *CLB19*, une protéine PPR E+ interagit avec une protéine DYW

renommée DYW2 et OTP100, une protéine PPR de type « Pure ». Dans ce cas précis, CLB19 n'ayant pas de motif DYW suspecté comme porteur de la fonction enzymatique, nécessite une interaction avec une protéine de type DYW. DYW2 apporterait cette fonction. Les résultats de TAP-TAG ne permettent pas de conclure si les interactions entre CLB19 et OTP100, d'une part, et CLB19 et DYW2, d'autre part, sont directes, même si cette dernière interaction est plus probable. La fonction d'OTP100 reste pour le moment inconnue, mais est en cours d'étude au laboratoire. L'interaction pourrait également s'effectuer de façon indirecte, via le ou les transcrits sur lesquels sont fixés CLB19. Dans ce cas OTP100 aurait une fonction différente de l'édition sur ces transcrits. Afin de tester cette hypothèse, la purification du complexe en absence d'ARN (après traitement avec une RNase) sera effectuée prochainement.

Dans le génome d'*Arabidopsis*, d'autres gènes codant pour des protéines similaires aux protéines DYW1 et DYW2 sont décrits. Il serait donc intéressant afin de valider notre nouveau modèle d'effectuer d'autres expériences de TAP-TAG avec celles-ci afin de voir avec quel type de protéine une éventuelle interaction est observée. Puis, dans un deuxième temps, étudier la manière dont elles interagissent et quels sont les domaines nécessaires pour ce type d'interactions.

III.5 Rôle et régulation des protéines PPR

La caractérisation d'un certain nombre de protéines PPR permet de montrer leur rôle primordial dans l'expression et la maturation des ARNm et des ARNt dans les organites. Cette famille joue donc un rôle essentiel dans l'expression de ces deux génomes et établit un lien étroit entre le noyau et les organites. On peut se demander pour quelle raison, chez les plantes, une telle complexité a été créée et conservée et implique autant de protéines de type PPR. Une des hypothèses avancées, et déjà évoquée, est que des mutations intervenant dans le génome des organites sont plus facilement rectifiées par l'élaboration de nouveaux facteurs codés par le noyau corrigeant de telles erreurs. Si cela est vrai, alors les PPR ne peuvent pas être considérés comme des facteurs de régulation et devraient agir d'une manière plus ou moins constitutive.

On peut également proposer que ces PPR jouent au contraire un rôle de régulation très fine contrôlée par leur état d'activité et leur expression. Cette régulation des PPR pourrait s'établir à deux niveaux. La première au niveau de la transcription et de la régulation post-transcriptionnelle et la deuxième au niveau de l'activité de la protéine.

III.5.1 Régulation transcriptionnelle et post-transcriptionnelle :

La majorité des gènes *PPR* est faiblement exprimée et a une durée de demi-vie très courte (Narsai, Howell et al. 2007) qui peut être expliquée par l'absence d'introns. En conséquence, cette durée de vie courte est un facteur limitant de la traduction des *PPR*. Cependant ce n'est pas parce qu'ils sont faiblement exprimés qu'ils ne sont pas différenciellement exprimés. En effet selon les conditions environnementales et selon les apports nutritifs, certains gènes *PPR* sont plus ou moins exprimés. Ainsi comme nous l'avons montré certains gènes vont être surexprimés en présence de sucre. Ce n'est évidemment pas la seule condition où l'on observe un changement dans l'expression. De nombreuses données de transcriptome ont été accumulées depuis ces dernières années par l'intermédiaire des puces CATMA et AFFYMETRIX. Ces données sont en cours d'analyse par Richard Berthomé à l'URGV afin de faire un « clustering » des *PPR* selon les conditions et ainsi essayer d'établir un lien entre

leur expression, leur localisation et leur fonction. Cette analyse pourra éventuellement prédire la fonction de certaines protéines PPR pour lesquelles celle-ci reste pour le moment inconnue.

Un autre moyen de limiter la traduction et de réguler l'expression est la dégradation de ces ARN via la voie TGS (Transcriptional Gene Silencing) ou PTGS (Post-Transcriptional Gene Silencing) qui est induite par les petits ARN de type microARN ou siARN. Chaque petit ARN est spécifique d'un gène et certains petits ARN sont également décrits pour des gènes PPR chez *Arabidopsis* et également chez le peuplier (Klevebring, Street et al. 2009). Ainsi miR161 code un microARN ciblant 20 gènes *PPR* et mais dont seule la cible At1g63130 a été validée (Fahlgren, Howell et al. 2007). Deux autres microARN, miR400 et miR158 sont prédits pour avoir plusieurs gènes *PPR* comme cibles mais aucune validation n'a été faite pour ces deux derniers (Rhoades, Reinhart et al. 2002; Sunkar and Zhu 2004).

Il est possible également que la traduction soit limitée via de l'épissage alternatif. En effet, en plus des phénomènes décrits précédemment dans les séquences codantes, des introns sont également décrits pour certains gènes *PPR* dans les parties 5' et 3' UTR. Si ces introns ne sont pas épissés, ils peuvent avoir une action sur la stabilité des ARNm comme il est montré pour le gène *AUF1* chez l'homme codant une protéine intervenant dans la reconnaissance d'éléments AU-rich et qui induit la déstabilisation des ARNm via la voie NMD (Banihashemi, Wilson et al. 2006). L'intron en 3' UTR de ce gène possède ces éléments et sa rétention conduit donc à une autorégulation de son expression. Je n'ai pas regardé si de tels éléments étaient présents dans les introns en 3' UTR pour les gènes *PPR*, mais il est fort possible que ceux-ci aient un impact sur l'expression.

Enfin l'épissage alternatif peut aussi créer des codons de terminaison de la traduction prématurés chez les *PPR*, ces transcrits seront alors pris en charge par la voie de dégradation NMD. Cet épissage alternatif s'il est régulé va donc avoir un impact sur l'expression des *PPR*.

III.5.2 Régulation post-traductionnelle :

Une autre voie de régulation peut avoir lieu au niveau de la protéine. En effet un certain nombre d'étapes au niveau post-traductionnel influencent l'activité de la protéine. En plus des phénomènes de glycosylation ou de myristoylation, des étapes de phosphorylation peuvent avoir lieu. Ainsi, il est fréquent que l'activité d'une protéine soit dépendante de son état de phosphorylation. Du fait de la faible expression générale des protéines *PPR*, peu de peptides phosphorylés spécifiques de protéines *PPR* ont pu être détectés par des approches de spectrométrie de masse systématiques. Cependant 7 protéines *PPR* sont décrites comme phosphorylées dans la base de données PepBase (<http://pepbase.iab.keio.ac.jp/phospho/msb/>) et 69 dans PhosPhat 3.0 (<http://phosphat.mpimp-golm.mpg.de/db.html>). La majorité de ces *PPR* n'a qu'un seul site phosphorylé de type Serine, Threonine ou Tyrosine, mais il est également possible de trouver de 2 à 4 sites pour certaines. Parmi les 69 protéines décrites, on trouve CRR4, CRR21, PPR596, PPR2, DOT4 et OTP52, qui sont des protéines dont la fonction dans l'édition, l'épissage ou la biogénèse des ribosomes a été décrite. Cette phosphorylation pourrait réguler l'activité de ces protéines importées dans les organites en fonction des besoins et des demandes des organites et de la cellule.

L'autre étape de régulation pour laquelle très peu d'éléments sont connus est une éventuelle implication des protéines HSP. Comme nous l'avons montré par TAP-TAG, la purification des protéines interagissant avec OTP43 est fortement enrichie en HSP60. Dans ce type d'expérience, ces protéines sont très souvent considérées comme des faux positifs, mais vu la quantité trouvée dans notre extrait, il est possible que cette interaction soit réelle. De plus, les premières données de double hybride montrent également des interactions entre des

protéines PPR et des HSP. Plusieurs hypothèses peuvent être proposées. Il est possible que ces protéines soient nécessaires pour le repliement et la configuration finale des protéines importées mais il est aussi envisageable qu'elles aient un rôle de régulateur de la quantité fonctionnelle des protéines actives. Il faut remarquer ici que les résultats obtenus concernant les protéines HSP ouvrent également de nouvelles pistes quant à l'expression de protéines PPR en système bactérien. Par exemple, il serait intéressant de tester l'effet d'une co-expression de protéine chaperone et d'OTP43 sur l'efficacité d'expression de cette PPR qui n'a jamais pu être obtenue en système bactérien.

Enfin, une dernière régulation peut être envisagée par l'épissage alternatif des transcrits *PPR*. En effet celui ci peut conduire à la synthèse de différentes isoformes caractérisées par la délétion de motifs PPR plus ou moins importante selon le type d'épissage. Cette modification a certainement un impact sur les capacités physico-chimiques de ces protéines puisque c'est au niveau de ces motifs que la reconnaissance avec l'ARN doit avoir lieu et il est possible que par ce biais une régulation de la fonction soit mise en place, avec un effet potentiel de compétition entre certaines isoformes (cf partie III.2). C'est le cas par exemple chez l'homme où la protéine CD45 qui joue un rôle dans l'activation des lymphocytes T lors d'une réponse immunitaire à une infection est active ou non selon l'isoforme d'épissage alternatif (Oberdoerffer, Moita et al. 2008). Ainsi lorsque les lymphocytes T sont au repos, la forme la plus longue est exprimée et comporte neuf exons, en revanche lorsque les lymphocytes T sont activés, les ARNm CD45 sont épissés alternativement et les exons 4, 5 et 6 sont excisés. Cette modification crée ainsi une protéine plus courte qui ne peut alors activer les lymphocytes T. Ce phénomène de régulation empêche de déclencher une réponse immunitaire disproportionnée, et l'on peut imaginer qu'une compétition puisse avoir lieu entre différentes isoformes dites actives et non-actives dans le cas des PPR.

Depuis la découverte de la famille PPR il y a plus de dix ans, de nombreuses avancées ont été faites dans l'élucidation du rôle de ces protéines. Aujourd'hui environ 1/10ième de ces protéines chez les plantes ont une fonction connue et joue un rôle majeur et très spécifique dans la transcription, la maturation et la traduction des ARN dans les organites. Cependant, même si de nombreux défauts ont été et seront certainement encore découverts chez des mutants ppr, on ne connaît pas encore comment ces protéines interagissent.

Un enjeu clef de la compréhension de la famille PPR sera donc de connaître la structure des protéines par cristallographie ainsi que la structure de l'interaction avec l'ARN afin de comprendre comment cette interaction s'établit et pourquoi elle est spécifique de la séquence de l'ARN.

La fonction de ces PPR codées par le génome nucléaire en fait des facteurs clefs pour le fonctionnement des organites, et créant ainsi un lien étroit entre le noyau, cerveau central de la cellule et les deux organites, moteurs énergétiques.

CHAPITRE IV : MATERIELS ET METHODES

IV.1 Matériels

IV.1.1 Souches bactériennes

La souche DH5 α a été utilisée pour tous les clonages de type GATEWAY ou pGEMT.

Afin de multiplier les vecteurs d'entrée et de destination GATEWAY vides, la souche DB3.1 résistante à la cassette suicide *ccdb* a été utilisée.

La transformation du mutant *otp43* en vue de complémentation fonctionnelle avec les différentes isoformes a été réalisée par l'intermédiaire de la souche d'agrobactérie C58C1pMP90.

IV.1.2 Oligonucléotides utilisés au cours de ma thèse et non répertoriés dans les articles.

Oligonucléotides utilisés pour le clonage Gateway :

| | Nom | Séquence | Info |
|----------------|------------|---|---------------|
| projet TAP-TAG | CLB19Fw-S | AAAAAGCAGGCTCCACCATGGGTCTCCTTCCCCTCGTCTCG | start + Kosak |
| | CLB19Rev-E | AGAAAGCTGGGTCAGCATTGAGGAGATCACCAGC | end |
| | DYW1Fw-S | AAAAAGCAGGCTCCACCATGAGTTTTTGTCTCTTCTCTG | start + Kosak |
| | DYW1Rev-E | AGAAAGCTGGGTCCTCAATAATCTCCACAAGAGCAG | end |
| | OTP43Fw-S | AAAAAGCAGGCTCCACCATGAAACGTCTTTTAGCAAATCG | start + Kosak |
| | OTP43Rev-E | AGAAAGCTGGGTCAGCCTGAACCTCCGGAGAAGC | end |
| | OTP51Fw-S | AAAAAGCAGGCTCCACCATGACAAAATCCAACGGCCAC | start + Kosak |
| | OTP51Rev-E | AGAAAGCTGGGTCGTCAATTCAGAAGCACATTTTAC | end |
| | attB1 | GGGGACAAGTTTGTACAAAAAAGCAGGCT | PCR2 |
| | attB2 | GGGGACCACTTTGTACAAGAAAGCTGGGTC | PCR2 |

Oligonucléotides utilisés pour la mutagenèse dirigée :

| | Nom | Séquence | Info |
|------------|--------------------|---------------------------|--------------------------------------|
| mutagenèse | OTP43_Acceptor1rev | TGCCAGCTTTGAAAAAACCTTTG | mutagenesis acceptor splicing site 1 |
| | OTP43_Acceptor1Fw | CAAAGTTTTTTCAAAGCTGGGCAG | mutagenesis acceptor splicing site 1 |
| | OTP43_Acceptor2Rev | GCTTGATTCCTTGTCAAAAAGAAGC | mutagenesis acceptor splicing site 2 |
| | OTP43_Acceptor2Fw | GCTTCTTTGACAAGGAATCAAGC | mutagenesis acceptor splicing site 2 |

Oligonucléotides utilisés pour la PCR quantitative :

| Oligonucléotides utilisés pour la PCR quantitative | |
|--|----------------------|
| Nom | Séquence |
| qOTP43_1Fw | GTGGTCGCTGGGAATCAAG |
| qOTP43_2Rev | TAGCCTGAACCTCCGGAGAA |
| qOTP43_3Fw | AATATGTCCACGACGCTTCC |
| qOTP43_4Rev | TAATCTGCCAGCTCTCTCG |
| qOTP43_5Rev | CAAAAGCTTGATTCCTCTCG |
| qOTP43_6Rev | AGAGAGCGCATACGGTGAAT |

Oligonucléotides utilisés pour la PPE :

| Nom | Séquence |
|------------|---------------------------|
| PPE-nad2-1 | GCGAGCAGAAGCAAGTTATT |
| PPE-rps3-1 | TTGACATCTTGATACAAAATTACCA |

Oligonucléotides utilisés pour la sonde biotinilée pour le *northern blot* :

| Nom | Séquence |
|----------|-----------------------|
| nad6-Fw | TATGCCGAAAGGTACGAAG |
| nad6-Rev | GTGAGTGGGTCAGTCGTCTCT |

MATERIELS ET METHODES

Oligonucléotides utilisés pour la détection d'isoformes d'épissage alternatif sur les 48 PPRs :

| PPR code | AGI code | start primer | stop primer |
|----------|-----------|--|---------------------------------------|
| PPR1 | At2g01860 | GATAGAACCATGATACTACACTGTCTCTGTC | GAAAGCTGGGTCTCAGTGCAAAGAGACATTGTTT |
| PPR2 | At1g61870 | GATAGAACCATGGCGTTACTCTCTCGAATCC | GAAAGCTGGGTCTCATTGGCGTAATGGCGCTTC |
| PPR3 | At2g34370 | GATAGAACCATGGTGAGACTAGTTTGACGCC | GAAAGCTGGGTCTTACCAAGTAATCTTTGCAGGAGC |
| PPR4 | At4g16390 | GATAGAACCATGTCATTTTCATCATCTATGTTCC | GAAAGCTGGGTCTACGCAGAAACTCCTCCAG |
| PPR5 | At3g25970 | GATAGAACCATGAAAGCGTTGTTAGCCTCTC | GAAAGCTGGGTCTTAAGCATGAAGTGAATCTGCATC |
| PPR6 | At4g31850 | GATAGAACCATGGTCGCGTTACTCTGCTC | GAAAGCTGGGTCTCAAGCTCTATTGGGAGCTG |
| PPR7 | At2g20540 | GATAGAACCATGGCGTTTCATGGAATCAGAG | GAAAGCTGGGTCTAAACTATGCCTATAAATGCAAGAG |
| PPR8 | At4g20090 | GATAGAACCATGCCAAATGCCAATCCC | GAAAGCTGGGTCTCAAGTACACAGGTTCTCCAG |
| PPR9 | At3g11460 | GATAGAACCATGATCGTCTGTTACAAGCTTTGTA | GAAAGCTGGGTCTCACCAGTAATCTTTGCACGAG |
| PPR10 | At5g52630 | GATAGAACCATGGCTCTGAATCTTCAGCTTTC | GAAAGCTGGGTCTCACCAGTATGCTTACAGAAGAAC |
| PPR11 | At5g62370 | GATAGAACCATGTTAAAGGCGAAAGCTTTATGTTAT | GAAAGCTGGGTCTCAAGTGTGTTGATTCTGGTTAAGC |
| PPR12 | At1g80550 | GATAGAACCATGGCTTTGTTGGCGGCACT | GAAAGCTGGGTCTATTCTACAGACAAGCGTTTC |
| PPR13 | At1g56690 | GATAGAACCATGAAGCGTTTAAAGCTAATCTTACG | GAAAGCTGGGTCTCACCAGTAGTCTCTACAAGAAC |
| PPR14 | At4g02750 | GATAGAACCATGGAATAAACAAGTTTCGTGCTC | GAAAGCTGGGTCTCACCAGTAGTCTCCACAAGAA |
| PPR15 | At1g09410 | GATAGAACCATGAAGTCAAGACTCTCCTACG | GAAAGCTGGGTCTCACCAGTAATCTTTACAGGAACA |
| PPR16 | At3g02010 | GATAGAACCATGAACTTTATTGCAACTCTAATGAAG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR17 | At2g02980 | GATAGAACCATGGCGATATCTTCACTCTCAC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR18 | At1g68930 | GATAGAACCATGGCGTCAAGTACTACTCTG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR19 | At3g24000 | GATAGAACCATGGCATTGCGTTTCCCGCC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR20 | At2g29760 | GATAGAACCATGGCTATCTTCTCCACAGCAC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR21 | At5g09950 | GATAGAACCATGACTAATGTGTCTCTTTCTTTT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR22 | At3g46790 | GATAGAACCATGTTCTGTCTCACCTCTCTC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR23 | At1g59720 | GATAGAACCATGGTTGTTCTCAATTTATCGTTTC | GAAAGCTGGGTCTCACCAGTAGTCTAACAAGAGC |
| PPR24 | At3g13770 | GATAGAACCATGTTAATCTGATGAGACTAATTCATC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR25 | At2g01510 | GATAGAACCATGCTGCGAACAACAGCCAT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR26 | At4g18750 | GATAGAACCATGGCGATGTTGGTGACGAATT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR27 | At5g65570 | GATAGAACCATGGCGGATATGAGCTACAC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR28 | At2g03880 | GATAGAACCATGAAGTCAAGTCAAGTCAAAAGATCAAA | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR29 | At5g15340 | GATAGAACCATGAAATGCTTATCTTACAAAAGTT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR30 | At1g71420 | GATAGAACCATGATTACAAGCTTATCTCAAATCTCT | GAAAGCTGGGTCTCACCAGTAGTCTTACAGGAGC |
| PPR31 | At5g13230 | GATAGAACCATGATCGTGTTCATGAGAATAATTCAT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR32 | At5g13270 | GATAGAACCATGACGATTCTCACTGTACAGTC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR33 | At5g14770 | GATAGAACCATGATAAGGATTTGGAATAACTATAAGG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR34 | At5g02860 | GATAGAACCATGGCGGACAAGCTACTCT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR35 | At5g12100 | GATAGAACCATGGTGACTCGCTTAAAGACTCG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR36 | At1g79540 | GATAGAACCATGAAACCACAGATGTTCTTTCTGA | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR37 | At2g19280 | GATAGAACCATGGGGTTTCAATCTTAACTCTC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR38 | At3g09040 | GATAGAACCATGATTTCCGAGTTCTCTTGACG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR39 | At5g55740 | GATAGAACCATGGCTTCTCTCTTTCAACAC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR40 | At3g16890 | GATAGAACCATGAGAGGCTTTGCTTTCTTAC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR41 | At1g71210 | GATAGAACCATGCTCCTCAGGCGGAGG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR42 | At4g26680 | GATAGAACCATGATTCGCATAAGCCTCGGAG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR43 | At1g74900 | GATAGAACCATGAAAGCTTTTATGCAAACTCGC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR44 | At1g18900 | GATAGAACCATGATACGCGTAAGCATATTAGT | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR45 | At5g46580 | GATAGAACCATGGCGACTGTTCTTACCACAG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR46 | At1g05750 | GATAGAACCATGGGTCTCCTTCCGCTCG | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR47 | At3g22150 | GATAGAACCATGGCTGGATCAGCTTACC | GAAAGCTGGGTCTCACCAGTAATCACCACAAGAAC |
| PPR48 | At3g13880 | GATAGAACCATGTTGTTACAGTTCCGAGCAAAG | GATAGAACCATGTTGTTCTGTTATGCCAACCAAAG |

Oligonucléotides utilisés pour le Génotypage :

| AGI | Nom | Séquence | Nom | Séquence |
|-----------|-----------------|------------------------|-----------------|------------------------|
| At1g74900 | SAIL_867_F07_LP | TGGTTACAGGAGACTCGGTTG | SAIL_867_F07_RP | CTTCTTCTGTTGCGACGTAC |
| At1g05750 | Salk_123752_LP | GAGGAAGCGACAATGGTTATG | Salk_123752_RP | TTGATTTCTTGGACGGCTATG |
| | Salk_104250_LP | ATGTGCATTTCCATTAGCAGC | Salk_104250_RP | ACCAACGGTGTATTATTAGGGC |
| At3g02330 | SALK_128453_LP | AAAAGAGGCAAAAGAAGCAATG | SALK_128453_RP | TACCTGTGCGAATCTTGCTTC |
| | SALK_027870_LP | TTGCACTGGTGGTTCTTTAGG | SALK_027870_RP | AAACCTTTTCCCACATTTCTG |
| | SALK_061252_LP | CGCAAAACATCTAGCGATAAGC | SALK_061252_RP | GTAGACAGCAAGACTTTTCG |
| | SALK_061253_LP | AGTTCTGTTATGACTCATGGCG | SALK_061253_RP | TCACACTTTGCGTACATGTCC |
| At3g09040 | CSHL_ET12921_LP | CACCATTATGTGCATAACCCC | CSHL_ET12921_RP | ACGTGTTTTGAGTTATTTCCC |
| At1g64310 | SALK_092788_LP | TTCACGAGGAAATTCAAACG | SALK_092788_RP | CAGAGAAAGTATCTCGTCCG |
| | SALK_069611_LP | GAACACGCAACTAAATCAGGC | SALK_069611_RP | CTGGTCTGCTTACAGACTCTG |
| At1g69350 | SALK_033399_LP | ACATGGAACTCAATGCTCTGC | SALK_033399_RP | GGGCCAAAGAAAAGAAAACAC |
| | SALK_033395_LP | TGTGGTCTATGTGCAAGAAGC | SALK_033395_RP | TGTTCTTTGGTCGATTTTGG |
| | SALK_033396_LP | ACATGGAACTCAATGCTCTGC | SALK_033396_RP | GGGCCAAAGAAAAGAAAACAC |
| At1g03540 | SALK_035148_LP | ACATTCATTGAACACTTGCC | SALK_035148_RP | GTTTACCGTTCATCTCTCCC |
| | SALK_066190_LP | AGAATTGTTCACCGTATTG | SALK_066190_RP | ACATTTCCATTGAACACTTGCC |
| | SALK_043866_LP | GCTCTGTCCAACAGTCTTTGC | SALK_043866_RP | TAGGGAGATGTTTCCATGGTG |
| At2g35030 | SALK_042973_LP | ACTGCTATGGTCTGATGGATTG | SALK_042973_RP | GTTCTGCTTCCCACCTTAAAC |
| | SALK_115308_LP | CATGGACGATGTTGATTACTG | SALK_115308_RP | TTTCAATCTACCAGCAGCAC |
| At5g06540 | SAIL_77_E03_LP | TGGTCTATGCTTTTCAAGACC | SAIL_77_E03_RP | AGGGATGTTACTTACAGCGG |
| At5g61990 | SALK_143317_LP | CATGGACTAGTCTTTGCGCAG | SALK_143317_RP | TACCATCGGATATGCAAGGAG |
| At3g09060 | SALK_144391_LP | CAACGATCTGTCAAAGCTGG | SALK_144391_RP | AAATTTGGTCCACAAATCAAG |
| At1g06580 | CS875502_LP | TCCTTGTCCAGTTACGGAATG | CS875502_RP | TTTCTTCCAAATGCAAGTAC |
| | SALK_022879C_LP | CCAAACACTTCAGCTTCAAAC | SALK_022879C_RP | AAATGGGAATTAGACCGGATG |
| | SALK_032203C_LP | ATAAGCCCATCTCTTTTTCG | SALK_032203C_RP | TGGAGATGTTGGGAATTTTAC |
| | SALK_032204_LP | ATAAGCCCATCTCTTTTTCG | SALK_032204_RP | TGGAGATGTTGGGAATTTTAC |

MATERIELS ET METHODES

| AGI | Nom | Séquence | Nom | Séquence |
|------------|-----------------|-----------------------------------|-----------------|-----------------------|
| At1g62670 | SALK_146198_LP | GTTAATTCACCTCGCAGATGC | SALK_146198_RP | TTACGGAAGATGGAGTGATGC |
| | SALK_146196_LP | GAAGCTTTTGTCCAATCTCCC | SALK_146196_RP | AAACATATGGATGATGCGCTC |
| At1g02060 | CS819524_LP | TTTATTCAACCGATTCCGTTG | CS819524_RP | GCTGTTGAAGTAGCGATCCTG |
| At2g01390 | SALK_038623_LP | TTGATGGAGTATCTTGTCCGCC | SALK_038623_RP | CCCCAAGAAGGCATTATCTC |
| | SALK_025648_LP | GGCTACGAATTCATTGACACAC | SALK_025648_RP | GATGAAGCAGAATCTGGTTGC |
| | SALK_049291_LP | CTGCCAAAGACAAGGCATAAG | SALK_049291_RP | TGATTACAGCTTGAAAATGGG |
| | SALK_049289_LP | CTGCCAAAGACAAGGCATAAG | SALK_049289_RP | TGATTACAGCTTGAAAATGGG |
| At2g32630 | SALK_043094_LP | CAAAAGCTGGCTTTTTAGCAG | SALK_043094_RP | CACCATCGCTGATCCTAACTC |
| | SALK_050020_LP | CAAAAGCTGGCTTTTTAGCAG | SALK_050020_RP | AAACCCTCTCAAACATTCCG |
| | SALK_043509_LP | CAAAAGCTGGCTTTTTAGCAG | SALK_043509_RP | AAACCCTCTCAAACATTCCG |
| At2g30780 | SALK_082869_LP | TTTTCAACTCAGAGACCTCGC | SALK_082869_RP | GAAGCAGTCTGCATTCTGTTT |
| At3g48810 | SALK_089911_LP | TTTGGGCAACATCCACTAGTC | SALK_089911_RP | GCAGCAAATGAAACTCAAGG |
| | SALK_009125C_LP | TCAC TTATCAACGGTTTTGCC | SALK_009125C_RP | AGCAGCTTCAGCAGATTCTTG |
| Salk lines | LBb1.3 | ATTTTGCCGATTTTCGGAAC | | |
| Sail lines | LB3 | TAGCATCTGAATTCATAACCAATCTCGATACAC | | |

IV.1.2 Lignées d'insertion

| AGI | T-DNA lines | Stock number ABRC |
|-----------|--------------|-------------------|
| At5g61990 | SALK_143317 | SALK_143317 |
| At3g09060 | SALK_144391 | SALK_144391 |
| At1g06580 | SAIL_633_D09 | CS875502 |
| | SALK_022879 | SALK_022879C |
| | SALK_032203 | SALK_032203C |
| | SALK_032204 | SALK_032204 |
| At1g62670 | SALK_146198 | SALK_146198 |
| | SALK_146196 | SALK_146196 |
| At1g02060 | SAIL_423_B07 | CS819524 |
| At2g01390 | SALK_038623 | SALK_038623 |
| | SALK_025648 | SALK_025648 |
| | SALK_049291 | SALK_049291 |
| | SALK_049289 | SALK_049289 |
| At2g32630 | SALK_043094 | SALK_043094 |
| | SALK_050020 | SALK_050020 |
| | SALK_043509 | SALK_043509 |
| At2g30780 | SALK_082869 | SALK_082869 |
| At3g48810 | SALK_089911 | SALK_089911 |
| At3g02330 | SALK_128453 | SALK_128453 |
| | SALK_027870 | SALK_027870 |
| | SALK_061252 | SALK_061252 |
| | SALK_061253 | SALK_061253 |
| At1g64310 | SALK_092788 | SALK_092788 |
| | SALK_069611 | SALK_069611C |
| At1g69350 | SALK_033399 | SALK_033399 |
| | SALK_033395 | SALK_033395 |
| | SALK_033396 | SALK_033396 |
| At1g03540 | SALK_035148 | SALK_035148 |
| | SALK_066190 | SALK_066190 |
| | SALK_043866 | SALK_043866 |
| At2g35030 | SALK_042973 | SALK_042973 |
| | SALK_115308 | SALK_115308 |
| At5g06540 | SAIL_77_E03 | CS803656 |
| At1g05750 | Salk_123752 | Salk_123752 |
| | Salk_104250 | Salk_104250 |

IV.2 Méthodes

IV.2.1 Méthodes relatives à la culture d'Arabidopsis

IV.2.1.1 Culture en terre

A Perth, les plantes ont été cultivées sur terreau (3/5 terreau, 1/5 vermiculite, 1/5 perlite), en chambre de culture confinée à 22°C et 60% d'humidité moyenne, 8H de jours 16H de nuit. A Evry, les plantes ont été cultivées sur terreau et également en chambre de culture confinée à 21°C et 45% d'humidité moyenne, 8H de jours 16H de nuit.

IV.2.1.2 Culture *in vitro*

En condition *in vitro*, les plantes ont été cultivées sur un milieu Murashige et Skoog appelé ½ MS. Pour 1L : 2,2g de MS (Sigma Aldrich™M5519 Murashige and Skoog Basal Medium), 0 à 3% de sucrose (m/v) ou 0 à 3% de glucose (m/v) et 8% d'agar type E (m/v) pour les cultures en milieu solide (A6674 sigma Aldrich©). Le pH est ajusté à 5,8 par ajout d'hydroxyde de potassium (KOH). Les conditions de culture de la chambre sont 25°C, 8 heures d'obscurité, 16h de luminosité.

IV.2.1.3 Préparation des graines d'Arabidopsis

IV.2.1.3.1 Vernalisation

Avant chaque semis, une vernalisation des graines est faite au préalable, la quantité de graines désirées est transférée dans un tube de 1,5 ml et 1 ml d'eau est ajouté. Le tube est alors conservé à 4°C pendant 48 heures avant semis. Cette vernalisation est aussi faite sur des graines venant d'être stérilisées en ajoutant également de l'eau mais stérile.

IV.2.1.3.2 Stérilisation des graines

Deux protocoles de stérilisation ont été employés selon le lieu où les expériences ont été faites.

A Perth, le protocole suivant a été utilisé :

Les graines sont incubées dans 1 ml d'éthanol 70 % (v/v) pendant 2 min, puis incubées dans 1 ml de la solution suivante : 5 % Javel (v/v)/ 0,1 % Tween20 (v/v) pendant 15 min en agitant toutes les 5 minutes. Les graines sont alors rincées 5 fois dans de l'eau stérile puis resuspendues une dernière fois dans 1 ml d'eau stérile. Les graines sont alors prêtes pour la vernalisation puis pour un semis en boîte de culture *in vitro*.

A Evry, le protocole suivant a été utilisé :

Une pastille de chlore est dissoute dans 50ml d'eau milliQ et 50 µl de tween 20 sont ajoutés. 1 ml de cette solution est ajouté à 9 ml d'éthanol technique 95 % (v/v). Les graines à stériliser sont alors incubées sous agitation légère dans 1 ml de cette nouvelle préparation pendant 7 min. 3 lavages à l'éthanol 96% sont ensuite effectués, suivi soit d'un séchage des graines sous la hotte soit de plusieurs rinçages à l'eau stérile en vue d'une vernalisation.

IV.2.2 Protocoles relatifs à l'ADN

IV.2.2.1 Extraction rapide d'ADN génomique

Ce protocole est utilisé en tube de 1,5 ml pour le génotypage de plantes.

| Tampon de Lyse : | Concentration finale |
|-------------------|----------------------|
| SDS | 0,5 % (m/v) |
| NaCl | 250 M |
| EDTA | 25 mM |
| Tris-HCl pH = 7,5 | 200 mM |

Une feuille est collectée et broyée dans de l'azote liquide pour chaque échantillon à génotyper. 400µl de Tampon de lyse sont alors ajoutés et le tout est mélangé sur un vortex pendant 15 secondes. Les tubes sont alors incubés pendant 15 minutes à 50°C puis centrifugés pendant 15 minutes à 5000 *G* à 4°C. 150 µl de surnageant sont transférés dans de nouveaux tubes de 1,5 ml et un volume de 150 µl d'isopropanol est ajouté. Le tout est alors mélangé sur un vortex pendant 15 secondes, puis incubé pendant 20 minutes à température ambiante et centrifugé à 4°C pendant 10 minutes à 10000 *G*. Les surnageants sont alors éliminés par renversement et les culots sont repris avec 125µl d'éthanol à 70%. Les échantillons sont alors centrifugés à 4°C pendant 5 minutes à 10000 *G*. Les surnageants sont de nouveau éliminés par renversement des tubes et les culots sont séchés en incubant les tubes ouverts dans une étuve à 60°C pendant 10 min. Une fois les culots secs, ceux ci sont repris dans 125µl de TE (10mM, 1mM pH 7.5) et 2µl de cette extraction seront utilisés pour une PCR.

IV.2.2.2 Extraction d'ADN génomique (Dneasy 96 Plant kit QIAGEN)

Ce protocole a été utilisé pour toutes les extractions d'ADN génomique en plaque 96 et toutes les centrifugations ont été faites à l'aide d'une centrifugeuse Eppendorf 5810 R.

Tampon AP1 (500 ml) :

| | |
|------------------|------------|
| Tris 1M pH8.0 | 50 ml |
| EDTA 0.5M | 50 ml |
| NaCl 2,5M | 100 ml |
| SDS 20% | 31,5 ml |
| PVP 40.000 | 5g |
| Sodium Bisulfite | 5g |
| Eau stérile | qsp 500 ml |

Tampon AP2 (200 ml) :

| | |
|------------------------------------|------------|
| Acétate de potassium (3M K, 5M Ac) | 120 ml |
| Acide acétique glacial | 23 ml |
| Eau | qsp 200 ml |

Tampon AP3 (150 ml) :

| | |
|--------------------|-----------|
| Guanidium Chloride | 37,25 g |
| Eau stérile | qsp 50 ml |
| Ethanol 96% | 100 ml |

Tampon AW (1 L) :

| | |
|----------------------|---------|
| Tris 1M pH 8.0 | 22,5 ml |
| EDTA 0,5 M | 200 µl |
| Potassium acétate 5M | 32 ml |
| Eau stérile | qsp 1L |

Le tampon AW/E est préparé en mélangeant 100 ml de solution AW et 170 ml d'éthanol 96%.

Une feuille est prélevée et placée dans un puits d'une plaque 96 contenant une bille d'acier. Après le prélèvement, la plaque est fermée à l'aide de bouchons et placée dans de l'azote liquide. Celle-ci est alors agitée dans un agitateur à peinture pendant 30 secondes puis centrifugée pour culoter tous les débris. Les bouchons sont retirés délicatement et 300 µl de tampon AP1 préchauffé à 65°C, contenant 125 mg/L de RNase A sont ajoutés dans chaque puit. Les bouchons replacés, la plaque est agitée énergétiquement et manuellement pendant 15 secondes de haut en bas, puis centrifugée pendant 5 secondes à 3000 rpm et à température ambiante. Les bouchons sont de nouveau retirés pour ajouter 100 µl de tampon AP2 puis refermés et la plaque est agitée énergétiquement pendant 15 secondes de haut en bas puis 5 secondes de retournement léger. La plaque, après avoir été centrifugé 5 secondes à 3000 rpm, est placée pendant 10 minutes à -20°C et est ensuite centrifugée pendant 10 minutes à 4100 rpm. 300 µl de surnageant sont ensuite transférés dans une nouvelle plaque et 1,5 volume de AP3/E sont ajoutés dans chaque puits. Après fermeture de la plaque par de nouveaux bouchons, celle ci est agitée énergétiquement pendant 15 secondes de haut en bas puis 5 secondes de retournement léger puis centrifugée 5 secondes à 3000 rpm. Une plaque de colonnes Dneasy 96 est placée sur une plaque 96 800 µl et 600 µl des échantillons sont transférés sur les colonnes. Ce montage est scellé avec une feuille « airpore tae sheet » et le tout est centrifugé pendant 5 minutes à 4100 rpm. La plaque 96 de 800 µl est vidée par retournement et la plaque de colonnes est replacée dessus. 600µl de tampon AW/E sont ajoutés à chaque colonne, la plaque est de nouveau scellée avec une feuille airpore et le montage est centrifugé 15 min à 4100 rpm. La plaque de colonnes est alors placée sur la plaque d'étude (nouvelle plaque 96 800µl), 100µl de tampon AE (10 mM TrisHCl ; 0,5mM EDTA ; pH9.0) sont ajoutés à chaque colonne et le tout est incubé 1 min à température ambiante. L'élution s'effectue en centrifugeant 3 minutes à 4100 rpm et la plaque de travail est alors conservée à -20°C et 2 µl seront utilisés pour la PCR de génotypage.

IV.2.2.3 Amplification d'ADN par PCR

IV.2.2.3.1 Méthode de PCR « classique »

Les réactions de PCR de type « classique » sont utilisées dans le cadre des réactions de test des colonies ou de génotypage.

Les réactions sont réalisées en utilisant une Taq Polymérase préparée au laboratoire d'Evry ou de Perth, dans un tampon adapté (Composition du tampon 10X : Tris-HCl 100mM pH8, MgCl₂ 25mM, KCl 500mM, Nonidet P40 0.5%) pour le test de colonies et une Taq Polymérase commerciale (Biolabs) dans certaines situations pour le génotypage.

Les réactions PCR ont été réalisées à l'aide d'un programme dont la température d'hybridation est adaptée aux différents couples d'oligonucléotides, et le temps d'élongation dépend de la taille du fragment à amplifier.

Protocole de PCR « classique »

Dans un volume final de 10 à 50 µl, 25 à 75 ng d'ADN sont utilisés comme matrice et sont mélangés avec du rouge de crésol 1X (solution 10 X : Rouge de crésol 2 mg/ml ; Saccharose 600 mg/ml), du tampon de Taq polymérase 1X, des dNTPs à 0,2 mM chacun, des oligonucléotides à 1 µM et de la Taq polymérase entre 1 à 5 unités (ou 0,2 µl de Taq polymérase produite dans le laboratoire).

La réaction d'amplification s'effectue dans un « thermocycler » en utilisant le programme suivant : une incubation de 5 minutes à 94°C permet une première dénaturation de la matrice ; Puis une dénaturation de 30 secondes à 94°C, suivie d'une hybridation de 30 secondes entre 50°C et 60°C selon les oligonucléotides utilisés et d'une élongation à 72°C d'une minute par kb à amplifier sont ensuite répétés 35 fois. La réaction est terminée par une élongation de 5 minutes à 72°C.

IV.2.2.3.2 Méthode de PCR « clonage »

Lorsque nous amplifions un fragment d'ADN par PCR afin de le cloner, les réactions de PCR sont réalisées comme décrit précédemment à la différence près que la Polymérase utilisée est une enzyme dite « fidèle » (avec une activité de relecture).

Différentes enzymes ont été utilisées telles que la iProof (Biorad) au laboratoire d'Evry et la DNA polymerase (Takara) au laboratoire de Perth, en utilisant les protocoles recommandés pour chaque Polymerase.

IV.2.2.3 PCR quantitative

Les réactions de PCR quantitatives ont été réalisées au laboratoire de Perth, dans la machine LightCycler 480 Real-time PCR system (Roche) à l'aide du master mix LightCycler 480 SYBR green 1 (Roche). Au laboratoire d'Evry, ces réactions ont été réalisées dans la machine Applied Biosystems 7900HT Sequence detection system à l'aide du MESA green qPCR Mastermix Plus for SYBR (Eurogentec). Les oligonucléotides utilisés figurent en Annexes des articles sur OTP43 et OTP51

IV.2.2.3 Electrophorèse

Les molécules d'ADN sont séparées sur gel d'agarose (concentration 0,8 à 2%) contenant du BET (Bromure d'éthidium) à une concentration de 0.5µg/ml dans un tampon de migration (TBE 1X ou TAE 1X) par application d'un champ électrique d'une tension variant entre 50 à 200 V suivant la taille du gel (7 à 30 cm de longueur). Pour aider au dépôt sur gel des échantillons d'ADN, du rouge de crésol est ajouté au mélange. La visualisation s'effectue sous UV.

TBE 1X, pH8 :

- 89 mM TRIS,
- 89 mM Acide borique,
- 2 mM EDTA

TAE 1X, pH8 :

- 89 mM TRIS,
- 89 mM Acide acétique,
- 2 mM EDTA

IV.2.2.4 Purification de produits PCR en vue d'un clonage ou d'un séquençage

IV.2.2.4.1 Utilisation du kit QIAquick (Qiagen)

Les produits de PCR destinés à être clonés ou séquencés, peuvent être directement extraits du gel d'électrophorèse purifiés et concentrés en utilisant le kit QIAquick et en suivant les indications du fournisseur.

IV.2.2.4.2 Purification au PEG

Alternativement, les produits PCR destinés à être clonés peuvent être précipités, purifiés et concentrés, par ajout de 5 volumes d'une solution de PEG/MgCl₂/TE (solution PEG/MgCl₂: 12mM MgCl₂, 12% PEG 8000 m/v, 6mM Tris, 0,6mM EDTA pH 7,5). Après centrifugation pendant 20 min à 14000 G à température ambiante, le culot est repris dans 20µL de TE (10mM Tris, 1mM EDTA pH 7,5).

IV.2.2.5 Clonage de produit PCR

IV.2.2.5.1 Clonage dans le vecteur pGEMT-Easy (Promega)

Les produits PCR purifiés peuvent être directement clonés dans le vecteur pGEMT-Easy en suivant les instructions du fournisseur (Promega), avant transformation de bactéries compétentes.

IV.2.2.5.2 Technologie de clonage GatewayTM (Invitrogen)

IV.2.2.5.2.1 Principe du clonage Gateway

Son principe repose sur les réactions de recombinaison permettant l'intégration et l'excision du phage lambda dans le génome d'*E. coli*. Les clonages sont réalisés en deux étapes. La première (réaction BP) consiste en la recombinaison directionnelle d'un gène d'intérêt avec un vecteur dit « donneur », dont le produit donne le vecteur dit « d'entrée ». Ce dernier sert de matrice à une seconde réaction (réaction LR) de recombinaison avec un vecteur de destination, aboutissant à un vecteur dit « d'expression ».

Les vecteurs donneurs et de destination utilisés au cours de ma thèse figurent dans la table en annexe tableau 2.

IV.2.2.5.2.2 PCR de clonage

A l'aide deux PCR successives, le fragment d'ADN à insérer est amplifié et les sites de recombinaisons *attB1* et *attB2* sont ajoutés. Les produits PCR obtenus sont purifiés par précipitation au PEG (comme décrit dans le paragraphe « purification au PEG »)

Protocole de PCR1 en vue d'un clonage Gateway

1 µl d'ADN (~50 ng/µl) est utilisé comme matrice et mélangé avec 1 µl de chaque oligonucléotides à 0,5 µM, de 0,2 µl de dNTP à 100mM, de 2 µl de tampon HF Biorad 5X, de 0,1 µl de Taq Iproof Biorad High Fidelity polymerase et complété avec de l'eau pour un volume final de 10 µl.

La réaction s'effectue dans un « thermocycler » en utilisant le programme suivant: 2 minutes à 98°C puis 10 cycles de 15 secondes à 98°C, de 30 secondes à 55°C et 30 secondes par kb à amplifier à 68°C et terminé par 5 minutes à 68°C.

Protocole de PCR2 en vue d'un clonage Gateway

Sur la réaction PCR1 sont ajoutés, 1 µl de l'amorce attB1 et attB2 à 10 µM, 0,8 µl de dNTP à 10 mM, 8 µl de tampon HF Biorad 5X, 0,2 µl de polymérase Iproof High Fidelity de chez Biorad et complété avec de l'eau pour un volume final de 40 µl.

La réaction s'effectue dans un « thermocycler » en utilisant le programme suivant: 2 minutes à 98°C puis 25 cycles de 15 secondes à 98°C, de 30 secondes à 55°C et 30 secondes par kb à amplifier à 72°C. La réaction est terminée par 5 minutes à 68°C.

IV.2.2.5.2.3 *Mutagenèse dirigée par PCR en vue d'un clonage dans un vecteur d'entrée du système Gateway*

Lors des expériences de complémentation du mutant *otp43* avec les différents isoformes d'épissage alternatif, j'ai voulu également compléter avec une forme ne pouvant plus épisser. Ainsi une mutagenèse ponctuelle sur les deux sites accepteurs AG a été réalisée par PCR, ainsi que l'ajout de sites de recombinaison Gateway.

Pour chaque site à muter, deux amorces en sens et antisens chevauchant le site ont été générées et sont représentées sur la figure 41 : 1-antisens (OTP43_Acceptor1rev) ; 2-sens (OTP43_Acceptor1Fw) ; 2-antisens (OTP43_Acceptor2rev) ; 3-sens (OTP43_Acceptor2Fw). Ces amorces ont été dessinées de façon à changer la séquence des deux sites accepteurs en AA au lieu de AG.

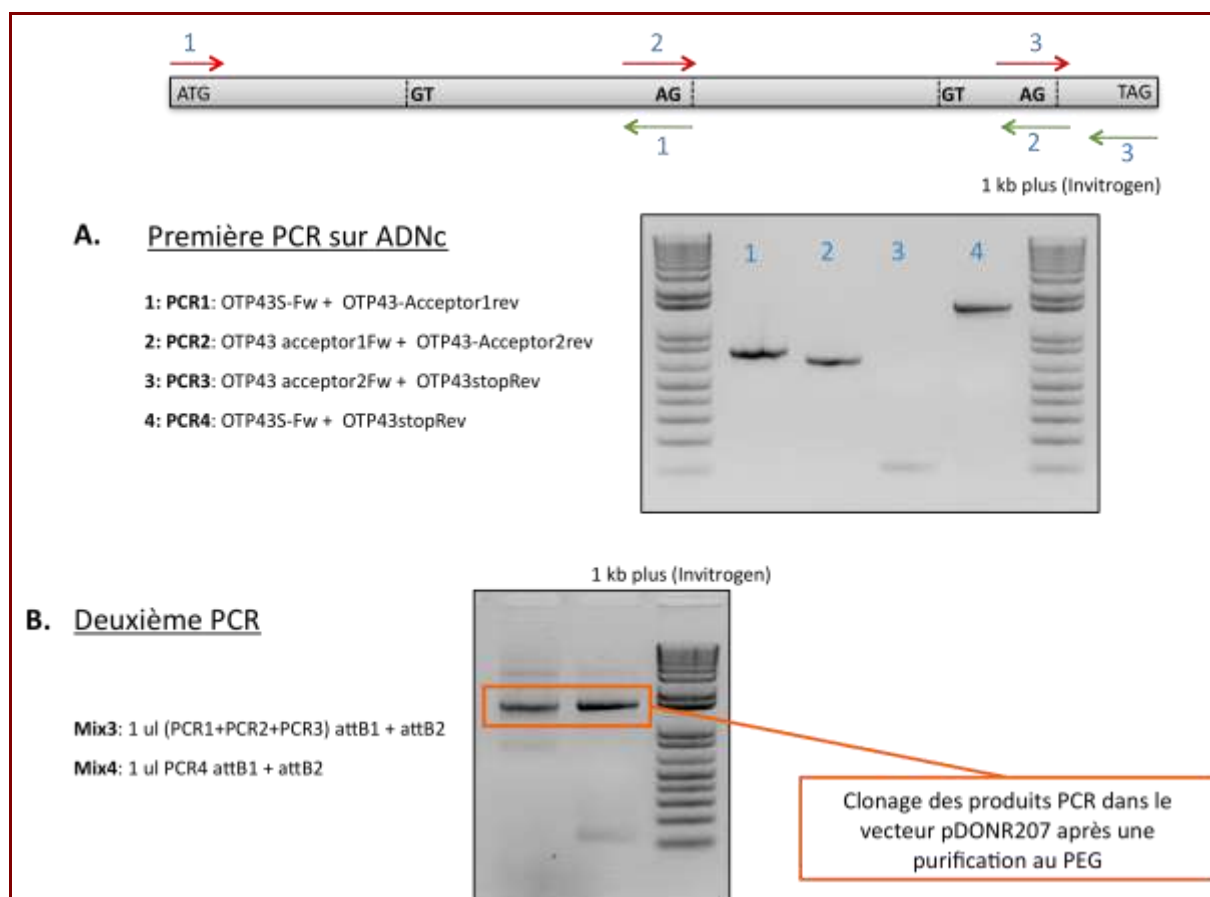


Figure 41 : Procédure de mutagenèse dirigée des sites accepteurs d'épissage d'*OTP43*.

Pour cela trois PCR indépendantes avec la Polymérase iProof de chez Biorad sont tout d'abord réalisées sur de l'ADNg: la première avec les amorces « start » de PPR43 (OTP43S-

Fw) et OTP43_Acceptor1rev, la deuxième avec les amorces OTP43_Acceptor1Fw avec OTP43_Acceptor2rev et la troisième en utilisant les amorces OTP43_Acceptor2Fw et l'amorce « stop » de PPR43 (OTP43stop_rev).

En parallèle, le clonage de la forme pleine longueur sans mutation est réalisé en utilisant les amorces « start » et « stop » de PPR43.

Par la suite, après vérification sur gel des amplifications, 1 µl de chacune des trois premières PCR est mélangé et utilisé comme matrice pour l'amplification avec les amorces attB1 et attB2, ces dernières permettant d'ajouter les sites de recombinaison Gateway.

Après vérification sur un gel contenant 1,2 % d'agarose dans du TAE, les produits PCR sont alors purifiés par la méthode PEG et clonés dans un vecteur pDONR207 par une réaction de BP clonase expliquée dans le paragraphe suivant.

IV.2.2.5.2.4 Réaction de BP clonase

La réaction de BP permet de cloner la séquence encadrée par les sites *attB1* et *attB2* dans un vecteur d'entrée contenant les sites *attP1* et *attP2*.

Protocole de réaction de BP clonase

2 µl de produit PCR purifié sont mélangés avec 100 ng de vecteur donneur (pDONR207 ou pDONR221), 1 µl de tampon de BP 5X, 0,4 µl de Clonase BP (Invitrogen) et complété à 5 µL avec du TE pH8. Le mélange réactionnel est incubé sur la nuit à 25°C. Puis la réaction est stoppée en ajoutant 0,5 µl de protéinase K (2 µg/µl) et en incubant 15 minutes à 37°C. La totalité de la réaction est utilisée pour la transformation de bactéries compétentes par choc thermique. Les colonies positives après PCR sont alors sélectionnées pour une minipréparation d'ADN. Les plasmides extraits, nommés vecteurs d' « entrée » seront utilisés pour la réaction LR.

IV.2.2.5.2.5 Réaction de LR clonase

La réaction de LR clonase va permettre de transférer les inserts clonés dans les vecteurs d' « entrée » dans un vecteur de destination entre les sites de recombinaison *attR1* et *attR2*

Protocole de réaction de LR clonase

100 ng de clone d'entrée sont ajoutés à 100ng de vecteur destination, 1 µl de tampon LR 5X, 0,4 µl de Clonase LR (Invitrogen) et complété à 5 µl final par du TE pH8. Le mélange réactionnel est incubé sur la nuit à 25°C. Puis la réaction est stoppée en ajoutant 0,5 µl de protéinase K (2 µg/µl) et en incubant 15 minutes à 37°C. La totalité de la réaction est utilisée pour la transformation de bactéries compétentes par choc thermique. Les colonies obtenues après sélection sur l'antibiotique adéquat sont testées par PCR.

IV.2.2.5.3 Transformation bactérienne, sélection et test des colonies transformées

Des cellules bactériennes de la souche DH5α compétentes (de l'ordre de 10^7 - 10^8 transformants par µg d'ADN) pour la transformation par choc thermique sont utilisées.

IV.2.2.5.3.1 Protocole de préparation de cellules DH5α Thermo-compétentes

Une colonie sur boîte est inoculée dans 10 ml de milieu LB et incubée à 37°C sous agitation (200 rpm) sur la nuit. Les 10 ml de la préculture sont ensuite ensemencés dans 500 ml de milieu SOB. La culture est alors mise sous agitation (200 rpm) à 19 °C jusqu'à l'obtention d'une DO à 600nm comprise entre 0,5 et 0,6, puis incubée 10 minutes dans la glace. La culture est alors centrifugée à 5000 G pendant 10 minutes à 4°C (en 2 flasques de

1L). Chaque culot est repris dans 80 ml de TB froid puis incubé 10 minutes dans la glace suivi d'une nouvelle centrifugation de 10 minutes à 5000 *G* à 4°C. Chaque culot est alors repris dans 20 ml de TB froid et 2,8 ml de DMSO sont ajoutés. Les bactéries sont alors aliquotées, congelées et conservées à -80°C.

Milieu SOB : 20g/L Bacto-tryptone, 5g/L Bacto-yeast extract, 10mM NaCl, 2,5mM KCl, 10mM MgCl₂, 10mM MgSO₄, pH ajusté à 7 avec du NaOH

TB : 10mM Pipes, 15mM CaCl₂, 250mM KCl, ajusté à pH 6,7 avec du KOH, 55mM MnCl₂, stérilisé par filtration (pas d'autoclave)

IV.2.2.5.3.2 Protocole de transformation par choc thermique de cellules DH5a compétentes

100µl de bactéries compétentes sont ajoutés à 10 pg de plasmide ou au mélange réactionnel de réaction BP ou LR. Le mélange est incubé 20 minutes dans un mélange glace/eau, puis une minute à 42°C dans un bain-marie (lorsque la réaction s'effectue en tube eppendorf) et tout de suite transféré dans un mélange glace/eau pendant 5 minutes. 900 µl de milieu SOC sont ensuite ajoutés et le mélange est incubé sous agitation pendant une heure à 37°C. 150 µl sont alors étalés sur boîte de milieu LB additionné de l'antibiotique approprié. Les boîtes sont incubées une nuit dans une étuve à 37°C et les colonies sont testées par PCR.

Milieu SOC : SOB + 20mM glucose

IV.2.2.5.4 Transformation d'Agrobactéries, sélection et test des colonies transformées

Des cellules bactériennes de la souche C58C1 pMP90 compétentes (de l'ordre de 10²-10³ transformants par µg d'ADN) pour la transformation par choc thermique sont utilisées.

IV.2.2.5.4.1 Protocole de préparation d'agrobactéries thermocompétentes

Une colonie sur boîte est inoculée dans 1,6 ml de milieu LB Agro et la préculture est incubée sous agitation (200 rpm) sur la nuit à 28°C. La préculture est ensuiteensemencée dans 40 ml de LB Agro et mise sous agitation (200 rpm) à 28°C jusqu'à l'obtention d'une DO à 600nm comprise entre 0,4 et 0,8. La culture est alors centrifugée à 5000 *G* pendant 15 minutes à 4°C. Le culot est repris dans 10mL 10 mM Tris-HCl pH 7,5 froid, puis de nouveau centrifugée à 5000 *G* pendant 15 minutes à 4°C. Le culot est enfin repris dans 10mL de Tris-HCl 10 mM pH 7,5 froid. Les agrobactéries sont alors aliquotées, congelées et conservées à -80°C.

LBAgro ; Bactotryptone : 10 g/l
Bacto Yeast Extract : 5 g/l
NaCl : 5 g/l
pH : 7,5

IV.2.2.5.4.2 Protocole de transformation par choc thermique de cellules C58C1 compétentes

200µl de bactéries compétentes sont ajoutés à 1µg de plasmide et incubés 5 minutes dans un mélange eau/glace, puis 5 minutes dans l'azote liquide et 5 minutes dans un bain-marie à 37°C. 800 µl de milieu LB Agro sont alors ajoutés et mis en incubation à 28°C sous agitation pendant deux heures. 300 µl sont ensuite étalés sur boîte de milieu LBAgro additionné de l'antibiotique approprié. Les boîtes sont placées dans une étuve à 28°C pendant deux jours et les colonies sont testées par PCR.

IV.2.2.6 Extraction et purification de plasmides (minipréparation d'ADN)

Selon les laboratoires, différents kits commerciaux ont été utilisés tels que :

-le kit SV Miniprep (Promega) au laboratoire d'Evry

-les kits : QIAprep Spin Miniprep Kit (Qiagen) et Plasmid Miniprep Kit (Invitrogen) au laboratoire de Perth.

La quantité de plasmide obtenue est systématiquement vérifiée par spectrométrie (mesure de la DO à 260nm).

IV.2.2.7 Séquençage d'ADN

Le séquençage des produits de PCR ou des ADN plasmidiques a été effectué par une entreprise externe, MacroGen (MacroGen, 908 World Meridian Venture Center, #60-24, Gasan-dong, Geumchun-gu, Seoul 153-781, Kore) ou GATC (GATC Biotech AG, Jacob-Stadler-Platz 7 78467 Konstanz, Germany).

IV.2.3 Méthodes relatives à l'ARN

IV.2.3.1 Extraction d'ARN total

L'ARN total est extrait à partir de quelques milligrammes d'un échantillon végétal (une feuille ou une plantule). L'échantillon prélevé et transféré dans un tube eppendorf de 2ml et est ensuite broyé après congélation dans l'azote liquide à l'aide d'une bille en acier et d'un agitateur prévu à cet effet. L'extraction en tant que telle est réalisée à l'aide du kit RNeasy plant minikit (Qiagen) en suivant les instructions du fournisseur.

IV.2.3.2 Traitement de décontamination de l'ADN contaminant

Les ARN totaux sont traités à la DNaseI (DNA-free™ DNase Treatment and Removal Reagents AM1906) suivant les instructions du fournisseur (Ambion®). Les ARN totaux sont quantifiés au spectrophotomètre par mesure de l'absorption à 260nm à l'aide d'un « nanodrop » et visualisés sur gel d'agarose 2 %, dans du tampon TAE ou TBE.

IV.2.3.3 Réaction de transcription « inverse »

Les ARNm des organites ayant une queue polyA sont, contrairement aux ARNm nucléaires, en phase de dégradation. C'est donc pour cette raison que la synthèse du premier brin d'ADN complémentaire d'ARN d'organites se fait à l'aide d'hexanucléotides aléatoires et non avec des oligonucléotides dT. En revanche, la synthèse du premier brin d'ADN complémentaire d'ARN nucléaires se fait à l'aide d'oligonucléotides dT. La synthèse d'ADN complémentaire est effectuée en utilisant la transcriptase inverse SuperscriptIII™ (11904-018, Invitrogen™) ou la reverse transcriptaseIII™ (18080-400, Invitrogen™) en suivant les instructions du fournisseur. La synthèse s'effectue à partir d'1 µg à 3 µg d'ARN décontaminé à la DNaseI.

IV.2.3.4 Northern blot

Le protocole de *northern blot* utilisé pour regarder l'accumulation des transcrits *nad6* dans les différents mutants de complexe 1 est le même que celui décrit dans l'article 2 sur OTP51.

Les amorces utilisées pour fabriquer la sonde biotinilée sont décrites dans le paragraphe IV.1.2.

IV.2.3.5 Protocole pour l'analyse PPE

Les ARN d'une plante sauvage et mutante sont extraits et des ADNc sont synthétisés. Alternativement, pour les témoins, 25pg d'ADN plasmidique (contenant la séquence du transcrit d'intérêt édité ou non édité) servent de matrice. Une PCR est effectuée en utilisant un couple d'amorces flanquant le site d'intérêt dans le transcrit. 2,5µl de PCR sont purifiés par 1µl d'enzyme ExoSAP-IT (USB 78200) et incubés à 37°C pendant 20 minutes pour éliminer les nucléotides et les oligonucléotides non incorporés. L'enzyme est inactivée à 80°C pendant 15 minutes. La réaction de PPE s'effectue par le biais d'un kit ThermoSequenase Cycle Sequencing kit (USB 78500).

Pour la réaction de l'extension PPE, 100ng de produit PCR purifié sont ajoutés à 2µl de tampon 10X, 0,5pmol d'amorce marquée en 5' à la 6-carboxyfluorescein (FAM), 1 unité de ThermoSequenase DNA polymérase, 100 nM de chaque nucléotides (mélange contenant trois dNTP+ un ddNTP) et complété à 20 µl avec de l'eau.

La réaction d'amplification est faite dans un « thermocycler » en utilisant le programme suivant : 1 minute à 94°C, suivie de 35 cycles de 30 secondes à 94°C, 30 secondes à 55°C et de 10 secondes à 68°C. La réaction est terminée par 3 minutes d'élongation à 68°C.

Le produit d'amplification est alors déposé sur un gel de séquençage à 12% acrylamide (Sequagel) et migré à 55W. Le gel est ensuite révélé sur un Typhoon Trio Imager (GE Healthcare, ex-Amersham), avec un filtre d'émission 520BP40. L'acquisition du signal fluorescent se fait à 488nm.

La quantification des bandes se fait grâce au logiciel ImageQuant TL 1D gel analysis (Amersham Bioscience). Les bandes correspondant aux produits épissés et non épissés sont manuellement déterminées, et le logiciel calcule un volume correspondant à l'intensité de fluorescence dans cette bande. Le bruit de fond est déterminé par la mesure d'une bande « absente » correspondant au produit épissé et non épissé, obtenu à partir des plasmides épissés et non épissés, respectivement.

IV.2.4 Méthodes propres à Arabidopsis

IV.2.4.1 Transformation d'Arabidopsis par « floral deeping »

Les plantes d'Arabidopsis sont transformées par trempage des boutons floraux dans une solution d'agrobactéries d'intérêt, selon la méthode décrite par Clough et Bent, (Clough and Bent 1998).

Protocole de transformation d'Arabidopsis :

Une culture de la souche d'agrobactérie d'intérêt est réalisée dans du LB-Agro jusqu'à une DO à 600 nm comprise entre 0,8 et 1,0 (le volume dépend de la quantité de plantes que l'on souhaite transformer). Les cellules sont ensuite centrifugées à 4000 G pendant 30 minutes et les culots sont resuspendus dans une solution 5% de sucrose, 0.02% de Silwet-L77 (volume identique au volume de culture). Les boutons floraux sont alors trempés pendant deux minutes dans la solution. Les plantes sont ensuite couvertes par un couvercle pendant 24 heures et cultivées jusqu'à l'obtention de graines.

IV.2.4.2 Etude du phénotype

L'analyse phénotypique des mutants consiste en l'étude de plusieurs caractères comme : le pourcentage de germination, la taille et l'aspect des germinations et des plantes adultes.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Asakura, Y. and A. Barkan (2006). "Arabidopsis orthologs of maize chloroplast splicing factors promote splicing of orthologous and species-specific group II introns." Plant Physiol **142**(4): 1656-63.
- Aubourg, S., N. Boudet, et al. (2000). "In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants." Plant Mol Biol **42**(4): 603-13.
- Baier, M., G. Hemmann, et al. (2004). "Characterization of mutants in *Arabidopsis* showing increased sugar-specific gene expression, growth, and developmental responses." Plant Physiol **134**(1): 81-91.
- Banihashemi, L., G. M. Wilson, et al. (2006). "Upf1/Upf2 regulation of 3' untranslated region splice variants of AUF1 links nonsense-mediated and A+U-rich element-mediated mRNA decay." Mol Cell Biol **26**(23): 8743-54.
- Barbazuk, W. B., Y. Fu, et al. (2008). "Genome-wide analyses of alternative splicing in plants: opportunities and challenges." Genome Res **18**(9): 1381-92.
- Barkan, A., M. Walker, et al. (1994). "A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms." Embo J **13**(13): 3170-81.
- Beick, S., C. Schmitz-Linneweber, et al. (2008). "The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts." Mol Cell Biol **28**(17): 5337-47.
- Bentolila, S., A. A. Alfonso, et al. (2002). "A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants." Proc Natl Acad Sci U S A **99**(16): 10887-92.
- Bentolila, S., A. L. Chateigner-Boutin, et al. (2005). "Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in *Arabidopsis*." Plant Physiol **139**(4): 2006-16.
- Bentolila, S., L. E. Elliott, et al. (2008). "Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*." Genetics **178**(3): 1693-708.
- Blatch, G. L. and M. Lassle (1999). "The tetratricopeptide repeat: a structural motif mediating protein-protein interactions." Bioessays **21**(11): 932-9.
- Brangeon, J., M. Sabar, et al. (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**(3): 269-80.
- Brown, G. G., N. Formanova, et al. (2003). "The radish Rfo restorer gene of *Ogura* cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats." Plant J **35**(2): 262-72.
- Cai, W., D. Ji, et al. (2009). "LPA66 is required for editing psbF chloroplast transcripts in *Arabidopsis*." Plant Physiol **150**(3): 1260-71.
- Chang, Y. F., J. S. Imam, et al. (2007). "The nonsense-mediated decay RNA surveillance pathway." Annu Rev Biochem **76**: 51-74.
- Chase, C. D. (2007). "Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions." Trends Genet **23**(2): 81-90.

REFERENCES

- Chateigner-Boutin, A. L., M. Ramos-Vega, et al. (2008). "CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts." *Plant J* **56**(4): 590-602.
- Chateigner-Boutin, A. L. and I. Small (2007). "A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons." *Nucleic Acids Res* **35**(17): e114.
- Chi, W., J. Ma, et al. (2008). "The pentatricopeptide repeat protein DELAYED GREENING1 is involved in the regulation of early chloroplast development and chloroplast gene expression in Arabidopsis." *Plant Physiol* **147**(2): 573-84.
- Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana." *Plant J* **16**(6): 735-43.
- Coffin, J. W., R. Dhillon, et al. (1997). "The Neurospora crassa cya-5 nuclear gene encodes a protein with a region of homology to the Saccharomyces cerevisiae PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein." *Curr Genet* **32**(4): 273-80.
- Cui, X., R. P. Wise, et al. (1996). "The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize." *Science* **272**(5266): 1334-6.
- Das, A. K., P. W. Cohen, et al. (1998). "The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions." *EMBO J* **17**(5): 1192-9.
- Davies, S. M., O. Rackham, et al. (2009). "Pentatricopeptide repeat domain protein 3 associates with the mitochondrial small ribosomal subunit and regulates translation." *FEBS Lett* **583**(12): 1853-8.
- de Longevialle, A. F., L. Hendrickson, et al. (2008). "The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana." *Plant J* **56**(1): 157-68.
- de Longevialle, A. F., E. H. Meyer, et al. (2007). "The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana." *Plant Cell* **19**(10): 3256-65.
- Delannoy, E., W. A. Stanley, et al. (2007). "Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles." *Biochem Soc Trans* **35**(Pt 6): 1643-7.
- Desloire, S., H. Gherbi, et al. (2003). "Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family." *EMBO Rep* **4**(6): 588-94.
- Diercks, T., E. Ab, et al. (2008). "Solution structure and characterization of the DNA-binding activity of the B3BP-Smr domain." *J Mol Biol* **383**(5): 1156-70.
- Ding, Y. H., N. Y. Liu, et al. (2006). "Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III." *Plant Cell* **18**(4): 815-30.
- Doolittle, W. F. and A. Stoltzfus (1993). "Molecular evolution. Genes-in-pieces revisited." *Nature* **361**(6411): 403.

REFERENCES

- Dutilleul, C., S. Driscoll, et al. (2003). "Functional mitochondrial complex I is required by tobacco leaves for optimal photosynthetic performance in photorespiratory conditions and during transients." *Plant Physiol* **131**(1): 264-75.
- Dutilleul, C., C. Lelarge, et al. (2005). "Mitochondria-driven changes in leaf NAD status exert a crucial influence on the control of nitrate assimilation and the integration of carbon and nitrogen metabolism." *Plant Physiol* **139**(1): 64-78.
- Ellis, T. P., K. G. Helfenbein, et al. (2004). "Aep3p stabilizes the mitochondrial bicistronic mRNA encoding subunits 6 and 8 of the H⁺-translocating ATP synthase of *Saccharomyces cerevisiae*." *J Biol Chem* **279**(16): 15728-33.
- Fahlgren, N., M. D. Howell, et al. (2007). "High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes." *PLoS One* **2**(2): e219.
- Filichkin, S. A., H. D. Priest, et al. (2010). "Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*." *Genome Res* **20**(1): 45-58.
- Fukui, K., H. Kosaka, et al. (2007). "Nuclease activity of the MutS homologue MutS2 from *Thermus thermophilus* is confined to the Smr domain." *Nucleic Acids Res* **35**(3): 850-60.
- Galtier, N. and M. Gouy (1995). "Inferring phylogenies from DNA sequences of unequal base compositions." *Proc Natl Acad Sci U S A* **92**(24): 11317-21.
- Garmier, M., A. J. Carroll, et al. (2008). "Complex I dysfunction redirects cellular and mitochondrial metabolism in *Arabidopsis*." *Plant Physiol* **148**(3): 1324-41.
- Giege, P. and A. Brennicke (1999). "RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs." *Proc Natl Acad Sci U S A* **96**(26): 15324-9.
- Gilbert, W. (1978). "Why genes in pieces?" *Nature* **271**(5645): 501.
- Gillman, J. D., S. Bentolila, et al. (2007). "The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus." *Plant J* **49**(2): 217-27.
- Hajdukiewicz, P. T., L. A. Allison, et al. (1997). "The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids." *EMBO J* **16**(13): 4041-8.
- Hammani, K., K. Okuda, et al. (2009). "A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites." *Plant Cell* **21**(11): 3686-99.
- Hashimoto, M., T. Endo, et al. (2003). "A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in *Arabidopsis*." *Plant J* **36**(4): 541-9.
- Hattori, M., H. Miyake, et al. (2007). "A Pentatricopeptide repeat protein is required for RNA processing of *clpP* Pre-mRNA in moss chloroplasts." *J Biol Chem* **282**(14): 10773-82.
- Hattori, M. and M. Sugita (2009). "A moss pentatricopeptide repeat protein binds to the 3' end of plastid *clpP* pre-mRNA and assists with mRNA maturation." *Febs J* **276**(20): 5860-9.
- Heazlewood, J. L., K. A. Howell, et al. (2003). "Mitochondrial complex I from *Arabidopsis* and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits." *Biochim Biophys Acta* **1604**(3): 159-69.

REFERENCES

- Hilson, P., J. Allemeersch, et al. (2004). "Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications." Genome Res **14**(10B): 2176-89.
- Hilson, P., I. Small, et al. (2003). "European consortia building integrated resources for Arabidopsis functional genomics." Curr Opin Plant Biol **6**(5): 426-9.
- Ikeda, T. M. and M. W. Gray (1999). "Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria." Mol Cell Biol **19**(12): 8113-22.
- Jenkins, B. D., D. J. Kulhanek, et al. (1997). "Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors." Plant Cell **9**(3): 283-96.
- Johnson, X., K. Wostrikoff, et al. (2010). "MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of *rbcl* mRNA in *Chlamydomonas* and Arabidopsis." Plant Cell **22**(1): 234-48.
- Jonietz, C., J. Forner, et al. (2010). "RNA PROCESSING FACTOR2 is required for 5' end processing of *nad9* and *cox3* mRNAs in mitochondria of Arabidopsis thaliana." Plant Cell **22**(2): 443-53.
- Kanamaru, K. and K. Tanaka (2004). "Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants." Biosci Biotechnol Biochem **68**(11): 2215-23.
- Kazama, T. and K. Toriyama (2003). "A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice." FEBS Lett **544**(1-3): 99-102.
- Keren, H., G. Lev-Maor, et al. (2010). "Alternative splicing and evolution: diversification, exon definition and function." Nat Rev Genet **11**(5): 345-55.
- Kim, E., A. Magen, et al. (2007). "Different levels of alternative splicing among eukaryotes." Nucleic Acids Res **35**(1): 125-31.
- Kim, S. R., J. I. Yang, et al. (2009). "Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria." Plant J **59**(5): 738-49.
- Klein, R. R., P. E. Klein, et al. (2005). "Fertility restorer locus Rf1 [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12." Theor Appl Genet **111**(6): 994-1012.
- Klevebring, D., N. R. Street, et al. (2009). "Genome-wide profiling of populus small RNAs." BMC Genomics **10**: 620.
- Kode, V., E. A. Mudd, et al. (2005). "The tobacco plastid *accD* gene is essential and is required for leaf development." Plant J **44**(2): 237-44.
- Koizuka, N., R. Imai, et al. (2003). "Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish." Plant J **34**(4): 407-15.
- Konishi, T., K. Shinohara, et al. (1996). "Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme." Plant Cell Physiol **37**(2): 117-22.
- Koonin, E. V. (2006). "The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?" Biol Direct **1**: 22.

REFERENCES

- Kotera, E., M. Tasaka, et al. (2005). "A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts." *Nature* **433**(7023): 326-30.
- Koussevitzky, S., A. Nott, et al. (2007). "Signals from chloroplasts converge to regulate nuclear gene expression." *Science* **316**(5825): 715-9.
- Kurihara-Yonemoto, S. and H. Handa (2001). "Low temperature affects the processing pattern and RNA editing status of the mitochondrial cox2 transcripts in wheat." *Curr Genet* **40**(3): 203-8.
- Lahmy, S., F. Barneche, et al. (2000). "A chloroplastic RNA-binding protein is a new member of the PPR family." *FEBS Lett* **480**(2-3): 255-60.
- Lareau, L. F., M. Inada, et al. (2007). "Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements." *Nature* **446**(7138): 926-9.
- Lee, B. H., H. Lee, et al. (2002). "A mitochondrial complex I defect impairs cold-regulated nuclear gene expression." *Plant Cell* **14**(6): 1235-51.
- Lee, C., A. S. Tibbetts, et al. (2009). "Yeast AEP3p is an accessory factor in initiation of mitochondrial translation." *J Biol Chem* **284**(49): 34116-25.
- Li, Y., K. K. Lee, et al. (2006). "Establishing glucose- and ABA-regulated transcription networks in Arabidopsis by microarray analysis and promoter classification using a Relevance Vector Machine." *Genome Res* **16**(3): 414-27.
- Li, Y., C. Smith, et al. (2007). "Signaling from an altered cell wall to the nucleus mediates sugar-responsive growth and development in Arabidopsis thaliana." *Plant Cell* **19**(8): 2500-15.
- Liere, K. and P. Maliga (1999). "In vitro characterization of the tobacco rpoB promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters." *EMBO J* **18**(1): 249-57.
- Lightowers, R. N. and Z. M. Chrzanowska-Lightowers (2008). "PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression." *Biochem J* **416**(1): e5-6.
- Loiselay, C., N. J. Gumpel, et al. (2008). "Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in Chlamydomonas reinhardtii chloroplasts." *Mol Cell Biol* **28**(17): 5529-42.
- Lurin, C., C. Andres, et al. (2004). "Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis." *Plant Cell* **16**(8): 2089-103.
- Lynch, M., B. Koskella, et al. (2006). "Mutation pressure and the evolution of organelle genomic architecture." *Science* **311**(5768): 1727-30.
- Maier, U. G., A. Bozarth, et al. (2008). "Complex chloroplast RNA metabolism: just debugging the genetic programme?" *BMC Biol* **6**: 36.
- Malek, O. and V. Knoop (1998). "Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort." *Rna* **4**(12): 1599-609.
- Malik, H. S. and S. Henikoff (2000). "Dual recognition-incision enzymes might be involved in mismatch repair and meiosis." *Trends Biochem Sci* **25**(9): 414-8.

REFERENCES

- Mancebo, R., X. Zhou, et al. (2001). "BSF binds specifically to the bicoid mRNA 3' untranslated region and contributes to stabilization of bicoid mRNA." Mol Cell Biol **21**(10): 3462-71.
- Manthey, G. M. and J. E. McEwen (1995). "The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*." Embo J **14**(16): 4031-43.
- Meierhoff, K., S. Felder, et al. (2003). "HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs." Plant Cell **15**(6): 1480-95.
- Meyer, E. H., T. Tomaz, et al. (2009). "Remodeled respiration in *ndufs4* with low phosphorylation efficiency suppresses Arabidopsis germination and growth and alters control of metabolism at night." Plant Physiol **151**(2): 603-19.
- Mili, S. and S. Pinol-Roma (2003). "LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs." Mol Cell Biol **23**(14): 4972-82.
- Mita, S., K. Suzuki-Fujii, et al. (1995). "Sugar-inducible expression of a gene for beta-amylase in *Arabidopsis thaliana*." Plant Physiol **107**(3): 895-904.
- Modrek, B. and C. Lee (2002). "A genomic view of alternative splicing." Nat Genet **30**(1): 13-9.
- Moore, B., L. Zhou, et al. (2003). "Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling." Science **300**(5617): 332-6.
- Mootha, V. K., P. Lepage, et al. (2003). "Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics." Proc Natl Acad Sci U S A **100**(2): 605-10.
- Moreira, D. and H. Philippe (1999). "Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family." Trends Biochem Sci **24**(8): 298-300.
- Moreno, J. I., K. S. Buie, et al. (2009). "Ccm1p/Ygr150cp, a pentatricopeptide repeat protein, is essential to remove the fourth intron of both COB and COX1 pre-mRNAs in *Saccharomyces cerevisiae*." Curr Genet **55**(4): 475-84.
- Naithani, S., S. A. Saracco, et al. (2003). "Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of *Saccharomyces cerevisiae*." Mol Biol Cell **14**(1): 324-33.
- Nakagawa, N. and N. Sakurai (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**(6): 772-83.
- Nakamura, T., K. Meierhoff, et al. (2003). "RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA." Eur J Biochem **270**(20): 4070-81.
- Nakamura, T. and M. Sugita (2008). "A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity." FEBS Lett **582**(30): 4163-8.

REFERENCES

- Narsai, R., K. A. Howell, et al. (2007). "Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*." *Plant Cell* **19**(11): 3418-36.
- O'Toole, N., M. Hattori, et al. (2008). "On the expansion of the pentatricopeptide repeat gene family in plants." *Mol Biol Evol* **25**(6): 1120-8.
- Oberdoerffer, S., L. F. Moita, et al. (2008). "Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPL." *Science* **321**(5889): 686-91.
- Ohta, H., A. Ogino, et al. (2010). "Fertility restoration by *Ifr1* in rice with BT-type cytoplasmic male sterility is associated with a reduced level, but not processing, of *atp6-orf79* co-transcribed RNA." *Plant Cell Rep* **29**(4): 359-69.
- Okuda, K., A. L. Chateigner-Boutin, et al. (2009). "Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts." *Plant Cell* **21**(1): 146-56.
- Okuda, K., K. Hammani, et al. (2010). "The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts." *Plant J* **61**(2): 339-49.
- Okuda, K., F. Myouga, et al. (2007). "Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing." *Proc Natl Acad Sci U S A* **104**(19): 8178-83.
- Okuda, K., T. Nakamura, et al. (2006). "A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing." *J Biol Chem* **281**(49): 37661-7.
- Pfalz, J., O. A. Bayraktar, et al. (2009). "Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts." *EMBO J* **28**(14): 2042-52.
- Pfalz, J., K. Liere, et al. (2006). "pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression." *Plant Cell* **18**(1): 176-97.
- Prikryl, J., K. P. Watkins, et al. (2008). "A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis." *Nucleic Acids Res* **36**(16): 5152-65.
- Puchta, O., M. Lubas, et al. (2010). "DMR1 (CCM1/YGR150C) of *Saccharomyces cerevisiae* encodes an RNA-binding protein from the pentatricopeptide repeat family required for the maintenance of the mitochondrial 15S ribosomal RNA." *Genetics* **184**(4): 959-73.
- Rackham, O., S. M. Davies, et al. (2009). "Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells." *Nucleic Acids Res* **37**(17): 5859-67.
- Reddy, A. S. (2007). "Alternative splicing of pre-messenger RNAs in plants in the genomic era." *Annu Rev Plant Biol* **58**: 267-94.
- Rhoades, M. W., B. J. Reinhart, et al. (2002). "Prediction of plant microRNA targets." *Cell* **110**(4): 513-20.
- Robbins, J. C., W. P. Heller, et al. (2009). "A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast *accD* transcript." *RNA* **15**(6): 1142-53.
- Rook, F., F. Corke, et al. (2001). "Impaired sucrose-induction mutants reveal the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling." *Plant J* **26**(4): 421-33.

REFERENCES

- Salone, V., M. Rudinger, et al. (2007). "A hypothesis on the identification of the editing enzyme in plant organelles." FEBS Lett **581**(22): 4132-8.
- Samson, F., V. Brunaud, et al. (2004). "FLAGdb++: a database for the functional analysis of the Arabidopsis genome." Nucleic Acids Res **32**(Database issue): D347-50.
- Schmitz-Linneweber, C. and I. Small (2008). "Pentatricopeptide repeat proteins: a socket set for organelle gene expression." Trends Plant Sci **13**(12): 663-70.
- Schmitz-Linneweber, C., R. Williams-Carrier, et al. (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**(10): 2791-804.
- Schmitz-Linneweber, C., R. E. Williams-Carrier, et al. (2006). "A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA." Plant Cell **18**(10): 2650-63.
- Shadel, G. S. (2004). "Coupling the mitochondrial transcription machinery to human disease." Trends Genet **20**(10): 513-9.
- Sheen, J., L. Zhou, et al. (1999). "Sugars as signaling molecules." Curr Opin Plant Biol **2**(5): 410-8.
- Small, I., N. Peeters, et al. (2004). "Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences." Proteomics **4**(6): 1581-90.
- Small, I. D. and N. Peeters (2000). "The PPR motif - a TPR-related motif prevalent in plant organellar proteins." Trends Biochem Sci **25**(2): 46-7.
- Sterky, F. H. and N. G. Larsson (2008). "Complex I: a complex gateway to the powerhouse." Cell Metab **7**(4): 278-9.
- Sung, T. Y., C. C. Tseng, et al. (2010). "The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria." Plant J.
- Sunkar, R. and J. K. Zhu (2004). "Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis." Plant Cell **16**(8): 2001-19.
- Takenaka, M., D. Verbitskiy, et al. (2010). "Reverse genetic screening identifies five E-class PPR-proteins involved in RNA editing in mitochondria of Arabidopsis Thaliana." J Biol Chem.
- Tasaki, E., M. Hattori, et al. (2010). "The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmFc transcript." Plant J.
- Tavares-Carreón, F., Y. Camacho-Villasana, et al. (2008). "The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial COX1 mRNA in yeast." J Biol Chem **283**(3): 1472-9.
- Till, B., C. Schmitz-Linneweber, et al. (2001). "CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin." RNA **7**(9): 1227-38.
- Tillich, M., H. T. Funk, et al. (2005). "Editing of plastid RNA in Arabidopsis thaliana ecotypes." Plant J **43**(5): 708-15.
- Tseng, C. C., T. Y. Sung, et al. (2010). "Editing of accD and ndhF chloroplast transcripts is partially affected in the Arabidopsis vanilla cream1 mutant." Plant Mol Biol **73**(3): 309-23.

REFERENCES

- Uyttewaal, M., N. Arnal, et al. (2008). "Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for *Ogura* cytoplasmic male sterility." *Plant Cell* **20**(12): 3331-45.
- Uyttewaal, M., H. Mireau, et al. (2008). "PPR336 is associated with polysomes in plant mitochondria." *J Mol Biol* **375**(3): 626-36.
- Van Leene, J., E. Witters, et al. (2008). "Boosting tandem affinity purification of plant protein complexes." *Trends Plant Sci* **13**(10): 517-20.
- Verbitskiy, D., A. Zehrmann, et al. (2010). "The PPR protein encoded by the LOVASTATIN INSENSITIVE 1 gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*." *Plant J* **61**(3): 446-55.
- Vidal, G., M. Ribas-Carbo, et al. (2007). "Lack of respiratory chain complex I impairs alternative oxidase engagement and modulates redox signaling during elicitor-induced cell death in tobacco." *Plant Cell* **19**(2): 640-55.
- Wakasugi, T., T. Tsudzuki, et al. (2001). "The genomics of land plant chloroplasts: Gene content and alteration of genomic information by RNA editing." *Photosynth Res* **70**(1): 107-18.
- Wang, E. T., R. Sandberg, et al. (2008). "Alternative isoform regulation in human tissue transcriptomes." *Nature* **456**(7221): 470-6.
- Wang, Z., Y. Zou, et al. (2006). "Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing." *Plant Cell* **18**(3): 676-87.
- Weisshaar, B. and G. I. Jenkins (1998). "Phenylpropanoid biosynthesis and its regulation." *Curr Opin Plant Biol* **1**(3): 251-7.
- Wiebauer, K., J. J. Herrero, et al. (1988). "Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals." *Mol Cell Biol* **8**(5): 2042-51.
- Williams, P. M. and A. Barkan (2003). "A chloroplast-localized PPR protein required for plastid ribosome accumulation." *Plant J* **36**(5): 675-86.
- Williams-Carrier, R., T. Kroeger, et al. (2008). "Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand." *Rna* **14**(9): 1930-41.
- Xu, F., C. Ackerley, et al. (2008). "Disruption of a mitochondrial RNA-binding protein gene results in decreased cytochrome b expression and a marked reduction in ubiquinol-cytochrome c reductase activity in mouse heart mitochondria." *Biochem J* **416**(1): 15-26.
- Xu, F., C. Morin, et al. (2004). "The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA." *Biochem J* **382**(Pt 1): 331-6.
- Yamazaki, H., M. Tasaka, et al. (2004). "PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*." *Plant J* **38**(1): 152-63.
- Yu, Q. B., Y. Jiang, et al. (2009). "AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript accD RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*." *Plant J* **59**(6): 1011-23.

REFERENCES

- Zehrmann, A., J. A. van der Merwe, et al. (2008). "Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*." Mitochondrion **8**(4): 319-27.
- Zehrmann, A., D. Verbitskiy, et al. (2009). "A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*." Plant Cell **21**(2): 558-67.
- Zhou, W., Y. Cheng, et al. (2008). "The *Arabidopsis* gene YS1 encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth." Plant J.
- Zoschke, R., M. Nakamura, et al. (2010). "An organellar maturase associates with multiple group II introns." Proc Natl Acad Sci U S A **107**(7): 3245-50.
- Zsigmond, L., G. Rigo, et al. (2008). "*Arabidopsis* PPR40 connects abiotic stress responses to mitochondrial electron transport." Plant Physiol **146**(4): 1721-37.

ANNEXES :

Tableau 1 : Sondes affymetrix pour chaque gène *PPR*

| AGI code | Probe set id | AGI code | Probe set id | AGI code | Probe set id |
|-----------|--------------|-----------|--------------|-----------|--------------|
| At1g01970 | 261622_at | At2g02750 | 267473_at | At3g62540 | 251211_s_at |
| At1g02060 | 264167_at | At2g02980 | 266794_at | At4g01030 | 255601_at |
| At1g02150 | 264177_at | At2g03380 | 265708_at | At4g01570 | 255588_at |
| At1g02370 | 259444_at | At2g03880 | 263362_at | At4g01990 | 255557_at |
| At1g02420 | 259414_at | At2g04860 | 263623_at | At4g02750 | 255446_at |
| At1g03100 | 263143_at | At2g06000 | 266028_at | At4g02820 | 255449_at |
| At1g03510 | 264819_at | At2g13600 | 263719_at | At4g04370 | 255329_at |
| At1g03540 | 264817_at | At2g15630 | 257395_at | At4g08210 | 255153_at |
| At1g03560 | 264816_at | At2g15690 | 265474_at | At4g11690 | 254880_at |
| At1g04840 | 261172_at | At2g15820 | 265484_at | At4g13650 | 254725_at |
| At1g05600 | 263200_at | At2g16880 | 266529_at | At4g14170 | 245595_at |
| At1g05670 | 263230_at | At2g17033 | 263550_at | At4g14190 | 245597_at |
| At1g05750 | 261310_at | At2g17210 | 263418_at | At4g14820 | 245580_at |
| At1g06140 | 260952_at | At2g17670 | 264591_at | At4g14850 | 245582_at |
| At1g06150 | 260951_at | At2g18520 | 265931_at | At4g15720 | 245508_at |
| At1g06270 | 260792_at | At2g18940 | 266951_at | At4g16390 | 245494_at |
| At1g06580 | 262627_at | At2g19280 | 267331_at | At4g16470 | 245498_at |
| At1g06710 | 260826_at | At2g20540 | 263364_at | At4g18750 | 254639_at |
| At1g07590 | 261088_at | At2g20710 | 265429_at | At4g18840 | 254588_at |
| At1g07740 | 261463_at | At2g21090 | 264011_at | At4g19195 | 254614_at |
| At1g08070 | 260620_at | At2g22070 | 263353_at | At4g19220 | 254556_at |
| At1g08610 | 264806_at | At2g22410 | 264047_at | At4g19440 | 254581_at |
| At1g09190 | 264251_at | At2g25580 | 265901_at | At4g19900 | 254552_at |
| At1g09220 | 264258_at | At2g26790 | 266853_at | At4g20090 | 254498_at |
| At1g09410 | 264503_at | At2g27610 | 266255_at | At4g20740 | 254483_at |
| At1g09680 | 264666_at | At2g27800 | 266241_s_at | At4g20770 | 254486_at |
| At1g09820 | 264713_at | At2g28050 | 266164_at | At4g21170 | 254457_at |
| At1g09900 | 264689_at | At2g29760 | 266668_at | At4g21300 | 254401_at |
| At1g10270 | 264452_at | At2g30780 | 267568_at | At4g25270 | 254051_at |
| At1g10330 | 264455_at | At2g31400 | 263242_at | At4g26680 | 253979_at |
| At1g10910 | 260465_at | At2g32630 | 267086_at | At4g26800 | 253932_at |
| At1g11290 | 262509_at | At2g33680 | 267445_at | At4g28010 | 253847_at |
| At1g11630 | 262822_at | At2g33760 | 267455_at | At4g30700 | 253584_at |
| At1g11710 | 262810_at | At2g34370 | 267047_at | At4g30825 | 253589_at |
| At1g11900 | 264392_at | At2g34400 | 266998_at | At4g31070 | 253576_at |
| At1g12620 | 255938_at | At2g35030 | 267420_at | At4g31850 | 253495_at |
| At1g12700 | 255935_at | At2g36240 | 263958_at | At4g32430 | 253430_at |
| At1g12770 | 261194_at | At2g36730 | 265195_at | At4g32450 | 253432_at |
| At1g13040 | 257465_at | At2g36980 | 263887_at | At4g33170 | 253365_at |
| At1g13630 | 256095_at | At2g37310 | 266002_at | At4g33990 | 253319_at |
| At1g13800 | 259449_at | At2g37320 | 266003_at | At4g35130 | 253170_at |
| At1g14470 | 261472_at | At2g38420 | 267031_at | At4g35850 | 253116_at |
| At1g15480 | 262586_at | At2g39230 | 267011_at | At4g36680 | 246249_at |
| At1g15510 | 261762_at | At2g39620 | 266973_at | At4g37170 | 246246_at |
| At1g16830 | 256107_at | At2g40240 | 263822_at | At4g37380 | 253092_at |
| At1g17630 | 260736_at | At2g40720 | 266102_at | At4g38010 | 253019_at |
| At1g18485 | 261715_at | At2g41080 | 267065_at | At4g38150 | 253027_at |
| At1g18900 | 259461_at | At2g42920 | 265267_at | At4g39530 | 252899_at |
| At1g19290 | 260654_at | At2g44880 | 266810_at | At5g01110 | 251128_at |
| At1g19520 | 260663_at | At2g45350 | 245129_at | At5g02860 | 250987_at |

ANNEXES

| | | | | | |
|-----------|-------------|-----------|-------------|-----------|-------------|
| At1g19520 | 260670_at | At2g46050 | 266602_at | At5g04810 | 250856_at |
| At1g20230 | 261277_at | At3g01580 | 259183_at | At5g06400 | 250725_at |
| At1g20300 | 255888_at | At3g02010 | 259003_at | At5g06540 | 250678_at |
| At1g22830 | 264749_at | At3g02330 | 256325_at | At5g08310 | 246052_at |
| At1g22960 | 264723_at | At3g02490 | 258502_at | At5g08490 | 250521_at |
| At1g25360 | 245630_at | At3g02650 | 258474_at | At5g08510 | 250523_at |
| At1g26460 | 261014_at | At3g03580 | 259218_at | At5g09450 | 245888_at |
| At1g26500 | 261003_at | At3g04130 | 258592_at | At5g09950 | 250508_at |
| At1g26900 | 263684_at | At3g04750 | 258817_at | At5g11310 | 250388_at |
| At1g28690 | 262763_at | At3g04760 | 258804_at | At5g12100 | 250303_at |
| At1g29710 | 246632_at | At3g05240 | 259355_at | At5g13230 | 250280_at |
| At1g30610 | 263220_at | At3g08820 | 258991_at | At5g13270 | 250283_at |
| At1g31430 | 256484_at | At3g09040 | 259206_at | At5g13770 | 250257_at |
| At1g31790 | 246262_at | At3g09060 | 259208_at | At5g14080 | 250215_at |
| At1g31840 | 246258_at | At3g09650 | 258696_at | At5g14770 | 246586_at |
| At1g31920 | 246313_at | At3g11460 | 259285_at | At5g14820 | 251211_s_at |
| At1g33350 | 256532_at | At3g12770 | 257685_at | At5g15010 | 246572_at |
| At1g43010 | 259622_at | At3g13150 | 257188_at | At5g15280 | 250164_at |
| At1g50270 | 262465_at | At3g13160 | 257193_at | At5g15300 | 250166_at |
| At1g52620 | 262144_at | At3g13770 | 256775_at | At5g15340 | 250169_at |
| At1g52640 | 262142_at | At3g13880 | 257607_at | At5g15980 | 246478_at |
| At1g55630 | 264540_at | At3g14330 | 258355_at | At5g16420 | 250115_at |
| At1g55890 | 260595_at | At3g14580 | 258088_at | At5g16640 | 250106_at |
| At1g56570 | 245680_at | At3g14730 | 258120_at | At5g16860 | 246458_at |
| At1g56690 | 245673_at | At3g15130 | 256886_at | At5g18390 | 250040_at |
| At1g59720 | 262907_at | At3g15200 | 256884_at | At5g18475 | 250011_at |
| At1g60770 | 264913_at | At3g15590 | 257291_at | At5g18950 | 249952_at |
| At1g61870 | 264286_at | At3g15930 | 257796_at | At5g19020 | 249965_at |
| At1g62260 | 264739_at | At3g16010 | 258334_at | At5g24830 | 246964_at |
| At1g62350 | 260631_at | At3g16610 | 258428_at | At5g25630 | 246901_at |
| At1g62590 | 265106_s_at | At3g16710 | 258422_at | At5g27110 | 246808_at |
| At1g62670 | 262687_at | At3g16890 | 257928_at | At5g27270 | 246818_at |
| At1g62680 | 262688_at | At3g18110 | 258149_at | At5g27460 | 246770_at |
| At1g62720 | 262690_at | At3g18970 | 256919_at | At5g28460 | 246173_s_at |
| At1g62915 | 261104_at | At3g20730 | 257979_at | At5g37570 | 249624_at |
| At1g63070 | 259684_at | At3g21470 | 258174_at | At5g38730 | 249529_at |
| At1g63080 | 259692_at | At3g22150 | 256792_at | At5g39350 | 249470_at |
| At1g63130 | 259689_x_at | At3g22670 | 258317_at | At5g39680 | 249462_at |
| At1g63150 | 259696_at | At3g23020 | 257765_at | At5g39710 | 249464_at |
| At1g63330 | 265106_s_at | At3g23330 | 258292_at | At5g39980 | 249436_at |
| At1g63400 | 261554_at | At3g24000 | 256906_at | At5g40415 | 249412_at |
| At1g64100 | 262328_at | At3g25060 | 257103_at | At5g41170 | 249341_at |
| At1g64310 | 259798_at | At3g25210 | 257838_at | At5g42310 | 249247_at |
| At1g64580 | 262007_at | At3g25970 | 258066_at | At5g42450 | 249222_at |
| At1g68930 | 259698_at | At3g26540 | 257616_at | At5g43790 | 249113_at |
| At1g68980 | 259700_at | At3g26630 | 257827_at | At5g44230 | 249088_at |
| At1g69350 | 260348_at | At3g26780 | 258254_at | At5g46100 | 248935_at |
| At1g71060 | 259746_at | At3g28640 | 257042_s_at | At5g46460 | 248874_at |
| At1g71210 | 259899_at | At3g28660 | 257042_s_at | At5g46580 | 248854_at |
| At1g71420 | 259949_at | At3g29230 | 257744_at | At5g46680 | 248871_at |
| At1g71460 | 259945_at | At3g29290 | 256613_at | At5g47360 | 248783_at |
| At1g71490 | 259951_at | At3g42630 | 252759_at | At5g47460 | 248803_at |
| At1g73400 | 245723_at | At3g46610 | 252473_s_at | At5g48730 | 248669_at |
| At1g73710 | 260062_at | At3g46790 | 252496_at | At5g48910 | 248640_at |
| At1g74400 | 260224_at | At3g47530 | 252420_at | At5g50280 | 248547_at |

ANNEXES

| | | | | | |
|-----------|-------------|-----------|-------------|-----------|-----------|
| At1g74580 | 260216_at | At3g47840 | 252386_at | At5g50390 | 248498_at |
| At1g74600 | 260217_at | At3g48250 | 252355_at | At5g50990 | 248483_at |
| At1g74630 | 260219_at | At3g48810 | 252332_at | At5g52630 | 248315_at |
| At1g74750 | 262219_at | At3g49240 | 252305_at | At5g52850 | 248307_at |
| At1g74850 | 262207_at | At3g49710 | 252245_at | At5g55740 | 248075_at |
| At1g74900 | 262224_at | At3g49740 | 252247_at | At5g56310 | 248012_at |
| At1g77010 | 264946_at | At3g50420 | 252216_at | At5g57260 | 247952_at |
| At1g77170 | 264475_s_at | At3g51320 | 252151_at | At5g59200 | 247737_at |
| At1g77360 | 246383_at | At3g53170 | 251972_at | At5g59600 | 247732_at |
| At1g77405 | 246388_at | At3g53360 | 251965_at | At5g59900 | 247701_at |
| At1g79080 | 264133_at | At3g53700 | 251936_at | At5g60960 | 247607_at |
| At1g79490 | 262941_at | At3g54980 | 251841_at | At5g61370 | 247550_at |
| At1g79540 | 262938_at | At3g56550 | 251693_at | At5g61400 | 247548_at |
| At1g80270 | 260331_at | At3g57430 | 251631_at | At5g61800 | 247518_at |
| At1g80550 | 260273_at | At3g58590 | 251583_at | At5g61990 | 247504_at |
| At1g80880 | 257475_at | At3g59040 | 251496_at | At5g65560 | 247141_at |
| At2g01360 | 266339_s_at | At3g60050 | 251451_s_at | At5g65570 | 247142_at |
| At2g01390 | 266339_s_at | At3g60980 | 251346_at | At5g65820 | 247160_at |
| At2g01510 | 266341_at | At3g61360 | 251313_at | At5g66500 | 247099_at |
| At2g01740 | 265865_at | At3g61520 | 246173_s_at | At5g66520 | 247100_at |
| At2g01860 | 263593_at | At3g62470 | 251211_s_at | At5g67570 | 247004_at |

Tableau 2 : Vecteurs Gateway utilisés pour les clonages.

| Nom | Type | résistance |
|------------|--|-------------------|
| pDONR207 | vecteur donneur | gentamycine |
| pDONR221 | vecteur donneur | kanamycine |
| pGWB2 | vecteur d'expression en plante sous le contrôle d'un promoteur 35S | kanamycine |