



**UNIVERSITÉ
D'EVRY-VAL-D'ESSONNE**



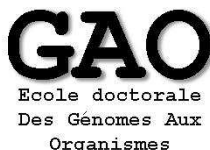
THESE DE DOCTORAT

**ETUDE DU ROLE DES PROTEINES DE LA FAMILLE PPR
DOUBLEMENT ADRESSEES A LA MITOCHONDRIE ET AU
PLASTE DANS L'EDITION DE L'ARN CHEZ *Arabidopsis thaliana***

MAURICIO LOPEZ OBANDO

2014





THESE DE DOCTORAT

Préparé à l'Unité de Recherche en Génomique Végétale URGV

Spécialité : Biologie cellulaire et moléculaire

ETUDE DU ROLE DES PROTEINES DE LA FAMILLE PPR DOUBLEMENT ADRESSEES A LA MITOCHONDRIE ET AU PLASTE DANS L'EDITION DE L'ARN CHEZ *Arabidopsis thaliana*

Présentée par Mauricio LOPEZ OBANDO

pour l'obtention du grade de docteur de l'université d'Evry Val d'Essonne

Soutenue le 06 Mai 2014 devant le Jury composé de :

Philippe GIEGE	(DR)	CNRS, IBMP Strasbourg	Rapporteur
Olivier VALLON	(DR)	CNRS, IBPC Paris	Rapporteur
Peter ROGOWSKY	(DR)	INRA, ENS Lyon	Examineur
Bénédicte STURBOIS	(PR)	UEVE, URGV Evry	Examinatrice
Claire LURIN	(DR)	INRA, URGV Evry	Co-directrice de thèse
Richard BERTHOME	(CR)	INRA, LIMP Toulouse	Co-directeur de thèse

REMERCIEMENTS

Tout d'abord je voudrais exprimer mes remerciements à mes directeurs du travail doctoral, Claire Lurin et Richard Berthomé, pour leur accueil, leur encadrement, leur orientation et leur contribution intellectuelle dans cette aventure des organites et les PPR.

Un grand merci aux membres, anciens et actuels, de l'équipe GFA de l'URGV ; Etienne Delannoy, Laure Hurtevin, Clément Boussardon, Alexandra Avon, Andeol F. de Longevialle, Rachel Fourdin, José Caius, Dario Monachello, Sandrine Balzergue, Stéphanie Pateyron, Ludivine Soubigou-Taconnat et Eddy Blondet pour leur gentillesse pendant ces années parcourues ensemble au laboratoire.

Merci à tous les membres de l'URGV, ceux des différentes équipes de recherche mais aussi ceux des équipes administrative et d'appui technique.

Un merci particulier au Pr. Mizuki Takenaka pour sa disponibilité et sa collaboration avec mon travail de recherche.

Je remercie Dominique Gagliardi et Sébastien Aubourg, membres de mon comité de thèse, pour leurs apports et conseils.

Je remercie également les membres du jury, Philippe Giegé, Olivier Vallon, Peter Rogowsky, Bénédicte Sturbois pour avoir accepté d'évaluer mon travail doctoral.

Je tiens à remercier aux institutions de l'état Français représenté par l'Université d'Evry Val d'Essonne, le Ministère de la Recherche et de l'Enseignement Supérieur, l'INRA et le CNRS pour leurs soutiens financiers pendant ce travail doctoral.

Finalement, la suivante phrase réunit ma pensée de gratitude envers mes proches de coeur « la reconnaissance ou gratitude est le désir ou le zèle d'amour par lequel nous nous efforçons de faire du bien à celui qui nous en a fait en vertu d'un pareil sentiment d'amour envers nous » Spinoza.

Toute ma reconnaissance à ma chère famille Corbeau; JF, Sylviane, Coco, Titi, papis, mamies et ma belle aimée Bilou, votre appui a été fondamental et ma gratitude reste immuable et certaine.

Mis amad@s madre, hermano, tías y abuelas, mi entrañable amor os rinde tributo. Vuestra confianza y apoyo ha sido el pilar de mi espíritu.

Simplemente Gracias!



ABBREVIATIONS

Aa : acides aminés
ADN : Acide DésoxyriboNucléique
ADNc : ADN complémentaire
AGI : Initiative pour le génome d'Arabidopsis
amiR: micro ANR artificiel
ARN : Acide RiboNucléique
ARNm : ARN messenger
ARNi : ARN interférent
ARNt : ARN de transfert
ARNr : ARN ribosomique

BLAST : Basic Local Alignment and Search Tool
BiFC: Complémentation de fluorescence bimoléculaire

CMS : Cytoplasmic Male Sterility (stérilité mâle cytoplasmique)

dNTP : désoxyribonucléotide
Ds-Red2 : Red Fluorescent Protein
DYW : domaine C-terminal de certaines protéines PPR

miARN : microARN
MEF : Facteur d'édition mitochondrial
MORF : Facteur multiple d'édition de l'ARN des organites
MS : milieu Murashige et Skoog

NCBI : National Center for Biotechnology Information
NLS : séquence de localisation nucléaire
NUMTS; Séquence d'ADN mitochondrial insérée dans le génome nucléaire

NUPTS; Séquence d'ADN plastidial insérée dans le génome nucléaire

ORF : Open Reading Frame (Cadre ouverte de lecture)
OTP : Organellar Transcript Processing

PCR : Polymerase Chain Reaction
PEP : Plastidial Encoded Polymerase
PLS : sous-famille de protéines PPR
PP : Processing Peptidases
PPR : Pentatricopeptide Repeat

RPOT : ARN polymérases de type phagique
RRM : RNA Recognition Motifs
RT-qPCR : réaction de transcription inverse suivie d'une PCR quantitative

TAIR: The Arabidopsis Information Resource
TIC : Complexe d'import de la membrane interne du chloroplaste
TIM : Complexe d'import de la membrane interne mitochondriale
TOC : Complexe d'import de la membrane externe du chloroplaste
TOM : Complexe d'import de la membrane externe mitochondriale

URGV : Unité de Recherche en Génomique Végétale

TABLE DE MATIERES

LISTE DE FIGURES	5
LISTE DE TABLEAUX	7
PREAMBULE	9
INTRODUCTION	10
Chapitre I	10
1. Origines et évolution des organites.....	11
1.1. Origines des mitochondries et des plastes	11
1.2. Réduction de la taille du génome et transferts génétiques au cours de l'évolution.	14
1.3. Le génome des mitochondries des plantes terrestres : taille, structure et contenu génétique	17
1.4. Le génome des plastes des plantes terrestres : taille, structure et contenu génétique.....	21
Chapitre II	23
2. Translocation Des Protéines Chez Les Organites	24
2.1. Caractéristiques générales du système d'import des protéines dans les organites.....	24
2.2. Mécanismes post-traductionnels et co-traductionnels intervenant dans les systèmes de translocation des protéines vers les organites	26
2.3. Les signaux d'adressage de nature protéique pour l'import des protéines chez les mitochondries et les plastes	27
2.4. Les complexes de translocation des mitochondries et des plastes	32
2.4.1. Les complexes mitochondriaux TOM et TIM	32
2.4.2. Les complexes chloroplastiques TOC et TIC.....	34
2.5. Le double adressage des protéines chez les organites des plantes terrestres.	36
Chapitre III	41
3. L'expression Génique Chez Les Organites Des Plantes Terrestres	42
3.1. La transcription chez les organites des plantes terrestres.....	43
3.1.1. Les complexes de transcription chez les organites	43
3.2. Le métabolisme des transcrits chez les organites des plantes terrestres.....	45
3.2.1. L'édition des transcrits des organites : aspects généraux et évolutifs.....	46
3.2.2. Le mécanisme d'édition C en U de l'ARN des organites de plantes: Principes généraux	49
3.2.3. Les protéines du complexe agissant sur l'édition de l'ARN des organites	51

3.2.4.	Les protéines MORF et leur implication dans les complexes d'édition.	51
3.3.	La maturation par clivage des transcrits des organites	55
3.3.1.	L'épissage d'introns chez les transcrits des organites	56
3.4.	La Famille Des Protéines A Domaines Pentatricopeptide Repeat.	60
3.4.1.	Généralités sur la famille des protéines PPR.	60
3.4.2.	Evolution de la famille PPR	61
3.4.3.	Les protéines PPR participant à la transcription des organites.....	65
3.4.4.	Rôles des protéines PPR au sein des complexes d'édition.....	67
3.4.5.	Les protéines PPR participent à l'épissage des introns des organites	70
3.5.	Mode de reconnaissance des acides nucléiques par les protéines PPR	75
3.5.1.	La structure des protéines PPR et les implications dans le mode de reconnaissance de l'ARN	79
Présentation et Objectifs de la Thèse		87
RESULTATS		89
Chapitre IV		89
4.	La Double Localisation Des Protéines PPR Chez <i>Arabidopsis Thaliana</i>	90
4.1.	La localisation subcellulaire des protéines PPR : Article I	91
4.2.	La localisation subcellulaire des protéines PPR : Actualisation de données.....	110
4.3.	Analyse de la double localisation des protéines PPR.....	111
4.3.1.	La double localisation et l'évolution des protéines PPR	116
4.3.2.	La double localisation et les réseaux d'interactions des protéines PPR	120
Chapitre V		121
5.	Analyse fonctionnelle de protéines PPR ayant une double localisation.	122
5.1.	Article II	123
5.2.	Projet d'Article III ^ψ	167
5.3.	Activités complémentaires du travail de thèse	210
5.3.1.	Approches mises en place lors du travail de thèse pour contourner l'embryon létalité des mutants affectés dans DYW2 et OTP100 et déterminer leur fonction moléculaire.....	210
5.3.1.1.	Stratégie d'interférence artificielle inductible à l'aide de microARN.....	210
5.3.1.2.	Stratégie de complémentation spécifique du stade graine	214
5.3.1.3.	Stratégie d'hémi-complémentation des mutants <i>dyw2</i> et <i>otp100</i>	216
5.3.1.4.	Etude de l'interaction entre les protéines DYW2 et OTP100 par complémentation de fluorescence bimoléculaire ou BiFC.	222

5.3.2. Caractérisation fonctionnelle de la protéine OTP91.....	223
DISCUSSION.....	228
Chapitre VI	228
6. Bilan critique des résultats et éléments additionnels de discussion.....	229
6.1. Limites de l'approche d'intégration de données pour la double localisation des protéines de la famille PPR.	229
6.2. Les limites des approches pour contourner l'embryon létalité des gènes DYW2 et OTP100..	230
6.3. Quels facteurs régulent la localisation des protéines PPR ?	231
6.4. La double localisation des PPR est un phénomène ancien mais dynamique entre les espèces. 233	
6.5. L'expression des génomes des organites est-elle coordonnée par des PPR doublement localisées ?	236
7. METHODES ET MATERIELS	239
7.1. Matériels Biologiques	239
7.1.1. Espèces végétales	239
7.1.2. Mutants d'insertion chez <i>A. thaliana</i>	239
7.1.3. Souches bactériennes	240
7.2. Méthodes de biologie moléculaire	240
7.2.1. Clonage à l'aide d'enzymes de restriction	240
7.2.1.1. Digestion de l'ADN par enzyme de restriction	241
7.2.1.2. Ligation de l'ADN.....	242
7.2.2. Clonage Gateway™ (Invitrogen™)	243
7.2.2.1. Amplification par PCR du fragment à cloner.....	243
7.2.2.2. Purification du produit PCR.....	245
7.2.2.3. Recombinaison BP (Invitrogen™)	245
7.2.2.4. Recombinaison LR (Invitrogen™)	246
7.2.3. Clonage pGEMT-easy	247
7.2.3.1. Ligation.....	247
7.2.4. Clonage de microARN artificiels (amiR).	248
7.2.5. Extraction d'ADN.....	249
7.2.5.1. Extraction rapide d'ADN génomique (ADNg) d' <i>A. thaliana</i>	249
7.2.5.2. Extraction d'ADNg en plaque 96 puits	250
7.2.6. Extraction d'ADN plasmidique	252
7.2.7. Quantification de l'ADNg ou l'ADNp	253

7.2.8.	Amplification d'ADN par la réaction de polymérisation en chaîne (PCR).	254
7.2.9.	PCR quantitative ou en temps réel (qPCR)	256
7.2.9.1.	Dessin des Amorces de qPCR	257
7.2.10.	Electrophorèse sur gel agarose	257
7.2.11.	Séquençage de produits de PCR ou de plasmides	258
7.2.12.	Extraction d'ARN totaux	258
7.2.12.1.	Extraction à l'aide du kit RNeasy plant Minikit (QIAGEN)	259
7.2.12.2.	Extraction au Trizol® «RNA Now »™ (Ozyme)	259
7.2.13.	Traitement de l'ARN total à la DNase	260
7.2.14.	Précipitation des ARN	261
7.2.15.	Quantification des l'ARN totaux	261
7.2.15.1.	Préparation de la gamme	261
7.2.15.2.	Préparation des échantillons	262
7.2.15.3.	Préparation du Ribogreen	262
7.2.16.	Test de la qualité et de l'intégrité des ARN issus d'une extraction	262
7.2.16.1.	Préparation des échantillons	263
7.2.17.	Réaction de transcription inverse	263
7.3.	Méthodes relatives aux bactéries	264
7.3.1.	Souche d'E. coli	264
7.3.1.1.	Préparation de bactéries thermo-compétentes d'E. coli	264
7.3.1.2.	Transformation de bactéries thermo-compétentes et sélection	264
7.3.2.	Souche d'A. tumefaciens	265
7.3.2.1.	Préparation de bactéries C58C1 thermo-compétentes	265
7.3.2.2.	Transformation de bactéries C58C1 thermo-compétentes	266
7.4.	Méthodes relatives aux plantes	266
7.4.1.	Culture d'A. thaliana en serre et chambre de culture	266
7.4.2.	Stérilisation des graines	267
7.4.3.	Culture in vitro	267
7.4.4.	Transformation stable de plantes d'A. thaliana	268
7.4.5.	Croisement des plantes d'A. thaliana	268
7.4.6.	Induction de l'expression des amiR sous chambre de culture	269
7.5.	Etude de la localisation subcellulaire des protéines	270
7.5.1.	Expression transitoire chez N. benthamiana	270
7.5.2.	Localisation des protéines dans les lignées stables d'A. thaliana	271

7.6. Test d'interaction protéine-protéine par complémentation biomoléculaire de fluorescence (BiFC) chez <i>A. thaliana</i>	271
ANNEXE	273
REFERENCES BIBLIOGRAPHIQUES.....	291

LISTE DE FIGURES

Figure 1. Les super groupes présents chez les eucaryotes et les évènements de symbiogenèse pouvant avoir eu lieu au cours de l'évolution des plastes.	13
Figure 2. Transferts de matériel génétique au sein des cellules des plantes.....	16
Figure 3. Contenu en gènes des génomes mitochondriaux d'une sélection de plantes terrestres.	19
Figure 4. Taille et nombre de gènes de génomes mitochondriaux d'une sélection de plantes terrestres.....	20
Figure 5. Taille et nombre de gènes de génomes chloroplastiques d'une sélection de plantes terrestres.....	22
Figure 6. Signaux d'adressage protéiques utilisés pour la translocation des protéines codées par le génome nucléaire vers différents compartiments des organites.	31
Figure 7. Sites d'édition des transcrits des organites dans différentes espèces de plantes terrestres.....	48
Figure 8. Réseau d'interactions protéine-protéine des protéines MORF au sein des plantes terrestres.....	55
Figure 9. Structure secondaire et tertiaire d'un intron de type II.	58
Figure 10. Classification de la famille PPR.....	61
Figure 11. Nombre de gènes de la famille PPR rencontrés chez les différents organismes vivants.....	64
Figure 12. Structure du motif PPR et du complexe d'élongation de la polymérase humaine mitochondriale POLRMT.....	80
Figure 13. Structure de la protéine PPR PRORP1 et modèle du complexe PRORP-ARNt.....	81
Figure 14. Structure de la protéine PPR10 et de son homodimère.....	83
Figure 15. Mode de reconnaissance de l'ARN par les protéines PPR.	85
Figure 16. Distribution actualisé de la localisation des protéines de la famille PPR.	110
Figure 17. Distribution de la localisation des protéines PPR au sein des sous-familles (P et PLS) et des sous-groupes (P ; P-D ; PLS ; PLS-E ; PLS-E-DYW).....	112
Figure 18. Distribution de la localisation des protéines PPR en fonction du comportement évolutif des gènes correspondants chez les angiospermes.	117

Figure 19. Réseau d'interactions protéine – protéine impliquant des protéines PPR ayant une potentielle ou double localisation.	120
Figure 20. Positions des microARN d'interférence artificielle « amiR » sélectionnés dans les gènes <i>DYW2</i> et <i>OTP100</i>	211
Figure 21. Système d'induction des amiR sur terreau et phénotype macroscopique des plantes après 5 jours d'induction.	212
Figure 22. Expression du transcrit <i>DYW2</i> dans les lignées amiRaa et amiRba.....	213
Figure 23. Expression du transcrit <i>OTP100</i> dans les lignées amiR1a et amiR2a.	214
Figure 24. Phénotype macroscopique de descendants T2 de plantes hétérozygotes pour les mutations <i>dyw2-1</i> ou <i>otp100-1</i> et porteuses de l'ADN-T d'hémi-complémentation.	219
Figure 25. Analyse de l'expression du transgène et du gène endogène <i>DYW2</i> dans les plantes hémi-complémentées P103 et P104.....	221
Figure 26. Structure de la protéine OTP91 et phénotype du mutant <i>otp91-2</i>	226
Figure 27. Éditeur chloroplastique du mutant homozygote <i>otp91-2</i>	227
Figure 28. Stratégie utilisée pour générer et amplifier des amiR.	249

LISTE DE TABLEAUX

Tableau 1. Bilan des différentes analyses en génétique inverse et des sites d'édition affectés chez les mutants <i>morf</i> (Adapté de Takenaka <i>et al.</i> , 2012, Bentolila <i>et al.</i> , 2013).....	53
Tableau 2. Code combinatoire de reconnaissance des acides nucléiques par les protéines PPR (Adapté de Barkan <i>et al.</i> , 2012 et Yagi <i>et al.</i> , 2013).....	78
Tableau 3. Protéines de la famille PPR d' <i>Arabidopsis thaliana</i> ayant une potentielle double localisation dans la mitochondrie et dans le plaste.....	114
Tableau 4. Nombre de plantes et génotypes identifiés dans la première génération de transformants (T1) exprimant les protéines DYW2 ou OTP100 sous le contrôle du promoteur <i>PLEC2</i>	215
Tableau 5. Nombre de plantes et génotypes identifiés dans la première génération de T1 exprimant les protéines de fusion Tp:HA:DYW2 ou Tp:HA:OTP100 sous le contrôle du promoteur <i>P35S</i>	217
Tableau 6. Ségrégation de l'insertion ADN-T d'hémi-complémentation à la génération T2 issue de plantes T1 hétérozygotes pour les mutations <i>dyw2-1</i> et <i>otp100-1</i>	218
Tableau 7. Génotypes identifiés lors du croisement de lignées hémi-complémentées <i>dyw2-1</i> mitochondriales et chloroplastiques.	221
Tableau 8. Génotypage de plantules issues de deux siliques de deux plantes hétérozygotes pour l'insertion <i>otp91-2</i> après croissance sur milieu MS contenant 1% de saccharose.	224
Tableau 9. Protéines d' <i>Arabidopsis thaliana</i> connues pour être localisées à la fois dans les mitochondries et les plastides disponibles dans de publications.	274
Tableau 10. Protéines d' <i>Arabidopsis thaliana</i> probablement doublement localisées dans les mitochondries et les plastides.	277
Tableau 11. Protéines d' <i>Arabidopsis thaliana</i> pouvant être doublement localisées dans les mitochondries et les plastides.	278
Tableau 12. Liste des protéines PPR impliquées dans l'édition de l'ARN des organites chez <i>A. thaliana</i>	280
Tableau 13. Liste de microRNA artificiels obtenus pour les gènes <i>DYW2</i> et <i>OTP100</i>	281

Tableau 14. Séquences des amorces utilisées pour les approches d'hémi-complémentation, complémentation embryon spécifique et des amiR.....	283
Tableau 15. Séquences des amorces pour clonage BP des séquences codantes de gènes utilisées dans cette thèse.....	284
Tableau 16. Séquences des amorces pour évaluer l'édition de 34 sites majeurs du chloroplaste	285
Tableau 17. Lignées T1 d'hémi-complémentation des mutants <i>dyw2-1</i> et <i>dyw2-2</i>	286
Tableau 18. Lignées T1 d'hémi-complémentation des mutants <i>otp100-1</i> et <i>otp100-2</i>	287
Tableau 19. Liste de sites d'édition prédits pour être les cibles de la protéine OTP91 utilisant le code de reconnaissance des PPR (Données fournis par M. Takenaka)	288
Tableau 20. Liste de vecteurs de clonage utilisées dans cette thèse.....	288
Tableau 21. Vecteurs de destination pGWB15 et pGWB13 remodifiées pour permettre l'hémicomplémentation et la complémentation embryon spécifique à l'aide des promoteurs LEC.....	289
Tableau 22. Vecteurs utilisées pour la stratégie de microRNA artificielles inductibles.	290

PREAMBULE

Cette thèse s'inscrit dans le contexte de la localisation et l'analyse fonctionnelle de la famille de protéines Pentatricopeptide Repeat (PPR). Particulièrement, elle analyse le phénomène de double localisation mitochondriale et chloroplastique au sein de la famille et se focalise sur la caractérisation de quatre membres ayant une fonction potentielle dans l'édition des ARN des organites chez *Arabidopsis thaliana*. Pour cela, l'introduction de cette thèse a été divisée en trois chapitres apportant des informations pour la discussion des résultats. Le chapitre I présente l'origine et évolution des organites, le chapitre II traite de la translocation des protéines chez les organites et le chapitre III présente l'état de l'art de l'expression génique chez les organites, notamment du rôle des protéines PPR dans les différents mécanismes de maturation des ARN. Au sein de ce dernier chapitre une attention particulière a été portée au processus d'édition. Enfin, les aspects structuraux de reconnaissance des acides nucléiques par les protéines PPR sont abordés de manière précise.

INTRODUCTION

Chapitre I

Origines Et Evolution Des Organites

Origines des mitochondries et des plastes

Réduction de la taille du génome des organites et transfert génétique au cours de l'évolution

Le génome mitochondrial des plantes terrestres : taille, structure et contenu en gènes

Le génome plastidial des plantes terrestres : taille, structure et contenu en gènes

1. Origines et évolution des organites

1.1. Origines des mitochondries et des plastes

La vie est apparue sur terre il y a environ 3,5 milliards d'années. Au cours de cette période, les processus évolutifs ont permis de créer de la biodiversité permettant l'adaptation mais aussi la complexification des organismes vivants. La cellule représente l'unité de base de la vie, tout être vivant étant constitué d'une cellule isolée ou d'une association de cellules. L'origine des cellules reste un sujet très controversé, même s'il est admis que deux principaux groupes d'organismes, les procaryotes et les eucaryotes, permettent de faire la distinction entre les êtres vivants. Les premiers sont unicellulaires et leur matériel génétique n'est pas enfermé dans un noyau. Les seconds ont leur matériel génétique enfermé dans un noyau et ils possèdent des organites cellulaires. Plus particulièrement, dans la classification scientifique des espèces, le domaine Eucaryote est subdivisé en super groupes: Excavates, Unikonts ou Amorphea, SAR (Stramenopiles, Alveolata, and Rhizaria) et Plantae ou Archeplastida (lignée verte; algues rouges, algues vertes auxquelles se raccrochent les végétaux terrestres) (Adl *et al.*, 2012). Notamment, le super groupe Archeplastida se caractérise par la présence de mitochondries et de plastes primaires dans les cellules. Au sein de ce super groupe, les organites ont une origine similaire et ont suivi une évolution spécifique dont je donnerai ici une vue d'ensemble.

L'origine des mitochondries et des plastes a été très longtemps controversée. A ce jour, en se basant sur des données phylogénétiques et structurales, la communauté scientifique propose que les mitochondries et les plastes sont issus de deux événements endosymbiotiques séquentiels ayant eu lieu il y a plusieurs centaines de millions d'années (Gray M. 1999, Dyll *et al.*, 2004). Cette origine a initialement été proposée au XX siècle par les scientifiques russes Konstantin Sergeyevich Mereschkovsky, Andrey Sergeyevich Famintsyn et Boris Kozopolynsky ainsi que par le scientifique nord américain Ivan Emmanuel Wallin (Margulis L. 2010). Ignorée pendant plusieurs décades, cette hypothèse a de nouveau été étayée par les travaux de Lynn Margulis à la fin des années 1960 (Archivald J. 2012). Selon cet auteur et ses

INTRODUCTION

prédécesseurs, la symbiogenèse constitue un point de rupture, au cours de l'évolution, conférant des innovations dans les processus cellulaires. Ainsi, l'émergence des eucaryotes et du super groupe Archeplastida seraient le fruit des événements de symbiogenèse majeurs. En se basant sur des données d'observation de fossiles et des estimations phylogénétiques, le premier événement, aboutissant à la formation des cellules eucaryotes et à l'origine des mitochondries, aurait eu lieu il y a 1200 à 2500 millions d'années (Ma) (Parfrey *et al.*, 2011). Il aurait impliqué une cellule pré-eucaryote et une protéobactérie (Parfrey *et al.*, 2011, Snih & Matzke 2013). Malgré l'acceptation d'un événement d'endosymbiose primaire à l'origine de l'émergence des cellules eucaryotes, la nature des cellules ancestrales et les processus impliqués font toujours l'objet de débats (Thiergart *et al.*, 2012, Gray M. 2012). En utilisant d'autres types d'analyses, Snih & Matzke suggèrent que cette endosymbiose pourrait avoir eu lieu il y a 909 à 1551 Ma.

Quelle que soit l'origine de la cellule eucaryote et de l'ancêtre mitochondrial, il est admis par la communauté des scientifiques étudiant les végétaux que ce type de cellule a subi un second événement de symbiogenèse, donnant les plastes primaires, à l'origine des cellules de la lignée verte (Keeling *et al.*, 2010, Green 2011). Parfrey *et al.* (2011) proposent que ces endosymbioses aient eu lieu il y a 1428 à 1670 Ma contre 896 à 1613 Ma pour Snih & Matzke (2013). Encore une fois, ces différences s'expliquent principalement par le type d'analyses réalisées dans chaque étude. Cependant, de nombreuses études s'accordent et proposent que les plastes de la lignée verte seraient issus d'une cyanobactérie (Archibald J. 2009, Keling P. 2010, Criscuolo & Gribaldo 2011, Green B. 2011, Keling P. 2013). Malgré la connaissance de l'origine cyanobactérienne des plastes primaires, l'embranchement monophylétique ou polyphylétique est encore controversé. Cependant la découverte récente d'un endosymbiote d'origine cyanobactérienne chez l'espèce *Paulinella chromatophora* amoeba (Nowack and Grossman 2012, Nakayama and Archibald 2012, Body *et al.*, 2012) conduisant à un plaste primaire favorise un scénario polyphylétique.

L'évolution des plastes au sein des cellules eucaryotes est plus complexe que celle des mitochondries du fait d'événements d'endosymbiose réitératifs avec des eucaryotes possédant déjà des plastes primaires. Ces endosymbioses secondaires et tertiaires sont à l'origine de

INTRODUCTION

l'apparition de plastes secondaires et tertiaires, à plus de deux membranes, dans plusieurs groupes taxonomiques. Ces évènements d'endosymbiose, les différents plastes et les groupes taxonomiques concernés sont présentés dans la figure 1. *Arabidopsis thaliana* étant une plante terrestre possédant uniquement des plastes primaires, seules les caractéristiques des plastes primaires seront présentées dans ce manuscrit.

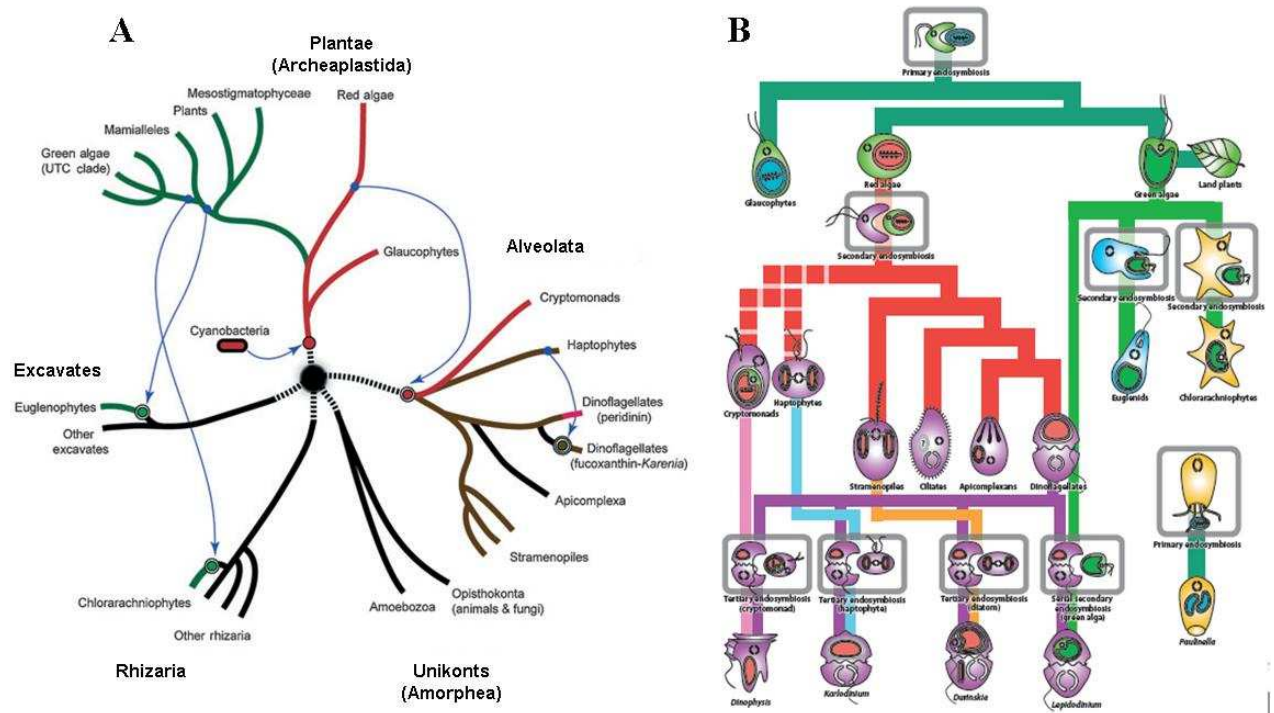


Figure 1. Les super groupes présents chez les eucaryotes et les évènements de symbiogenèse pouvant avoir eu lieu au cours de l'évolution des plastes.

A. Relations entre les règnes de l'embranchement des eucaryotes et les processus d'endosymbiose. Les flèches indiquent le sens de la filiation des plastes entre un endosymbiote photosynthétique et un hôte eucaryote non photosynthétique. (Adapté de Green B. 2011). B. Origine polyphylétique des plastes au sein de différents groupes taxonomiques (adapté de Keeling P. 2013)

L'origine endosymbiotique des mitochondries et des plastes a probablement conduit ces organites à évoluer de façon similaire, expliquant en partie, certaines caractéristiques communes. Parmi les principales, on observe la rétention et la réduction de leur génome, le transfert d'informations génétiques vers le noyau et/ou l'autre organite et la nécessité d'importer une quantité importante de protéines codées par le génome nucléaire pour permettre l'expression de leur génome et leur fonctionnement. Ces caractéristiques seront présentées dans les paragraphes suivants.

1.2. Réduction de la taille du génome et transferts génétiques au cours de l'évolution.

A ce jour, plusieurs milliers de génomes d'organites de différents règnes ont été séquencés. Ces génomes présentent, entre organismes, un contenu génétique et une structure génomique variable. Cependant, de façon générale, une réduction du contenu génétique du génome de l'endosymbiote est observée suite à l'endosymbiose. Cette réduction peut être liée à une simple perte de matériel génétique, à un transfert génétique vers le noyau de la cellule hôte ou, moins fréquemment, entre organites dans le cas d'organismes possédant un autre endosymbiote (Timmis *et al.*, 2004, Kleine *et al.*, 2009, Selosse M. 2009).

Plusieurs hypothèses ont été avancées pour expliquer la perte de matériel génétique des organites, parmi elles, une redondance fonctionnelle entre le génome des endosymbiotes et celui de la cellule hôte. Par contre, le transfert de matériel génétique est un phénomène plus complexe participant à la formation des génomes de la cellule hôte et des organites au cours de leur coévolution. Ce processus est toujours en cours dans des organismes ayant des organites issus de processus endosymbiotiques et en particulier dans les organismes de sous-règnes appartenant aux Archeplastida (Timmis *et al.*, 2004, Kleine *et al.*, 2009, Selosse M. 2009, Michalovova *et al.*, 2013).

Le transfert génétique entre compartiments cellulaires est un mécanisme peu décrit mais des progrès majeurs ont été réalisés dans la compréhension de ce processus au cours des 20 dernières années. Actuellement, il est connu que ce type de transfert peut avoir lieu à la fois entre les organites et le noyau des cellules mais aussi d'organite à organite. En détails et comme le présente la figure 2, il est maintenant admis que le transfert de matériel génétique des organites vers le noyau est le processus préférentiel. Les premiers travaux effectués dans ce domaine ont montré que l'intégration de matériel génétique provenant des organites dans les génomes nucléaires des cellules de levures, humaines et de plantes était très fréquent (pour revue voir Kleine *et al.*, 2009). Ce type de transferts conduit à des inserts d'ADN plastidial (NUPTS; nuclear integrants of plastid DNA) et d'ADN mitochondrial (NUMTS; nuclear integrants of mitochondrial DNA) dans le noyau. Ensuite et en fréquence décroissante, ce

INTRODUCTION

transfert s'effectue des plastes vers les mitochondries, du noyau vers la mitochondrie et, uniquement dans quelques rares cas, des mitochondries vers les plastes (Kleine *et al.*, 2009, Wang *et al.*, 2012, Iorizzo *et al.*, 2012, Staub *et al.*, 2013). Cependant Staub *et al.* (2013) ont récemment suggéré que le transfert des séquences d'ADN mitochondrial vers le plaste serait un processus plus commun que supposé.

Concernant le mécanisme d'intégration, il a été suggéré que l'ADN provenant des organites serait inséré dans le génome nucléaire au cours du processus de réparation des cassures d'ADN double-brin par recombinaison non-homologue, préférentiellement dans les régions fortement transcrites de l'euchromatine. L'étude des NUMTS de mammifères suggère également que l'ADN des organites s'intègre dans les régions proches des rétrotransposons et dans les régions riches en oligomères A+T (Tsuji *et al.*, 2012, Wang & Timmis 2013).

Chez les plantes, l'analyse des NUMTS et NUPTS dans différentes espèces a montré qu'à la fois de larges et de petits fragments d'ADN plastidial et mitochondrial sont présents dans le génome nucléaire de toutes les espèces, les grandes insertions étant associées à des événements plus récents (Michalovova *et al.*, 2013, Yoshida *et al.*, 2013). Par ailleurs, il a été observé que les NUMTS et NUPTS de grande taille ont une tendance à être localisées dans des régions centromériques alors que les insertions plus courtes sont dispersées dans tout le génome. L'apparition de ces fragments plus courts et leur distribution le long du génome semblent résulter d'un processus de fragmentation en relation avec l'évolution du génome nucléaire de chaque espèce. Ainsi, Michalovova *et al.* (2013) ont suggéré un modèle dans lequel le dynamisme de la mobilité des éléments transposables de chaque espèce pouvait rendre compte de la fragmentation et de la dispersion des NUMTS et NUPTS dans le génome des plantes.

La nature des séquences nucléotidiques à l'origine des insertions reste cependant encore controversée. Les données actuelles suggèrent que l'insertion directe de séquences d'ADN des organites dans le génome nucléaire serait le mécanisme préférentiel. Ainsi des travaux destinés à analyser ces événements de transferts génétiques grâce à des systèmes utilisant des plantes transplastomiques de *Nicotiana tabacum* et des gènes rapporteurs, ont

INTRODUCTION

permis de montrer que ni l'édition, ni l'épissage de l'ARN des organites ne sont requis pour l'insertion du matériel génétique dans le génome nucléaire (Sheppard *et al.*, 2011, Fuentes *et al.*, 2012). Par exemple, Fuentes *et al.* (2012) ont montré que les NUPTS semblaient s'intégrer dans le génome nucléaire principalement sous forme de fragments d'ADN et que la présence d'introns dans la séquence des organites ne constituait pas un obstacle à l'intégration. Néanmoins, ces mêmes auteurs ont suggéré que des transferts par le biais d'ARN et/ou ADNc pourraient également se produire mais de façon beaucoup moins fréquente.

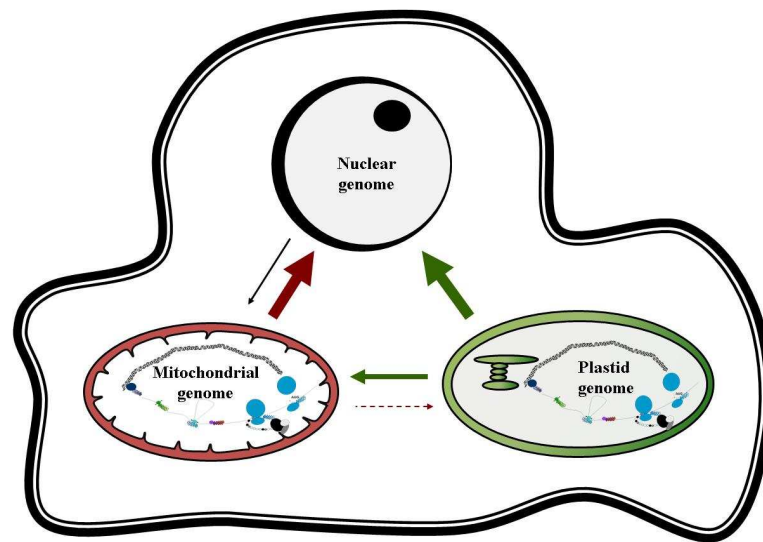


Figure 2. Transferts de matériel génétique au sein des cellules des plantes.

Cette figure présente de façon schématique l'importance relative (épaisseur des flèches) de chaque type de transfert entre compartiments cellulaires. Les plastides sont représentés en vert, les mitochondries en rouge et le noyau en noir. Le sens des flèches indique la direction du transfert et l'épaisseur représente son intensité. Les traits en pointillés précisent que de très rares cas de transfert ont été détectés entre mitochondries et plastides.

A ce jour, la compréhension du processus de transfert génétique entre génomes reste partielle et de nombreuses questions restent en suspens. Cependant, comme il a été suggéré auparavant, il apparaît évident que le transfert de matériel génétique des organites vers le noyau est un facteur important à considérer dans le contexte de l'identification de l'origine de nouveaux gènes nucléaires (Sheppard & Timins 2009, Lloyd & Timins 2011).

1.3. Le génome des mitochondries des plantes terrestres : taille, structure et contenu génétique

Jusqu'à présent (Février 2014), les données disponibles au NCBI (National Center for Biotechnology Information) et celles de la littérature permettent d'avoir accès à environ 90 génomes mitochondriaux séquencés de la division des *Viridiplantae*, dont 23 ont été obtenus à partir d'organismes appartenant à la subdivision des *Chlorophytes* (algues vertes) et 67 à partir de *Streptophytes* (à laquelle appartient les plantes terrestres). Les génomes mitochondriaux de la lignée verte présentent une très grande hétérogénéité dans leur taille, certains ayant une taille identique à celle des génomes mitochondriaux animaux (environ 10 à 20 Kbp) et d'autres pouvant être jusqu'à 100 fois plus grands (Kubo & Newton 2008, Sloan *et al.*, 2012, Gualberto *et al.*, 2013). Dans le chapitre qui va suivre, je présenterai une vue d'ensemble des données pertinentes les plus récentes concernant la taille, la structure et le contenu génétique des génomes mitochondriaux connus de la division des *Streptophytes*.

La plupart des génomes mitochondriaux de *Streptophytes* actuellement séquencés ont une taille comprise entre 0,2 à 0,7 Mbp, mais ce n'est pas une règle générale dans cette division. Par exemple, le génome mitochondrial de *Mesostigma viride*, une espèce basale récemment rattachée aux *Streptophytes*, a une taille de seulement 42 Kbp (Turmel *et al.*, 2002), alors que celui de l'angiosperme *Silene conica*, atteint 11,3 Mbp (Sloan *et al.*, 2012). Cette variabilité de taille est aussi observée entre espèces du même genre, comme par exemple dans les genres *Oryza*, *Beta*, *Silene* et *Brassica* (Darracq *et al.*, 2011, Sloan *et al.*, 2012, Wang *et al.*, 2012), mais aussi à l'intérieur des espèces, comme c'est le cas pour *Zea mays*, *Beta vulgaris* et *Triticum aestivum* (Allen *et al.*, 2007, Darracq *et al.*, 2011, Liu *et al.*, 2011). Du point de vue évolutif, l'augmentation de la taille des génomes mitochondriaux semble coïncider avec la divergence du sous-règne des *Embryophytes*, à l'origine des plantes terrestres. Ce fait vient supporter l'hypothèse selon laquelle le processus d'adaptation à l'environnement terrestre pourrait constituer l'une des forces évolutive ayant stimulé l'expansion du génome mitochondrial des plantes. Comme l'indiquent les analyses comparatives entre espèces citées précédemment, ces différences de tailles sont principalement liées à l'augmentation de la quantité de séquences non codantes au cours de l'évolution.

Tout comme la taille, la structure du génome mitochondrial est un caractère important à prendre en considération dans le contexte de l'histoire évolutive des génomes mitochondriaux des plantes. En se basant sur l'analyse des génomes séquencés, une première hypothèse, sujette à controverse depuis plusieurs années, a suggéré que le génome mitochondrial des plantes est composé d'une structure circulaire principale ayant la capacité de se réarranger, donnant ainsi naissance à une structure multipartite composée de petits génomes circulaires. Cependant, des études utilisant diverses stratégies de séquençage ont remis en question cette hypothèse (Kubo & Newton 2008, Handa H. 2008, Mower *et al.*, 2012, Sloan D. 2013). En effet, elles montrent que la structure du génome mitochondrial des plantes est très variable et peut adopter des structures linéaires et circulaires comme cela a été observé chez la Silène et le concombre (Sloan D. 2013, Gualberto *et al.*, 2013). Sloan D. (2013) souligne d'ailleurs que ces données sont en accord avec celles obtenues pour d'autres eucaryotes chez lesquels la structure linéaire du génome semble être la règle (Bendich 1993, Burger *et al.*, 2003).

En ce qui concerne le contenu en gènes, comme c'est le cas pour de nombreuses espèces, le génome mitochondrial des espèces de la division des *Streptophytes* contient deux classes de gènes : les gènes codant pour des ARN structuraux et des gènes à phase ouverte de lecture (ORF: Open Reading Frame) codant potentiellement pour des protéines. Chacune de ces deux classes est subdivisée en deux sous-classes. Pour la première classe, on distingue la sous-classe des gènes codant les ARN ribosomiaux (ARNr) et celle des gènes codant les ARN de transfert (ARNt). Pour la seconde classe, on trouve la sous-classe des gènes conservés codant pour des protéines de fonction connue et conservée et celle des gènes codant pour des protéines de fonction non conservée (souvent inconnue). Dans cette dernière sous-classe, les gènes prédits codent généralement pour des ORF spécifiques d'une espèce ou ayant une conservation faible par rapport à d'autres espèces. A l'exception de la sous-classe des gènes codant les ARNr, le contenu en gènes des autres sous-classes est très variable entre espèces. En particulier, au sein du sous-règne des *Embryophytes*, les génomes portent entre 2 à 26 gènes codant pour des ARNt et de 25 à 41 gènes correspondants à des ORF conservées (Sloan *et al.*, 2012, Chang *et al.*, 2013). Ces différences en contenu en gènes des génomes mitochondriaux constituent un indicateur de voies évolutives différentes empruntées par les

INTRODUCTION

clades de plantes. Ainsi, en se basant sur la sous-classe des gènes conservés codant pour des protéines de fonction connue et la classe des gènes codant des ARN structuraux, la figure 3 présente la vision actuelle de l'évolution des génomes mitochondriaux de plusieurs plantes terrestres.

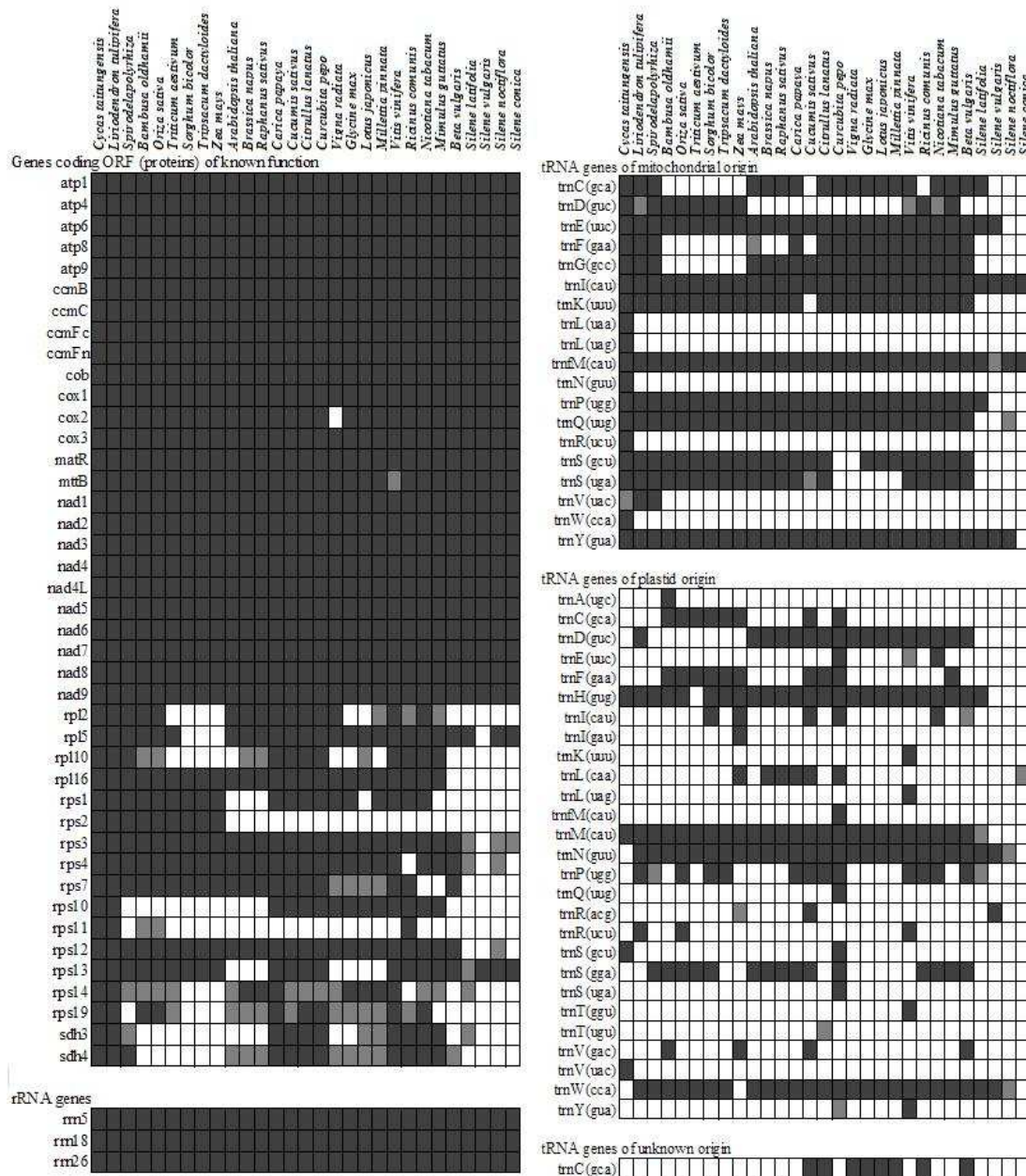


Figure 3. Contenu en gènes des génomes mitochondriaux d'une sélection de plantes terrestres.

Le contenu en gènes des génomes mitochondriaux de 29 espèces de plantes terrestres est indiqué. La couleur noire indique la présence du gène fonctionnel dans le génome, le gris la présence d'un pseudogène et le blanc l'absence du gène dans le génome de l'espèce en question (Adapté de Sloan *et al.*, 2012, Chang *et al.*, 2013).

INTRODUCTION

L'étude du contenu en gènes présenté figure 3 montre qu'un processus a entraîné une perte de gènes chez plusieurs espèces d'angiospermes après la divergence entre angiospermes et gymnospermes. La fréquence de ces pertes est variable selon les espèces et peut être très faible comme c'est le cas pour le tulipier de virginie (*Liriodendron tulipifera*) (Richardson *et al.*, 2013). De plus, il semble que l'acquisition de gènes codant pour des ARNt par le génome mitochondrial à travers du processus de transfert d'ADN du plaste ait pu avoir lieu à différentes périodes, la première étant juste après la divergence entre angiospermes et gymnospermes. Même si des différences significatives existent dans la taille des génomes mitochondriaux et leur contenu en gènes entre les plantes terrestres, ces deux caractéristiques ne sont pas corrélées. En effet, par exemple, alors que *Silene conica* et *Silene vulgaris* possèdent un contenu en gènes similaire, elles montrent une grande différence dans les tailles de leur génome : celle de *S. conica* étant de 11300 Kpb alors que celle de *S. vulgaris* est de 427 Kpb (Figure 4).

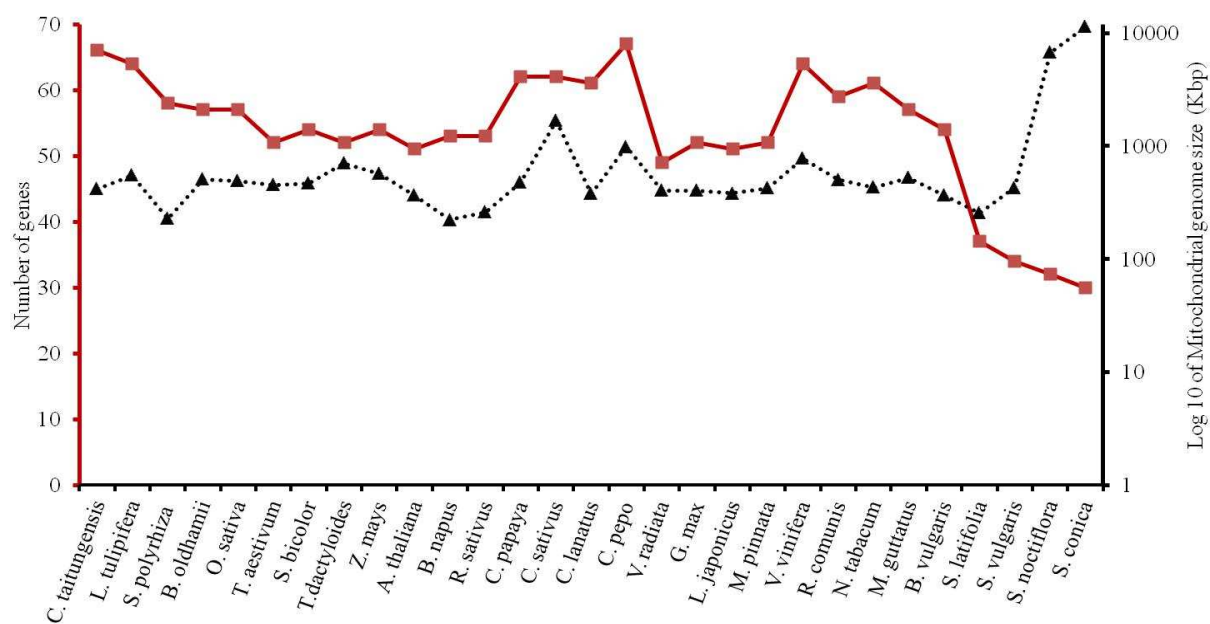


Figure 4. Taille et nombre de gènes de génomes mitochondriaux d'une sélection de plantes terrestres.

Différentes espèces de plantes terrestres représentant plusieurs clades sont indiquées en abscisse. Le nombre total de gènes présents dans leur génome mitochondrial (codant pour des ORF de fonction connue et des ARNt et ARNr) est indiqué par la courbe rouge et la légende sur l'axe de gauche. La taille des différents génomes mitochondriaux est indiquée par la courbe noire pointillée et l'axe de droite ayant une échelle logarithmique en base 10.

1.4. Le génome des plastes des plantes terrestres : taille, structure et contenu génétique

De la même façon que pour les mitochondries, de nombreuses séquences de génomes de plastes sont disponibles. A ce jour, 429 séquences de génomes de la division *Viridiplantae* sont accessibles dans la base de données organites du NCBI, 27 correspondant à des génomes de plastes de la subdivision des *Chlorophytes* et 402 de la subdivision des *Streptophytes*. Les génomes des plastes du sous-règne des *Embryophytes* représentent d'ailleurs une part importante de ces génomes séquencés. La plupart des génomes chloroplastiques des *Streptophytes* ont une taille comprise entre 100 et 200 Kbp. On observe cependant des plantes présentant un génome plastidial de taille réduite, comme par exemple celui de l'orchidée souterraine *Rhizanthella gardneri* qui ne fait que 60 Kbp et celui de la plante parasite *Conopholis americana* de 45 Kbp (Delannoy *et al.*, 2011, Wicke *et al.*, 2013). De façon générale, la taille des génomes des chloroplastes des plantes terrestres est moins variable entre espèces que celle des génomes mitochondriaux. Comme le montre la figure 5, la taille de ces génomes n'a pas beaucoup changé chez les *Streptophytes* autotrophes au cours de l'évolution (Bock 2007). Cependant une réduction des génomes chloroplastiques s'observe dans des plantes où la photosynthèse est peu ou pas requise comme cité précédemment chez l'orchidée souterraine ou chez les plantes parasites (Wicke *et al.*, 2013).

Bien qu'ils soient généralement représentés sous la forme de génomes circulaires, il est connu que les génomes plastidiaux présentent des structures complexes et des formes linéaires (Bendich 2004). Les génomes circulaires ont une structure quadripartite constituée de deux régions, répétées et inversées, IR_A et IR_B, séparant une région longue unique (LSC: long single copy region) celle-ci contenant la plupart des gènes codant des ORF, et une petite région unique (SSC : small single copy region) (Bock 2007, Wicke *et al.*, 2011).

La plupart des génomes chloroplastiques des plantes terrestres, à l'exception de ceux des plantes hétérotrophes, comprennent entre 70 et 110 gènes codant pour des protéines et 40 à 60 gènes codant pour des ARN structuraux, quelques uns d'entre eux étant organisés en structures de type opéron (Figure 5). La perte de gènes a été observée dans différentes espèces

de plantes et les gains ont été décrits comme des événements rares (Wicke *et al.*, 2011, Straub *et al.*, 2013). Les plantes hétérotrophes ou non photosynthétiques présentent très peu de gènes fonctionnels en raison de l'absence du processus photosynthétique (Delannoy *et al.*, 2011, Wicke *et al.*, 2013). L'analyse de différents génomes d'espèces de la famille des *Orobanchaceae* a montré qu'il existe des différences dans le degré de réduction des génomes en fonction de l'histoire évolutive de chaque espèce, un contenu minimum de 16 gènes codant des ORF et 19 codant des ARN structuraux restant commun à tous les génomes (Wicke *et al.*, 2013). La plupart de ces gènes communs sont requis pour la traduction et n'ont pas été relocalisés dans le génome nucléaire. Dans ces génomes, la réduction de l'information génétique s'accompagne d'une disparition de la plupart des introns et d'une réduction du pourcentage global en GC, en revanche la densité en zones répétées reste similaire à celle des génomes plastidiaux des plantes photosynthétiques (Wicke *et al.*, 2013).

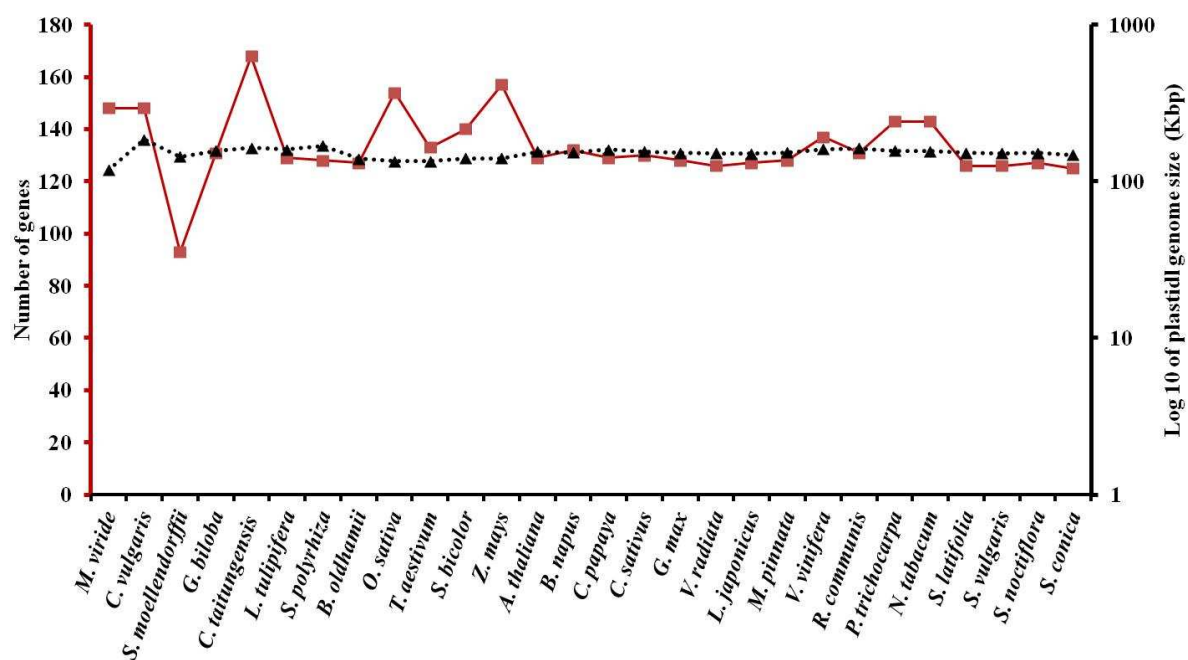


Figure 5. Taille et nombre de gènes de génomes chloroplastiques d'une sélection de plantes terrestres.

Différentes espèces de plantes terrestres représentant plusieurs clades sont indiquées en abscisse. Le nombre total de gènes présents dans leur génome plastidial (codant pour des ORF de fonction connue ou des ARNt et ARNr) est indiqué par la courbe rouge et la légende sur l'axe de gauche. La taille des différents génomes mitochondriaux est indiquée par la courbe noire pointillée et l'axe de droite ayant une échelle logarithmique en base 10.

Chapitre II

Translocation Des Protéines Chez Les Organites

Caractéristiques générales du système d'importation des protéines dans les organites

Mécanismes post-traductionnels et co-traductionnels du système d'importation des protéines dans les organites

Les signaux d'adressages de nature protéique pour l'import des protéines chez les mitochondries et les plastes

Complexes de translocation des mitochondries et des plastes

Complexes TOM et TIM mitochondriaux

Complexes TOC et TIC plastidiaux

Convergence dans la translocation des protéines chez les plastes et les mitochondries des plantes: Le double adressage des protéines chez les organites

2. Translocation Des Protéines Chez Les Organites

2.1. Caractéristiques générales du système d'import des protéines dans les organites

La structure et la compartimentation des différents génomes des cellules de plantes nécessitent l'existence de mécanismes contrôlant l'acheminement correct de produits ou de molécules aux endroits où ils seront actifs ou métabolisés. Comme décrit dans le chapitre précédent, le transfert de matériel génétique est un des processus fondamental de l'endosymbiose qui a conduit à une réduction des génomes des endosymbiotes ancestraux et à une intégration de matériel génétique dans le génome nucléaire, contribuant à sa diversification génétique (Selosse M. 2009). Aussi, au cours de l'évolution, les échanges ont contribué à l'élimination de redondances fonctionnelles mais aussi à la mise en place de nouvelles voies métaboliques. Ces échanges ont nécessité le développement de mécanismes permettant le fonctionnement des organites et leur contrôle dans la cellule hôte. Dans ce contexte, le génome nucléaire actuel code pour plus de 3500 protéines prédites pour fonctionner dans les organites (Tanz *et al.*, 2012). Ces protéines doivent être importées dans les organites pour y exercer leurs activités. Des mécanismes permettant leur acheminement dans les organites ont ainsi du être mis en place et adaptés par la cellule hôte. Chez les organites des cellules de plantes, plusieurs processus de translocation ont été identifiés, permettant la translocation des protéines à travers les membranes externes et/ou internes.

Comme le soulignent de très bonnes revues, au cours des dernières décennies, de nombreuses études ont été réalisées pour comprendre les mécanismes de translocation de protéines dans les organites (Neupert & Herrman 2007, Jarvis 2008, Balsera *et al.*, 2009, Li & Chiu 2010, Schleiff & Becker 2011, Stojanovski *et al.*, 2013, Duncan *et al.*, 2013, Dudek *et al.*, 2013). Ces revues présentent les principaux éléments des complexes de translocation et les mécanismes permettant l'identification des protéines à importer, leur maturation et leur acheminement vers leur destination finale. Je présenterai ici un résumé de l'état des connaissances concernant la translocation des protéines dans les organites.

Historiquement, avant 1990, la plupart des travaux se sont focalisés sur l'identification des caractéristiques des protéines déterminant leur import dans les organites. Ces études ont montré que le processus de translocation impliquait probablement des facteurs cytosoliques et que les protéines adressées aux organites étaient acheminées après traduction grâce à un peptide signal d'adressage, situé le plus souvent dans la partie N-terminale de la séquence protéique. De plus, il était suggéré que les protéines adressées étaient reconnues par des récepteurs, maturées par une peptidase après leur import et que le mécanisme de translocation nécessitait de l'énergie (Pfanner *et al.*, 1988, Pfanner & Neupert 1990). Bien que certains composants des systèmes de translocation aient été découverts avant 1990 (principalement chez les champignons), c'est pendant la dernière décennie du 20^{ième} siècle que les connaissances de base concernant les mécanismes et la composition des complexes de translocation ont été obtenues. Au cours de cette période, plusieurs travaux ont permis d'identifier de nombreuses protéines et complexes, essentiellement par des approches génétiques et biochimiques (pour revue voir Neupert 1997, Keegstra & Cline 1999). C'est au cours de cette décennie que les complexes de translocation des membranes externe et interne de la mitochondrie, TOM (Translocation of the Outer Mitochondrial membrane) et TIM (Translocation of the Innner Mitochondrial membrane complexes), respectivement, et les complexes correspondant chez le chloroplastes, TOC et TIC (Translocation of the Outer and Innner Chloroplast complexes) ont été partiellement définis. Après cette période, de nouvelles connaissances et des détails concernant ces mécanismes ont été principalement obtenus par l'utilisation de technologies 'omiques' et d'outils de génomique comparative.

Ainsi, de nos jours, il est admis que les organites disposent de multiples complexes de translocation qui sont finement coordonnées pour permettre l'adressage correct des protéines dans leur compartiment cellulaire. Dans la section suivante, je décrirai plus en détails les connaissances acquises concernant le mécanisme d'import, les signaux d'adressage spécifiant la localisation subcellulaire des protéines et la composition des complexes de translocation permettant leur acheminement.

2.2. Mécanismes post-traductionnels et co-traductionnels intervenant dans les systèmes de translocation des protéines vers les organites

Depuis le début des études réalisées sur les systèmes de translocation, il a souvent été proposé que des mécanismes post-traductionnels et co-traductionnels y participent (Chua & Schmidt 1979). Les mécanismes post-traductionnels impliquent des facteurs cytosoliques qui reconnaissent et fixent ces protéines traduites. Ils assurent l'acheminement de ces protéines non structurées vers le système de translocation permettant la reconnaissance des peptides d'adressage pour initier le processus d'import. Les mécanismes co-traductionnels, quant à eux, impliquent la reconnaissance d'une séquence d'adressage dans l'ARN par des protéines se fixant à ces séquences d'ARN. Ces complexes riboprotéiques vont s'associer à des complexes ribosomiaux localisés dans la membrane externe des organites où la traduction est couplée au processus de translocation.

Chez les plastes, pendant de nombreuses années, le mécanisme post-traductionnel a été considéré comme étant le principal mécanisme expliquant l'import des protéines. En effet, en l'absence de donnée prouvant l'existence du mécanisme co-traductionnel chez les plastes, le mécanisme post-traductionnel apparaissait être le principal voir même le seul mécanisme utilisé par la cellule végétale pour transférer les protéines du cytoplasme vers le stroma de l'organite (Li & Chiu 2010, Schleiff & Becker 2011). Néanmoins, la mise en évidence à proximité de l'enveloppe plastidiale de *Chlamydomonas Reinhardtii* de complexes ribosomiaux contenant des séquences d'ARN d'origine nucléaire codant pour des protéines plastidiales a permis de proposer que l'import de certaines protéines dans les plastes pourrait impliquer des mécanismes co-traductionnels (Uniacke & Zerges 2009). De plus, des travaux récents suggèrent que certaines protéines codées par le génome des plastes utiliseraient un mécanisme de type co-traductionnel pour l'import dans l'enveloppe des thylacoïdes (pour revue Weis *et al.*, 2013).

De la même façon, après de nombreuses années, au cours desquelles ces mécanismes ont été très controversés, il est maintenant admis que les protéines mitochondriales codées par le génome nucléaire sont importées dans la matrice par le biais de mécanismes post-

traductionnels et co-traductionnels; l'import post-traductionnel restant le mécanisme principal pour acheminer la plupart des protéines mitochondriales codées par le génome nucléaire. Le mécanisme co-traductionnel a été mis en évidence dans l'import de protéines dans les mitochondries de levures (Weis *et al.*, 2013). Ce mécanisme a été suggéré pour la première fois au cours d'analyses montrant qu'une quantité importante d'ARN messagers (ARNm), codant pour des protéines mitochondriales, était localisée dans des complexes ribosomiaux à proximité des membranes externes des mitochondries (Marc *et al.*, 2002). Par la suite, des travaux ont montré que ces ARNm étaient reconnus par des protéines de la famille des Pumillo-Fem (Puf) se fixant aux ARN et facilitant leur localisation à la périphérie des mitochondries (Saint-Georges *et al.*, 2008). Enfin, ces dernières années, des travaux ont montré que certaines protéines nécessitent une traduction et une translocation simultanée pour pouvoir être correctement adressées (Yogev *et al.*, 2007). Chez les plantes, le mécanisme de co-traduction n'a pas encore été décrit. Cependant, les similitudes entre les systèmes de translocations mitochondriaux suggèrent que ce mécanisme pourrait être conservé chez les plantes (Weis *et al.*, 2013).

2.3. Les signaux d'adressages de nature protéique pour l'import des protéines chez les mitochondries et les plastes

Comme nous l'avons vu précédemment, les signaux d'adressages ARN ou peptidiques reconnus par des facteurs cytosoliques sont indispensables à l'acheminement correct de protéines à proximité des complexes de translocation. Les premières études dans le domaine ont montré qu'une séquence d'acides aminés (aa) localisée en région N-terminale des protéines était importante pour la translocation. La caractérisation de ces séquences, nommées ici « signaux d'adressage N-terminaux » (aussi connus sous le nom de Séquence d'adressage à la Matrice – MTS -, ou peptide de transit - PT - ou préséquence) n'a pas permis d'identifier de séquence consensus parmi les protéines mitochondriales et plastidiales de différentes espèces. Cependant, certaines propriétés biochimiques et structurales contribuant à la localisation mitochondriale ou plastidiale ont été identifiées. Selon les espèces, le signal d'adressage N-terminal mitochondrial peut avoir une longueur variant de 18 à 100 aa. Il est constitué de

régions possédant des structures secondaires de type hélice α chargées positivement. De plus, la plupart des signaux d'adressage N-terminaux contient un site de clivage. Ainsi, ces peptides sont reconnus et coupés par des peptidases (PP; Processing Peptidases) permettant la génération des protéines fonctionnelles matures après leur passage dans les complexes de translocation (Dudek *et al.*, 2013, Huang *et al.*, 2009).

L'analyse de nombreux signaux d'adressage N-terminaux mitochondriaux d'*Arabidopsis* et de riz a montré que ces peptides avaient une longueur, une composition en acides aminés et des zones de clivages par les peptidases variables. Par exemple, chez *Arabidopsis*, la taille moyenne du peptide d'adressage est de 50 aa mais elle peut varier de 19 et 109 aa selon les protéines (Huang *et al.*, 2009). De plus, même si les signaux d'adressage N-terminaux d'*Arabidopsis* et de riz présentent généralement un enrichissement en Arginine (R), Sérine (S), Alanine (A) et Leucine (L), ceux de riz ont une proportion moins importante de Sérine et sont souvent plus riches en Alanine (Huang *et al.*, 2009). Récemment, il a été suggéré que la nature physico-chimique et la structure de la séquence protéique située après le peptide d'adressage pouvait également influencer le processus de translocation des protéines dans les mitochondries (Doyle *et al.*, 2013). Cependant, les auteurs de cette étude ont signalé que chez les plantes cette incidence semblait mineure, en raison de la grande taille des signaux d'adressages localisés en région N-terminale des protéines.

Les signaux d'adressage aux plastides sont également peu conservés : une grande variabilité de la longueur des séquences, de leur composition en aa et de leur site de clivage est ainsi observée. Néanmoins, de façon générale, ces peptides N-terminaux possèdent peu de résidus acides et génèrent, comme dans le cas des protéines adressées aux mitochondries, des structures secondaires dans les membranes ou dans des environnements mimant des membranes (Balsera *et al.*, 2009, Li & Teng *et al.*, 2013). En général, la longueur moyenne nécessaire à une translocation optimale dans le plaste est de 60 aa. Cependant, la longueur des présquences varie de 13 à 150 aa selon des protéines et les espèces (Huang *et al.*, 2009, Bionda *et al.*, 2010). Les premiers travaux menés sur la composition en aa des signaux d'adressage N-terminaux ont montré qu'ils étaient enrichis en aa hydroxylés comme la Sérine (S) et la Thréonine (T). Cependant, ces résultats ont été remis en question par de nombreuses

INTRODUCTION

études et il est maintenant admis que les préséquences des protéines adressées aux plastides chez *Arabidopsis* sont seulement enrichies en Sérine alors que celles des protéines de riz sont enrichies en Alanine mais pas en Sérine ni en Thréonine (Zybailov *et al.*, 2008).

Récemment, un modèle de composition modulaire des signaux d'adressage N-terminaux plastidiaux en domaines a été proposé. Dans ce dernier, connu sous le nom de modèle bimodal, le signal d'adressage est composé de deux principaux blocs modulaires, ou domaines, séparés par un domaine de liaison de 22 à 25 aa. Dans ce modèle, le premier domaine permettrait l'interaction avec des protéines chaperones associées aux protéines motrices du système d'import dans le stroma du plaste alors que le deuxième domaine interagirait avec les protéines du complexe TOC (Chotewutmontri *et al.*, 2012). Plusieurs récepteurs du complexe TOC montrant une régulation différentielle au cours du développement et/ou des affinités différentes selon la nature du peptide de transit ont été identifiés récemment (cf ci-dessous). En se basant sur le modèle bimodal, il a été proposé une reconnaissance sélective et d'ordre multiple dépendant du récepteur (M&M; Multi-selection and Multi-order plastid translocation model) (Li & Teng *et al.*, 2013). Dans ce modèle, les déterminants de la translocation d'une protéine seraient à la fois la nature des différents domaines du peptide de transit et la composition des récepteurs du complexe TOC présents dans les plastides (à un état développemental spécifique). Les auteurs du modèle M&M ont suggéré qu'il pourrait aussi fonctionner au niveau des mitochondries puisque la sous unité γ de la F1-ATPase présente plusieurs motifs au niveau de son peptide d'adressage (Lee *et al.*, 2012, Li & Teng *et al.*, 2013). Ces modèles d'interaction entre signaux d'adressage N-terminaux et complexes de translocation pourraient partiellement expliquer la variabilité observée dans la longueur et la composition en aa des ces signaux d'adressage (Chotewutmontri *et al.*, 2012).

D'autres signaux, différents de ceux localisés en région N-terminale des protéines et situés à l'intérieur ou à l'extrémité C-terminale, sont également importants et nécessaires à l'adressage de certaines protéines aux mitochondries et aux plastides. Ces signaux ont principalement été caractérisés dans le cas de protéines mitochondriales des champignons, modèle très étudié car utilisé pour caractériser la biogenèse des mitochondries (pour revue

INTRODUCTION

voir Dudeck *et al.*, 2013, Stojanovski *et al.*, 2012). Par exemple, l'intégration de protéines de type tonneau β (comme TOM40), principalement dans la membrane externe mitochondriale, est réalisée par l'utilisation d'un signal, appelé signal β , localisé dans le dernier feuillet β du précurseur de la protéine. Ce signal est composé de résidus polaires, préférentiellement d'une lysine (L) ou une glutamine (E), suivis d'une glycine (G) et d'acides aminés hydrophobes. Il a été montré que les protéines portant ce signal traversent la membrane externe et sont intégrés depuis l'espace inter membranaire grâce à l'activité d'un complexe appelé SAM50 (figure 6). D'autres exemples de signaux internes ont été observés dans le cadre de l'intégration de transporteurs de métabolites dans la membrane interne des mitochondries. Ces signaux sont reconnus au cours des étapes successives de translocation. Tout d'abord, ces protéines traversent la membrane externe en utilisant le complexe TOM. Une fois dans l'espace inter membranaire, elles sont reconnues par des petites protéines chaperonnes TIC qui les conduisent vers le complexe TIM22 où, par un mécanisme nécessitant le potentiel de membrane ($\Delta\psi$), elles sont intégrées dans la membrane interne de la mitochondrie (Stojanovski *et al.*, 2012).

Dans le cas des plastes, même si des signaux d'adressage de type β n'ont pas encore été détectés, on suppose que ce type d'adressage existe pour certaines protéines comme le suggèrent Schleiff & Becker (2011). Enfin, les mécanismes de translocation impliqués dans l'adressage aux plastes de nombreuses protéines ne possédant pas de signaux d'adressage N-terminaux restent mal connus. L'intervention de signaux internes ou localisés en région C-terminale de ces protéines n'est donc pas à exclure. Pour résumer, la figure 6 présente schématiquement l'ensemble des signaux d'adressage peptidiques identifiés dans les protéines adressées aux mitochondries et aux plastes à travers la membrane externe (OM, outer membrane), interne (IM, inner membrane), l'espace inter-membranaire et la matrice des mitochondries ou le stroma des plastes.

Des signaux N-terminaux à caractère ambigu dirigeant les protéines vers la mitochondrie et le plaste ont également été identifiés chez les organismes possédant ces organites (cf ci-dessous partie « double adressage »). Il a été suggéré que les signaux d'adressage ambigus seraient reconnus par les systèmes d'import « classiques » des

INTRODUCTION

mitochondries et des plastes (Carrie & Small 2013). En réalité, très peu d'indications existent quant à la nature des récepteurs impliqués dans leur translocation. Le seul résultat publié concerne la mitochondrie pour laquelle il a été montré que, *in vitro*, le récepteur AtTOM20-4 interagit avec un peptide N-terminal de 60 acides aminés de la protéine threonyl-ARNt synthétase d'Arabidopsis (AtThrRS) (Ye *et al.*, 2012). Dans cette étude, il a également été montré que ce peptide N-terminal a une certaine tendance à former des structures hélicoïdales en solution aqueuse. De plus, de façon intéressante, les auteurs ont identifié deux régions du peptide impliqués de façon dynamique dans l'interaction avec AtTOM20-4. L'ensemble de ces résultats a permis de proposer que, dans ce cas, le signal d'adressage ambigu serait reconnu de la même façon qu'un signal mitochondrial classique.

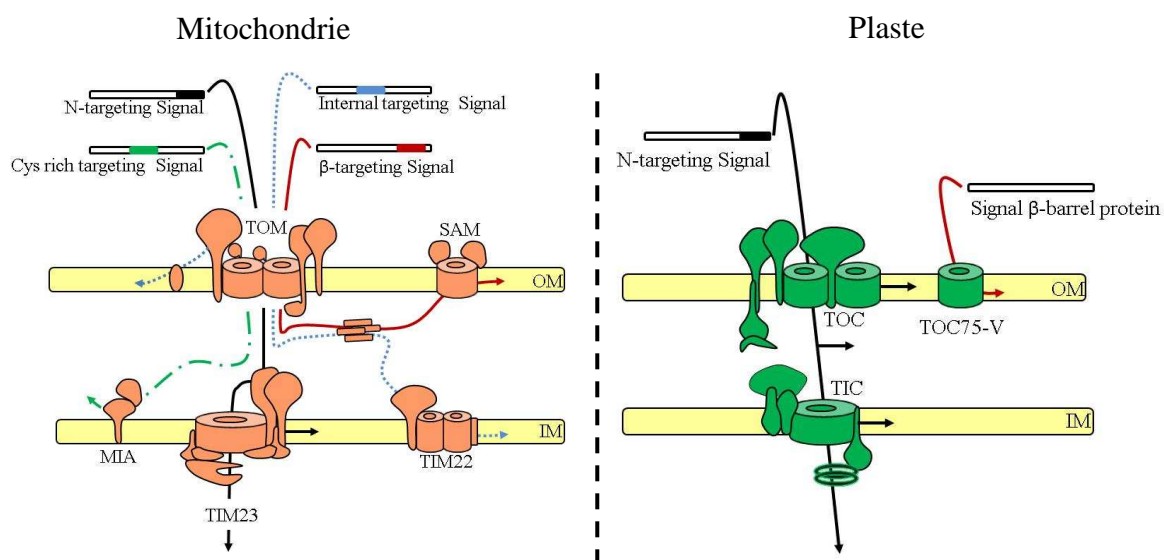


Figure 6. Signaux d'adressage protéiques utilisés pour la translocation des protéines codées par le génome nucléaire vers différents compartiments des organites.

Les principaux complexes de translocation des protéines dans la membrane des mitochondries (orange) et des plastes (vert) sont représentés de manière schématique. OM : membrane externe, IM : membrane interne, TOM : Complexe de translocation de la membrane externe de la mitochondrie, TIM : Complexe de translocation de la membrane interne de la mitochondrie, TOC : Complexe de translocation de la membrane externe des plastes, TIC : Complexe de translocation de la membrane interne des plastes (adapté de Dudeck *et al.*, 2013, Schleiff & Becker 2011).

2.4. Les complexes de translocation des mitochondries et des plastes

L'import des protéines dans les mitochondries et les plastes est un processus consommateur d'énergie médié par des complexes protéiques situés dans les membranes externes et internes de ces organites. Comme mentionné précédemment, ces complexes de translocation sont appelés TOM et TIM dans la mitochondrie et TOC et TIC dans le plaste. Les complexes TOM et TOC sont les points d'entrée principaux des protéines codées par le génome nucléaire et adressées respectivement à la mitochondrie et aux plastes. Les complexes TIM et TIC permettent la translocation des protéines dans la membrane interne ou vers, respectivement, la matrice ou le stroma de ces organites. Les 4 complexes sont composés de protéines présentant des similarités structurelles : protéines récepteurs et protéines formant des canaux dans les membranes externes et internes. Une présentation complète de l'adressage et la localisation des protéines dans les organites nécessite celle des systèmes d'import. Dans la partie suivante, je décrirai donc la composition et les fonctions principales des complexes de translocation des organites ainsi que leurs spécificités chez les plantes.

2.4.1. Les complexes mitochondriaux TOM et TIM

Le complexe mitochondrial TOM des plantes comporte 7 sous-unités comme chez la plupart des espèces vivantes. Néanmoins, seulement 4 d'entre-elles sont similaires à celles décrites chez d'autres espèces (pour une revue récente voir Duncan *et al.*, 2013). Ainsi, comme chez les animaux et la levure, TOM40 est la protéine centrale formant le pore et les petites protéines TOM5, TOM6 et TOM7 sont importantes pour la formation et la stabilité du complexe (Neupert & Herrmann 2007). Les trois autres protéines TOM9 (ou TOM22 de plante), TOM20 et OMP64 (probable TOM70 de plantes) sont spécifiques des plantes. La protéine TOM9 végétale ne possède pas la région cytoplasmique décrite chez l'animal et la levure comme importante dans la reconnaissance du signal d'adressage N-terminal (Macasev *et al.*, 2000). Les deux dernières protéines, TOM20 et TOM70, forment l'antenne réceptrice des complexes TOM chez les animaux et les champignons. A la différence des animaux et des champignons, chez lesquels une seule protéine TOM20 est codée par le génome, chez les plantes, 1 à 4 copies du gène *TOM20* sont généralement identifiées dans chaque génome.

INTRODUCTION

Parmi elles, TOM20-1 est la seule isoforme qui n'ait pas été clairement associée à une fonction mitochondriale (Werhahn *et al.*, 2001, Lister *et al.*, 2007, Paul *et al.*, 2013). Les protéines TOM20 sont donc considérées comme des protéines intégrées au complexe TOM végétal et non comme des protéines périphériques comme chez les animaux et les champignons. La protéine OM64 semble remplir la fonction de TOM70 chez les végétaux (Lister *et al.*, 2007). Cette protéine est très proche des récepteurs de type TOC64 impliqués dans l'import des protéines chloroplastiques mais son interaction transitoire avec TOM40 semble indiquer une fonction plus accessoire au sein du complexe (Qbadou *et al.*, 2006, Lister *et al.*, 2007, Duncan *et al.*, 2013).

Chez toutes les espèces, deux principaux types de complexes d'import ont été identifiés dans la membrane interne des mitochondries (complexes TIM) : les complexes TIM23 et TIM22. Les complexes TIM23 permettent l'import de toutes les protéines matricielles et de la plupart des protéines de la membrane interne tandis que les complexes TIM22 assurent l'import de seulement certaines protéines de la membrane interne (Neupert & Herrmann 2007). Les protéines TIM23, TIM17 et TIM50 composent le « cœur » des complexes TIM23. Lorsqu'elles sont associées avec la protéine TIM21, elles forment le complexe appelé « TIM23-SORT ». En parallèle, le complexe intégrant les protéines TIM44, la protéine « heat shock » 70 (mtHsp70) et les chaperonnes TIM14, TIM16 et Mge1 est appelé « TIM23-PAM » (parfois aussi nommé TIM23-MOTOR). Le complexe TIM23-SORT permet l'intégration de protéines dans la membrane interne des mitochondries grâce au potentiel de membrane ($\Delta\psi$). Ce complexe peut ainsi être couplé aux complexes III et IV de la chaîne respiratoire par l'intermédiaire de TIM21 stimulant le potentiel de membrane nécessaire à l'intégration des protéines. De plus, la protéine TIM21 est impliquée dans la formation des « super complexes » TOM-TIM23.

Le transit des protéines depuis le complexe TOM vers l'espace matriciel est effectué par le complexe TIM23-PAM grâce à l'activité de la protéine mtHsp70, à la régulation par les protéines chaperonnes et de façon énergie dépendante (Stojanovski *et al.*, 2012, Sikor *et al.*, 2013). Chez les plantes, en addition de cette nature dynamique des complexes TIM23, la structure des complexes TIM23 est encore compliquée par le fait que les protéines TIM23 et

TIM17 sont toutes deux codées par de petites familles multi-géniques (Murcha *et al.*, 2007). Ces spécificités des complexes TOM et TIM chez les plantes pourraient répondre aux spécificités d'import des protéines requises suite à l'endosymbiose à l'origine des chloroplastes. En effet, des changements importants des complexes d'import ont été nécessaires afin d'assurer un adressage correct des protéines vers leur organe cible.

2.4.2. Les complexes chloroplastiques TOC et TIC

Le complexe chloroplastique TOC est un complexe dynamique composé de nombreuses sous-unités : récepteur, canal et protéines accessoires (pour revue voir Andrés *et al.*, 2010, Shi & Teng 2013). La protéine formant le canal, appelée TOC75, est une protéine ayant une structure en tonneau β . Bien que TOC75 soit codée par un gène unique chez *Pisum sativum* (le pois), chez d'autres espèces, une petite famille multi-génique pourrait coder pour cette sous-unité (Paul *et al.*, 2013). Récemment, une étude phylogénétique a permis de proposer l'existence de deux sous-familles à l'intérieur de cette petite famille, ce résultat suggérant un rôle divergent entre ces protéines dans l'import (Topel *et al.*, 2012).

Plusieurs petites familles de gènes codent pour les récepteurs décrits dans la membrane externe des chloroplastes. Parmi elles, les deux familles TOC159 et TOC34 codent pour des protéines ayant une architecture similaire avec, en particulier, des protéines à domaines GTPases. La famille de récepteurs de la membrane externe des plastes de type TOC159 est, selon les espèces, composée de 1 à 4 paralogues (Kalanon & McFadden 2008, Lung & Chuong 2012, Shi & Theg *et al.*, 2013). La famille TOC34 est composée de 1 à 3 protéines dans les différents génomes de plantes (Kalanon & McFadden 2008, Shi & Theg 2013). La dernière famille de récepteurs de la membrane externe est plus controversée, il s'agit de la famille TOC64. Elle est composée de 1 à 3 membres par génome (Aronsson *et al.*, 2007, Shi & Theg 2013, Paul *et al.*, 2013). L'import des protéines dans les plastes serait modulé par l'abondance des récepteurs dans la membrane externe et leur capacité à former des complexes de translocation actifs (Bischof *et al.*, 2011).

La multiplicité des récepteurs TOC semble indiquer de probables interactions moléculaires et fonctionnelles entre les protéines mais pour l'instant aucun modèle précis n'a pu être proposé. Le complexe TIC semble être encore plus diversifié en raison des très nombreuses protéines impliquées dans l'import des protéines au travers de la membrane interne des plastides. Ainsi, selon la littérature, 13 protéines ou petites familles de protéines ont été identifiées dans ce processus : TIC20, TIC21, TIC22, TIC32, TIC40, TIC55, TIC56, TIC62, TIC100, TIC110, TIC214, HSP70 et HSP93 (Kovács-Bogdán *et al.*, 2010, Kikuchi *et al.*, 2013). Par ailleurs, il a été montré que plusieurs complexes TIC pourraient co-exister. Par exemple, les protéines TIC20-I, TIC56, TIC100 et TIC214 forment un complexe TIC spécifique. TIC214 étant une protéine codée par le gène chloroplastique *ycf1* alors que *TIC56* et *TIC100* sont des gènes nucléaires à copie unique (Kikuchi *et al.*, 2013). La complexité du mécanisme est augmentée car plusieurs protéines TIC forment de petites familles multi-géniques dans certains génomes végétaux. En résumé, les familles de protéines TIC32, TIC55 et TIC62 sont des régulateurs de potentiel calcique ou redox associés à l'activité de translocation. Les protéines de type TIC20 et TIC21 sont plus centrales dans les complexes TIC, elles sont considérées comme étant directement associées au canal d'import des complexes. Dans les génomes des plantes, de 2 à 9 gènes *TIC20* et jusqu'à 4 gènes *TIC21* ont été identifiés (Kalanon & McFadden 2008, Shi & Teng 2013). La protéine TIC110, codée généralement par un seul gène par génome, a été également impliquée dans les complexes TIC mais récemment sa présence dans le complexe a été contestée (Kikuchi *et al.*, 2013). La dernière famille de protéines TIC est la famille *TIC22* généralement composée de deux membres par génome (Kasmati *et al.*, 2013).

Finalement, en intégrant toutes ces données, un modèle récent a été proposé dans lequel différents types de complexes TOC et TIC sont assemblés en fonction du type de plaste, du type cellulaire et du stade de développement permettant la plasticité et la spécificité du système d'import. Cette hypothèse a permis à Li et Teng (2013) de proposer un modèle de reconnaissance des signaux N-terminaux d'adressage par les complexes de translocation (cf ci-dessus partie « signaux d'adressage »).

2.5. Le double adressage des protéines chez les organites des plantes terrestres.

L'identification de la localisation d'une protéine dans la cellule est une donnée importante pour comprendre le réseau et la dynamique dans lesquels la protéine s'insère pour assurer sa fonction. Pour cette raison, plusieurs méthodologies ont été développées afin de localiser les protéines, chacune permettant de proposer une localisation probable et l'intégration de plusieurs outils permettant de définir précisément la localisation des protéines (Millar *et al.*, 2009). Un intérêt particulier est apparu récemment pour des protéines montrant des localisations multiples dans la cellule. Ces localisations cellulaires complexes sont le reflet de mécanismes cellulaires développés pendant l'évolution pour répondre à des nécessités toujours plus grandes tout en restant flexible et sans augmenter de façon démesurée le nombre de gènes. L'importance de ce phénomène est par exemple illustré chez la levure par le grand nombre de protéines mitochondriales montrant une localisation dans un autre compartiment cellulaire (Dinur-Mills *et al.*, 2008, Ben-Menachem *et al.*, 2011). Des protéines montrant plusieurs localisations dans la mitochondrie ont été également caractérisées en détails. L'analyse des mécanismes de localisation des protéines a montré que des mécanismes très différents pouvaient être impliqués dans l'adressage des protéines à leurs différentes destinations (pour une revue complète voir Yogeve & Pines 2011). Deux grands types de mécanismes peuvent être ainsi mis en jeu : soit deux protéines différentes (ayant des signaux d'adressage différents) sont produites, soit un seul produit de traduction porte plusieurs signaux d'adressage. Le premier type de mécanismes peut lui-même être divisé en trois sous-mécanismes. Les protéines peuvent être codées par des gènes différents codant pour des isoprotéines avec de petites différences influençant uniquement leur localisation. Un gène unique peut conduire à la synthèse de plusieurs ARNs (différents sites d'initiation de la transcription ou d'épissage des ARN) ayant soit une localisation différente ou codant pour des protéines différentes. Enfin, un même gène et un même ARN peuvent coder pour plusieurs protéines ayant des sites d'initiation différents (mais en phase) permettant la synthèse ou non de signaux d'adressage N-terminaux (Yogeve & Pines 2011). Dans le cas du deuxième type de mécanisme (impliquant un seul produit de traduction portant plusieurs signaux d'adressage), quatre sous-mécanismes peuvent être distingués. Le premier implique la présence d'un signal

d'adressage unique qui permet la localisation de la protéine dans plusieurs compartiments, ce type de signal est appelé signal d'adressage « ambigu » car il est reconnu par plusieurs systèmes d'import permettant ainsi un adressage multiple. Un deuxième type de mécanisme utilise des signaux d'adressage indépendants dans la séquence de la protéine permettant chacun la localisation dans un compartiment spécifique. Dans ce cas, l'accessibilité des signaux peut influencer l'efficacité d'adressage des protéines. Un autre mécanisme proposé est la translocation inverse de protéines. Dans ce cas, la seconde localisation est obtenue lorsqu'une fraction des protéines transitant par le système d'import est bloquée, maturée par des peptidases dans sa région N-terminale puis rejetée par le complexe d'import de façon rétrograde. Le dernier sous-mécanisme proposé correspond au tri des protéines au travers de structures membranaires permettant le partage des protéines entre différents compartiments cellulaires (Yogev & Pines 2011).

Bien que les différents mécanismes proposés expliquent les adressages multiples dans la cellule, des modifications post-traductionnelles, le type cellulaire, le stade de développement et les conditions environnementales peuvent également influencer l'adressage des protéines. Par exemple, la composition et l'abondance des complexes d'import des plastides pouvant être modulées dans les différents types cellulaires et au cours du développement, il n'est pas surprenant que ces facteurs aient un impact sur la localisation de certaines protéines.

Afin de simplifier la suite du manuscrit, et comme c'est le phénomène qui m'a intéressé au cours de ma thèse, j'utiliserai désormais le terme « double adressage » en référence à des protéines ayant une localisation double dans les mitochondries et les plastides. Comme mentionné dans les chapitres précédents, les plastides et les mitochondries partagent un certain nombre de caractéristiques en raison, en particulier, de leur origine endosymbiotique. Bien qu'un très grand nombre de protéines nucléaires soient très spécifiquement adressées à l'un ou l'autre des deux organites, il existe une catégorie de protéines adressées à la fois vers les plastides et les mitochondries. Ainsi, dans une revue récente, il a été rapporté que plus de 100 protéines de plantes ont été montrées comme localisées dans les deux organites (Carrie & Small 2013, Xu *et al.*, 2013, Fuss *et al.*, 2013, Baudisch *et al.*, 2014). En particulier, 88 protéines d'*A. thaliana* ont été identifiées à la fois dans les mitochondries et les plastides

(Annexe, Tableau 9). De plus, une analyse des données de la base « *SUB-cellular location database for Arabidopsis proteins* (SUBA) » a permis à Carrie et Small de proposer que 277 protéines pourraient être potentiellement localisées dans les deux organites chez *A. thaliana*. La liste des protéines potentiellement doublement adressées n'étant pas publiée dans cet article et du fait de l'importance d'une telle liste pour mon travail de thèse, j'ai établi ma propre sélection de protéines candidates pour une double localisation dans les mitochondries et les plastides (Tableaux 10 et 11 de l'annexe). Les protéines ont été initialement sélectionnées sur la base de l'intégration des protéomes plastidiaux et mitochondriaux d'*A. thaliana* prédits. Néanmoins, seules les protéines présentant des données expérimentales de localisation plastidiale et mitochondriale ont été retenues pour former ces listes (cf Annexe, Tableau 10 et 11). Il est important de noter que ces tables n'incluent pas les résultats obtenus pour les protéines de la famille PPR (Pentatricopeptide Repeat) car l'analyse de la localisation des membres de cette famille est le sujet d'un chapitre dans la section des résultats et sera présenté en détails.

Comme observé précédemment (Carrie *et al.*, 2009a, Carrie & Small 2013), les protéines pour lesquelles le double adressage a été confirmé sont enrichies dans certaines catégories fonctionnelles comme le métabolisme de l'ADN, des ARN et la traduction, toutes étant essentielles et similaires dans les deux organites. Il a été suggéré par Carrie et Small (2013) que ces biais ne seraient pas le résultat de systèmes d'import particuliers mais plus probablement la conséquence de la fonction biologique des protéines. En effet, des protéines impliquées dans des processus biologiques partagés entre la mitochondrie et les plastides sont statistiquement plus souvent doublement adressées. Cette observation a permis également aux auteurs de spéculer que le double adressage pourrait être un processus limitant la variabilité génétique. En effet, le double adressage permettrait la diminution du nombre de gènes nucléaires mais surtout augmenterait la pression de sélection sur ces gènes afin de maintenir la fonction dans les deux organites. Même si l'origine et les causes du double adressage sont inconnues, une analyse récente de protéines orthologues chez trois espèces éloignées de *Viridiplantae* (*Oriza sativa*, *Physcomitrella patens* et *Arabidopsis thaliana*), a montré que le double adressage est un processus ancien et dynamique (Xu *et al.*, 2013). Ainsi, l'ancienneté de son apparition est illustrée par le fait que 11 des 16 protéines analysées ont montré une

conservation du double adressage chez les trois espèces étudiées, à l'inverse, le dynamisme du processus est caractérisé par la perte probable de la capacité de double adressage de certaines protéines.

Le mécanisme impliquant des signaux d'adressage ambigus (ciblant les protéines dans les deux organites) a été particulièrement étudié au cours des 10 dernières années pour deux raisons principales. D'une part, la plupart des aminoacyl-ARNt synthétases doublement adressées portent ce type de signaux (Berglund *et al.*, 2009). D'autre part, la comparaison entre la perception et l'import de ces protéines doublement adressées et de celles ayant un import canonique grâce à des signaux spécifiques a permis de mieux comprendre les systèmes d'import des deux organites. L'analyse de délétions dans les signaux ambigus des aminoacyl-ARNt synthétases a montré que ces signaux ne sont pas structurellement identiques (Berglund *et al.*, 2009). En particulier, les séquences permettant l'adressage dans chacun des deux organites peuvent être chevauchantes dans certaines protéines alors qu'elles sont distinctes dans d'autres. De plus, il a été montré que dans l'import de l'ARNt nucléotidyl-transférase non seulement le signal d'adressage ambigu mais aussi la stabilité de la protéine sont nécessaires à un import correct dans les deux organites (Leibovitch *et al.*, 2013). Pour ces raisons, les prédictions de double adressage sont très difficiles. Néanmoins, un outil de prédiction de signaux d'adressage ambigus permettant de proposer une localisation double des protéines dans les mitochondries et les plastides a été développé et actualisé récemment (Mitschke *et al.*, 2009, Fuss *et al.*, 2013). Même si ces prédictions sont encore limitées en raison de leur difficulté et du petit nombre de données de référence pour entraîner le logiciel, c'est le seul outil pour prédire les signaux d'adressage ambigus permettant un double adressage aux organites. Dans la récente actualisation et validation du logiciel de prédiction, Fuss *et al.* (2013) ont observé qu'il existe des différences de résultats expérimentaux de localisation en fonction du système d'expression utilisé. Pour cette raison, ils ont développé des logiciels spécifiques à chaque espèce et suggèrent l'utilisation de systèmes expérimentaux homologues pour la localisation des protéines. Par ailleurs, et de façon intéressante pour mon sujet de thèse, ils ont observé deux protéines PPR, l'une d'*Arabidopsis thaliana* et l'autre de *Physcomitrella patens*, ayant une double localisation. En particulier, la protéine de *P. patens* conserve sa double localisation dans les trois systèmes d'expression testés (*A. thaliana*, *N.*

tabacum et *P. patens*) tandis que celle d'*A. thaliana* montre une localisation double uniquement lorsqu'elle est testée en système homologue (Fuss *et al.*, 2013).

Le nombre de protéines présentant un double adressage est en constant croissance depuis sa découverte il y a un peu moins de 20 ans. A cette effet, une étude récente a suggéré, par des données prédictives, que plusieurs centaines de protéines végétales pourraient avoir une double localisation (Baudisch *et al.*, 2014). Ce travail se base sur l'extrapolation de résultats de localisation obtenus pour 16 protéines ayant une origine endosymbiotique probable représentant 4 catégories (A, B, C, Z), ces catégories sont caractérisées par la différence entre les scores de prédiction mitochondriale et plastidiale obtenus à l'aide du software TargetP (ΔMC). Les protéines de la catégorie A ont une ΔMC inférieure à 0,1, la catégorie B une ΔMC comprise entre 0,1 et 0,2, la catégorie C une ΔMC comprise entre 0,2 et 0,3 et la dernière catégorie Z une ΔMC supérieure à 0,8. Par fusion traductionnelle avec la protéine eYFP, 7 des 8 séquences d'adressage testées appartenant à des protéines des catégories A et B (possédant une faible différence ΔMC) ont montré une double localisation. De plus, 3 protéines des groupes C ou Z ont également été localisées, en fusion avec la eYFP, à la fois dans les mitochondries et les plastes. Ces localisations ont été évaluées dans un système transitoire d'expression utilisant des feuilles de pois et confirmées par des expériences d'import sur organites purifiés. Ces données suggèrent que la quantité des protéines avec une double localisation est sous-estimée.

Chapitre III

L'expression Génique Chez Les Organites des plantes Terrestres

La transcription chez les organites des plantes terrestres

Les complexes de transcription chez les organites

Le métabolisme des transcrits chez les organites

L'édition des transcrits des organites : aspects générales et évolutifs

Le mécanisme d'édition C en U de l'ARN des organites de plantes: Principes généraux

Les protéines du complexe agissant sur l'édition de l'ARN des organites

Les protéines MORF et leur implication dans les complexes d'édition.

La maturation par clivage des transcrits des organites

L'épissage d'introns chez les transcrits des organites

La Famille Des Protéines A Domaines Pentatricopeptide Repeat.

Généralités sur la famille des protéines PPR.

Evolution de la famille PPR

Les protéines PPR participant à la transcription des organites.

Rôles des protéines PPR au sein des complexes d'édition

Les protéines PPR participent à l'épissage des introns des organites

Mode de reconnaissance des acides nucléiques par les protéines PPR

La structure des protéines PPR et les implications dans le mode de reconnaissance de l'ARN

3. L'expression Génique Chez Les Organites Des Plantes Terrestres

La transcription, la maturation des transcrits et leur traduction sont les trois étapes clés du processus d'expression des gènes d'un génome. Cependant, même si l'ensemble des processus conduisant à l'expression des gènes sont partagés pour tous les génomes, des spécificités existent en fonction de l'organisme ou de l'organite dans lequel elle s'inscrit. Chez les eucaryotes, il existe une grande diversité des mécanismes d'expression des gènes des organites qui est en relation avec l'histoire évolutive de l'organisme et de l'organite (Schuster & Stern 2009, Barbrook *et al.*, 2010, Stern *et al.*, 2010, Small *et al.*, 2013, Hammani & Giegé 2014). En effet, l'expression des génomes mitochondriaux animaux implique des mécanismes différents de celle des mitochondries des plantes terrestres (Small *et al.*, 2013). Cependant, chez les organites des plantes terrestres, la mitochondrie et le plaste, les mécanismes d'expression des gènes sont relativement similaires, en partie parce qu'ils partagent une même histoire évolutive au sein de la cellule végétale (Stern *et al.*, 2010, Barkan A. 2011, Small *et al.*, 2013, Hammani & Giegé 2014). Lors de la co-évolution des génomes nucléaires et des organites la plupart des facteurs impliqués dans l'expression des génomes sont passés sous contrôle du génome nucléaire tandis qu'une faible proportion est restée sous contrôle des génomes des organites. Ainsi, la régulation des différentes phases nécessaires à l'expression des génomes des organites et à leur biogenèse implique une coordination du fonctionnement des génomes de la même cellule (Woodson & Chory 2008).

En raison de la complexité de la littérature disponible, et des objectifs spécifiques de ce travail de thèse, je présenterai plus particulièrement les données acquises en matière de transcription, d'édition et d'épissage des gènes des organites des plantes terrestres dans les sections suivantes de ce chapitre. En fin de chapitre, je présenterai en détails la famille des protéines à domaines Pentatricopeptide Repeat (PPR), facteurs impliqués dans l'ensemble de ces processus.

3.1. La transcription chez les organites des plantes terrestres

La transcription chez la mitochondrie et le plaste présentent des similitudes et des différences, les processus transcriptionnels étant généralement plus complexes chez le plaste que chez la mitochondrie (Liere *et al.*, 2011). En particulier, la transcription dans la mitochondrie semble être un processus peu contrôlé ne présentant pas de forte modulation, suggérant un rôle mineur de la transcription dans la régulation de l'expression des gènes (Giegé *et al.*, 2000, Holec *et al.*, 2006). Les unités transcriptionnelles des organites ressemblent plus aux unités transcriptionnelles des organismes procaryotes qu'à celles du génome nucléaire des eucaryotes. En particulier, la plupart des gènes sont transcrits en séquences primaires contenant deux ou plusieurs gènes (Bock R. 2007). Cependant, à la différence des procaryotes, les transcrits des organites sont soumis à des processus post-transcriptionnels complexes et variés avant d'être pris en charge par la machinerie de traduction (Binder & Brenicke 2003, Stern 2010, Small *et al.*, 2013). Bien que les complexes transcriptionnels de la mitochondrie ou du plaste ne soient pas totalement caractérisés, il est admis que la transcription est réalisée par l'activité de complexes contenant des ARN polymérases de type phagique (RPOT), codées par le génome nucléaire et adressées aux organites (aussi appelées complexes NEP pour Nuclear Encoded Polymerase) (Liere *et al.*, 2011). Chez les plastides, l'activité transcriptionnelle est aussi assurée par une ARN polymérase appelée PEP (pour Plastid Encode Polymerase) contenant à la fois des sous-unités codées par le noyau et par le génome du plaste (pour revue voir Liere *et al.*, 2011, Yagi *et al.*, 2012).

3.1.1. Les complexes de transcription chez les organites

Les polymérases RPOT font partie d'une famille de gènes nucléaires chez les plantes, le nombre de gènes la constituant étant variable suivant les espèces. Par exemple, il a été rapporté que les espèces *Arabidopsis thaliana* et *Physcomitrella patens* possèdent trois gènes, *Oriza sativa* en a deux et *Selaginella moellendorffii* n'en possède qu'un (Liere *et al.*, 2011). L'étude de localisation subcellulaire des ces protéines montre qu'elles peuvent être localisées exclusivement dans un des deux organites ou doublement adressées (Hedtke *et al.*, 2000, Hricova *et al.*, 2006, Kuhn *et al.*, 2009, Xu *et al.*, 2013). Chez *A. thaliana*, parmi les trois

INTRODUCTION

RPOT polymérases, une première est adressée aux mitochondries (RPOTm), une autre aux plastides (RPOTp) et la troisième est doublement adressée dans les deux organites (RPOTmp) (Hedtke *et al.*, 2000, Xu *et al.*, 2013). De plus, il a été montré que RPOTm est une protéine essentielle pour la survie de la plante, suggérant un rôle majeur de cette protéine dans la transcription de gènes mitochondriaux (Kuhn *et al.*, 2009, Tan *et al.*, 2010). De la même manière, la polymérase plastidiale RPOTp a été montrée comme étant nécessaire pour l'activité transcriptionnelle dans le plaste, puisque sa mutation produit un phénotype de réduction de taille et de chlorose qui sont associés à des défauts structuraux des chloroplastes et à une diminution de la transcription des gènes plastidiaux (Hricova *et al.*, 2006, Swiatecka-Hagenbruch *et al.*, 2008). En ce qui concerne la protéine RPOTmp, sa mutation entraîne un phénotype moins sévère que celui observé dans les mutants *rpoTm* et *rpoTp*. L'étude du mutant nul *rpoTmp* a permis de montrer que la protéine RPOTmp est nécessaire à la transcription optimale de certains gènes mitochondriaux et plastidiaux. Tous ces éléments suggèrent que les activités des trois polymérases ne sont pas redondantes (Courtois *et al.*, 2007; Swiatecka-Hagenbruch *et al.*, 2008, Kuhn *et al.*, 2009). Actuellement, il n'existe pas d'évidence claire montrant que les polymérases RPOT interagissent avec d'autres facteurs. Cependant, il a été suggéré que ces protéines nécessitent probablement l'activité de facteurs auxiliaires *in vivo* (Liere *et al.*, 2011).

Chez le plaste, une deuxième activité transcriptionnelle est présente et correspond aux complexes de type PEP. L'architecture du complexe PEP a été initialement supposée ressembler à celles des complexes polymérases de type bactérien mais des analyses récentes indiquent qu'il aurait une architecture différente et plus complexe (Pfalz & Pfannschmidt 2013). Le complexe PEP serait une adaptation du complexe transcriptionnel provenant de la machinerie de l'endosymbionte à l'origine du plaste. La composition du complexe PEP n'est pas totalement connue, mais deux fractions de solubilités différentes peuvent être physiquement séparées à partir d'extraits plastidiaux ayant une activité PEP. En particulier, une de ces fractions, connue sous le nom de « Chromosomes Transcriptionnellement Actifs » (TAC : Transcriptionally Active Chromosomes), se trouve dans des structures compactes, appelées nucléoides, attachées aux membranes, et contenant des molécules d'ADN et d'ARN (Pfalz *et al.*, 2006). Plusieurs études ont montré que ces structures possédaient une

composition variée. Steiner *et al.* (2011) ont suggéré que des facteurs intervenant dans la transcription, comme par exemple les facteurs sigma, participent probablement d'une manière transitoire au complexe. Néanmoins, des études montrent que différents mutants de facteurs sigma *sig1*, *sig2*, *sig6* présentent des phénotypes associées à la transcription de gènes chloroplastiques et il a été montré que certains d'entre eux peuvent interagir avec l'ADN plastidial (Lerbs-Mache S. 2011, Hanaoka *et al.*, 2012). En parallèle, l'étude des deux fractions PEP a permis d'identifier la composition de base d'un complexe constitué de cinq sous unités codées par le génome plastidial (*rpoA*, *rpoB*, *rpoC2* et deux variantes de *rpoC1*) et de 10 protéines associées à la PEP : les protéines PAP (PAP: PEP-Associated Proteins) qui sont codées par le génome nucléaire, néanmoins aucun facteur sigma n'a pu être identifié par cette approche (Steiner *et al.*, 2011). Leur association transitoire au complexe est une explication avancée pour expliquer qu'ils n'aient pas pu être identifiés biochimiquement. Dans l'étude de l'activité transcriptionnelle et de ses implications dans la biogenèse du plaste, Pfalz & Pfannschmidt (2013) ont récemment proposé un modèle dans lequel le complexe transcriptionnel est un domaine d'un macro-complexe constitué de plusieurs domaines de fonctions différentes, interconnectés au sein du nucléoïde. Ce modèle étant proposé chez les dicotylédones et certaines monocotylédones.

A l'exception de quelques unités transcriptionnelles, la plupart des unités plastidiales possèdent des promoteurs qui peuvent être reconnus par les complexes NEP et PEP (Zhelyazkova *et al.*, 2012). Il a par ailleurs été observé que les complexes NEP et PEP sont actifs dans tous les types de plastes (Liere *et al.*, 2011). Il est donc probable que la participation de chaque complexe à la transcription des gènes chloroplastiques est coordonnée en partie par des changements d'affinité pour les promoteurs.

3.2. Le métabolisme des transcrits chez les organites des plantes terrestres

La transcription est suivie par la maturation des séquences d'ARN qui détermine la formation des transcrits fonctionnels utilisés par la machinerie de traduction ou pour sa dégradation. Chez les organites de plantes, les processus post-transcriptionnels qui permettent d'obtenir des transcrits codant les protéines des organites font intervenir des modifications des

bases nucléotidiques, l'épissage d'introns lorsqu'ils sont présents, la coupure des transcrits primaires, la maturation des extrémités 5' et 3' et la dégradation des molécules d'ARN. Plusieurs revues décrivant l'ensemble de ces processus ont été publiées récemment (Stern *et al.*, 2010, de Longevialle *et al.*, 2010, Barkan A. 2011, Knoop V. 2011, Castandet & Araya 2011, Stoppel *et al.*, 2013, Takenaka *et al.*, 2013). Aussi, je présenterai principalement dans les sections suivantes les données les plus récentes et pertinentes par rapport à mon travail de thèse.

3.2.1. L'édition des transcrits des organites : aspects généraux et évolutifs

La modification de bases nucléotidiques dans la séquence d'une molécule d'ARN est une signature très répandue dans tous les organismes. Plusieurs types de modifications d'ARN sont connues comme des délétions, des insertions, des substitutions et des conversions de bases dans des séquences d'ARN faisant intervenir jusqu'à plus de 100 types de modifications chimiques (Grosjean *et al.*, 2005). A partir de la fin des années 80, un nouveau concept associé à la modification de bases dans les séquences de l'ARN a été décrit : « l'édition de l'ARN ». Ce concept, impliquant initialement des insertions ou délétions de bases dans les séquences des transcrits codant pour des protéines, a rapidement évolué pour intégrer des conversions et substitutions de nucléotides. Pour différencier le concept de modification de bases nucléotidiques de l'édition de l'ARN, H. Grosjean a précisé que le terme d'édition de l'ARN correspond aux transformations qui ont pour conséquences d'induire des changements dans la perception de l'information génétique portée par l'ARN alors que la modification de l'ARN n'induit qu'un changement purement structural (Grosjean *et al.*, 2005). Bien que plusieurs définitions du concept d'édition de l'ARN aient été données au cours des trois dernières décades - la plus récente étant celle de V. Knoop (Knoop 2011) - celle proposée par Grosjean, traduit à mon avis de la meilleure façon le concept d'édition dans sa globalité. Par conséquent, il servira de référence dans les paragraphes qui suivent.

Le phénomène d'édition de l'ARN a été initialement découvert dans la deuxième moitié des années 80, où il a été montré que des insertions de nucléotides dans l'ARN du gène mitochondrial *cox2* étaient nécessaires pour l'obtention d'un transcrit mature et fonctionnel

chez *Trypanosomma brucei* (Benne *et al.*, 1986). Cette découverte a été rapidement suivie par des travaux similaires, impliquant des délétions, conversions et substitutions de nucléotides dans les transcrits, principalement d'organites, chez différents organismes (Grosjean *et al.*, 2005, Knoop 2011). Dans une analyse de tous les processus d'édition de l'ARN chez les organismes, une distribution évolutive des phénomènes d'édition de l'ARN a été proposée (Knoop 2011, Gray 2012). Cette distribution semble être corrélée avec une machinerie protéique particulière à chaque processus d'édition (Gray 2012).

Chez les plantes, l'édition des ARN a été détectée pour la première fois dans des transcrits de gènes mitochondriaux de façon indépendante par trois équipes. Cette édition correspondant à des conversions de cytosine en uracile (C en U) (Gualberto *et al.*, 1989, Hiesel *et al.*, 1989, Covello & Gray 1989). L'histoire liée à ces études simultanées a été décrite dans trois articles lors du 20^{ième} anniversaire de la découverte de l'édition, montrant la richesse du processus scientifique (Grienenberger 2009, Gray 2009, Brennicke 2009). Par la suite, il a été mis en évidence que des conversions d'uracile en cytosine (U en C) pouvaient aussi avoir lieu dans des transcrits mitochondriaux, mais de façon plus rare que l'édition de C en U (Schuster *et al.*, 1990, Gualberto *et al.*, 1990). Un peu plus tard, le premier phénomène d'édition rapporté chez le chloroplaste a concerné le changement de C en U nécessaire pour produire le codon d'initiation du transcrit du gène *rpl2* chez le Maïs (Hoch *et al.*, 1991). On sait désormais que les conversions de C en U sont relativement communes chez les organites des plantes. La quasi-totalité de la littérature montre que l'édition de l'ARN se produit essentiellement dans les transcrits des organites, mais quelques études suggèrent que des transcrits d'origine nucléaire pourraient être soumis à l'édition. Cependant, ces études restent controversées et nécessitent d'être confirmées (Ebhardt *et al.*, 2009, Meng *et al.*, 2010). D'autre part, il a été montré que la fréquence d'édition chez les plantes est variable d'une espèce à l'autre. En général, on trouve un nombre plus important de sites d'édition chez la mitochondrie par rapport au plaste (Figure 7). Quantitativement, l'édition de l'ARN par conversion de C en U est le phénomène le plus répandu chez les deux organites (Chateigner-Boutin & Small 2010, Grewe *et al.*, 2011, Hecht *et al.*, 2011, Knoop 2011, Fujii & Small 2011, Takenaka *et al.*, 2013b), suivi par la conversion de U en C, processus d'édition un peu plus fréquent dans certaines espèces des bryophytes et fougères (Yoshinaga *et al.*, 1997,

INTRODUCTION

Kugita *et al.*, 2003, Grewe *et al.*, 2011), puis par l'édition de A en I, minoritaire (Delannoy *et al.*, 2009). Ces travaux ont permis de suggérer que chaque type de conversion est indépendant et suit des processus évolutifs relativement différents (Gray 2012).

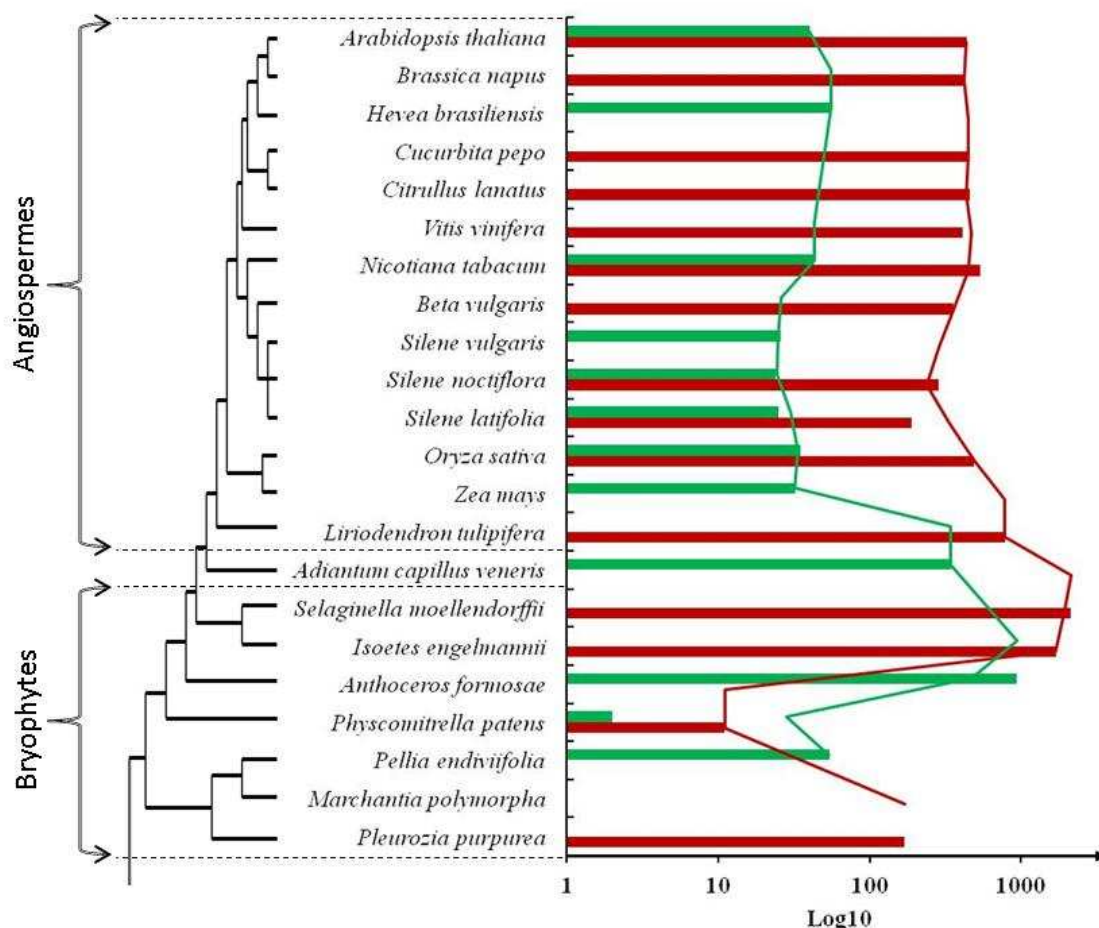


Figure 7. Sites d'édition des transcrits des organites dans différentes espèces de plantes terrestres.

Les barres en vert indiquent le nombre d'événements d'édition détectés dans les plastides tandis que celles correspondant aux mitochondries sont en rouge. Les données sont issues de différents articles (Hecht *et al.*, 2011, Sloan 2012, Grosche *et al.*, 2012, Lenz & Knoop 2013, Richardson *et al.*, 2013, Ruwe *et al.*, 2013). Les données d'édition plastidiale chez les Silènes, observées par Sloan *et al.*, 2012, sont en partie prédites d'une espèce par rapport aux autres.

L'édition de C en U est le processus le mieux compris du point de vue évolutif et mécanistique. Même si l'origine de ce processus n'est pas totalement définie, l'édition de C en U chez les organites de plantes semble s'être étendue après la divergence des embryophytes. Cette affirmation est basée sur le fait que, jusqu'à présent, aucun site d'édition n'a pas été détecté chez les algues vertes (Rudinger *et al.*, 2012 citant Steinhauser *et al.*, 1999). Les

données actuelles montrent que l'édition est un processus dynamique puisque des apparitions ou pertes de sites d'édition sont possibles au cours de l'évolution (Fujii & Small 2011, Richardson *et al.*, 2013). Malgré ce dynamisme, les données actuelles suggèrent un modèle évolutif qui est en relation avec l'adaptation de plantes au milieu terrestre (Fujii & Small 2011, Rudinger *et al.*, 2012). Par ailleurs, il a été observé qu'après une probable phase de croissance importante du nombre de sites d'édition chez les lycophytes, une tendance à la perte de sites d'édition semble avoir été favorisée chez les spermatophytes (Sloan *et al.*, 2010, Salmans *et al.*, 2010, Fujii & Small 2011, Rudinger *et al.*, 2012, Sloan *et al.*, 2012, Richardson *et al.*, 2013). De nos jours, il est important de noter que le peu de données disponibles sur les nombres de sites d'édition et la variabilité importante observée au sein des bryophytes ne permet pas de définir clairement un scénario évolutif pour le groupe des embryophytes. Dans ce contexte, on trouve cependant un cas particulier, *Marchantia polymorpha*, espèce appartenant à la classe des Marchantiopsida, dans laquelle aucun site d'édition n'a été détecté (Salone *et al.*, 2007, Rüdinger *et al.*, 2008, Rüdinger *et al.*, 2012). Il a été suggéré que l'absence totale de site d'édition chez cette espèce serait la conséquence d'une perte secondaire puisque des espèces proches, appartenant à la classe Jungermanniopsida, font toujours de l'édition (Grosche *et al.*, 2012).

3.2.2. Le mécanisme d'édition C en U de l'ARN des organites de plantes: Principes généraux

Actuellement, la composition des complexes d'édition reste partiellement connue et il n'est pas exclus que des facteurs encore inconnus soient impliqués dans ce processus. Malgré cela, plusieurs mécanismes expliquant le processus biochimique d'édition de l'ARN des organites des plantes ont été suggérés au fil du temps. Initialement trois mécanismes généraux probables ont été proposés :

1. Par délétion de la base cytidine à partir de son squelette de phosphate suivie d'une insertion de l'uridine.
2. Par glycosilation, changeant la nucléobase cytosine par l'uracile.

3. Par modification de groupe amide du carbone C4 de la nucléobase cytosine impliquant un désamination ou transamination.

Des expériences de marquage isotopiques, *in vitro* et sur organites purifiés, ont montré que ni le squelette phosphate de l'ARN ni la nucléobase subissait des excisions ou modifications (Rajasekhar & Mulligan 1993, Yu & Schuster 1995, Blanc *et al.*, 1995, Hirose & Siguira 2001). Ces données ont donc exclu les deux premiers mécanismes, laissant seulement la désamination ou transamination comme probables processus biochimiques responsables de l'édition de C en U. La transamination est un processus nécessitant des molécules accepteuses avec des groupes cétones. Takenaka *et al.* (2013b) ont suggéré que ce mécanisme est moins probable car les expériences d'édition sur organites purifiés allaient à l'encontre de l'implication de telles molécules (Takenaka & Brennicke 2003). Dans ce contexte, il est actuellement admis par la communauté scientifique que le mécanisme préférentiel d'édition passe par la désamination de la base cytosine.

Du point de vu mécanistique, il a été montré que le processus de l'édition des transcrits, dans différents organismes, implique l'activité de facteurs protéiques regroupés au sein d'un complexe. Par exemple le complexe d'édition chez *T. brucei*, appelé éditosome 20S, est une macro structure protéique de haut poids moléculaire (Goringer H. 2011). Ceci s'explique probablement par la nature complexe d'une multitude d'insertions et de délétions, nécessaires pour obtenir des transcrits fonctionnels dans la mitochondrie de cet organisme.

Des expériences biochimiques préliminaires effectuées chez le plaste ont suggéré que des facteurs, agissant en *trans* et possédant une spécificité importante, étaient responsables de l'édition des ARN des organites des plantes (Chaudhuri *et al.*, 1996). Plus tard, des expériences d'édition *in vitro* ont montré que les facteurs responsables de l'édition avaient une nature protéique et que la participation de séquences d'ARN guides en *trans* ne semblait pas requise (Hirose & Siguira 2001). Les données actuelles montrent l'implication de plusieurs types de protéines dans l'édition de l'ARN des organites mais la composition du complexe d'édition de C en U n'est pas totalement résolue. Il est clair qu'il n'existe pas un complexe unique agissant sur tous les sites d'édition des organites. Cependant les multiples complexes

qui peuvent exister semblent avoir une structure similaire dans laquelle des facteurs de spécificité de liaison à l'ARN et des facteurs responsables de l'activité catalytique sont impliqués. Malgré la multitude des données disponibles sur l'édition, la nature de la (des) protéine(s) responsable(s) de l'activité enzymatique reste un point majeur devant être élucidé.

3.2.3. Les protéines du complexe agissant sur l'édition de l'ARN des organites

A ce jour, seules trois types de protéines ont été associées au phénomène d'édition chez les plantes. Ainsi, certains membres de la famille de protéines Pentatricopeptide Repeat (PPR) et les Multiple Organellar RNA editing Factors (MORF) aussi connues comme RNA editing-Interacting Proteins (RIP) sont nécessaires à l'édition C en U de transcrits des deux organites (Takenaka *et al.*, 2012, Bentolila *et al.*, 2013, Takenaka *et al.*, 2013b). D'autre part, des protéines à motifs de liaison à l'ARN ou RRM (RNA Recognition Motif) sont impliquées dans l'édition, mais d'autres travaux sont nécessaires pour déterminer s'ils le font d'une manière directe ou indirecte (Tillich *et al.*, 2009, Sun *et al.*, 2013). La fonction des protéines PPR au sein des complexes d'édition sera détaillée dans une section qui leur est dédiée (cf section 3.4.4). De manière générale, il est accepté que les protéines PPR interviennent dans la spécificité de reconnaissance de l'ARN et il a été proposé qu'elles pourraient également participer à l'activité enzymatique d'édition (Salone *et al.*, 2007).

3.2.4. Les protéines MORF et leur implication dans les complexes d'édition.

Chez *A. thaliana*, où elle a été découverte, la famille MORF est composée de dix gènes (*MORF10*, étant potentiellement codée par le pseudogène *ATIG53260*). Cette famille est présente chez plusieurs Angiospermes mais elle n'a pas été détectée chez *P. patens* (Takenaka *et al.*, 2012, Takenaka *et al.*, 2013). Ces protéines sont considérées comme des facteurs généraux permettant une probable stabilisation des complexes d'édition (Takenaka *et al.*, 2012). En effet, les analyses de génétique inverse par analyse de mutants d'insertion ou par « RNA interference » (RNAi) ont montré que les sites d'édition mitochondriaux et

chloroplastiques d'*Arabidopsis* étaient affectés en grand nombre dans de nombreux mutants *morf* (cf Table 1 et Takenaka *et al.*, 2012, Bentolila *et al.*, 2013).

Bien que la localisation subcellulaire de toutes les protéines MORF ne soit pas totalement déterminée, les données de prédiction d'adressage suggèrent une localisation dans les organites pour l'ensemble de la famille. Par ailleurs, des données protéomiques montrent une localisation plastidiale des protéines MORF2 et MORF9, une localisation mitochondriale pour MORF3 et MORF5, et une double localisation pour MORF8 (Takenaka *et al.*, 2012, Bentolila *et al.*, 2013). La double localisation subcellulaire de cette dernière protéine fusionnée à la GFP, observée en microscopie confocale, est la seule donnée expérimentale disponible par des expériences de microscopie de fluorescence pour l'ensemble de la famille (Bentolila *et al.*, 2012). Elle complète la double localisation initialement détectée lors des analyses protéomiques et confirmée par des défauts d'édition observés dans les deux organites du mutant *morf8* (Bentolila *et al.*, 2012, Bentolila *et al.*, 2013).

En se basant sur le nombre de sites d'édition affectés dans les différents mutants (table 1), il est proposé que les protéines MORF1, MORF2, MORF3, MORF8 et MORF9 seraient des facteurs majeurs de l'édition tandis que MORF4, MORF5, MORF6, MORF7 et MORF10 auraient un rôle plus mineur dans ce processus. Plus précisément, selon Bentolila *et al.* (2013), le mutant *morf8* serait le plus fortement affecté dans l'édition, suivi par les mutants *morf3*, *morf1*, *morf2* et *morf9* qui présentent respectivement un nombre de sites d'édition affectés décroissant (Voir table 1). La proportion de sites affectés dans le mutant *morf8* est plus importante chez la mitochondrie (70%) que chez le plaste (22%) et il a aussi été observé que plus de 40% des sites d'édition mitochondriaux affectés dans le mutant *morf8* le sont aussi dans les mutants *morf1* et/ou *morf3*. De façon similaire, plus de 70% des sites plastidiaux sont affectés dans les mutants *morf2* et/ou *morf9* (Bentolila *et al.*, 2013).

INTRODUCTION

Tableau 1. Bilan des différentes analyses en génétique inverse et des sites d'édition affectés chez les mutants *morf* (Adapté de Takenaka *et al.*, 2012, Bentolila *et al.*, 2013)

Gène (AGI V10)	Nom	Lignées évaluées		Total de sites d'édition affectés	
		VIGS	Mutants	Plaste	Mitochondrie
<i>AT4G20020</i>	<i>MORF1 RIP8</i>	✓	<i>morf1-1</i> <i>morf1-2</i>	0	46 ^ψ -74
<i>AT2G33430</i>	<i>MORF2 RIP2</i>	✓	<i>morf2-1</i>	22-33 ^ψ	21
<i>AT3G06790</i>	<i>MORF3 RIP3</i>		<i>morf3-1</i> <i>morf3-2*</i>	2-7	46 ^ψ -201
<i>AT5G44780</i>	<i>MORF4 RIP4</i>		<i>morf4-1</i> <i>morf4-2*</i>	0	1 ^ψ -19
<i>AT1G32580</i>	<i>MORF5 RIP5</i>	✓	<i>morf5-1</i>	0	12-25
<i>AT2G35240</i>	<i>MORF6 RIP6</i>		<i>morf6-1</i>		1 ^ψ
<i>AT1G72530</i>	<i>MORF7 RIP7</i>	✓			6
<i>AT3G15000</i>	<i>MORF8 RIP1</i>		<i>morf8*</i>	8	474
<i>AT1G11430</i>	<i>MORF9 RIP9</i>	✓	<i>morf9-1</i>	24-33 ^ψ	1
<i>AT1G53260</i>	<i>MORF10</i> <i>RIP10</i>	✓		1	0

* Les noms de ces lignées ont été changés. Ils étaient originalement appelés *rip* dans Bentolila *et al.* (2013) ^ψ: Total de sites d'édition affectés et reportés dans Takenaka *et al.* (2012)

L'étude des interactions entre certaines protéines de cette famille par double hybride chez la levure a montré que les protéines MORF peuvent interagir entre elles. Ainsi MORF1 interagit fortement avec elle-même et les autres MORF testées (MORF2, MORF3 et MORF9) tandis que MORF3 présente un profil d'interaction différent. Elle ne se dimérise pas et semble interagir faiblement avec MORF2 et MORF9 et fortement avec MORF1. MORF2 et MORF9 ont un profil d'interaction similaire (Takenaka *et al.*, 2012). D'autre part, il a été démontré que les MORF peuvent aussi interagir en double hybride avec des facteurs d'édition de type PPR, la figure 8 présente les interactions connues à l'heure actuelle. Les profils d'interaction sont divers. Par exemple la protéine PPR MITOCHONDRIAL RNA EDITING FACTOR9 (MEF9) montre une interaction avec plusieurs MORF alors que MEF10 interagit plus particulièrement avec MORF8 et MORF2 (Takenaka *et al.*, 2012, Hartel *et al.*, 2013). Suite à

l'analyse de l'interaction par double hybride entre différentes combinaisons d'ORF et de fragments tronqués de MORF8 et de la protéine PPR REQUIERED FOR AccD RNA EDITING 1 (RARE1), un modèle d'interaction a été proposé (Bentolila *et al.*, 2012). Ainsi, l'interaction entre les protéines PPR et MORF ferait intervenir les domaines PPR des protéines de cette famille et la partie N-Terminale des protéines MORF. L'absence de domaine catalytique dans les protéines MORF et leurs interactions avec les domaines de liaison à l'ARN des protéines PPR suggèrent que les facteurs MORF auraient un rôle potentiel dans la stabilisation de la liaison du complexe avec l'ARN. Cependant, un rôle dans la réaction catalytique d'édition n'est pas à exclure. Ce modèle reste à confirmer avec d'autres combinaisons de couples PPR-MORF. Ces données précédentes confirment la présence de protéines MORF au sein de complexes d'édition et leur association aux protéines de type PPR.

Des données contradictoires entre les publications décrivant l'influence des MORF sur les sites d'édition doivent cependant être clarifiées. Par exemple, les données obtenues dans le cadre de la caractérisation de mutants, publiées par Takenaka *et al.* (2012), montrent que chez *morf2-1* et *morf9-1*, presque tous les 34 sites d'édition plastidiaux analysés sont affectés tandis que les expériences de RNAi de Bentolila *et al.* (2013) sur *MORF2* et *MORF9* montrent que 22 et 24 des 36 sites analysés étaient respectivement affectés. Ceci s'explique très probablement par la nature du matériel végétal étudié. En effet, les mutants d'insertion sont caractérisés par une absence ou une forte réduction de l'accumulation des transcrits correspondant au gène tandis que l'approche RNAi entraîne souvent une réduction moins forte et partielle de l'accumulation des transcrits. Par ailleurs, Bentolila *et al.* (2013) ont également mis en évidence des défauts d'édition dans des compartiments où les facteurs MORF ne sont pas détectés. Ainsi, le RNAi ciblant *MORF2* et *MORF9* a un impact sur l'édition mitochondriale (alors que les deux protéines sont probablement uniquement plastidiales) et les lignées mutantes affectées dans la protéine mitochondriale MORF3 montrent des défauts d'édition plastidiale. Dans le cas de *MORF2* et *MORF9*, la spécificité de l'extinction par RNAi n'a pas été démontrée et une explication simple pourrait être que des MORF mitochondriaux pourraient être touchés. Le cas des lignées mutantes *morf3* reste énigmatique, avant toute conclusion, il faudrait vérifier que la protéine MORF3 n'est réellement pas localisée dans les plastes.

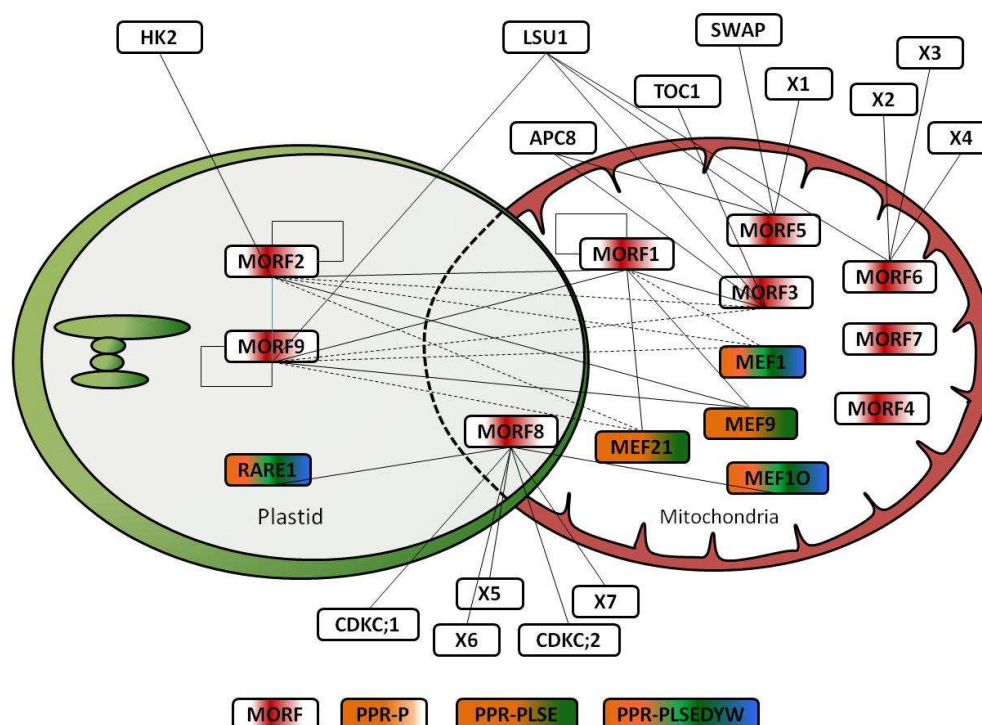


Figure 8. Réseau d'interactions protéine-protéine des protéines MORF au sein des plantes terrestres.

Les données ont été obtenues à partir de la littérature et de la base de données SUBA version 3. La totalité des interactions ont été identifiées en utilisant le système double hybride chez la levure. Les lignes en pointillés indiquent une faible interaction. Les protéines MORF1, MORF2 et MORF9 interagissent également avec elles-mêmes formant ainsi un homodimère. Les protéines X1 à X7 sont des protéines de fonction inconnue. Les gènes correspondant à ces protéines sont les suivants (version TAIR 10 de l'annotation du génome d'*Arabidopsis*): X1: *AT1G06510*, X2: *AT4G14740*, X3: *AT4G39050*, X4: *AT5G13810*, X5: *AT3G27960*, X6: *AT4G10840*, X7: *AT5G04840*. CDKC ;1 : protéine faisant partie du complexe Kinase cycline dépendante C1. CDKC ;2 : protéine faisant partie du complexe Kinase cycline dépendante C2.

3.3. La maturation par clivage des transcrits des organites

La maturation des transcrits codés par le génome des organites, nécessaire avant que ces derniers puissent être utilisés par la machinerie de traduction, fait intervenir plusieurs processus tels que l'épissage d'introns, le clivage des transcrits primaires et la génération des extrémités 5' et 3' matures.

3.3.1. *L'épissage d'introns chez les transcrits des organites*

Andéol Falcon de Longevialle a abordé en détail ce sujet dans son manuscrit de thèse, soutenue en 2010. Je ne présenterai donc dans cette section que les aspects généraux et les avancés réalisées dans ce domaine au cours des trois dernières années. Dans la section suivante j'exposerai des éléments particuliers aux PPR (cf section 3.4.5).

A la différence des organismes à l'origine des organites, l'épissage d'introns chez les organites est un processus actif catalysé *in trans* par l'activité de facteurs protéiques (Bonen *et al.*, 2008). Plusieurs gènes des organites des plantes possèdent des introns, avec une conservation variable entre les espèces. Ces introns appartiennent à deux classes principales observées chez les procaryotes comme chez les eucaryotes; les classes I et II avec une prépondérance de la classe II. Des analyses évolutives suggèrent que les introns de la classe II seraient les ancêtres des introns qui sont reconnus par le complexe d'épissage des gènes nucléaires (spliceosome) des eucaryotes (Rogozin *et al.*, 2012). Par ailleurs, Rogozin *et al.* (2012) ont suggéré que le spliceosome serait une modification évolutive de l'épissage des introns de classe II qui trouve probablement son origine dans le processus d'eucaryogénèse impliquant l'ancêtre de la mitochondrie.

Les introns de classe II sont principalement classés en trois sous classes majeures, les sous-classes IIA, IIB et IIC, seules les sous-classes IIA et IIB étant présentes chez les organites de plantes (de Longevialle *et al.*, 2010 citant Bonen L. 2008, Pyle M. 2010). Bien que les séquences primaires ne soit pas conservées, des analyses de la structure secondaires des introns de classe II ont montré qu'ils forment une structure conservée dans laquelle six domaines, de DI à DVI, sont distingués, comme le montre la figure 9A (Bonen L. 2008, Marcia *et al.*, 2013) Récemment, les structures tridimensionnelles de l'intron de classe II auto-épissable d'*Oceanobacillus iheyensis* ont été résolues à différentes étapes de la réaction d'épissage (Pyle M. 2010, Marcia *et al.*, 2013). Ces travaux décrivent une structure compacte composée de plusieurs super-hélices et précisent que son repliement requiert des centres métalliques monovalents (K^+), nécessaires au repliement initial du domaine I. Une fois la structure tertiaire du domaine I établie, elle coordonne le repliement des autres domaines à

INTRODUCTION

l'aide des centres bivalents (Mg^{+2}) pour finalement arriver à la structure tertiaire active présentée dans la figure 9B et C, formant un ribozyme. La mise en place de cette structure permet le bon positionnement des séquences du ribozyme et en particulier de son centre catalytique. Ainsi, la réaction d'épissage est obtenue par une attaque nucléophile du phosphate de l'ARN qui est facilitée par des centres métalliques bivalents (Mg^{+2}) et monovalents (K^{+}), permettant une coordination des molécules du centre réactionnel, l'ensemble suivant un mécanisme de substitution nucléophile 2 (SN_2) (pour plus de détails voir Marcia *et al.*, 2013). Pendant la réaction d'épissage, des changements conformationnels rapides du centre catalytique s'opèrent pour permettre le repositionnement des structures autorisant finalement la jonction entre les exons. Marcia *et al.* (2013) suggèrent que les données structurales similaires entre l'épissage des introns de type II et de celles du spliceosome confirment le lien évolutif entre ces processus. De plus ces auteurs proposent que les protéines associées au spliceosome auraient un rôle de stabilisation de la structure de l'ARN mais qu'elles ne semblent pas participer à la réaction d'épissage, ceci indiquant que, comme pour l'épissage des introns II, c'est l'ARN qui est le responsable catalytique de la réaction chimique. Il semble donc possible d'extrapoler cette interprétation à l'épissage des introns II des organites.

Plusieurs familles de protéines facilitent l'épissage des introns des gènes des organites. C'est le cas des protéines des familles Chloroplast RNA splicing and ribosome Maturation (CRM), Plant Organellar RNA Recognition (PORR), nuclear MATurases (nMAT), mitochondrial Transcription tERmination Factor (mTERF), PPR ou encore Accumulation of Photosystem One (APO dérivé de la caractérisation de son premier membre APO1 mais connu aussi comme Domain of Unknown Function 794 (DUF794)) (Stern *et al.*, 2010, de Longevialle *et al.*, 2010, Khrouchtchova *et al.*, 2012, Zmudjak *et al.*, 2013, Hammani & Barkan 2014). Bien que l'unique protéine de la famille OctotricoPeptide Repeat (OPR) détecté dans le génome d'*A. thaliana* est impliquée dans le métabolisme de l'ARN ribosomal 16S plastidial (Kleinknect *et al.*, 2014), il semble que ces protéines pourraient aussi être impliquées dans l'épissage des introns des organites car elles sont présentes dans le complexe de *trans*-épissage de l'intron 1 du gène *psbA* chez *C. reinhardtii* (Jacobs *et al.*, 2013).

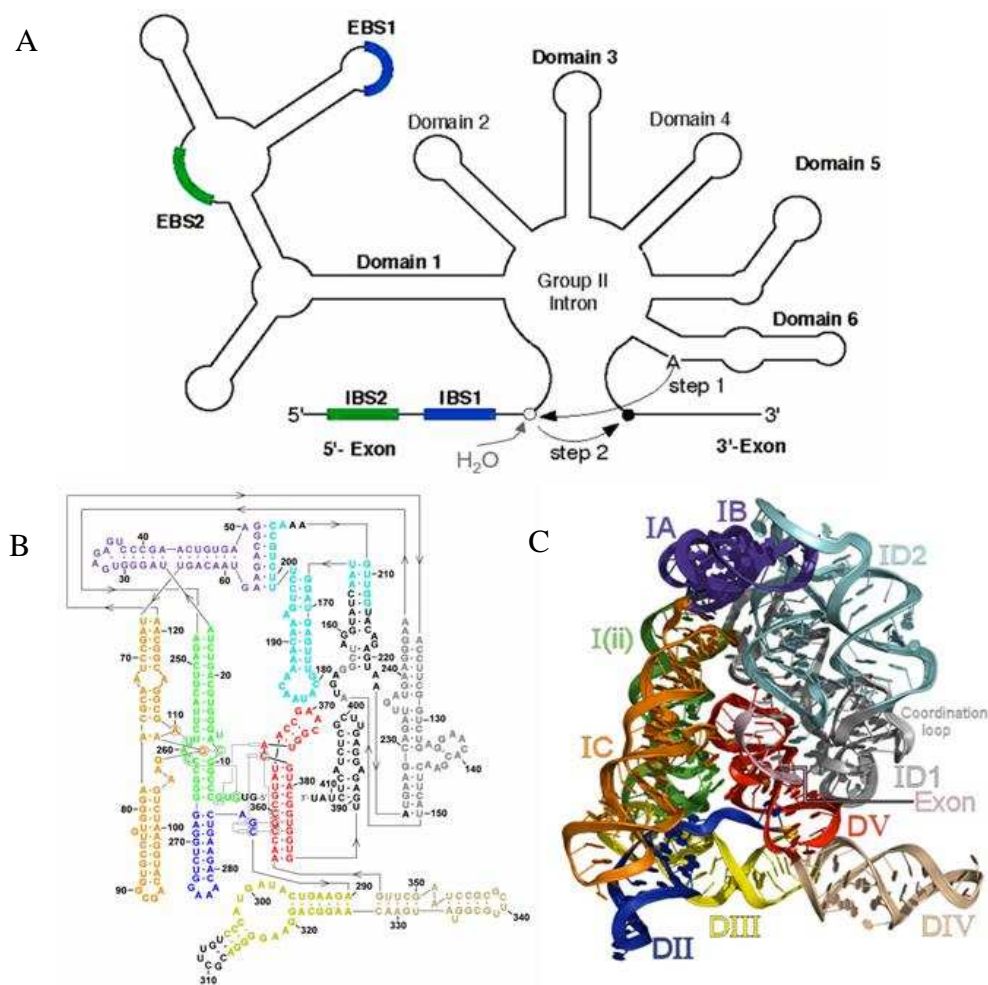


Figure 9. Structure secondaire et tertiaire d'un intron de type II.

A) La structure secondaire des introns de type II est schématisée, indiquant les domaines I à VI, les régions et les aspects importants pour la réaction d'épissage. B) Représentation de la structure tertiaire de l'intron de type IIC d'*Oceanobacillus iheyensis* et des sites de contact au sein de la structure dans son repliement tridimensionnel. Les couleurs indiquent les domaines et régions de l'intron : DI formé par plusieurs sous-domaines (Vert, violet, orange, gris, cyan), DII (Bleu), DIII (Jaune), DIV (rose) et DVI (noir). C) Structure tridimensionnelle de l'intron de type IIC d'*Oceanobacillus iheyensis* avec le même code couleur que pour la planche B. Le domaine VI n'est pas présenté dans cette structure. (toutes les figures proviennent de la page web du laboratoire de A. Pyle <http://www.pylelab.org/research/RNAtertiary/groupii.html>).

Les analyses montrent qu'il n'existe pas un seul modèle commun pour l'ensemble des facteurs d'épissage. Cependant, les résultats de phénotypage moléculaire permettent de préciser que les facteurs de la famille des protéines PPR influencent l'épissage d'un ou de quelques introns spécifiques tandis que les protéines à domaines CRM et nMAT influencent l'épissage de nombreux introns. Le faible nombre de protéines mTERF, PORR et APO

actuellement caractérisées ne permet pas de conclure sur leurs spectres d'activité. Actuellement, peu de travaux décrivent des interactions protéine-protéine entre des facteurs d'épissage. Deux complexes d'épissage ont été mis en évidence chez les chloroplastes, un premier commun aux plantes terrestres *A. thaliana* et *Zea mays* et un deuxième chez l'algue verte *C. reinhardtii* (Khrouchtchova *et al.*, 2012, Jacobs *et al.*, 2013). En effet, les analyses de Khrouchtchova *et al.* (2012) montrent que l'association de trois facteurs, la protéine à domaines PPR THYLACOID ASSEMBLY 8 (THA8), la protéine à domaine PORR WHAT'S THIS FACTOR 1(WTF1) et la protéine à domaine Ribonuclease III RNC1, est spécifique à l'épissage de l'intron du gène *trnA*. D'autre part, la mise en évidence d'un méga-complexe d'épissage de l'intron 1 du gène *psbA* chez le plaste de *C. reinhardtii* va en faveur de l'existence de plusieurs complexes chez les plantes. Ainsi, l'implication de différentes protéines sur un même intron observée par analyse de phénotypes moléculaires, le fait qu'on ne retrouve pas les mêmes protéines influençant ces épissages et la nature différente des complexes protéiques mis en évidence montrent qu'il existe probablement plusieurs complexes d'épissage des transcrits chez les organites. Cependant, l'implication d'un macro-complexe de type spliceosome chez les organites reste un sujet de débat entre les différentes équipes travaillant sur ce sujet (Barkan A. 2011, Jacobs *et al.*, 2013). Il est important de noter que l'expansion de familles de protéines différentes au sein de clades, comme par exemple les PPR chez les plantes terrestres, les protéines à domaine OPR chez l'algue *C. reinhardtii* et la divergence de certaines membres d'autres familles comme les protéines à domaine CRM entre ces espèces, suggère que les complexes d'épissage des plantes terrestres pourraient être différents des ceux des algues vertes. Ceci peut également être mis en relation avec l'évolution des génomes des organites et des processus de maturation des ARN, ces derniers semblant avoir divergé au moment de l'adaptation des plantes à l'environnement terrestre.

3.4. La Famille Des Protéines A Domaines Pentatricopeptide Repeat.

3.4.1. Généralités sur la famille des protéines PPR.

Bien que certaines protéines appartenant actuellement à la famille PPR aient été déjà partiellement caractérisées dans différents organismes (Manthey *et al.*, 1995, Fisk *et al.*, 1999), c'est seulement à partir des années 2000 et grâce au séquençage du génome nucléaire de l'espèce modèle *A. thaliana* que la famille de protéines PPR a été mise en évidence (Aubourg *et al.*, 2000, Small & Peeters, 2000). Ce nom a été proposé du fait de la similarité du motif PPR avec le motif *TetratricoPeptide Repeats* des protéines TPR. La signature PPR a été établie par la présence d'un motif dégénéré de 35 acides aminés, motif pouvant être répété entre 2 et 26 fois selon la protéine. Plus tard, la classification de la famille PPR a été établie grâce à une analyse bioinformatique de ses membres chez *A. thaliana* (Lurin *et al.*, 2004, Rivals *et al.*, 2006). Cette étude a permis de proposer une classification plus détaillée sur la base de trois variantes du motifs PPR, le motif P canonique de 35 acides aminés, un motif plus long L et un motif plus court S (du mot anglais « Short ») et de trois motifs protéiques additionnels en région C-terminale, les motifs E, E+ et DYW. Ainsi les membres de la famille ont été séparés en deux grandes sous-familles, la sous-famille P et la sous-famille PLS, cette dernière étant divisée en quatre sous-groupes. Plus tard, une distribution similaire des sous-familles au sein des angiospermes a été observée (O'Toole *et al.*, 2008, Fujii & Small 2011). Après l'étude de Lurin *et al.* (2004), la classification de la famille a peu changé mais il est important de noter qu'au sein de la sous-famille P, deux sous-groupes peuvent aussi être distingués. Ce léger changement donne une classification de la famille PPR qui, pour *A. thaliana*, est présentée dans la figure 10. Dans ce travail de thèse, les domaines E et E+ seront considérés comme constituant un seul domaine appelé E, comme l'ont suggéré O'Toole *et al.* (2008) auparavant. En fait, en se basant sur des alignements de séquences, ce domaine pourrait probablement être constitué de deux ou trois sous-domaines distincts. A ce jour, une étude a démontré que le domaine E ne peut pas être échangé dans des protéines PLS-E agissant dans différents organites, suggérant une fonction différente de ces régions (Chateigner-Boutin *et al.*, 2013). Cependant, ces données sont préliminaires, et toutes les hypothèses permettant d'élucider la fonction du domaine E n'ont pas été totalement analysées.

INTRODUCTION

Aussi, le manque de connaissance sur la fonction de ces sous-domaines justifie ce regroupement en un seul sous-groupe PLS-E.

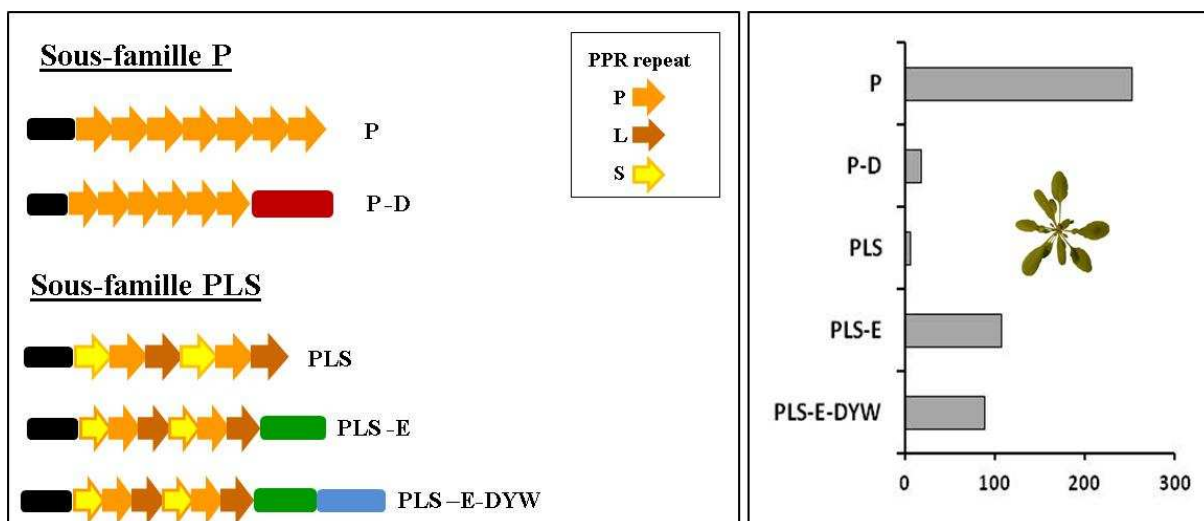


Figure 10. Classification de la famille PPR.

Les protéines de la sous-famille P comportent uniquement des domaines de type P. La sous-famille P est divisée en deux sous-groupes en fonction de la présence (P-D) ou l'absence d'un domaine additionel en C-terminal de la protéine. Les protéines de la sous-famille PLS portent des domaines P, L et S alternés et sont classées dans trois sous-groupes : PLS, PLS-E et PLS-E-DYW. Le sous-groupe PLS-E est composé par la fusion de sous-groupes PLS-E et PLS-EE+ originellement proposés. A droite de la figure, la répartition des 473 protéines PPR identifiées chez *A. thaliana* dans les différents sous-groupe est indiquée (Adapté de Lurin *et al.*, 2004 et Rivals *et al.*, 2006 et actualisé avec la version TAIR10)

3.4.2. Evolution de la famille PPR

Actuellement, il est admis que cette famille de protéines est présente au sein de la plupart des organismes eucaryotes et dans quelques procaryotes. La figure 11 présente le nombre de protéines PPR identifiées à ce jour par espèce (Lurin *et al.*, 2004, O'toole *et al.*, 2008, Schmitz-Linneweber and Small, 2008, de Longevialle A. 2010, Fujii & Small 2011, Tourasse *et al.*, 2013, www.phytozome.org, Schallenberg-Rüdinger *et al.*, 2013b). De manière générale, et à l'exception des plantes terrestres, le nombre de protéines constituant cette famille est relativement modeste chez la plupart des eucaryotes comme le montre la figure 11 où chaque organisme possède généralement entre 10 à 30 membres. Cependant, le séquençage et l'annotation de génomes nucléaires d'embryophytes, y a révélé une expansion importante de

la famille, pouvant aller de 100 à plus de 800 membres selon l'espèce (O'toole *et al.*, 2008, Schmitz-Linneweber *and* Small, 2008, Fujii & Small 2011). Ce constat a permis de suggérer que les membres de la famille PPR ont contribué à l'adaptation des plantes au milieu terrestre (Fujii & Small 2011). Certains auteurs ont aussi proposé que l'expansion des membres de la famille PPR, particulièrement celles de ceux impliqués dans l'édition, pourrait être la conséquence d'un processus évolutif de construction neutre (CNE) à l'origine de l'émergence de l'édition et non d'un processus évolutif de sélection positive (Maier *et al.*, 2008, Gray 2012). En particulier, Gray (2012) a suggéré que l'expansion des protéines PPR impliquées dans l'édition pourrait « relaxer » les processus de mutagenèse dans les génomes des organites, c'est-à-dire permettre que la mutation puisse apparaître.

Une évaluation claire de l'évolution de la famille au sein des embryophytes ne sera possible qu'après l'analyse d'un plus grand nombre de génomes d'espèces de bryophytes, de lycophytes, de fougères et de gymnospermes. Cependant, il est possible de proposer, à l'aide des données actuelles, qu'après une première expansion des gènes PPR chez tous les embryophytes, d'autres expansions plus spécifiques de la famille ont pu avoir lieu chez certaines espèces. Cette proposition repose sur les différences importantes observées entre le nombre de gènes codant pour des PPR chez *P. patens*, *S. moellendorffii* et les Angiospermes. Néanmoins, la possibilité de pertes localisées de gènes pourrait aussi expliquer ces données.

En particulier, chez les angiospermes, la taille de la famille est comprise entre 400 et 700 gènes avec une valeur moyenne proche de 500 membres (Figure 11). L'étude de la conservation des gènes PPR dans trois espèces distantes *A. thaliana*, *O. sativa* et *P. patens* a montré que la conservation était très importante chez les deux angiospermes *A. thaliana* et *O. sativa* mais aussi qu'une clade particulière semblait suivre un schéma évolutif très spécifique de chaque espèce (O'toole *et al.*, 2008, Fujii *et al.*, 2011). Ces résultats préliminaires ont été confirmés par la suite dans une analyse plus importante à l'échelle des génomes de plantes séquencés disponibles à la date de publication (Schmitz-Linneweber *and* Small, 2008, Fujii & Small 2011).

De façon intéressante, j'ai pu constater en utilisant une analyse récente cherchant à élucider les gènes en copie unique sur 20 génomes au sein des angiospermes (De Smet *et al.*, 2013), que parmi les 2986 gènes trouvés en copie unique dans ces génomes, plus de 7% (217) entre eux sont des gènes PPR. Ces PPR se répartissent proportionnellement entre tous les sous-groupes constituant la famille. Bien que De Smet *et al.* (2013) aient trouvé un enrichissement des catégories fonctionnelles associées à l'embryogénèse, aux chloroplastes et aux processus de base impliqués dans survie de la cellule, le nombre des gènes par catégorie des familles des protéines n'a pas été discuté par ces auteurs, probablement parce qu'ils se sont essentiellement focalisés sur l'enrichissement en terme d'ontologie de gènes (GO) ce qui ne leur a pas permis d'identifier les PPR. Les données ci-dessus montrent donc un enrichissement significatif en gènes PPR en copie unique dans les génomes des angiospermes ($p < 10^{-40}$ si l'on utilise un test statistique hypergéométrique avec les données de De Smet *et al.* (2013) et un nombre de gènes par génome de plantes analysées de l'ordre de 25000 à 30000 gènes, parmi lesquels 400 à 700 coderaient pour des PPR). Ainsi, en tenant compte des multiples duplications des génomes des angiospermes, il semblerait qu'il existe une pression de sélection favorisant la conservation des gènes uniques et/ou une résistance à la duplication de plus de 45% des gènes PPR au sein des angiospermes, comme l'ont proposé De Smet *et al.* (2013) pour l'ensemble de 2986 gènes identifiés. La nature de ce processus évolutif reste à être identifiée. Cependant ces données viennent confirmer des résultats précédents montrant une conservation importante des gènes PPR chez les angiospermes (O'toole *et al.*, 2008, Fujii & Small 2011).

Les protéines PPR participent à la maturation des ARN en intervenant dans la plupart des mécanismes post-transcriptionnels. Ainsi, de nombreux travaux ont permis de montrer qu'elles participent aux processus de transcription, d'édition, d'épissage et de maturation des transcrits et de leur traduction (Schmitz-Linneweber *and* Small, 2008; de Longevialle *et al.*, 2010; Takenaka *et al.*, 2013, Hammani *et al.*, 2013). A l'heure actuelle, plus d'une centaine de protéines PPR ont été caractérisées fonctionnellement dans différents organismes, dont plus de 80 chez les plantes. Dans ces travaux, divers phénotypes associés aux défauts moléculaires touchant les événements de maturation des ARN ont été décrits. Ces défauts conduisent à des phénotypes macroscopiques divers comme l'embryon-létalité, la stérilité Male Cytoplasmique

INTRODUCTION

(CMS), défauts de maturation de la graine, retard de croissance, photosynthèse affectée, défaut dans la caryogamie et de la signalisation rétrograde, etc... mais aussi absence de phénotype macroscopique évident. Ces phénotypes macroscopiques sont généralement expliqués par la nature de l'ARN cible touché et non par le type de maturation concernée.

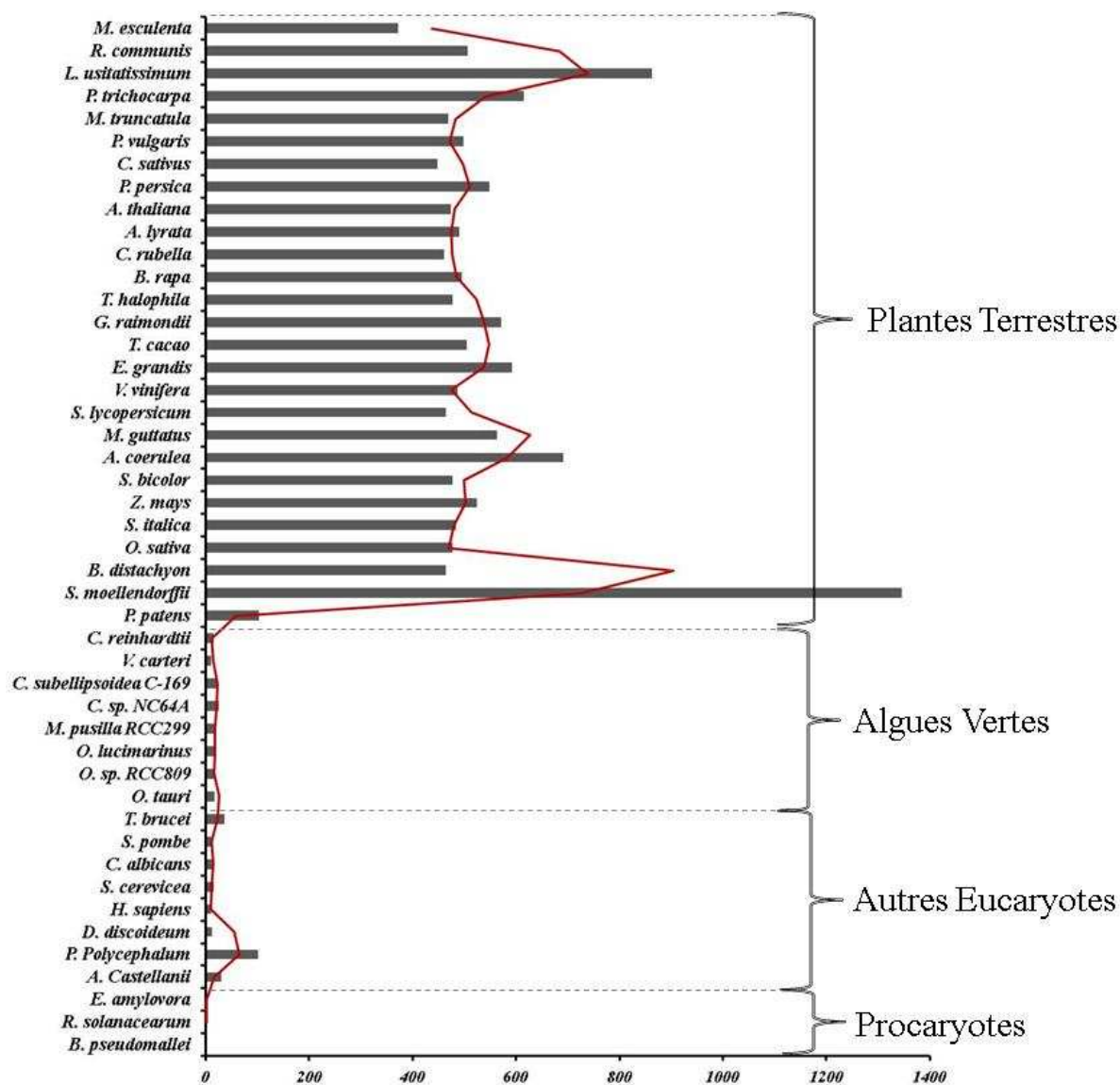


Figure 11. Nombre de gènes de la famille PPR rencontrés chez les différents organismes vivants.

Les données ont été obtenues à partir de la littérature et des séquences de génomes de la division Viridiplantae présents sur le site www.phytozome.org version 9 du Joint Genome Institut (JGI).

3.4.3. Les protéines PPR participant à la transcription des organites.

L'implication des protéines PPR dans la transcription mitochondriale chez les plantes est un sujet controversé. Avant la définition du domaine PPR, une étude chez le blé avait indiqué que la protéine p63 (nom lié à sa masse moléculaire de 63 KDa) pouvait se lier à l'ADN de la séquence promotrice du gène *cox2* et activer sa transcription (Ikeda & Gray 1999). La séquence de cette protéine est proche de celle de protéines paralogues codées par les gènes *AT1G15480*, *AT1G80270* et *AT3G15590* d'*A. thaliana*, appartenant à la famille des protéines PPR du sous-groupe P. Parmi ces protéines, seule la fonction de la protéine codée par *AT1G80270*, nommée PENTATRICOPEPTIDE REPEAT 596 (PPR596), a été partiellement caractérisée, la mutation du gène entraînant l'altération de la croissance de la plante (Doniwa *et al.*, 2010). Bien que les analyses fonctionnelles soient partielles et peu concluantes, ces données ne montrent pas d'effet sur la transcription des gènes mitochondriaux. De plus, il a aussi été montré que ces trois protéines se liaient de manière non spécifique aux séquences promotrices et qu'elles n'interagissaient pas avec la polymérase RPOTm ni avec RPOTmp (Liere *et al.*, 2011).

La participation de domaines PPR dans la transcription mitochondriale est mieux décrite chez les PPR humaines, particulièrement en interaction avec la RPOTm. La RPOTm humaine (connue aussi sous les noms POLRMT ou mtRNAP) possède deux domaines PPR dans son extrémité N-terminale. Ces deux domaines ne sont pas présents dans la polymérase de type phagique T7. Malgré que ce faible nombre de domaines PPR, cette protéine a été considérée comme une des sept protéines PPR actuellement identifiées chez l'Homme (pour revue voir Rackham & Filipovska 2012, Lightowlers & Chrzanowska-Lightowlers 2013). Les travaux permettant la résolution de la structure de la protéine RPOTm humaine ont été les premiers à confirmer la conformation du domaine PPR en hélice-boucle-hélice (Ringel *et al.*, 2011). Par ailleurs, la récente résolution de sa structure au sein du complexe d'élongation met en évidence que le mode d'action de la protéine diffère considérablement de celle de la polymérase du phage T7. De plus, ces travaux montrent que c'est face aux domaines PPR que la séquence naissante d'ARN apparaît (Schwinghammer *et al.*, 2013). Cet élément suggère que les domaines PPR de RPOTm peuvent participer à la stabilisation de la séquence d'ARN

INTRODUCTION

naissante. Il est important de noter que la présence de domaines PPR semble être une caractéristique particulière à la RPOTm humaine, par exemple aucune des RPOT des organites de plantes ne porte de domaine PPR. Il faut remarquer que la région N-terminale, incluant les domaines PPR, est mal conservée entre les RPOT mitochondriales des différents organismes. Cette région de la protéine RPOTm semble être impliquée dans un couplage de la transcription avec la traduction chez les champignons (RPO41) (Deshpande & Patel 2012) qui pourraient ainsi porter une région N-terminale structurellement et fonctionnellement similaire à celle de la RPOTm humaine.

Une deuxième protéine PPR humaine, La protéine LEUCINE-RICH REPEAT PPR CASSETTE (LRPPRC), a été proposée pour participer à la transcription mitochondriale. Il a été trouvé que cette protéine pouvait stimuler la transcription et interagir avec la RPOTm (pour revue voir Lightowlers & Chrzanowska-Lightowlers 2013). Mais une récente analyse conteste ces résultats et montre que la protéine LRPPRC ne stimule pas la transcription *in-vitro* dans les conditions analysées et n'interagit pas avec la RPOTm (Harmel *et al.*, 2013) ce qui remet en question la participation de cette protéine dans la transcription.

Chez le plaste, seulement deux protéines PPR appartenant au sous-groupe P ont été montrées comme impliquées dans la transcription. Tout d'abord la protéine appelée PLASTID TRANSCRIPTIONALLY ACTIVE CHROMOSOME 2 (pTAC2) [aussi connue comme PEP ASSOCIATED PROTEIN 2 (PAP2)] a été identifiée par des approches biochimiques associée avec le cœur du complexe de transcription PEP (Pfalz *et al.*, 2006, Steiner *et al.*, 2011, cf. section 3.1.). Cette protéine fait partie d'une clade spécifique constituée de 8 membres de protéines PPR, comportant un domaine SMALL MUTS-RELATED (SMR) de type endonucléase situé en région C-terminale de la protéine, en plus de 16 domaines PPR prédits (Liu *et al.*, 2013). L'analyse des mutants *ptac2* a montré que l'activité de pTAC2 est importante pour la survie de la plante. En effet, le mutant homozygote chez *A. thaliana*, *At-ptac2*, ne peut germer qu'après l'ajout d'une source carbonée sur milieu gélosé. Par contre, chez le Maïs, les mutants *Zm-ptac2* peuvent se développer sans source carbonée jusqu'aux premiers stades de la plantule (Williams-Carrier *et al.*, 2013). Les plantules obtenues montrent un phénotype peu pigmenté de couleur ivoire ou chlorotique. Les analyses de l'activité

transcriptionnelle chloroplastique ont montré que *At-ptac2* et *Zm-ptac2* présentent un défaut dans l'expression des gènes associés à l'activité transcriptionnelle de la PEP (Pfalz *et al.*, 2006, Williams-Carrier *et al.*, 2013). Même si sa fonction moléculaire précise n'est pas claire au sein du complexe PEP, tous les éléments précédents permettent de penser que pTAC2 agirait directement dans la transcription des gènes chloroplastiques. La protéine DELAYED GREENING 1 (DG1) est une deuxième PPR qui a aussi été associée à la transcription du chloroplaste dans les étapes précoces du développement de la plante, particulièrement au niveau des cotylédons (Chi *et al.*, 2008, Chi *et al.*, 2010). Bien que la protéine DG1 n'ait pas été détectée dans le complexe PEP, Chi *et al.* (2010) ont montré que le mutant *dgl* présentait une réduction de la transcription chloroplastique associée à la PEP. Plus tard, il a été montré que DG1 pouvait interagir, principalement par sa partie C-terminale, avec le facteur de transcription sigma SIG6. Les mutants *sig6* et *soldat8* affectant la fonction du gène *SIG6* présentent un phénotype similaire à *dgl* dans les étapes précoces du développement de la plante (Coll *et al.*, 2009, Chi *et al.*, 2010).

3.4.4. Rôles des protéines PPR au sein des complexes d'édition

Les données actuelles montrent qu'au sein des complexes d'édition, les protéines PPR contribuent à la spécificité de reconnaissance de l'ARN et, bien que ce soit encore controversé, il a été proposé que certaines PPR pourraient aussi contribuer à l'activité catalytique (Salone *et al.*, 2007, voir aussi section 3.2.1). Ainsi, plus de 60 protéines PPR impliquées dans l'édition des transcrits des organites ont été identifiées chez les plantes terrestres (cf. Annexe Tableau 12). Par exemple, chez la mousse *P. patens*, les 11 sites d'édition mitochondriaux ont été associés à l'activité de huit protéines PPR (Ichinose *et al.*, 2013, Schallenberg-Rüdinger *et al.*, 2013). Par ailleurs, à l'exception de la protéine AtPPR_596, tous les facteurs d'édition à motifs PPR identifiés chez les plantes terrestres appartiennent à la sous-famille PLS.

Les données biochimiques ont montré que les domaines PPR des facteurs d'édition chez *A. thaliana*, *O. sativa* et *P. patens* pouvaient se lier aux séquences d'ARN à proximité de sites d'édition ciblés (Okuda *et al.*, 2006, Tasaki *et al.*, 2010, Hammani *et al.*, 2011, Toda *et*

INTRODUCTION

al., 2012, Okuda & Shikanai 2012). Par exemple, la protéine CHLOROREPIRATORY REDUCTION 4 (CRR4) a une affinité spécifique pour la séquence d'ARN autour de son site d'édition cible sur le transcrit *ndhD* (de -25 nt à +10 nt) (Okuda *et al.*, 2006). D'autre part, les protéines ORGANELLAR TRANSCRIPT PROCESSING87 (OTP87), CRR22 et OTP82, intervenant dans l'édition de multiples cibles, peuvent aussi se lier avec une affinité importante aux séquences des ARN autour de ces sites d'édition (Hammani *et al.*, 2011, Okuda & Shikanai 2012). Dans le cas de CRR22, les données suggèrent que la reconnaissance de chacun des sites dépend d'une longueur de séquence d'ARN différente, variable selon les cibles (Okuda & Shikanai 2012).

Comme décrit précédemment, la sous famille PLS est caractérisée par la présence dans la partie C-terminale de chacun des ses membres de domaines E ou E et DYW. Bien que la participation dans l'édition des domaines E et DYW ne soit pas encore totalement comprise, des données évolutives et fonctionnelles montrent qu'ils sont associées à l'activité d'édition de C en U. Une discussion détaillée a été abordée dans le travail de thèse de Veronique Salone (2008) et récemment de Clément Boussardon (2013) au sein de l'équipe. Ici, je présenterai brièvement les sujets et données majeurs avec une vision d'ensemble actualisée. Les analyses génétiques impliquant des mutations dans le domaine E et DYW sont un point en faveur de ce rôle. Ainsi, il a été démontré que ces mutation, en particulier les délétions du domaine E, affectent l'édition comme cela été observé pour les protéines CRR4, CRR22 et CRR28 chez le chloroplaste et les protéines MITOCHONDRIAL EDITION FACTOR9 (MEF9) et MEF12 chez la mitochondrie (Okuda *et al.*, 2007, Okuda *et al.*, 2009, Takenaka M. 2010, Hartel *et al.*, 2013). En considérant les derniers résultats, deux hypothèses sont actuellement proposées pour tenter d'expliquer le rôle du domaine E, l'une n'excluant pas l'autre. La première propose que ce domaine participe à la reconnaissance des nucléotides très proches du site d'édition. Elle repose sur le fait que le dernier domaine PPR semble s'aligner entre trois et cinq nucléotides avant le site d'édition (Yagi *et al.*, 2013, Takenaka *et al.*, 2013) et au fait que ce domaine est prédit pour avoir une structure de type PPR. Dans la deuxième hypothèse, le domaine E permettrait le recrutement ou l'interaction d'autres facteurs du complexe, probablement avec l'acteur responsable de l'activité catalytique d'édition.

La fonction du domaine DYW est moins comprise et beaucoup plus controversée. Tout d'abord, les données vont en faveur de l'hypothèse que ce domaine est en partie responsable de l'activité catalytique d'édition. Ainsi, les analyses phylogénétiques montrent que ce domaine est associé à la présence ou l'absence d'édition dans les espèces actuellement analysées (Salone *et al.*, 2007, Rüdinger *et al.*, 2012). De plus, ce domaine contient une signature peptidique constituée d'acides aminés conservés, similaire à celle observée dans les protéines cytidine désaminases (Salone *et al.*, 2007). Dans ce sens, il a été aussi trouvé que la protéine DYW1, ne possédant que le domaine DYW, a une affinité pour les atomes métalliques de zinc (Zn). Or, c'est aussi un facteur important pour l'activité des cytidines désaminases (Boussardon 2013, Hayes *et al.*, 2013). Par ailleurs, il a été montré que toutes les protéines agissant sur l'édition mitochondriale de *P. patens* appartiennent aux PPR du sous-groupe DYW, ce qui implique que le domaine DYW est présent depuis l'origine du phénomène d'édition chez les plantes terrestres (Schallenberg-Rüdinger *et al.*, 2013). Ce constat suggère également que les protéines PLS-E-DYW pourraient être les ancêtres de protéines du type PLS-E, ces dernières ayant été générées par délétion du domaine DYW et dans une certaine mesure d'une partie du domaine E (Rivals *et al.*, 2006). Enfin, le fait qu'une mutation ponctuelle dans ce domaine DYW chez la protéine MEF1 entraîne une perte d'édition de ses cibles constitue un argument supplémentaire en faveur de l'hypothèse de ce rôle du domaine DYW.

Cependant, certains résultats suggèrent également que le domaine DYW n'est pas le responsable direct de l'activité catalytique d'édition. En effet, jusqu'à présent, aucune expérience *in vitro* a permis de détecter une activité catalytique d'édition pour des protéines à domaine DYW (Nakamura & Sugita 2008, Okuda *et al.*, 2009, Boussardon C. 2013) Par ailleurs, des formes tronquées des protéines CRR22 et CRR28, sans le domaine DYW, peuvent compléter les mutants correspondants (Okuda *et al.*, 2009). De plus, il a été observé que des protéines avec le domaine DYW comme la protéine CRR2 chez *Arabidopsis* et la protéine PpPPR_43 chez *P. patens* ne participent pas à l'édition (Hashimoto *et al.*, 2003, Ichinose *et al.*, 2012). Dans le cas de CRR2, une activité endoribonucléase a été signalée pour le domaine DYW ce qui permet de maturer le transcrit polycistronique *rsp12-rps7-ndhB*, et de séparer les transcrits *rps7* et *ndhB* (Hashimoto *et al.*, 2003, Okuda *et al.*, 2009). Par ailleurs, la

protéine PpPPR₄₃, appartenant au sous-groupe PLS-E-DYW de *P. patens*, participe aux processus d'épissage mitochondrial et non à l'édition. En effet, le mutant perte de fonction correspondant ne présente aucun défaut d'édition dans les 11 sites d'édition mitochondriaux actuellement connus mais montre un défaut d'épissage de l'intron 3 du gène *cox1* (Ichinose *et al.*, 2012).

Néanmoins, ces résultats contradictoires impliquant le domaine DYW sont discutables et certaines explications peuvent être suggérées. Face à la perte d'activité d'un domaine DYW d'une protéine de type PLS-E-DYW, il est possible que les complexes impliqués dans le processus d'édition des plantes recrutent alternativement d'autres facteurs DYW, secondaires *en trans*, assurant ainsi l'édition de sites indispensables au bon fonctionnement des organites. Cette possibilité se base sur le fait que les domaines DYW de CRR22 et CRR28 sont interchangeablement fonctionnellement et que l'interaction entre la protéine PLS-E CRR4 et la protéine DYW1 permet l'édition d'un site dans le transcrit *ndhD* (nucléotide 116494 du génome chloroplastique) (Okuda *et al.*, 2009, Boussardon *et al.*, 2012). Le domaine DYW pourrait aussi avoir, dans certaines conditions, une double fonctionnalité. Par exemple, il a été observé *in vitro* que les domaines DYW d'OTP85 et de CRR22 ont une activité endoribonucléase (Nakamura & Sugita 2008, Okuda *et al.*, 2009). Or, le mutant *otp85* montre une perte d'édition dans un site du transcrit *ndhD* (116494) et les mutants *crr22* présentent un défaut d'édition dans les sites *rpoB* (25779), *ndhB* (96419) et *ndhD* (116281). Finalement, les nouvelles technologies de séquençage des transcrits par « RNAseq » (Ruwe *et al.*, 2013, Bentolila *et al.*, 2013) ont récemment permis d'identifier de nouveaux sites d'éditeurs, qui n'avaient pas pu être détectés précédemment. Ces dernières données suggèrent que le nombre de sites d'édition actuellement connus sont très probablement sous-estimés et qu'il n'est pas à exclure que des protéines comme CRR2 ou PpPPR43 de *P. patens* puissent finalement être impliquées dans l'édition.

3.4.5. Les protéines PPR participent à l'épissage des introns des organites

L'implication de protéines PPR dans l'épissage des introns des organites des plantes terrestres a été caractérisée principalement par l'analyse des mutants. Actuellement, 11

INTRODUCTION

protéines PPR ont été décrites comme impliquées dans l'épissage de respectivement 9 et 4 introns dans les chloroplastes et les mitochondries. A l'exception d'OTP70 et PpPPR43, ces protéines appartiennent toutes à la sous-famille P, certaines possédant des domaines ou régions conservés particuliers à l'extrémité C-terminale. L'analyse des mutants correspondants a montré que ces facteurs PPR ont une spécificité pour l'épissage d'un ou deux introns, les défauts moléculaires entraînant généralement des phénotypes macroscopiques importants comme la diminution de la capacité photosynthétique et/ou le ralentissement de la croissance. Bien qu'il soit montré que plusieurs protéines peuvent agir sur l'épissage du même intron, peu de données biochimiques sont actuellement disponibles sur la composition des complexes d'épissage et/ou sur les interactions protéine-protéine entre ces facteurs. A ce jour, et comme je l'ai décrit précédemment (cf section 3.3.1.), un seul complexe impliquant la protéine PPR THYLACOID ASSEMBLY 8 (THA8) a été partiellement déterminé chez le plaste de plantes terrestres (Khrouchtchova *et al.*, 2012, Hammani & Barkan 2014).

L'implication de protéines PPR comme facteurs d'épissage chloroplastique a été initialement suggéré pour la protéine HIGH-CHLOROPHYLL-FLUORESCENCE 152 (HCF152), celle-ci agissant probablement sur l'épissage de l'intron du gène *petB* chez *A. thaliana* (Meierhoff *et al.*, 2003, Nakamura *et al.*, 2003). Nakamura *et al.* (2003) ont montré que HCF152 pouvait s'associer à trois régions du transcrit polycistronique *psbB-psbT-psbH-petB-petD* (opéron *psbB*), la première étant la région intergénique *psbH-petB*, la seconde l'intron du gène *petB* et la troisième l'intron du gène *petD*. La fonction de HCF152 n'est pas totalement définie car les données actuelles ne permettent pas de distinguer les relations entre la maturation, la stabilisation et l'épissage. Dans une revue récente sur la maturation de l'opéron *psbB*, la protéine HCF152 n'a pas été considérée comme un facteur d'épissage (Stoppel *et al.*, 2013). Cependant ces auteurs ont basé leurs affirmations sur les résultats de chevauchement des transcrits matures monocistroniques de *psbH* et *petB* chez le Maïs (Pfalz *et al.*, 2009) et non sur les données fonctionnelles relatives au mutant ou à la protéine. Une autre protéine PPR, PPR4, participerait au *trans*-épissage de l'intron 1 du gène *rps12* chez *Z. mays* (Schmitz-linneweber *et al.*, 2006). Cette protéine a été décrite comme possédant un domaine RRM dans sa partie N-terminale en plus des 16 domaines PPR. Le phénotype

INTRODUCTION

macroscopique du mutant correspondant se traduit par une chlorose et un arrêt de développement après le stade plantule, conséquence indirecte de l'altération du processus de la traduction du plaste. La protéine PPR5 chez le Maïs est aussi impliquée dans l'épissage plastidial de l'intron du gène codant *trnG*. La mutation du gène correspondant provoque un défaut de traduction, ce qui explique son phénotype chlorotique et de létalité après le stade plantule chez le *Z. mays* et l'embryon létalité chez *A. thaliana* (Beick *et al.*, 2008). En effet, il a été montré que PPR5 se fixe sur une région spécifique de l'intron. Ce résultat suggère que cette PPR stabilise le transcrit non épissé, probablement en créant une barrière de protection contre des attaques des nucléases (Williamns-Carrier *et al.*, 2008). Une quatrième protéine PPR, OTP51, est nécessaire à l'épissage de l'intron 2 du transcrit du gène *ycf3* de fonction actuellement inconnue (de Longevialle *et al.*, 2008). OTP51 possède 7 domaines PPR, et dans sa partie C-terminale, deux motifs conservés appelés LAGLIDADG, supposés conférer l'activité catalytique endonucléase. Son activité d'épissage de l'intron de *ycf3* est conservée chez *A. thaliana* et *Z. mays* (Khrouchtchova *et al.*, 2012). Cependant, il a été également détecté chez le mutants *otp51* d'*A. thaliana* une réduction de l'épissage de l'intron du transcrit *atpF* et probablement aussi de *trnV* et *trnK* (de Longevialle *et al.*, 2008).

Deux nouveaux facteurs d'épissage chloroplastique de type PPR ont été décrits après la revue de de Longevialle *et al.* (2010). Ils permettent de préciser de nouvelles caractéristiques des facteurs d'épissage des organites et ces travaux sont présentés ci-dessous. Le premier facteur correspond à la protéine THA8 (Khrouchtchova *et al.*, 2012). Elle possède quatre domaines PPR prédits dans sa partie C-terminale avec une région conservée à son extrémité N-terminale, présentant un repliement prédit similaire aux domaines PPR. La structure de la protéine et de son plus proche, THA8-LIKE2, chez *A. thaliana* confirment que malgré l'absence de prédiction de domaines PPR dans la région N-terminale, celles-ci peut avoir une conformation hélice-boucle-hélice similaire au domaine PPR (Ban *et al.*, 2013,). THA8 est impliquée dans l'épissage de l'intron 2 du gène *ycf3* et l'intron du gène *trnA* chez le *Z. mays* et *A. thaliana* (Khrouchtchova *et al.*, 2012). L'épissage des ces introns nécessite aussi les protéines OTP51 et APO1 pour l'intron de *ycf3* et de WTF1 et RNC1 pour l'intron du gène *trnA*. Khrouchtchova *et al.* (2012) ont confirmé, à partir d'expériences de co-immunoprécipitation en utilisant des extraits chloroplastiques, que THA8, WTF1 et RNC1

INTRODUCTION

prennent part à un même complexe de haut poids moléculaire contenant de l'ARN et en particulier les transcrits *trnA* et *ycf3*. Cependant, l'interaction directe de THA8 avec ces protéines n'a pas été évaluée. De plus, des expériences de « pulldown » n'ont pas permis de valider l'interaction de THA8 avec les protéines OTP51 ou APO1. Ces auteurs ont aussi montré par des expériences *in vitro* que la protéine recombinante MBP-THA8 (rTHA8) est capable de s'associer faiblement à l'intron 2 du transcrit *ycf3* mais seulement lorsque celui-ci est très concentré. Le deuxième facteur est OTP70. Cette protéine est une PPR de la sous famille PLS possédant, à son extrémité C-terminale, la partie la plus courte du domaine E des protéines de cette sous-famille jusqu'à présent analysées. Cette protéine est impliquée dans l'épissage de l'intron du gène *rpoC1* mais sa mutation affecte également l'édition de l'unique site d'édition d'*Arabidopsis* chez le transcrit de *rpoC1* (Chateigner-Boutin *et al.*, 2011). Chateigner-Boutin *et al.* (2011) ont montré que la fonction principale d'OTP70 serait associée à l'épissage de l'intron et ont suggéré que l'augmentation de l'édition chez *otp70* était probablement la conséquence d'un changement dans la cinétique d'épissage, ce qui influencerait l'édition du transcrit *rpoC1*. Des analyses destinées à déterminer le rôle du domaine E d'OTP70 suggèrent qu'il n'est pas nécessaire à la complémentation du mutant. Cependant, la relation entre l'épissage et l'édition chez le mutant *otp70* n'est pas claire et plusieurs hypothèses expliquant ces résultats sont envisageables:

- I. La protéine est impliquée directement dans les deux processus, réprimant d'un côté l'édition et stimulant l'épissage de l'autre,
- II. La protéine est impliquée seulement dans l'épissage, entraînant une augmentation des transcrits non épissés et par conséquent une cinétique plus lente d'épissage ce qui permet une meilleure efficacité de la machinerie d'édition,
- III. La protéine réprime l'édition de *rpoC1* ce qui conduit à un meilleur taux de l'épissage de l'intron de *rpoC1*.

Les auteurs font aussi observer que chez *O. sativa*, le gène *rpoC1* n'a pas d'intron et que le site d'édition n'existe plus. Un orthologue putatif au gène codant OTP70 a également été identifié dans cette espèce, indiquant une probable divergence fonctionnelle de la protéine

INTRODUCTION

chez les monocotylédones. Alternativement, la protéine pourrait avoir un rôle additionnel non détecté pour l'instant.

Chez la mitochondrie, quatre protéines PPR ont été montrées comme nécessaires à l'épissage d'introns, toutes appartenant à la sous-famille P. Tout d'abord, la protéine OTP43 a été impliquée dans le *trans*-épissage de l'intron 1 du gène *nad1*, le mutant *otp43* présentant une absence totale ou non détectable du complexe I (de Longevialle *et al.*, 2007). Des plantes homozygotes *otp43* ne germent pas directement sur terreau mais il est possible de les faire germer *in vitro* avec un apport d'une source carbonée. Le développement de ces plantes montre un phénotype macroscopique sévère caractérisé par une réduction de la vitesse de croissance et des feuilles courbées. Plus tard, un crible de mutants montrant une croissance racinaire normale lorsqu'ils sont exposés au Buthionine Sulfoximine (BSO), un inhibiteur de la synthèse de glutathion, a permis d'identifier une deuxième protéine PPR. Cette protéine appelée BSO-INSENSITIVE ROOTS 6 (BIR6) est impliquée dans une diminution importante de l'épissage de l'intron 1 du gène *nad7* du complexe I de la chaîne respiratoire (Koprivova *et al.*, 2010). En jours courts, sans inhibiteur et sur terre, les mutants *bir6* montrent un ralentissement de la croissance par rapport au sauvage. Ce retard développemental a été corrélé à la réduction de l'activité du complexe I, provoquant la réduction de la sensibilité au BSO et le phénotype de résistance au BSO chez les mutants. La troisième PPR impliquée dans l'épissage mitochondrial a été également identifiée au cours d'un crible de mutants, cette fois ci, hypersensibles à l'acide abscisique (ABA). Cette protéine, appelée ABA OVERLAY SENSITIVE 5 (ABO5), est impliquée dans l'épissage de l'intron 3 du gène *nad2* du complexe I (Liu *et al.*, 2010). Les analyses du mutant *abo5* montrent qu'il présente une sensibilité accrue à l'ABA au moment de la germination et tout au long de la croissance racinaire. Comme les deux précédents facteurs d'épissage, le mutant *abo5* présente une réduction de l'épissage et un retard de croissance par rapport au sauvage. En absence d'analyse de l'état du complexe I et de son activité, l'origine du phénotype observé reste à être déterminée car la seule faible réduction de l'épissage semble ne pas être l'unique facteur responsable de la chute de l'activité du complexe, en comparaison aux autres mutants d'OTP43 et BIR6.

3.5. Mode de reconnaissance des acides nucléiques par les protéines PPR

Les premières évidences de l'activité des protéines PPR ont été obtenues à partir de l'analyse de mutants. Dans ces études il a été observé que la perte de fonction des gènes PPR perturbait l'accumulation ou la maturation de transcrits de gènes des organites. En se basant sur des études préliminaires, dès la classification de la famille PPR, il a été proposé un modèle dans lequel la protéine PPR pouvait lier des séquences d'ARN grâce à la présence de motifs peptidiques répétés (Small and Peters 2000, Lurin *et al.*, 2004). Ces observations ont été vérifiées plus tard par diverses analyses biochimiques montrant que les protéines PPR étaient capables de lier l'ARN de manière spécifique pour agir sur leurs cibles (Nakamura *et al.*, 2003, Okuda *et al.*, 2006, Williams-Carrier *et al.*, 2008, Pfalz *et al.*, 2009, Tasaki *et al.*, 2010, Prikryl *et al.*, 2011). A partir de ces données, plusieurs équipes ont cherché à déchiffrer le mode de reconnaissance de l'ARN par les domaines PPR, aspect qui sera abordé dans les paragraphes suivants.

Les premières analyses prédictives de la structure du domaine PPR ont montré que ces domaines avaient tendance à former une structure secondaire en deux hélices α antiparallèles (hélice 1 ou A et hélice 2 ou B) jointes par une région non structurée formant probablement une boucle (Small and Peters 2000). La présence des motifs répétés en tandem du domaine PPR dans la séquence protéique a permis de suggérer que le mode de reconnaissance pouvait suivre un modèle similaire à celui des protéines avec des répétitions hélicoïdales répétées en tandem comme les protéines PUF (Delannoy *et al.*, 2007). Ainsi, Delannoy *et al.* (2007) ont proposé que les protéines PPR formaient probablement des structures tertiaires en forme de super hélice et qu'une base nucléotidique serait probablement reconnue par un domaine PPR, respectant une stœchiométrie 1:1. Plus tard, des analyses de comparaison entre les séquences cibles d'ARN de plusieurs PPR impliquées dans l'édition a suggéré que les protéines PPR pouvaient faire la distinction entre les bases puriques et pyrimidiques (Hammani *et al.*, 2009). Des analyses bioinformatiques sur les domaines PPR de la clade des restaurateurs de stérilité, domaines qui ont été obtenus suivant un modèle structural similaire à celui du modèle de PPR de PROSITE (<http://prosite.expasy.org/>), ont montré une pression sélective sur trois acides aminés dans la position 1, 3 et 6 au sein du domaine PPR (Fujii *et al.*, 2011). Fujii *et al.*

(2011) ont donc suggéré un modèle de reconnaissance dans lequel ces résidus étaient importants pour l'identification de la séquence de l'ARN, ces derniers étant les points de contact déterminant la spécificité d'interaction du domaine PPR avec la base nucléotidique. Du fait de l'absence de données structurales, le modèle de Fujii *et al.* (2011) a été basé sur la prédiction de la conformation secondaire du domaine PPR en deux hélices α antiparallèles. Cette prédiction structurale permettait un positionnement adéquat de résidus 1, 3 et 6 faisant face au solvant et à une probable interaction avec la séquence nucléotidique.

En même temps, une étude biochimique préliminaire effectuée avec différents couples de domaines PPR tronqués de la protéine HCF152 a renforcé les données bioinformatiques de Fujii *et al.* (2011) (Kobayashi *et al.*, 2011). Ainsi, la mutagenèse de certains résidus dans un des couples de domaine PPR analysés a permis à Kobayashi *et al.* (2011) de suggérer qu'en plus de trois résidus auparavant décrits, deux autres résidus, aux positions 10 et 14 (considérant toujours le modèle structural du domaine PPR de PROSITE) étaient importants pour définir l'interaction à la séquence nucléotidique. Ces mêmes auteurs ont aussi signalé que la charge positive du résidu 14 était corrélée avec l'affinité d'interaction à la séquence d'ARN. Ils ont également précisé que le résidu 6 avait un rôle plus déterminant dans la fonction du domaine PPR, probablement dans la spécificité de la reconnaissance de la base nucléotidique, car son changement affectait sévèrement l'affinité à l'ARN. Dans ce contexte, le croisement des données de mutations ponctuelles affectant la fonction des certaines protéines PPR et les modélisations de la structure du domaine PPR prédites par analyses informatiques, ont permis de proposer un modèle des interactions des domaines PPR avec l'ARN (Nakamura *et al.*, 2012). Ce modèle montrait que les positionnements des acides aminés de l'hélice A formaient une probable surface d'interaction entre la protéine et l'ARN cible.

Les études citées précédemment ont été suivies d'une plus vaste analyse bioinformatique, incluant des procédures de représentation statistique, couplées à des analyses biochimiques. Celles-ci ont permis d'élaborer un premier code de reconnaissance de l'ARN par les domaines PPR. En premier lieu, des analyses biochimiques sur la reconnaissance de la séquence d'ARN cible de la protéine PPR10, appuyées par les données des acides aminés du

domaine PPR obtenues précédemment par Fujii *et al.* (2011), ont permis de générer une corrélation entre la présence de deux acides aminés et le type d'acide nucléique reconnu dans la séquence de l'ARN (Barkan *et al.*, 2012). Ainsi, un premier acide aminé situé à la position 6 d'un domaine PPR et un deuxième résidu situé à la position 1 du domaine PPR suivant (appelé 1') sont déterminants pour la reconnaissance de la cible de PPR10. Barkan *et al.* (2012) ont utilisé et étendu cette règle pour montrer qu'en général les positions 6 et 1' des domaines PPR de type P et S peuvent s'aligner avec les séquences cibles connues des différentes protéines PPR. Ces données ont permis d'établir un code combinatoire aux positions 6 et 1' qui permettait d'expliquer les préférences de reconnaissance sur certains acides nucléiques, comme indiqué dans la table 2. Bien que ce code combinatoire ne permette pas de déterminer totalement à lui seul la spécificité exacte de reconnaissance, il confirme que les positions 6 et 1' ont une influence majeure dans la distinction du nucléotide.

En particulier pour les facteurs d'édition, Barkan *et al.* (2012) ont observé et suggéré, en tenant compte des données biochimiques obtenues lors de travaux antérieurs, que le dernier nucléotide reconnu par les domaines PPR était situé autour du quatrième nucléotide en amont du site à éditer. Ces résultats ont ouvert la voie à la capacité de prédiction des cibles putatives de facteurs d'édition PPR, en évaluant tous les sites d'édition des transcrits des organites. Malgré les progrès réalisés, ces auteurs n'ont pas obtenu de résultats satisfaisants dans le positionnement et l'implication du domaine PPR de type L, ce qui les a conduits à suggérer que ces domaines participaient aux mécanismes permettant de relâcher des contraintes structurales qui pouvaient subvenir lors de l'interaction.

En analysant la reconnaissance des facteurs PPR agissant sur l'édition, deux autres publications simultanées ont mis en évidence que les domaines de type L participent aussi à la reconnaissance des séquences cibles de l'ARN (Yagi *et al.*, 2013, Takenaka *et al.*, 2013a). Les travaux de Yagi *et al.* (2013) ont révélé trois acides aminés importants dans le modèle PPR, aux positions 1, 4 et ii, participant à l'interaction avec les acides nucléiques de la séquence cible d'ARN. Ces résidus correspondent aux positions 3, 6 et 1' du modèle de PROSITE, les deux derniers résidus étant ceux identifiés par Barkan *et al.* (2012). Ces auteurs ont donc montré que différentes combinaisons des trois résidus (voir table 2) contribuaient à une

INTRODUCTION

meilleure prédiction des cibles d'une protéine PPR chez *P. patens* ou *A. thaliana*. Par ailleurs, ils montrent également que le résidu 6 est un facteur important pour la distinction entre les bases puriques et pyrimidiques. Ils ont aussi suggéré que le dernier nucléotide reconnu par le motif PPR des facteurs d'édition est le 4^{ème} en amont du site à éditer. Au même moment, par une approche incluant les domaines PPR de type L, L2 et S2, Takenaka *et al.* (2013a) ont amélioré la prédiction de l'ARN ciblé par les facteurs d'édition. Ces auteurs ont suggéré que le positionnement du dernier domaine PPR sur la séquence cible pouvait être variable pour certaines protéines, ce qui avait été suggéré auparavant pour la protéine CRR22 grâce aux données biochimiques (Okuda & Shikanai 2012). Par ailleurs, ils ont aussi observé que les domaines PPR proches de l'extrémité C-terminale étaient les plus importants pour la spécificité de la reconnaissance de l'ARN cible.

Tableau 2. Code combinatoire de reconnaissance des acides nucléiques par les protéines PPR (Adapté de Barkan *et al.*, 2012 et Yagi *et al.*, 2013).

Position d'acides aminés			Spécificité de reconnaissance des nucléotides			
Position 3	Position 6	Position 1'	A	G	C	U
-	T	D	+	+++	+	+
V	T	D	-	++++	-	-
I	T	D	-	++	++	++
-	T	N	+++	+	+	+
F	T	N	++++	-	-	-
V	T	N	++	+	+	-
-	N	D	+	+	++	+++
Y	N	D	-	-	-	++++
F	N	D	+	+	++	+++
V	N	D	+	-	++	+++
-	N	N	+	+	+++	++
I	N	N	-	-	+++	+++
V	N	N	+	-	++	+
F	N	N	++	++	-	-
-	S	N	+++	-	+	+
F	S	N	+++	+	+	+
-	N	S	+	-	+++	+
V	N	S	-	-	+++	+
-	S	S	++++	-	+	-
-	S	D	+	+++	+	+
F	P	D	-	-	-	++++
-	T	T	-	++++	-	-

3.5.1. La structure des protéines PPR et les implications dans le mode de reconnaissance de l'ARN

Cinq structures cristallographiques de protéines PPR ont été déterminées récemment : celle de l'ARN polymérase mitochondriale humaine (mtRANP ou POLRMT) (Ringel *et al.*, 2011), celle de la protéine PROteinaceous RNase P1 (PRORP1) (Howard *et al.*, 2012), celles des protéines THA8-LIKE2 et THA8 (Ban *et al.*, 2013, Ke *et al.*, 2013) et celle de PPR10 (Yin *et al.*, 2013). Par ailleurs, deux complexes ribonucléoprotéiques PPR-RNA ont été récemment résolus dans le cas des protéines PPR10 (Yin *et al.*, 2013) et THA8 (Ke *et al.*, 2013). L'ensemble des résultats de structure des protéines PPR ont confirmé la structure secondaire du domaine PPR, prédit précédemment, pour les variantes S et P et L (Figure 12A). De plus, des régions non prédites comme étant des domaines PPR mais proches, ont été aussi assignées structurellement comme des PPR.

Ce n'est qu'à partir de la résolution de la structure de l'ARN polymérase mitochondriale (mtRANP ou POLRMT) chez l'homme qu'a été confirmée la structure secondaire du domaine PPR en deux hélices α antiparallèles, cette protéine présentant deux domaines PPR consécutifs au sein de sa séquence (voir figure 12B) (Ringel *et al.*, 2011). Par ailleurs, le complexe d'élongation de POLRMT a montré que la chaîne d'ARN naissante fait face aux domaines PPR comme le montre la figure 12C (Schwinghammer *et al.*, 2013). Cette conformation de la polymérase POLRMT dans son complexe d'élongation suggère que les domaines PPR peuvent contribuer à la stabilisation de la chaîne naissante d'ARN ce qui est en accord avec l'affinité de ces domaines pour ces molécules.

Les premiers indices de la relation structure-fonction d'une PPR avec son ARN cible ont été obtenus avec la détermination de la structure cristallographique de la protéine PROteinaceous RNase P1 (PRORP1). Cette protéine est considérée comme responsable de l'activité RNase P pour la maturation des structures de type ARN de transfert (ARNt) chez les organites (Gobert *et al.*, 2010). Le modèle structural de PRORP1 a montré que malgré la dégénérescence de la séquence primaire des acides aminés du domaine PPR, sa conservation structurale est plus importante, comme il avait été suggéré des années auparavant (Small &

INTRODUCTION

Peeters 2000, Delanoy *et al.*, 2007). En effet, la structure de PRORP1 a permis d'identifier 5 domaines et demi de type PPR, trois d'entre eux déjà prédits au sein de la séquence des acides aminés de la protéine mais les 2 et demi restants n'ont été identifiés qu'à partir de la détermination de leur structure (voir figure 13A) (Howard *et al.*, 2012). La structure a aussi permis d'identifier deux domaines structurels additionnels. Un premier domaine déjà prédit à partir de la séquence de la protéine a été confirmé dans l'extrémité C-terminale. Il présente un repliement similaire à un domaine de type métallonucléase et une affinité pour des centres métalliques. Plus particulièrement, ce domaine est supposé être le responsable de l'activité enzymatique de PRORP1 nécessitant de la présence d'ions Mg^{+2} ou Mn^{+2} pour son activité (Howard *et al.*, 2012). Le deuxième est un domaine central connectant les domaines PPR avec le domaine métallonucléase. Il a une affinité pour le zinc qui stabilise probablement sa structure. De plus, il interagit avec les domaines PPR et le domaine métallonucléase pour établir la conformation en V de la protéine et permet probablement d'orienter ces domaines pour optimiser l'activité catalytique de la protéine (Howard *et al.*, 2012).

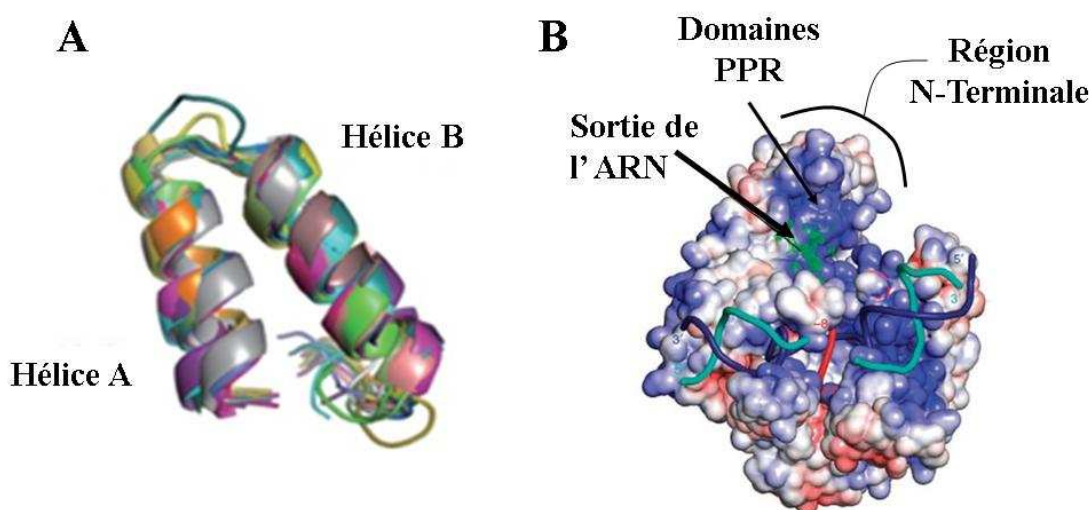


Figure 12. Structure du motif PPR et du complexe d'élongation de la polymérase humaine mitochondriale POLRMT.

A) Structure du domaine PPR en hélice-boucle-hélice. B) Complexe d'élongation de la POLRMT. La structure de ce complexe est identifiée avec les codes Protein Data Bank PDB 4BOC. Cette structure montre que la séquence naissante de l'ARN (brin rouge) peut sortir par la région N-terminale de la protéine et ainsi probablement impliquer l'activité des domaines PPR. (Adapté Yin *et al.*, 2013, Schwinghammer *et al.*, 2013)

INTRODUCTION

Lors de la publication de la structure de PRORP1, et grâce à des analyses biochimiques et structurales d'un deuxième article, un mode de reconnaissance de l'ARN a été suggéré pour PRORP1 et pour PRORP2, la protéine la plus proche chez *A. thaliana* (Gobert *et al.*, 2013). Ainsi, il a été montré que pour la reconnaissance des éléments de type ARNt, les protéines PRORP utilisent un mode liaison similaire à celui des complexes ribonucleoprotéiques de type RNase P où la présence de certains nucléotides dans les boucles D et T Ψ C des ARNt sont des facteurs déterminants pour l'interaction et le bon positionnement du complexe (voir figure 13B). A cet effet, Gobert *et al.* (2013) ont suggéré que les domaines PPR des protéines PRORP pourraient reconnaître ces nucléotides et ainsi contribuer à établir un positionnement spatial optimal permettant la maturation de l'ARNt par le domaine métallonucléase. Comme les précédentes structures, les structures cristallographique de la protéine THA8 et THA8-LIKE2 montrent une conformation en hélice-boucle-hélice correspondant aux cinq domaines PPR constituant ces protéines (Ban *et al.*, 2013, Ke *et al.*, 2013).

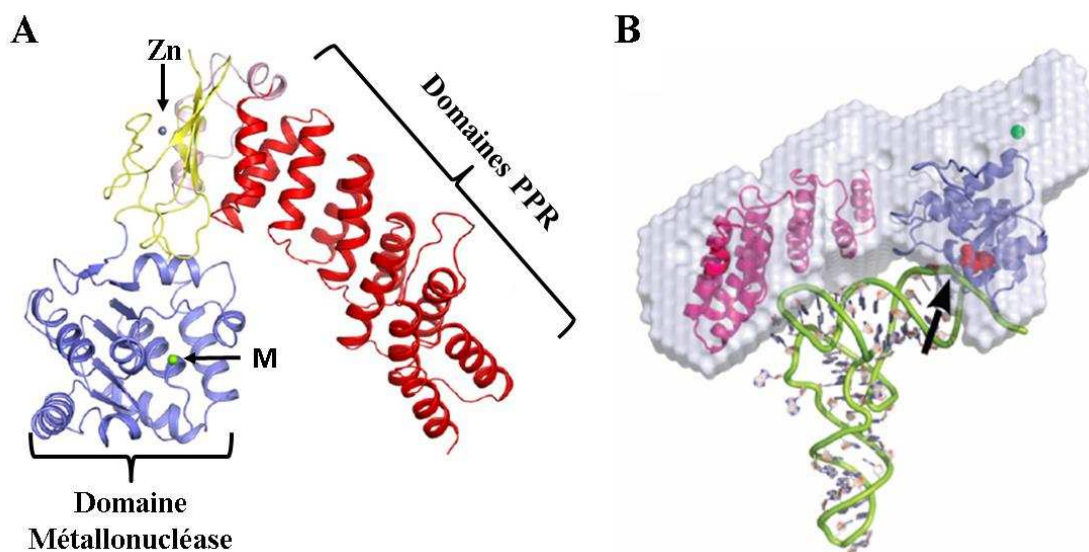


Figure 13. Structure de la protéine PPR PRORP1 et modèle du complexe PRORP-ARNt. A) Structure cristallographique correspondant à PRORP1. La structure présente 5,5 domaines PPR en rouge, le domaine central capable de lier le Zinc (Zn) en jaune, le domaine métallonucléase en bleu. Les centres métalliques M sont signalés par une flèche (M: Mg⁺² ou Mn⁺²). Les structures de cette protéine sont identifiées avec les codes Protein Data Bank PDB 4G23 à 4G26 (<http://www.wwpdb.org/>) B) Modèle de reconnaissance de l'ARNt par les protéines PRORP. L'ARNt en vert et les domaines des protéines PRORP sont présentés (Adapté de Howard *et al.*, 2012, Gobert *et al.*, 2013).

Ces structures précédemment citées ont été obtenues à partir de protéines PPR un peu particulières composées d'un faible nombre de répétitions de domaines PPR. Cependant, la structure de la protéine PPR10 déterminée récemment rend compte de la structure d'une PPR type appartenant au sous-groupe P (Yin *et al.*, 2013). La connaissance approfondie de ses fonctions moléculaires en font une protéine candidate exceptionnelle pour la compréhension des relations structure-fonction au sein des protéines PPR, particulièrement de celles du sous-groupe P. D'ailleurs, en se basant sur les résultats obtenus pour la protéine PPR10, un modèle de fonction des protéines PPR du sous-groupe P a été précédemment suggéré (Prikryl *et al.*, 2011). Il propose que les protéines PPR pourraient avoir une fonction de protection de l'ARN contre les attaques exo-ribonucléotidiques mais aussi d'activation de la traduction car elles faciliteraient le relâchement structural de l'ARN, permettant ainsi l'accès aux ribosomes.

La structure de PPR10 montre que cette protéine possédant 19 domaines PPR forme deux spirales d'une super hélice tournant vers la droite à laquelle s'ajoute une coiffe de trois petites hélices α dans la partie N-terminale et une hélice α dans la partie C-terminale. Les distributions spatiales des hélices a et b de chaque domaine PPR montrent que celles-ci constituent les couches intérieure et extérieure, respectivement, dans la structure de la super hélice (figure 14A). De plus il a été observé que la protéine PPR10 peut former un homodimère, dans lequel les deux protéines sont étroitement liées formant deux spirales antiparallèles où l'extrémité N-terminale d'une des protéines est en contact avec la partie C-terminale de l'autre (Figure 14B) (Yin *et al.*, 2013).

Une meilleure compréhension et la confirmation de l'interaction ribonucléoprotéique a été obtenue grâce à la détermination de la structure cristallographique des complexes THA8-ARN et PPR10-ARN. Comme beaucoup de résultats biochimiques le montrent, les protéines PPR ont une affinité importante pour les séquences d'ARN simple brin. Les données cristallographiques l'ont effectivement confirmé (Ke *et al.*, 2013, Yin *et al.*, 2013). Dans le cas du complexe THA8-RNA, il a été observé que THA8 pouvait se lier à des séquences présentes dans l'intron 2 de *yef3*, une des ses cibles précédemment décrites (cf. section 3.4.5.). Cette liaison de THA8 à l'ARN induit une dimérisation et une oligomérisation de THA8. Ainsi, des structures quaternaires avec l'ARN se forment par l'intermédiation des domaines

INTRODUCTION

PPR 4 et 5 de la partie C-terminale d'une protéine THA8 et les domaines PPR 1 et 2 de la partie N-terminale d'une autre protéine THA8 (figure 15A). Ces interactions provoquent des changements structuraux autorisant la formation de surfaces de contact entre ces deux protéines THA8. Ke *et al.* (2013) ont observé que les domaines 4 et 5 déterminent la reconnaissance de l'acide nucléique G suivant le code précédemment suggéré de la position 6 (Thréonine T172) du domaine 4 et la position 1' (Acide Aspartique D203) du domaine 5. Bien que Ke *et al.* (2013) aient montré que THA8 utilise ce code de reconnaissance, il est clair que son mode d'action semble spécifique mais probablement extrapolable à des protéines PPR possédant peu de motifs répétés.

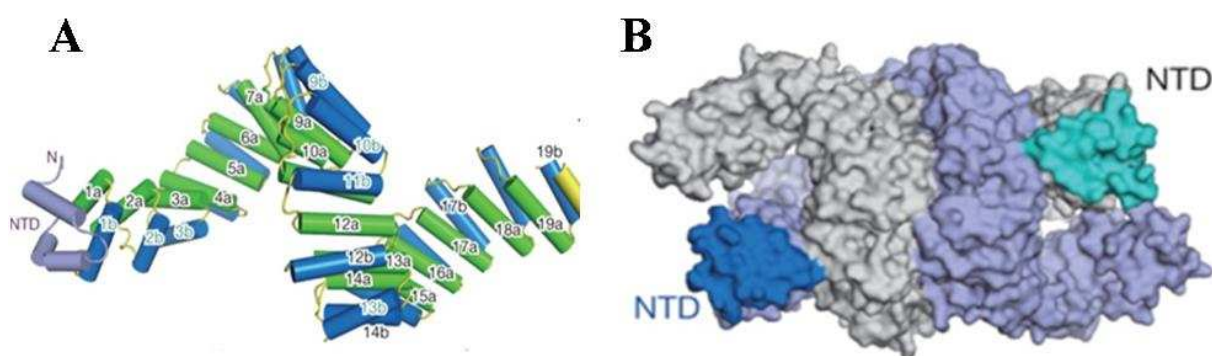


Figure 14. Structure de la protéine PPR10 et de son homodimère.

A) Structure du monomère de la protéine PPR10. Les 19 domaines PPR de PPR10 sont représentés sous forme cylindrique. Les cylindres verts correspondent aux hélices a et les bleus aux hélices b. N : extrémité N-terminale de la protéine à partir de laquelle les motifs sont numérotés. NTD: Coiffe N-terminal B) Configuration spatiale de l'homodimère de PPR10. Le dimère est formé de l'association antiparallèle d'un premier monomère en gris et d'un second monomère en violet, chacune des coiffes N-terminales étant colorée en bleu et en cyan, respectivement (Adapté de Yin *et al.*, 2013).

Le complexe PPR10-ARN a été obtenu avec une des deux séquences cibles de PPR10, c'est à dire 18 nucléotides du transcrit *psaJ* (Barkan *et al.*, 2012). Comme pour THA8, la résolution du complexe ribonucléoprotéique montre que le complexe de PPR10-*psaJ* forme un homodimère. De plus, la liaison à l'ARN cible entraîne des changements conformationnels du dimère PPR10 (dPPR10). Néanmoins, l'orientation spatiale des protéines dans le complexe diffère considérablement de celle du complexe de THA8, probablement en raison du plus grand nombre de domaines PPR dans PPR10. Dans ce complexe dPPR10-*psaJ*, Yin *et al.* (2013) ont observé que les 4 premiers nucléotides (5'-GUAU-3') étaient reconnus par les

domaines 3 à 7 de la partie N-terminale d'un des monomères tandis que les nucléotides 15 et 16 (5'-UU-3') étaient reconnus par les domaines 16 et 17 de la partie C-terminale du deuxième monomère. Ces domaines montrent un positionnement face aux nucléotides suivant le code de reconnaissance des acides aminés 3, 6 et 1' précédemment suggéré (redéfinis comme les positions 2, 5 et 35) (voir figure 15B). Pour les bases 5 à 14 (5'-UCUUUAAUUA-3') et les bases 18 et 19 la structure ne montre pas une coordination des domaines PPR en relation au code. Cependant le squelette phosphate des bases 15 à 18 interagit avec des résidus polaires des domaines 15 à 19 par l'intermédiaire de molécules d'eau (Yin *et al.*, 2013).

En se basant sur le code, la structure du complexe dPPR10-*psaJ* montre que les acides aminés en position 6 (T178, N213, S249) et 1' (T208, D244, S279) peuvent interagir avec les bases où entre eux pour les reconnaître sélectivement tandis que l'acide aminé en position 3 (R175, V210, F246, V281) se trouve intercalé entre les nucléotides pouvant ainsi interagir ou non avec l'ARN (Figure 15B). L'acide aminé en position 6 est le facteur déterminant de la reconnaissance spécifique des bases. De plus, ce complexe montre que des acides aminés polaires ou de charge positive peuvent interagir avec le squelette de phosphate, stabilisant ainsi la liaison à l'ARN. Cette dernière interaction peut faire intervenir d'autres positions comme celle observée pour l'arginine R433 qui se trouve en position 15 du domaine 10. Ceci conforte les analyses informatiques et de mutagenèse pour les protéines HCF152 (Kobayashi *et al.*, 2011, Nakamura *et al.*, 2012) et PGR3 (Fujii *et al.*, 2013) suggérant que la charge positive de l'acide aminé en position 14 à l'extrémité C-terminale de l'hélice A est importante pour la liaison à l'ARN. Toutes ces données suggèrent que cette partie du domaine PPR pourrait interagir avec le squelette de phosphate de l'ARN, chargé négativement. Ces interactions électrostatiques permettraient une stabilisation de l'interaction protéine-ARN.

Ces données montrent certaines similitudes avec la reconnaissance de l'ARN par des protéines hélicoïdales à domaines répétés telles que les protéines PUF. Ainsi, le positionnement d'une hélice interne fait face à l'ARN et des interactions entre l'ARN et des acides aminés clés des hélices permettent la spécificité de reconnaissance des séquences. Néanmoins, le mécanisme de reconnaissance reste spécifique.

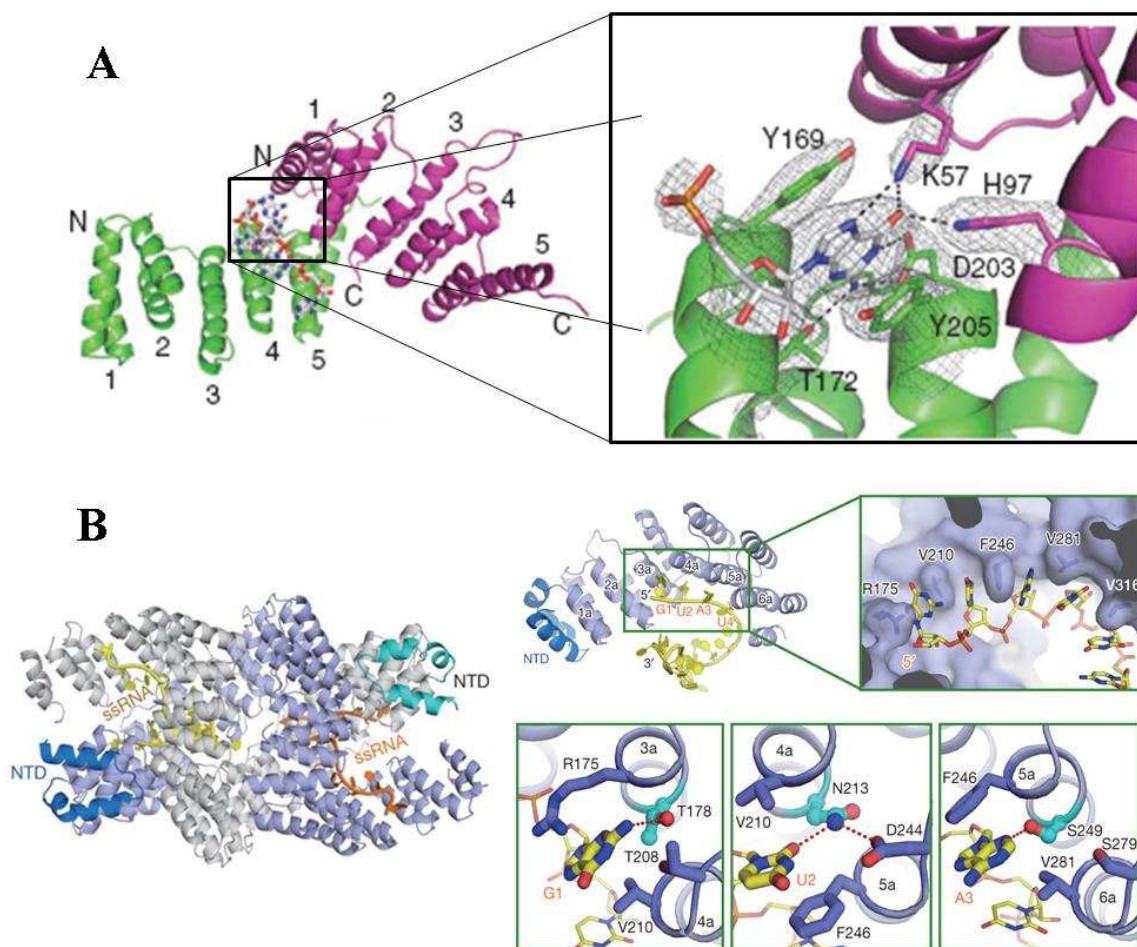


Figure 15. Mode de reconnaissance de l'ARN par les protéines PPR.

A) Reconnaissance de l'ARN par le dimère des protéines THA8. Le monomère de THA8 (en vert) utilise les acides aminés T172 et D203 des domaines PPR 4 et 5 respectivement pour l'interaction avec le nucléotide G tandis que l'autre monomère (en violet) utilise les acides aminés K57 et H97 des domaines PPR 1 et 2. B) Reconnaissance de l'ARN cible *PSAJ* par le dimère de la protéine PPR10. Les séquences d'ARN simple brin (en jaune et orange) interagissent avec des acides aminés spécifiques de domaines PPR de dPPR10. Les nucléotides 1 à 4 de la cible *psaJ* ont une coordination avec les acides aminés suivant le code de reconnaissance (position 6: position 1'-nucléotide) T178:T208-G1, N213:D244-U2, S249:S279-A3 précédemment suggéré (Adapté de Ke *et al.*, 2013, Yin *et al.*, 2013).

Il faut noter que les résultats de la structure du complexe PPR10-*psaJ* dimérique diffèrent légèrement des résultats de reconnaissance de *psaJ* par PPR10 publiés par Barkan *et al.* (2012). Ces différences peuvent s'expliquer par le modèle utilisé par ces auteurs. Ainsi, Barkan *et al.* (2012) ont considéré que PPR10 a une affinité plus grande pour l'ARN lorsqu'il est à l'état de monomère, ce monomère étant considéré comme la forme la plus proche de l'état physiologique (Prikryl *et al.*, 2011, Barkan *et al.*, 2012). En ce sens Barkan *et al.* (2012)

ont maximisé la reconnaissance de l'ARN par PPR10 ce qui les a conduit à décaler certains motifs PPR, générant ainsi un espace vide entre les domaines PPR par rapport aux positionnements des nucléotides de la séquence d'ARN. Yin *et al.* (2013) ont également détecté que PPR10 présentait une plus grande capacité de liaison à l'ARN lorsqu'elle est à l'état monomérique et pour de faibles concentrations en ARN. De plus, ils ont observé que le dimère de PPR10 n'est pas présent lorsque la séquence cible du transcrit *atpH* se lie à PPR10. Ceci suggère que la structure du dimère dPPR10-*psaJ* n'est pas représentative d'un état actif de PPR10. A l'état de monomère, les données structurales ne permettent pas de définir l'implication des domaines PPR 8 à 15 dans la reconnaissance de l'ARN. Malgré ces divergences, il est clair que les acides aminés 3, 6 et 1' des domaines PPR sont déterminants pour la reconnaissance de l'ARN cible, confirmant les analyses informatiques et biochimiques. De plus, ces résultats montrent aussi que des acides aminés chargés positivement ou polaires, dans d'autres positions de l'hélice a peuvent contribuer à la stabilisation de l'interaction PPR-ARN. La résolution d'autres structures cristallographiques de complexes PPR-RNA sera nécessaire pour mieux comprendre la fonctionnalité des PPR. Plus particulièrement, l'identification et la caractérisation des membres des complexes d'épissage, d'édition et leur étude structurale constituent un des défis majeurs à venir dans le but de d'élucider les rôles et les fonctions des protéines PPR

Présentation et Objectifs de la Thèse

Pendant la dernière décade, l'étude des mécanismes impliqués dans le contrôle de l'expression des gènes des organites chez les plantes a fait l'objet de nombreux travaux de par le monde. Dans ce domaine, un intérêt croissant a été porté à la famille PPR. En effet, comme cela a été détaillé dans l'introduction, les protéines PPR participent depuis la transcription jusqu'à la traduction aux mécanismes régulant l'expression des gènes des organites. Dès le début de la caractérisation de la famille PPR, plusieurs équipes françaises ont participé à la compréhension des rôles de ces protéines au sein de la cellule végétale. Ce travail de thèse s'inscrit dans la continuité des travaux de recherche sur les protéines PPR menées au sein de l'équipe *Génomique Fonctionnelle d'Arabidopsis* de l'Unité de Recherche en Génomique Végétale (URGV).

A mon arrivée dans l'équipe, un projet d'étude systématique de la localisation de protéines PPR d'*Arabidopsis thaliana* ayant une prédiction de localisation subcellulaire ambiguë avait été initié. Des résultats suggéraient que certaines de ces protéines PPR étaient doublement localisées dans la mitochondrie et le chloroplaste. Ce projet était mené en utilisant 2 approches. La première était basée sur l'expression transitoire dans des feuilles de *N. benthamiana* de fusions traductionnelles constituées de la région N-terminale (100 acides aminés) de ces protéines, contenant potentiellement le signal d'adressage, fusionnée à une protéine fluorescente. Dans la seconde approche, des expériences identiques ont été réalisées pour un certain nombre de protéines PPR sélectionnées en fusionnant cette fois-ci les séquences codantes pleine longueur à la protéine de fluorescence. A cette époque, PRORP1, la première protéine PPR ayant une double localisation mitochondriale et chloroplastique venait d'être publiée. Cette protéine montrait également une fonctionnalité dans les deux organites (Gobert *et al.*, 2010). Les résultats préliminaires du projet de localisation de l'équipe, renforcés par les analyses de la fonction de la protéine PRORP1, ont été à l'origine du questionnement sur l'importance et la fonction de ces protéines doublement localisées au sein de la famille PPR. Ils ont aussi permis de proposer une hypothèse de travail selon laquelle « *les protéines PPR doublement adressées à la mitochondrie et au chloroplaste participeraient à la coordination de l'expression des gènes entre les deux organites par le noyau de la cellule hôte* ». C'est dans ce cadre que mon sujet de thèse m'a été proposé. En raison du lien avec mon projet de thèse, la poursuite de la caractérisation du complexe

protéique contenant la protéine PPR CLB19 (CHLOROPLAST BIOGENESIS 19) m'a été également été confiée par la suite. Dans ce contexte, mon travail de thèse s'est focalisé plus particulièrement sur les protéines PPR doublement localisées dans les mitochondries et les chloroplastes et ayant potentiellement une activité de facteur d'édition. Plus précisément, mes objectifs de thèse ont été :

- ❖ D'identifier des protéines PPR doublement adressées au chloroplaste et à la mitochondrie au sein de la cellule végétale,
- ❖ D'analyser la fonction d'une sélection de quatre protéines PPR doublement localisées et ayant un rôle potentiel dans l'édition des ARN chez les organites d'*Arabidopsis thaliana*.

Ainsi, le premier volet des résultats de ce travail de thèse, présenté dans le chapitre IV, rassemble les résultats de l'analyse globale de la double localisation au sein de la famille des protéines PPR. Dans une deuxième partie des résultats, le chapitre V, présente l'analyse fonctionnelle de quatre protéines PPR présentant une double localisation, deux d'entre elles participant au complexe protéique impliquant la protéine CLB19.

RESULTATS

Chapitre IV

La Double Localisation Des Protéines PPR Chez *A. Thaliana*:

La localisation subcellulaire des protéines PPR: Article

La localisation subcellulaire des protéines PPR: Actualisation de données

Analyse de la double localisation des protéines PPR

La double localisation et l'évolution des protéines PPR

La double localisation et les réseaux d'interaction des protéines PPR

4. La Double Localisation Des Protéines PPR Chez *Arabidopsis Thaliana*

Les premiers résultats du projet de localisation systématique, en cours dans l'équipe au début de mon travail de thèse, montraient que certains membres de la famille PPR pouvaient avoir une la double localisation dans la mitochondrie et le chloroplaste. Ces résultats étaient issus d'observations de la localisation subcellulaire de fusions traductionnelles comprenant les 100 premiers acides aminés de la partie N-terminale de PPR, dont la prédiction d'adressage était ambiguë, fusionnés à la protéine fluorescente DsRed2. (cf section 4.1 Article I). Dans ce contexte, j'ai initié une approche complémentaire permettant l'analyse globale de la localisation des protéines PPR au sein de la famille. Cette approche, basée sur la méthode proposée par Millar *et al.* (2009) pour définir la localisation des protéines des plantes, utilise tous les types de données provenant d'études au cours desquelles différentes technologies ont été développées afin de préciser la localisation subcellulaire des protéines. Cette approche a permis d'intégrer les données de localisation produites par le groupe avec les données publiques concernant des protéines PPR disponibles dans différentes bases de données ou publications. Les résultats de cette intégration de données ont fait l'objet de la récente publication de l'équipe dans le numéro de septembre 2013 de la revue *RNA Biology* spécialement dédié aux protéines PPR (Colcombet *et al.*, 2013). Cependant, cet article traite de la localisation des protéines PPR dans son ensemble et tous les éléments relatifs à leur double localisation n'ont pas pu être discutés. Je présenterai donc une analyse détaillée et actualisée de la double localisation au sein de la famille PPR à la suite de la revue dans la section 4.3.

4.1. La localisation subcellulaire des protéines PPR : Article I

RNA Biology 10:9, 1–19; September 2013; © 2013 Landes Bioscience

RESEARCH PAPER

Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles

Jean Colcombet⁵, Mauricio Lopez-Obando⁵, Laure Heurtevin^{1,5}, Clément Bernard, Karine Martin, Richard Berthomé², and Claire Lurin*

Unité de Recherche en Génomique Végétale (URGV); UMR INRA/UEVE - ERL CNRS 91057; CP 5708; 91057 EVRY CEDEX, France

¹Current address: Génétique et Amélioration des Fruits et Légumes (GAFL); INRA - UR1052; 84143 MONTFAVET CEDEX, France; ²Laboratoire des Interactions Plantes Micro-organismes (LIPM); UMR INRA 441/CNRS 2594; CS 52627; 31326 CASTANET TOLOSAN CEDEX, France

*These authors contributed equally to this work.

Keywords: PPR proteins, organelle genome expression, *Arabidopsis thaliana*, targeting peptide, subcellular localization

Abbreviations: PPR, Pentatricopeptide Repeat; RFP, red fluorescent protein; ORF, open reading frame; TP, targeting peptide

458 genes coding for Pentatricopeptide Repeat (PPR) proteins are annotated in the *Arabidopsis thaliana* genome. Over the past 10 years, numerous reports have shown that many of these proteins function in organelles to target specific transcripts and are involved in post-transcriptional regulation. Therefore, they are thought to be important players in the coordination between nuclear and organelle genome expression. Only four of these proteins have been described to be addressed outside organelles, indicating that some PPRs could function in post-transcriptional regulations of nuclear genes.

In this work, we updated and improved our current knowledge on the localization of PPR proteins of *Arabidopsis* within the plant cell. We particularly investigated the subcellular localization of 166 PPR proteins whose targeting predictions were ambiguous, using a combination of high-throughput cloning and microscopy. Through systematic localization experiments and data integration, we confirmed that PPR proteins are largely targeted to organelles and showed that dual targeting to both the mitochondria and plastid occurs more frequently than expected. These results allow us to speculate that dual-targeted PPR proteins could be important for the fine coordination of gene expressions in both organelles.

Introduction

Plant nuclear genomes code for more than 99% of the 25000–30000 proteins required to build plant cells and tissues.¹ These proteins are addressed to various cell compartments to ensure specific cellular processes. Two other small genomes, formed by primary endo-symbiosis events, which led to the organelle formation, are found in mitochondria and plastids.² Throughout evolution, organelles have lost much of their original genomes by the transfer of genetic material to the nucleus. However, they have retained small genomes encoding key proteins and RNAs necessary for their biology. In *Arabidopsis*, 57 mitochondrial genes and 128 chloroplast genes have been annotated on the corresponding genomes (TAIRv10). The proteins encoded by these genes, acting together with nuclear imported proteins, play an important role in mitochondria and plastid functions.^{3,4} Many of the proteins encoded by genes transferred from organelles to the nucleus are important for organelle gene expression or metabolism and

need to be targeted back to their original compartment. In addition, many other nuclear encoded proteins have acquired functions in different steps of organelle biology. Overall, more than 3000 proteins encoded by the nuclear genome are predicted to be targeted to the organelles,⁵ creating a requirement for a coordinated regulation of nuclear and organelle gene expression and a precise control of protein addressing and import into the organelles. Several import systems exist in mitochondria and plastids where translocation is mediated mainly by co-translational and post-translational machineries. The main machineries are well known.^{6–8} They are named Translocase of the Outer/Inner Mitochondria membrane complexes (TOM/TIM) in mitochondria and Translocase of the Outer/Inner Chloroplast membrane complexes (TOC/TIC) in plastids. TOM/TIM and TOC/TIC account for the targeting of most organellar proteins. These two Translocase complexes share both similar structural conformations and import mechanisms with the recognition of a Targeting Peptide (TP) and the involvement of chaperones, receptor, and

*Correspondence to: Claire Lurin; Email: lurin@evry.inra.fr
Submitted: 03/29/13; Revised: 08/11/13; Accepted: 08/12/13
<http://dx.doi.org/10.4161/rna.26128>

pore type proteins.⁸ Despite these similarities, the mechanisms of translocation are specific to each Translocase. For example, the translocation into plastids requires GTP hydrolysis whereas it is not the case in mitochondria.⁸

Organelle physiological processes are under the control of proteins expressed from distinct genomes suggesting a tight and complex coordination in gene expression and, therefore, intracellular signaling pathways between cell compartments. Whereas nuclear genes are largely regulated at the transcriptional level, organelle genes are often constitutively expressed but tightly regulated at post-transcriptional levels.⁹ Imported nuclear proteins are necessary for a wide range of organellar transcriptional and post-transcriptional processes, including RNA transcription, RNA processing, RNA editing, RNA splicing, and translation. Among these nuclear factors, the large family of Pentatricopeptide Repeat (PPR) proteins are emerging more and more as central actors of the inter-compartmental coordination of gene expression.¹⁰ As expected for proteins involved in complex genome regulations, they define one of the largest families encoded by the nuclear genome with 458 members in *Arabidopsis*, 477 in rice, and up to 800 in *Selaginella moellendorffii*.^{11–14}

A typical PPR protein is constructed from a stretch (2–30) of 35-amino acid motifs (known as PPR motifs) often merged in N terminus with a targeting peptide thought to allow an organelle subcellular localization. Several studies confirmed that the targeting peptide is functional, suggesting that PPR proteins are massively targeted to mitochondria or plastids.^{10,11} Based on the PPR motif sequences and their relative serial organization, we proposed a classification of PPR proteins in two main subfamilies.¹¹ In *Arabidopsis*, the largest one, named the P-type subfamily, contains 255 PPR proteins harboring tandem repeats of a simple canonical PPR motif (the P-type motif). The second one is known as the PLS-type subfamily and contains the remaining 203 PPR proteins.¹¹ Their module-based structures but also biochemical and genetic data indicate that PPR proteins are able to interact in a sequence-specific way with organelle RNAs to assure various post-transcriptional functions.^{10,15} Recently, through computational and molecular biology approaches, a RNA recognition code was proposed for PPR proteins where two adjacent PPR motifs are able to recognize one specific nucleotide.^{16,17} The specificity of the base recognition is accomplished by the combination of three amino acids, two located in the first PPR motif (third and sixth positions) and the third at the first position in the subsequent PPR motif.^{16,17}

PPR proteins have largely been associated with transcriptional, post-transcriptional, and translational regulation of organellar expression.¹⁰ A growing number of PPR proteins have been shown to be required for editing. For example, CHLOROESTEROXYL REDUCTION 4 (CRR4) is necessary for editing of the chloroplast *ndbD* transcript¹⁸ and MITOCHONDRIAL RNA EDITING FACTOR1 (MEF1) is required for editing of three mitochondrial transcripts.¹⁹ *Arabidopsis* PPR proteins are also involved in splicing of organelle transcripts: ORGANELLAR TRANSCRIPT PROCESSING43 (OPT43) and OPT51 are necessary for the correct trans-splicing of *nad1* and cis-splicing of *yef3* transcripts,

respectively.^{20,21} Finally, PPR proteins are involved in translation processes. For example, CHLOROPLAST RNA PROCESSING 1 (CRP1) has been proposed to be a chloroplast translation regulator²² and PPR336 is associated with mitochondrial polysomes.²³ As expected with essential players in gene expression involved in respiration and photosynthesis, a large proportion of mutants in PPR genes are embryo or gametophyte lethal.^{11,24,25}

Despite the growing PPR literature indicating that PPR proteins function mainly in organelles, some members could also have targets in the nucleus or the cytoplasm. In *Arabidopsis*, four PPR proteins were shown to be localized out of organelles. Two of them, PROTEINACEOUS RNase P 2 (PRORP2) and PRORP3, are localized exclusively in the nucleus where they are needed to achieve RNase P activity.²⁶ The two others have a more complex subcellular localization with a dual targeting to both mitochondria and nucleus. The GLUTAMINE-RICH PROTEIN23 (GRP23) interacts in nucleus with RNA polymerase II but its nuclear and mitochondrial functions are not yet understood.²⁷ Similarly, Hammani and co-workers showed that PPR PROTEIN LOCALIZED TO THE NUCLEUS AND MITOCHONDRIA1 (PNM1) is involved in protein translation in mitochondria whereas it physically interacts with two proteins in the nucleus, NUCLEOSOME ASSEMBLY PROTEIN1 and the transcription factor TCP8.²⁸ In animals, one example of a PPR protein localized out of the organelles has also been reported but its localization is still a matter of debate. This PPR protein, named BICOID STABILIZATION FACTOR (BSF) in *Drosophila*, as well as Leucine-Rich Repeat Pentatricopeptide Repeat Cassette (LRPPRC) in humans, was localized in the cytoplasm and nucleus of early *Drosophila* embryo cells²⁹ with roles in transcription and RNA transport. Other authors showed the protein to be localized in mitochondria where it would be involved in mRNAs maturation, poly-adenylation, and translation.³⁰

Only a handful of PPR proteins were shown to function out of organelles. Many post-transcriptional processes are being shared by both the organelles and nucleus; therefore, this number may be underestimated. In order to identify new *Arabidopsis* PPR proteins addressed out of the organelles but also to improve our general knowledge on PPR targeting, we systematically investigated the subcellular localization of a third of the PPR family whose addressing prediction was ambiguous. We took advantage of a high-throughput cloning strategy combined with a transient expression system to elucidate whether the N terminus targeting peptides of candidate PPR proteins were functional to address the protein into organelles. We report in this work that, despite erroneous predictions of subcellular localization, most PPR proteins are addressed to one of the organelles and showed that a fraction of them, probably underestimated, are addressed to both mitochondria and plastids.

Results

Localization study of PPR proteins with ambiguous predictions of localization. We previously published a manually curated list of *Arabidopsis* PPR gene models.¹² When this work was initiated, the most accurate algorithms to predict subcellular

localization of plant proteins were TargetP v1.01³¹ and Predotar v1.03.³² Therefore, we used them to identify Arabidopsis genes coding for PPR proteins with ambiguous localization predictions (Table 1). TargetP was recently improved with the TargetP v1.1 version of the software. Among the 458 PPR genes, Predotar predicts that 244 and 92 PPR proteins are addressed, respectively, to mitochondria and plastids, whereas 122 PPRs would not have any organelle localization (Table 2). TargetP v1.1 gives similar results with 232 and 123 PPRs localized to mitochondria and plastids, respectively, and 103 PPRs without organelle localizations (Table 1). Taken together, 166 PPR proteins were predicted not to be addressed to either of the two main plant organelles by at least one of the two software (Predotar v1.03 and TargetP v1.01). Among them, 53 PPR proteins were not predicted to be addressed in the organelles by both algorithms. We chose to experimentally investigate the subcellular localization of those 166 PPR proteins as they were good candidates to have atypical functions out of the organelles.

Almost all the proteins addressed to organelles contain a targeting peptide in their N terminus extremity, which is cleaved during the transfer through the organelle membranes.⁸ A mitochondrial Targeting Peptide (mTP) is typically 40–50 amino acid long,^{33–36} whereas a chloroplast Targeting Peptide (cTP) is usually up to 60 amino acid long.³⁷ To assess the targeting peptide functionality, we systematically merged in frame the first 300 bp, coding for the first 100 amino acids of each candidate PPR ORFs to the Red Fluorescent Protein (RFP) coding sequence using the Gateway technology. The aim of this approach was to experimentally detect any mTP or cTP present in the first 100 amino acids but not recognized by the prediction software. Vector cloning based on Gateway recombination technology was successful for 162 genes (97%). After agro-infiltration of *Nicotiana benthamiana* leaves with these constructs and subsequent protoplasts preparation, we were able to detect RFP signals for 131 constructs (79%) using either epifluorescent or confocal microscope. All localization experiments were repeated at least three times and observed independently by two of the authors. Table 1 summarizes all predicted and experimental data obtained during this study. Presented in Figure 1A are examples of typical subcellular localizations observed using 300 bp constructs. In Figure 1, RFP fluorescence was visualized using a confocal microscope and compared with the distribution of the mitochondrion-specific probe MitoTracker Green and the chlorophyll autofluorescence. In the overlay panels, combined fluorescence from RFP (in red), MitoTracker (in green), and chlorophyll autofluorescence (in blue) appears in yellow when RFP signal co-localizes with MitoTracker staining indicating a localization of the fusion protein in mitochondria whereas it appears in violet when RFP signal is localized in plastids. It was detected that 68 and 31 300 bp-PPR constructs gave an exclusive mitochondrial and plastid localization, respectively, as exemplified by AT3G15130 and AT3G46610 in Figure 1A. Interestingly, 24 constructs exhibited a signal in both organelles (see for example, AT2G36240 and AT5G47460 in Fig. 1A) and nine constructs gave localizations out of organelles, appearing as typical nuclear and cytosolic signals (AT1G06150 and AT1G06580 in Fig. 1A). These localization results in the

nucleus and the cytosol of the protoplasts suggest that the first 100 amino acids of these proteins do not code for a functional peptide targeting to organelles and that the RFP fusion proteins are localized where the translation occurs (in the cytosol) and in the nucleus by passive diffusion of small proteins through nuclear pores.

As addressing signals could be outside the first 100 amino acids and because using the first 100 amino acids may induce addressing artifacts, we decided to investigate in more detail the subcellular localization of the 33 PPR proteins that did not show a simple single organellar localization. Out of the 24 PPR proteins localized in both organelles and the nine proteins appearing outside of the organelles, we successfully cloned the whole ORFs and created RFP fusions for 19. Subcellular localizations of these fusions were monitored using Agro-infiltrated *N. benthamiana*-derived protoplasts observed under epifluorescent and confocal microscope (examples in Fig. 1B). Results are summarized in Table 1. We confirmed the dual subcellular localization for six out of the 11 ORFs successfully expressed and encoding full-length proteins thought to be addressed in both organelles (see AT5G47460 in Fig. 1B for example). As for AT2G36240 (Fig. 1B), we showed a single localization in mitochondria for the other five. Among the nine PPR proteins thought to be out of the organelles on the base of the first 100 amino acids, seven ORFs were successfully cloned but no cytosolic localization was confirmed: the whole proteins fused to RFP were systematically addressed to one or both organelles, as exemplified by AT1G06150 and AT1G06580 in Figure 1B. Surprisingly, six of them were localized in both organelles (AT1G06150 in Fig. 1B). Altogether, 12 PPRs proteins were verified as being localized in both mitochondria and plastids using the full-length protein.

Integrative overview of the subcellular localizations of PPR protein family. In order to provide a general overview of the localization of the whole PPR protein family, we aggregated our results concerning one-third of the family, with all available data from published studies and accessible databases (Table 2).

We first re-examined the localization prediction of the 458 PPR proteins encoded in the Arabidopsis genome using six available bio-informatics prediction tools: Predotar,³² TargetP,³¹ iPSORT,³⁸ LocTree,³⁹ Multiloc,⁴⁰ and AtSubP.⁴¹ Despite using distinct algorithms, those tools largely provide similar results and Table 2 gives a single localization prediction aggregating the six software results following a rule emphasized in the caption of Table 2. Overall, 65% and 17% of the PPR proteins are predicted to function in mitochondria and in plastids, respectively. For 18%, the results are unclear either because a majority of the software was unable to provide an organellar prediction or because they provide overmuch diverging organellar predictions.

We also added the growing data coming from the proteomics identification of organelle proteins in Arabidopsis mitochondria (SUBA3⁵) and chloroplast (SUBA3,⁵ AT_Chloro,⁴² PPDB⁴³), also including localization data obtained from maize chloroplast (PPDB⁴³) and rice mitochondria,⁴⁴ according to the recent concept of orthoproteomics.⁴⁵ As published,¹² a very good level of orthology observed between the members of PPR families in *A. thaliana* and *Oryza sativa* suggests that both their function and their subcellular

Table 1. Subcellular localization study of 166 PPR proteins with ambiguous prediction data

PPR	Gene model		Prediction		Fluorescent signal		Conclusion
	TAIR v10	O'Toole	Target P	Predotar	Targeting Peptide	FL Protein	
At1g01970	AT1G01970.1		M	none	M	C	C
At1g02420	AT1G02420.1		ER	M	M/C	c.u.	m/c
At1g04840	AT1G04840.1		none	M	C	n.a.	C
At1g06670		AIPPR_1g06670	M	none	N/Ct	M/C	M/C
At1g06150		AIPPR_1g06150	C	none	N/Ct	M/C	M/C
At1g06580	AT1G06580.1		M	ER	N/Ct + M	M/C	M/C
At1g08810	AT1G08810.1		none	C	no signal	n.a.	-
At1g09190	AT1G09190.1		M	none	M	n.a.	M
At1g09410	AT1G09410.1		M	M	no signal	n.a.	pM
At1g09900	AT1G09900.1		none	C	M	n.a.	M
At1g10270	AT1G10270.1		M	M	M	n.a.	M
At1g10330	AT1G10330.1		ER	none	no signal	n.a.	-
At1g11290	AT1G11290.1		C	none	C	n.a.	C
At1g14470	AT1G14470.1		ER	none	C	n.a.	C
At1g15480		AIPPR_1g15480	M	ER	no signal	n.a.	-
At1g18485	AT1G18485.1		C	ER	N/Ct	c.u.	-
At1g19290	AT1G19290.1		M	ER	M	n.a.	M
At1g19720	AT1G19720.1		none	none	M	n.a.	M
At1g20230	AT1G20230.1		none	none	M	n.a.	M
At1g22830	AT1G22830.1		none	M	M	n.a.	M
At1g25360	AT1G25360.1		M	none	M/C	M/C	M/C
At1g31430	AT1G31430.1		ER	none	M	n.a.	M
At1g31790	AT1G31790.1		C	none	C	n.a.	C
At1g31840		AIPPR_1g31840	ER	none	M	n.a.	M
At1g31920	AT1G31920.1		none	none	no signal	n.a.	-
At1g33350	AT1G33350.1		none	C	M	n.a.	M
At1g50270	AT1G50270.1		none	none	M	n.a.	M
At1g53330	AT1G53330.1		M	none	c.u.	n.a.	-
At1g56670	AT1G56670.1		C	none	M	n.a.	M
At1g59720	AT1G59720.1		C	none	C	n.a.	C
At1g60770	AT1G60770.1		none	M	M	n.a.	M
At1g62290	AT1G62290.1		ER	M	M	n.a.	M
At1g62590	AT1G62590.1		M	none	M/C	M	M
At1g63330	AT1G63330.1		none	none	M	n.a.	M
At1g63400	AT1G63400.1		M	none	M/C	M	M
At1g64100		AIPPR_1g64100	C	none	M/C	c.u.	m/c
At1g68930	AT1G68930.1		M	none	M	n.a.	M
At1g69290		AIPPR_1g69290	ER	ER	M	n.a.	M
At1g71490	AT1G71490.1		none	C	M	n.a.	M
At1g73710	AT1G73710.1		M	none	C	n.a.	C
At1g74400	AT1G74400.1		none	M	M/C	M	M
At1g74580	AT1G74580.1		none	none	no signal	n.a.	-
At1g74630	AT1G74630.1		ER	C	no signal	n.a.	-
At1g76280		AIPPR_1g76280	M	ER	M	n.a.	M
At2g01360		AIPPR_2g01360	ER	ER	no signal	n.a.	-
At2g01740	AT2G01740.1		none	M	M	n.a.	M
At2g02750	AT2G02750.1		M	M	no signal	n.a.	pM
At2g04860	AT2G04860.1		none	M	M	n.a.	M
At2g06000	AT2G06000.1		none	M	C	n.a.	C
At2g13600	AT2G13600.1		none	none	M	n.a.	M
At2g15820	AT2G15820.1		C	none	C	n.a.	C
At2g15980	AT2G15980.1		none	none	no signal	n.a.	-
At2g16880	AT2G16880.1		none	none	no signal	n.a.	-
At2g20540	AT2G20540.1		none	none	M	n.a.	M
At2g21090	AT2G21090.1		M	none	M	n.a.	M
At2g22070	AT2G22070.1		none	C	M	n.a.	M
At2g26790	AT2G26790.1		M	ER	M	n.a.	M
At2g27610	AT2G27610.1		M	none	M	n.a.	M
At2g28050	AT2G28050.1		none	M	M	n.a.	M
At2g32630	AT2G32630.1		M	none	M	n.a.	M
At2g33680	AT2G33680.1		none	none	c.u.	n.a.	-
At2g33760	AT2G33760.1		none	C	C	n.a.	C
At2g34400	AT2G34400.1		none	M	M/C	c.u.	m/c
At2g35130		AIPPR_2g35130	ER	none	C	n.a.	C
At2g36240	AT2G36240.1		none	none	M/C	M	M
At2g36730	AT2G36730.1		none	none	no signal	n.a.	-
At2g37230	AT2G37230.1		M	none	M	n.a.	M
At2g37310	AT2G37310.1		none	none	C	n.a.	C
At2g39620		AIPPR_2g39620	M	ER	C	n.a.	C
At2g40720	AT2G40720.1		none	none	M	n.a.	M
At2g41080		AIPPR_2g41080	C	none	M/C	no signal	m/c

Table 1. Subcellular localization study of 166 PPR proteins with ambiguous prediction data (continued)

PPR	Gene model		Prediction		Fluorescent signal		Conclusion
	TAIR v10	O'Toole	Target P	Predotar	Targeting Peptide	FL Protein	
At2g41720	AT2G41720.1		none	C	C	n.a.	C
At2g44880	AT2G44880.1		ER	none	M/Ct	M/C	M/C
At2g45350	AT2G45350.1		none	C	c.u.	n.a.	-
At3g01580	AT3G01580.1		M	none	M	n.a.	M
At3g02010	AT3G02010.1		M	none	M	n.a.	M
At3g05240	AT3G05240.1		none	M	no signal	n.a.	-
At3g06920	AT3G06920.1		none	none	M	n.a.	M
At3g08820	AT3G08820.1		none	C	M/C	M	M
At3g09060	AT3G09060.1		none	M	M	n.a.	M
At3g09850	AT3G09850.1		C	none	C	n.a.	C
At3g12770	AT3G12770.1		none	none	M	n.a.	M
At3g14330	AT3G14330.1		none	M	M	n.a.	M
At3g15130	AT3G15130.1		none	M	M	n.a.	M
At3g15930	AT3G15930.1		none	M	M	n.a.	M
At3g16610	AT3G16610.1		M	none	M	n.a.	M
At3g18840	AT3G18840.2		C	none	M/Ct	c.u.	-
At3g20730		AtPPR_3g20730	ER	ER	M	n.a.	M
At3g21470	AT3G21470.1		none	none	M/C	c.u.	m/c
At3g23020	AT3G23020.1		none	C	c.u.	n.a.	-
At3g23330	AT3G23330.1		none	M	M/C	c.u.	m/c
At3g25970	AT3G25970.1		ER	none	M	n.a.	M
At3g26540	AT3G26540.1		none	C	M	n.a.	M
At3g28640	AT3G28640.1		M	none	no signal	n.a.	-
At3g28660	AT3G28660.1		M	none	no signal	n.a.	-
At3g29290		AtPPR_3g29290	M	none	no signal	n.a.	-
At3g42630	AT3G42630.1		none	M	C	n.a.	C
At3g46610	AT3G46610.1		none	ER	C	n.a.	C
At3g46790	AT3G46790.1		C	none	C	n.a.	C
At3g47530	AT3G47530.1		none	none	M/C	M/C	M/C
At3g47840	AT3G47840.1		M	none	C	n.a.	C
At3g48810	AT3G48810.1		ER	none	M	n.a.	M
At3g49240	AT3G49240.1		M	none	M/C	M/C	M/C
At3g49710	AT3G49710.1		none	none	M/Ct	M/C	M/C
At3g49740	AT3G49740.1		C	none	M	n.a.	M
At3g50420	AT3G50420.1		none	none	M/C	M/C	M/C
At3g53170		AtPPR_3g53170	none	none	C	n.a.	C
At3g56550	AT3G56550.1		none	C	no signal	n.a.	-
At3g57430	AT3G57430.1		C	ER	C	n.a.	C
At3g58590	AT3G58590.1		none	none	M	n.a.	M
At3g62890		AtPPR_3g62890	C	none	M/Ct	M/C	M/C
At4g01570	AT4G01570.1		C	none	C	n.a.	C
At4g02750	AT4G02750.1		M	none	M	n.a.	M
At4g04370	AT4G04370.1		none	C	M/C	c.u.	m/c
At4g06210	AT4G06210.1		none	none	M	n.a.	M
At4g11690	AT4G11690.1		ER	ER	M	n.a.	M
At4g13650	AT4G13650.1		none	M	no signal	n.a.	pM
At4g14820	AT4G14820.1		none	C	M	n.a.	M
At4g14850	AT4G14850.1		C	none	M/C	c.u.	m/c
At4g15720	AT4G15720.1		none	none	C	n.a.	C
At4g16470	AT4G16470.1		ER	M	M	n.a.	M
At4g18840		AtPPR_4g18840	none	C	M/Ct	C	C
At4g20090	AT4G20090.1		C	ER	M	n.a.	M
At4g20740	AT4G20740.1		C	none	C	n.a.	C
At4g21065	AT4G21065.1		ER	none	no signal	n.a.	-
At4g21880	AT4G21880.1		ER	M	M	n.a.	M
At4g22760	AT4G22760.1		ER	M	M	n.a.	M
At4g28010	AT4G28010.1		M	none	M	n.a.	M
At4g30700	AT4G30700.1		none	M	M/C	M/C	M/C
At4g33170	AT4G33170.1		C	none	M/C	c.u.	m/c
At4g37170	AT4G37170.1		M	none	M	n.a.	M
At4g38010	AT4G38010.1		none	none	no signal	n.a.	-
At5g03800	AT5G03800.1		C	none	M/C	c.u.	m/c
At5g04810	AT5G04810.1		C	none	C	n.a.	C
At5g06540	AT5G06540.1		ER	M	C	n.a.	C
At5g08310		AtPPR_5g08310	none	M	no signal	n.a.	-
At5g08490	AT5G08490.1		M	none	no signal	n.a.	-
At5g08510	AT5G08510.1		none	none	M	n.a.	M
At5g10690	AT5G10690.1		C	ER	C	n.a.	C
At5g14080		AtPPR_5g14080	M	none	M/C	c.u.	m/c
At5g15300	AT5G15300.1		none	none	no signal	n.a.	-
At5g16860	AT5G16860.1		none	M	no signal	n.a.	-

Table 1. Subcellular localization study of 166 PPR proteins with ambiguous prediction data (continued)

PPR	Gene model		Prediction		Fluorescent signal		Conclusion
	TAIR v10	O'Toole	Target P	Predotar	Targeting Peptide	FL Protein	
At5g18475	AT5G18475.1		C	none	M	n.a.	M
At5g18960	AT5G18960.1		M	none	M	n.a.	M
At5g21222	AT5G21222.1		none	none	M/C	c.u.	m/c
At5g25630		AtPPR_5g25630	none	none	no signal	n.a.	-
At5g27270	AT5G27270.1		C	none	C	n.a.	C
At5g37570	AT5G37570.1		none	M	no signal	n.a.	-
At5g38730	AT5G38730.1		none	C	no signal	n.a.	-
At5g39680	AT5G39680.1		none	none	M	n.a.	M
At5g40405	AT5G40405.1		none	none	M	n.a.	M
At5g43790	AT5G43790.1		none	none	M/C	c.u.	m/c
At5g46100	AT5G46100.1		none	M	no signal	n.a.	-
At5g46680	AT5G46680.1		none	M	M	n.a.	M
At5g47460	AT5G47460.1		M	none	M/C	M/C	M/C
At5g48910	AT5G48910.1		C	none	C	n.a.	C
At5g50990		AtPPR_5g50990	none	none	no signal	n.a.	-
At5g52630	AT5G52630.1		none	none	M	n.a.	M
At5g55840		AtPPR_5g55840	none	none	M	n.a.	M
At5g56310	AT5G56310.1		M	none	M	n.a.	M
At5g59600	AT5G59600.1		none	none	C	n.a.	C
At5g59900	AT5G59900.1		M	none	M	n.a.	M
At5g65570	AT5G65570.1		M	none	no signal	n.a.	-
At5g65820	AT5G65820.1		M	none	M	n.a.	M
At5g68520	AT5G68520.1		none	none	C	n.a.	C
At5g67570		AtPPR_5g67570	none	none	no signal	n.a.	-

Manually curated Arabidopsis PPR gene models were used.¹² Most of them are identical to TAIR v10 gene models but 22 models are different and are indicated with their AtPPR codes. Predictions of localization using Predotar v1.03 and Target P v1.1 software are listed. Experimental fluorescent signals observed in protoplasts expressing Targeting Peptide or Full-Length (FL) protein fused to RFP are shown. Two independent observations by two of the authors were done on at least three independent agro-infiltrations. For each PPR, a tentative conclusion is proposed with the following rules: (1) if available, the observation of FL-protein fusion is considered as the true localization, (2) if a mitochondrial or a chloroplastic localization was observed for the targeting peptide and no observation was recorded for the full-length protein, the result of TP is indicated as conclusion, (3) if a dual localization was observed and no observation was obtained with the full-length protein, the result of TP is indicated as probable in lowercase, (4) if no experimental observation was obtained, the predicted localization is indicated with a preceding "p", .M, mitochondria; C, chloroplasts; N/Ct, nucleus and cytoplasm; M/C, dual localization in mitochondria and chloroplasts; pM, predicted in mitochondria (conclusion column); pC, predicted in chloroplasts (conclusion column); m, probably in mitochondria (conclusion column); c, probably in chloroplasts (conclusion column); m/c, probably in mitochondria and chloroplasts (conclusion column); -, no conclusion; c.u., cloning unsuccessful; n.a., not attempted.

localization are largely conserved between species even between monocotyledons and dicotyledons. Overall, 83 (about 18%) of the Arabidopsis PPR proteins, or PPR orthologs in other species, were identified either in the plastidial or the mitochondrial proteomes, providing a very useful set of PPR protein localization data (Table 2). Three and five PPR proteins were identified during proteomics characterization of Arabidopsis nuclear and cytosolic proteins, respectively.⁴⁶⁻⁴⁹ Surprisingly, 28 PPR proteins were characterized in plasma membrane or vacuole extracts.⁵⁰⁻⁵⁶ Without functional characterization of any of these membrane PPR proteins, these observations cannot be solved. They may be due to intrinsic technical limitations of proteomics approaches; in contrast, their number may indicate unsuspected localizations and functions. However, proteome-based localizations validate many of the prediction results of bio-informatics software as 48 (71%) of them matched the available predictions (Table 3).

A growing number of PPR proteins were subjected to in planta functional characterization either in dedicated studies (see references in Table 2) or in systematic studies including the work reported here and three previous ones^{15,57} unpublished data in SUBA3 (Table 2). Authors usually characterized localizations by microscopy using fusions between PPR proteins or, if suspected, putative targeting peptides and a fluorescent reporter. Including the work reported here, 208 PPR localizations were experimentally determined using fluorescent fusion proteins,

largely correlating with both bio-informatics and proteomics approaches (Table 2 and 3). Among the 159 PPRs proteins for which both experimental localization data based on protein fusion and predictions using bio-informatics tools are available, 135 (85%) have a similar localization. In addition, among the 36 PPR proteins being both identified in sub-proteomes and subjected to experimental localization studies using fluorescent protein fusion, 30 (83%) were compatible. The last set of data comes from the identification of the molecular functions of PPR proteins using reverse genetics, providing very important data about their localization (Table 2). As largely reported in the literature, PPR proteins are involved in regulating gene expression by acting through direct interaction with specific RNAs. A literature survey indicates that molecular roles were assigned to 68 PPR proteins (Table 2), occurring in plastids (31), in mitochondria (34), in both mitochondria and plastids (1), or in the nucleus (2). These reverse genetics studies are very strong statements of PPR localization, which could be considered as true localization. When compared with this very high quality data set, our data as well as all data of fluorescent protein localization appeared as very highly correlated with 17 out of 18 (94%) and 53 out of 57 (93%) compatible localization, respectively (Table 3).

As concluded in Table 2 and taking into account all the above depicted approaches, we assigned all Arabidopsis PPR proteins a probable localization depending on the strength of the available

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins

AGI	Gene Annotation ¹	Localization				Conclusion ⁸	EMSA ⁹	Molecular Function (localization) ⁸	References
		Domains ²	Predictions ³	Proteomics ⁴	Experimentals ⁵				
At1g01970	PPR containing protein	P	M		C ¹	C			"this report"
At1g02080	PPR containing protein	P	M			pM			
At1g02160	PPR containing protein	P	M	C (Al ³⁺ , Zn ²⁺)		C			"AT_Chiron, Huang et al 2011, "PPDB"
At1g02370	PPR containing protein	P	M	M (Al ³⁺)		M			"Kodmann et al 2011"
At1g02420	PPR containing protein	P	M		mtc ⁶	mtc			"this report"
At1g03100	PPR containing protein	P	M	C1 (Al ³⁺)		pM			"Narsai et al 2012"
At1g03510	PPR containing protein	PLS-E	C			pC			
At1g03540	PPR containing protein	PLS-E	M			pM			
At1g03580	PPR containing protein	P	M			pM			
At1g04840	PPR containing protein	PLS-E-DYW	m	PM (Al ³⁺)	C ³	-			"Mits et al 2009, "this report"
At1g05050	EMB1101	P	M			pM	confirmed ⁷		"SeedGenes"
At1g05670	PPR containing protein	P	M		MC ⁴	mtc			"this report"
At1g07570	CLB1WPCD347	PLS-E	c		C ^{1,2}	C		Editing rpoA and rpoB (C ¹)	"Challegier-Boudin et al 2008, "in house SUBA3"
At1g08140	MEF3	PLS-E	M			M		Editing rpoA (M ¹)	"Vozobek et al 2012"
At1g09190	EMB1444	PLS-E	C		MC ⁴	mtc	potentia ⁸		"this report, "Cushing et al 2005"
At1g09270	PPR containing protein	P	M		PRC ³	-			"Narsai et al 2011"
At1g09380	PPR containing protein	P	M		MC ⁴	mtc			"this report"
At1g09710	MTSF1	P	m		M ¹	M		Processing and stability of rnaH (M ¹)	"Hall et al 2013"
At1g07390	PPR containing protein	P	M			pM			
At1g07740	PPR containing protein	P	M			pM			
At1g08370	QTF2	PLS-E-DYW	C		C ¹	C		Editing rpoA, rpoB (C ^{1,2})	"in house SUBA3, "Hammani et al 2009, "Claude et al 2010"
At1g09610	PPR containing protein	P	C			pC			
At1g09190	PPR containing protein	PLS-E	-		M ¹	m			"this report"
At1g09220	PPR containing protein	PLS-E	M			pM			
At1g09410	PPR containing protein	PLS-E-DYW	M			C			
At1g09580	PPR containing protein	P	M		M ¹	M			"Narsai et al 2011"
At1g09820	PPR containing protein	P	M			pM			
At1g09900	PPR containing protein	P	c	C (Zn ²⁺)	M ¹	mtc			"PPDB, "this report"
At1g10270	GFP23	P	M		N ¹ , M ²	MN	confirmed ⁷		"Ding et al 2006, "Narsai et al 2011, "this report, "SeedGenes"
At1g10330	PPR containing protein	PLS-E	m			pM			
At1g10910	EMB1103	P	M	C (Zn ²⁺)		C	confirmed ⁸		"PPDB, "SeedGenes"
At1g11280	CRR2	PLS-E-DYW	c		C ^{1,2}	C		Editing rpoB, rpoD, rpoE (C ¹)	"this report, "in house SUBA3, "Claude et al 2009"
At1g11830	PPR containing protein	P	M	M (Al ³⁺)		M			"Heastwood et al 2004"
At1g11710	PPR containing protein	P	M			pM			
At1g11860	PPR containing protein	P	M			pM			
At1g12300	PPR containing protein	P	M			pM			
At1g12630	PPR containing protein	P	M			pM			
At1g12700	RPF1	P	M		M ¹	M		Processing rnaH transcript (M ¹)	"Holme et al 2011"
At1g12775	EMB1586	P	M			pM	confirmed ⁷		"SeedGenes"
At1g13040	PPR containing protein	P	M	V (Al ³⁺)		pM			"Jaquard et al 2007"
At1g13410	PPR containing protein	PLS-E	m			pM			
At1g13632.1	PPR containing protein	P	-			-			
At1g13800	FAC19	P	M			pM	confirmed ⁷		"Yu et al 2011"
At1g14470	PPR containing protein	PLS	m		C ¹	C			"this report"
At1g15480	PPR containing protein	P	m	M (Al ³⁺) PM (Al ³⁺)		M			"Kodmann et al 2011, "Mits et al 2009"
At1g15510	AECB2 NAC1	PLS-E-DYW	M		C ^{1,2,3}	C		Editing accD and rnaH (C ^{1,2})	"in house SUBA3, "Yu et al 2009, "Tsang et al 2010"
At1g16480	pseudogene	PLS-E-DYW	M			pM			
At1g16830	PPR containing protein	P	M			pM			
At1g17630	PPR containing protein	PLS-E	M			pM			
At1g18485	PPR containing protein	PLS-E-DYW	C			pC			
At1g18900	PPR containing protein	P-Q	M			pM			
At1g19250	PPR containing protein	P	m		M ^{1,2}	M			"this report, "in house SUBA3"
At1g19320	NFD3	P	M	PM (Al ³⁺)		pM	potentia ⁸		"Zhang et al 2011, "Porterko et al 2006"
At1g19720	PPR containing protein	PLS-E-DYW	-	C (Al ³⁺ , Zn ²⁺)	M ¹	mtc			"Kong et al 2011, "PPDB, "this report"
At1g20230	PPR containing protein	PLS-E-DYW	-		M ¹	m			"this report"
At1g20300	PPR containing protein	P	M		M ¹	M			"Narsai et al 2011"
At1g22530	PPR containing protein	PLS-E	M		M ¹	M			"this report"
At1g22960	PPR containing protein	P	M			pM			
At1g25380	PPR containing protein	PLS-E-DYW	M		MC ⁴	mtc			"this report"
At1g26480	PPR containing protein	P	M	M (Al ³⁺ , Cu ²⁺) PM (Al ³⁺)		M			"Heastwood et al 2004, "Kodmann et al 2011, "Huang et al 2009, "Zhang et al 2011"
At1g26500	PPR containing protein	P	M			pM			
At1g26900	PPR containing protein	PLS-E	M			pM			
At1g28120	PPR containing protein	P	M			pM			
At1g28190	PPR containing protein	PLS-E	M			pM			
At1g29710	PPR containing protein	PLS-E-DYW	M			pM			
At1g30290	pseudogene	P	M			pM			
At1g30610	EMB2279	P	C	C (Zn ²⁺)		C	confirmed ⁸		"PPDB, "SeedGenes"
At1g31430	PPR containing protein	PLS-E	c		M ¹	m			"this report"
At1g31780	PPR containing protein	PLS	c		C ¹	C			"this report"
At1g31840	PPR containing protein	P	-		M ¹	C			"this report"
At1g31920	PPR containing protein	PLS-E-DYW	-	C (Zn ²⁺)		C			"PPDB"
At1g32415	PPR containing protein	PLS-E	M			pM			
At1g33330	PPR containing protein	PLS-E	m		M ¹	M			"this report"
At1g34180	QGR1	PLS-E-DYW	M		M ¹	M		Editing rnaH, rnaD, comC, comD (M ¹)	"Kim et al 2009"
At1g35010	PPR containing protein	P	M			pM			
At1g35860	PPR containing protein	PLS-E	M			pM			
At1g37580	DYW1	PLS-E-DYW	C		C ^{1,2}	C		Editing rnaD (C ¹)	"Boussard et al 2012, "in house SUBA3"
At1g50270	PPR containing protein	PLS-E	M		M ¹	M			"this report"
At1g51965	P	ADOS	M	C (Al ³⁺)	M ¹	M		Splicing rnaD intron3 (M ¹)	"AT_Chiron, Liu et al 2010"
At1g52020	PPR containing protein	P	M	C (Al ³⁺)		C			"PPDB"
At1g52040	PPR containing protein	P	M			pM			
At1g53330	CB_1285	P	M			pM	confirmed ⁷		"Kocakbek et al 2008"
At1g53600	PPR containing protein	PLS-E	M			pM			
At1g55630	PPR containing protein	P	M			pM			
At1g55920	PPR containing protein	P	M	M (Al ³⁺) PM (Al ³⁺)		M			"Heastwood et al 2004, "Kodmann et al 2011, "Mits et al 2009"
At1g66570	PON	PLS-E	-		M ^{1,2}	M			"Lak et al 2011, "this report"

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation ^a	Domains ^b	Predictions ^c	Proteomics ^d	Experimental ^e	Conclusion ^f	EMSA ^g	Molecular Function (localization) ^h	References
At1g26990	PPR containing protein	PLS-E-DYW	M			pM			
At1g29720	CRR28	PLS-E-DYW	s		M ^h , C ^h	C		Editing rcd8, rcd10 (C)	¹ Linn et al 2004, ² in house SUBA3, ³ this report, ⁴ Okada et al 2009
At1g50770	PPR containing protein	P	m	M (At ^h , Os ^h) PM (At ^h)	M ^h	M			¹ Headwood et al 2004, ² Klodmann et al 2011, ³ Yuang et al 2009, ⁴ Isawa et al 2008, ⁵ this report
At1g61870	PPR336	P	M	M (At ^h)	M ^h	M			¹ Headwood et al 2004, ² Lytelwood 2007, ³ Klodmann et al 2011, ⁴ Linn et al 2004, ⁵ this report, ⁶ Takenaka et al 2010
At1g62280	MEF9	PLS-E	-		M ^h	M		Editing nat7 (M)	
At1g62390	THA5-LIKE3	P-D	M			pM			
At1g62990	PPR containing protein	P	m		M ^h	M			¹ this report
At1g62970	RPF2	P	M		M ^h	M		Processing nad9 & cdx3 (M)	¹ Jorjatz et al 2010
At1g62990	PPR containing protein	P	M			pM			
At1g62720	ABG1	P	-		M ^h	M			¹ Yang et al 2011
At1g62910	PPR containing protein	P	M	PM (At ^h)		pM			¹ Nitta et al 2009
At1g62930	RPF3	P	C		M ^h	M		Processing comC (M)	¹ Jorjatz et al 2011
At1g63070	PPR containing protein	P	M			pM			
At1g63080	PPR containing protein	P	M			pM			
At1g63130	PPR containing protein	P	M			pM			
At1g63150	PPR containing protein	P	M			pM			
At1g63320	PPR containing protein	P	m			pM			
At1g63330	PPR containing protein	P	-		M ^h	m			¹ this report
At1g63490	PPR containing protein	P	-		M ^h	m			¹ this report
At1g64100	PPR containing protein	P	-	M (At ^h)	m ^h	M/c		Editing comF2 (M)	¹ Klodmann et al 2011, ² this report, ³ Chateigner-Boulin et al 2013, ⁴ in house SUBA3
At1g64310	OTF1	PLS-E	m		M ^h , M/C ^h	M/c		Editing comF2 (M)	
At1g64430	PPR containing protein	P	C			pC			
At1g64590	PPR containing protein	P	M			pM			
At1g66345	PPR containing protein	P	M			pM			
At1g66930	PPR containing protein	PLS-E-DYW	m		M ^h	M			¹ this report
At1g66980	PPR containing protein	P	M			pM			
At1g66990	PPR containing protein	P	M	PM (At ^h)	M ^h	M			¹ Li et al 2012, ² this report
At1g69390	PPR containing protein	PLS-E	M		M ^h	M			¹ in house SUBA3
At1g71090	PPR containing protein	P	M			pM			
At1g71210	PPR containing protein	P	M			pM			
At1g71430	PPR containing protein	PLS-E-DYW	M		C ₂ (At ^h)	C			¹ Hummel et al 2012
At1g71460	PPR containing protein	PLS	G	C (At ^h , Zn ^h) C ₂ (At ^h)		C			¹ PPDB, ² Hummel et al 2012
At1g71490	PPR containing protein	PLS-E	-		M ^h	m			¹ this report
At1g73490	PPR containing protein	P	M			pM			
At1g73710	PPR containing protein	P	-		C ^h	C			¹ this report
At1g74490	PPR containing protein	PLS-E	M		M ^h	M			¹ this report
At1g74590	PPR containing protein	P	-			-			
At1g74690	OTF7 & OsPFR1	PLS-E	C		M/C ^h	M/c		Editing nat7, alp1 (M)	¹ Hammann et al 2011
At1g74690	PPR containing protein	PLS-E-DYW	-			-			
At1g74730	PPR containing protein	P-D	C			pC			
At1g74880	PTAC2	P-D	C	C (At ^h , Zn ^h) N (At ^h)		C	confirmed ^h		¹ Kaufmann et al 2004, ² AT-Chloro, ³ Platz et al 2006, ⁴ Kong et al 2011, ⁵ Engelsson et al 2012, ⁶ YPOB, ⁷ Sakanaka et al 2013
At1g74900	DTP43	P	M		M ^h	M		Splicing nad7 intron1 (M)	¹ in house SUBA3, ² de Kraker et al 2007
At1g76290.1	PPR containing protein	P	M		M ^h	M			¹ this report
At1g77010	PPR containing protein	PLS-E	M			pM			
At1g77170	PPR containing protein	PLS-E	C			pC			
At1g77340	PPR containing protein	P	m			pM			
At1g77390	AFPR6	P	M		M ^h	M		Processing and translation stabilisation of psu3 (M)	¹ Stanewski et al 2012
At1g77490	PPR containing protein	P	C			pC			
At1g78080	PPR containing protein	P	C			pC			
At1g78490	EMB2217	P-D	M		M ^h	M	potential ^h		¹ Narsai et al 2011, ² SeedGenes
At1g78540	PPR containing protein	P	M			pM			
At1g80190	PPR containing protein	P	M			pM			
At1g80270.1	PPR_398	P	M	M ^h (At ^h , Os ^h) C (At ^h) PM (At ^h)	M ^h	M/c			¹ Klodmann et al 2011, ² Huang et al 2009, ³ Froehlich et al 2003, ⁴ Zhang et al 2011, ⁵ Narsai et al 2011
At1g80590	PPR containing protein	P	M			pM			
At1g80890	PPR containing protein	P	M			pM			
At2g01390	PPR containing protein	P	-						
At2g01390	PPR containing protein	P	M		M ^h	M	confirmed ^h		¹ Linn et al 2004, ² SeedGenes
At2g01510	PPR containing protein	PLS-E-DYW	M			pM			
At2g01740	PPR containing protein	P	M		M ^h	M			¹ this report
At2g01890	EMB975	P	-		C ^h	C	confirmed ^h		¹ Linn et al 2004, ² SeedGenes
At2g02190	PPR containing protein	P	C			pC			
At2g02790	PPR containing protein	PLS-E	M			pM			
At2g02990	DTP85	PLS-E-DYW	M		C ^h	C		Editing rcd8 (C)	¹ in house SUBA3, ² Hammann et al 2009
At2g03390	PPR containing protein	PLS-E	M			pM			
At2g03890	REME1	PLS-E-DYW	M	PM (At ^h)	M ^h , C ^h	M		Editing rcd8, rcd8 (M)	¹ Alexander et al 2004, ² Bertolotti et al 2010, ³ Linn et al 2004
At2g04990	PPR containing protein	PLS-E	M		M ^h	M			¹ this report
At2g05000.1	PPR containing protein	P	M		C ^h	C			¹ this report
At2g13430	PPR containing protein	P	M			pM			
At2g13690	SLD2	PLS-E	-		M ^h	M		Editing nat8, nat1, nat8L, nat7, nat1 (M)	¹ Zhu et al 2012, ² this report
At2g15630	PPR containing protein	P	M		M ^h	M			¹ Narsai et al 2011
At2g15690	PPR containing protein	PLS-E-DYW	M	M (Os ^h , C ₂ Zn ^h) C (Zn ^h)	M/C ^h	M/c			¹ Huang et al 2009, ² YPOB, ³ in house SUBA3
At2g15820	DTP51	P	M		C ^h	C		Splicing ycd3 intron2 (C)	¹ PPDB, ² this report, ³ de Longeville et al 2008
At2g15990	PPR containing protein	P	-			-			
At2g16690	PRORP2	P-D	-		M ^h	N		Processing tRNA and maturation of Rna (N)	¹ Gobert et al 2010, ² Gutmann et al 2012
At2g16990	PPR containing protein	P	m			pM			
At2g17033.1	PPR containing protein	P-D	M	C (Zn ^h)		C			¹ YPOB
At2g17140	PPR containing protein	P	m			pM			
At2g17210	PPR containing protein	PLS-E	-			-			
At2g17525	PPR containing protein	P	M			pM			
At2g17570.1	PPR containing protein	P	M			pM			
At2g18530	PPR containing protein	PLS	M	M (At ^h)		M			¹ Klodmann et al 2011
At2g18940	ZnPPR10	P	C	C (Zn ^h)		C		Translation stabilisation of alpha-ApoI & paa-FpD3 (C)	¹ Platz et al 2009

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation ^a	Domains ^b	Predictions ^c	Proteomics ^d	Experimental ^e	Conclusion ^f	EMBL ^g	Molecular Function (localization) ^h	References
At2g19280	PPR containing protein	P	M	C (At ^h)	M ^h	Mic			¹ PROB, ² Lurin et al 2004
At2g20540	MEF21	PLS-E	-		M ^h	M		Editing <i>cox3</i> (M ^h)	¹ Lurin et al 2004, ² this report, ³ Takenaka et al 2010
At2g20710.1	PPR containing protein	P	M			pM			⁴ this report
At2g21090	PPR containing protein	PLS-E	M		M ^h	M			⁵ this report
At2g20770	PPR containing protein	PLS-E-DYW	C		M ^h	m			⁶ Sung et al 2010
At2g20410	SLD1	PLS-E	M		M ^h	M		Editing <i>nad4</i> , <i>nad9</i> (M ^h)	⁷ Takenaka et al 2010, ⁸ Venkatesh et al 2012
At2g20580	MEF3	PLS-E-DYW	M		M ^h	M		Editing <i>nad5</i> , <i>nad6</i> (M ^h)	⁹ this report
At2g20790	PPR containing protein	P	M		M ^h	M			¹⁰ Hurrel et al 2012, ¹¹ this report
At2g21610	PPR containing protein	PLS-E-DYW	-	Cl (At ^h)	M ^h	M			¹² this report
At2g27800	PPR containing protein	P	M		M ^h	pM			¹³ AT_Chloro, ¹⁴ this report
At2g28950	PPR containing protein	P	M	C (At ^h)	M ^h	Mic			¹⁵ Lurin et al 2004, ¹⁶ in house SUBA3, ¹⁷ Harman et al 2000
At2g29780	OTF81	PLS-E-DYW	C		C ^h	C		Editing <i>psb27</i> (C ^h)	¹⁸ this report
At2g30100	PPR containing protein	P	C			pC			
At2g30780	PPR containing protein	P	M			pM			
At2g31490	GUN1	P-D	C		C ^h	C			¹⁹ Koussevitzky et al 2007
At2g32230	PRORP1	P-D	M	C (Zm ^h)	MIC ^h	M/C	confirmed ^g	Processing tRNA elements (pM/C)	²⁰ PROB, ²¹ Gobert et al 2010
At2g32830	PPR containing protein	P	M		M ^h	M			²² this report
At2g33280	PPR containing protein	PLS-E	-		C ^h	C			²³ this report, ²⁴ in house SUBA3
At2g33780	PPR containing protein	PLS-E-DYW	C		C ^h	C			²⁵ Lurin et al 2004
At2g34370	PPR containing protein	PLS-E-DYW	M		M ^h	M			²⁶ this report
At2g34490	PPR containing protein	PLS-E	M		mM ^h	mM			²⁷ in house SUBA3
At2g35030	PPR containing protein	PLS-E	M		M ^h	M			²⁸ PROB, ²⁹ this report
At2g35130	PPR containing protein	P	-	C (Zm ^h)	C ^h	C			³⁰ this report
At2g36240	PPR containing protein	P	M		M ^h	M			³¹ this report
At2g36730	PPR containing protein	PLS-E	M		M ^h	pM			³² this report
At2g36980	PPR containing protein	PLS-E	M		M ^h	pM			³³ this report
At2g37230	PPR containing protein	P	M	C (At ^h) (M/Os ^h)	M ^h	Mic			³⁴ AT_Chloro, ³⁵ PROB, ³⁶ Huang et al 2009, ³⁷ Zhang et al 2011, ³⁸ this report
At2g37310	PPR containing protein	PLS-E	M		C ^h	C			³⁹ this report
At2g37320	PPR containing protein	PLS-E	M			pM			
At2g38420	PPR containing protein	P	M			pM			
At2g39230	LDU	P	M			pM			
At2g39820	PPR containing protein	PLS-E	M		C ^h	C			⁴⁰ this report
At2g40240	PPR containing protein	P	M		M ^h	M			⁴¹ Hansal et al 2011
At2g40720	PPR containing protein	PLS-E	M		M ^h	M			⁴² this report
At2g41080	PPR containing protein	PLS-E-DYW	-		mM ^h	mM			⁴³ this report
At2g41720.1	EMB2554	P	-		C ^h	C	potential ^g		⁴⁴ this report, ⁴⁵ SeedGenes
At2g42820	PPR containing protein	PLS-E	C			pC			⁴⁶ this report
At2g44880	ANG11	PLS-E	-	PM (At ^h)	M ^h , MIC ^h	Mic		Editing <i>nad6</i> (M ^h)	⁴⁷ Mita et al 2009, ⁴⁸ Moriyama et al 2012, ⁴⁹ this report
At2g45350	CRR4	PLS-E	-	V (At ^h)		C		Editing <i>ndm2</i> (C ^h)	⁵⁰ Sapich et al 2004, ⁵¹ Volterri et al 2004
At2g46050	PPR containing protein	PLS-E	M			pM			
At2g48050	PPR containing protein	P	M			pM			
At2g51580	PPR containing protein	PLS-E	M	C (At ^h)	M ^h	Mic			⁵² Kong et al 2011, ⁵³ this report
At2g52010	PPR containing protein	PLS-E-DYW	M		M ^h	M			⁵⁴ Lurin et al 2004, ⁵⁵ this report
At2g52330	PPR containing protein	PLS-E	M		MIC ^h	mM			⁵⁶ in house SUBA3
At2g52490	PPR containing protein	P	M			pM			
At2g53030	PPR containing protein	P	M			pM			
At2g53380	PPR containing protein	PLS-E-DYW	-			-			
At2g54130.1	PPR containing protein	P	M			pM			
At2g54750	PPR containing protein	PLS-E	M			pM			
At2g54780	PPR containing protein	P	C	C (At ^h)		C			⁵⁷ PROB, ⁵⁸ Jefferson et al 2004
At2g55040	MEF19	PLS-E	M			M		Editing <i>cox20F</i> (M ^h)	⁵⁹ Takenaka et al 2010
At2g55340	PPR containing protein	PLS-E	M			pM			⁶⁰ PROB, ⁶¹ in house SUBA3, ⁶² Williams & Berken 2003, ⁶³ Lu et al 2011
At2g58430	EMB2790 (atPPR2)	P	C	C (Zm ^h)	C ^h	C		Translation stabilisation (C ^h)	⁶⁴ this report
At2g58920	PPR containing protein	P	-		M ^h	m			⁶⁵ this report
At2g57280	PPR containing protein	P	M			pM			⁶⁶ this report, ⁶⁷ in house SUBA3
At2g58830	PPR containing protein	PLS-E-DYW	M		M ^h	M			⁶⁸ this report
At2g59040	PPR containing protein	PLS-E	M			pM			⁶⁹ this report
At2g59050	PPR containing protein	P	M		M ^h	M			⁷⁰ this report
At2g59850	HCF152/OM33	P	C	C (Zm ^h)	C ^h	C		Processing <i>psbB</i> (C ^h)	⁷¹ PROB, ⁷² Meierhoff et al 2003, ⁷³ this report
At2g11490	MEF10	PLS-E-DYW	M		M ^h	M		Editing <i>nad7</i> (M ^h)	⁷⁴ Lurin et al 2004, ⁷⁵ Hartel et al 2013
At2g12770	MEP22	PLS-E-DYW	-	Cl (At ^h)	M ^h	M		Editing <i>nad9</i> (M ^h)	⁷⁶ Hurrel et al 2012, ⁷⁷ this report, ⁷⁸ Takenaka et al 2010
At2g13150	PPR containing protein	P	-			-			
At2g13180	PPR containing protein	P	M	M (At ^h)	M ^h	M			⁷⁹ Heardwood et al 2004, ⁸⁰ Klodmann et al 2011, ⁸¹ Taylor et al 2011, ⁸² in house SUBA3
At2g13770	PPR containing protein	PLS-E-DYW	M		M ^h	M			⁸³ Lurin et al 2004
At2g13880	OTF72	PLS-E	M		M ^h	M		Editing <i>psbE</i> (M ^h)	⁸⁴ Lurin et al 2004, ⁸⁵ Chateignier-Boulin et al 2013
At2g14330	CREF3	PLS-E-DYW	M		M ^h	mM		Editing <i>psbE</i> (C ^h)	⁸⁶ this report, ⁸⁷ Tagli et al 2013
At2g14580	PPR containing protein	P	M			pM			
At2g14730	PPR containing protein	PLS-E	-			-			
At2g15130	PPR containing protein	PLS-E-DYW	M		M ^h , MIC ^h	pM			⁸⁸ this report, ⁸⁹ in house SUBA3
At2g15290	PPR containing protein	P	M			pM			
At2g15390	PPR containing protein	P	M	M (At ^h)		M			⁹⁰ Buglioni et al 2004, ⁹¹ Klodmann et al 2011, ⁹² Taylor et al 2011
At2g15930	PPR containing protein	PLS-E	C		M ^h	m			⁹³ this report
At2g16010	PPR containing protein	P	M			pM			
At2g16010	PPR containing protein	PLS-E	M		M ^h	M			⁹⁴ this report
At2g16710	PPR containing protein	P	M			pM			
At2g16890	PPR40	P	M		M ^h	M			⁹⁵ Zaigner et al 2008, ⁹⁶ in house SUBA3
At2g18020	PPR containing protein	P	M			pM			
At2g18110	EMB1270	P-D	C	C (Zm ^h)		C	confirmed ^g		⁹⁷ PROB, ⁹⁸ SeedGenes
At2g18340	PPR containing protein	PLS-E	M			M			⁹⁹ AT_Chloro, ¹⁰⁰ Takenaka et al 2010
At2g18970	MEP20	PLS-E	M	C (At ^h)	M ^h	Mic		Editing <i>psaF</i> (M ^h)	¹⁰¹ this report
At2g20730	PPR containing protein	PLS-E	M		M ^h	M			¹⁰² this report
At2g21470	PPR containing protein	PLS-E	-		mM ^h	mM			¹⁰³ this report
At2g22150	MPR25	PLS-E	C	C (At ^h)	M ^h , C ^h	Mic		Editing <i>nad6</i> (M ^h)	¹⁰⁴ AT_Chloro, ¹⁰⁵ Tolia et al 2012, ¹⁰⁶ in house SUBA3
At2g23470	PPR containing protein	P	M			pM			

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation ¹	Domains ²	Predictions ³	Proteomics ⁴	Experimental ⁵	Conclusion ⁶	EMB ⁷	Molecular Function (localization) ⁸	References
Atg22670	PPR containing protein	P	M			pm			
Atg22680	YB1	PLS-E-DYW	C		C ^a	C		Editing yob ¹ (C ^a)	¹ Zhou et al 2008
Atg23390	PPR containing protein	P	m	C (At ^a)		c			¹ PPDB, ¹ Kieffner et al 2004
Atg24000	PPR containing protein	PLS-E-DYW	-		m/c ^a	mc			¹ this report
Atg24000	PPR containing protein	PLS-E-DYW	M		M ^a	M			¹ Lurin et al 2004
Atg24000	PPR containing protein	PLS-E	M			pm			
Atg24010	PPR containing protein	P	-			-			
Atg24070	PPR containing protein	PLS-E	-		M ^a	m			¹ this report
Atg24540	PPR containing protein	PLS	-		M ^a	m			¹ this report
Atg24630	PPR containing protein	PLS	C			pc			
Atg24780	MEF14	PLS-E-DYW	-	M (At ^a)		M		Editing muf1 (M ^a)	¹ Headehead et al 2004, ¹ Vorobsky et al 2011
Atg27780	EMB1237HAB	P-D	C	C (At ^a)	C ^a	C		Splicing yob3 intron 2 & muf1 (C ^a)	¹ PPDB, ¹ Wojtowicz et al 2012
Atg28640	PPR containing protein	PLS-E	-			c			
Atg28660	PPR containing protein	PLS-E	-			pc			
Atg29030	PPR containing protein	PLS-E	C	C (At ^a)	C ^a	C			¹ PPDB, ¹ in house SUBA3
Atg29090	EMB1076	P	M			pm	potentia ^a		¹ SeedGenes
Atg42630	PPR containing protein	P	M	C (At ^a , Zn ^a)	C ^a	C			¹ PPDB, ¹ Kieffner et al 2004, ¹ this report
Atg44610	PPR containing protein	P	-	C (Zn ^a)	C ^a	C			¹ PPDB, ¹ this report
Atg46790	CRR2	PLS-E-DYW	C		C ^a	C		Processing rbf2b (C ^a)	¹ this report, ¹ in house SUBA3, ¹ et al 2003, ¹ Headehead
Atg46870	THAB-LIKE2	P	M	C (At ^a) PM (At ^a)		c			¹ PPDB, ¹ Kieffner et al 2004, ¹ Mits et al 2008
Atg47530	PPR containing protein	PLS-E-DYW	m		MC ^a	mc			¹ this report
Atg47540	PPR containing protein	PLS-E	M		C ^a	c			¹ this report
Atg48290	BPR	P	C		M ^a	M		Splicing nad7 intron1 (M ^a)	¹ Kozmin et al 2010
Atg48810	PPR containing protein	P	m		M ^a	M			¹ this report, ¹ in house SUBA3
Atg49140	PPR containing protein	PLS-E-DYW	M	C (At ^a)		c			¹ AT_Chloro, ¹ PPDB
Atg49170	EMB2081	PLS-E-DYW	C		C ^a	C	confirmed ^a		¹ in house SUBA3, ¹ SeedGenes
Atg49640	EMB1796	P	M	M (At ^a) PM (At ^a) C (At ^a , Zn ^a)	MC ^a	MC	confirmed ^a		¹ et al 2008, ¹ Kieffner 2011, ¹ PPDB, ¹ Zhang et al 2011, ¹ this report, ¹ SeedGenes
Atg49710	PPR containing protein	PLS-E-DYW	-		MC ^a	mc			¹ this report
Atg49730	Zmtrnpp1 orthologous	P	-		M (Zn ^a)	M			¹ Gutierrez-munoz et al 2007
Atg49740	PPR containing protein	PLS-E	m		M ^a	M			¹ this report, ¹ in house SUBA3
Atg50420	PPR containing protein	PLS-E	m		MC ^a	mc			¹ this report
Atg51320	PPR containing protein	PLS-E	M			pm			
Atg53170	PPR containing protein	P	-	N (At ^a) C (Zn ^a)	C	n/c			¹ Pendle et al 2005, ¹ PPDB
Atg53360	PPR containing protein	PLS-E	M			pm			
Atg53700	MEG40	P	-	C (At ^a , Zn ^a)		C	confirmed ^a		¹ PPDB, ¹ Headehead et al 2005
Atg54080	PPR containing protein	P	M			pm			
Atg56030	PPR containing protein	P	M			pm			
Atg56690	PPR containing protein	PLS-E-DYW	m			pm			
Atg57430	OTPB4	PLS-E-DYW	C	PM (At ^a)	C ^a	C		Editing yob2, muf5, muf6 (C ^a)	¹ Li et al 2012, ¹ this report, ¹ in house SUBA3, ¹ Harman et al 2009
Atg58690	PPR containing protein	P	m		M ^a	M			¹ this report
Atg59440	PPR containing protein	P	C	C (Zn ^a)		C			¹ PPDB
Atg59500	PPR containing protein	P	M			pm			
Atg59560	PPR containing protein	P	M	M (At ^a)		M			¹ Headehead et al 2004
Atg59590	PPR containing protein	P	M			pm			
Atg51170	PPR containing protein	PLS-E-DYW	M			pm			
Atg51360	PPR containing protein	P	M			pm			
Atg51520	PPR containing protein	P	M			pm			
Atg52470	PPR containing protein	P	M			pm			
Atg52540	PPR containing protein	P	M			pm			
Atg52690	PPR containing protein	PLS-E-DYW	C		MC ^a , C ^a	mc			¹ this report, ¹ in house SUBA3
Atg53370	OTPB6	PLS-E-DYW	C		C ^a	C		Editing ore14 (C ^a)	¹ in house SUBA3, ¹ Harman et al 2009
Atg51030	PPR containing protein	PLS-E-DYW	M	C (Zn ^a)		c			¹ PPDB
Atg51400	PPR containing protein	P-D	M	PM (At ^a)		pm			¹ Mits et al 2009
Atg51570	PPR containing protein	P	M			c			¹ this report
Atg51890	PPR containing protein	P	M			pm			
Atg52790	PPR containing protein	PLS-E-DYW	M		M ^a	M			¹ Lurin et al 2004, ¹ this report
Atg52820	PPR containing protein	P	M		M ^a	M			¹ Harman et al 2011
Atg54370	PPR containing protein	PLS-E	m		MC ^a	MC			¹ this report, ¹ in house SUBA3
Atg54790	PPR containing protein	P	M			pm			
Atg58210	PPR containing protein	PLS-E	m		M ^a	M			¹ this report
Atg51690	PPR containing protein	P	m		M ^a	M			¹ this report
Atg51690	PPR containing protein	PLS-E-DYW	M			pm			
Atg51690	PPR containing protein	PLS-E-DYW	M			pm			
Atg51470	PPR containing protein	PLS-E	M			pm			
Atg51490	PPR containing protein	P	M			pm			
Atg51480	PPR containing protein	PLS-E-DYW	C		M ^a	m			¹ this report
Atg51490	LOI/MEF11	PLS-E-DYW	M		M ^a MC ^a	MC		Editing oox3, oox4, oox203 (M ^a)	¹ Tang et al 2010, ¹ this report, ¹ Vorobsky et al 2010
Atg51720	PPR containing protein	PLS-E-DYW	-		C ^a	c			¹ this report
Atg51930	SVU7 RNA binding P27	P-D	C	C (At ^a , Zn ^a)	C ^a	C			¹ AT_Chloro, ¹ PPDB, ¹ Lurin et al 2004, ¹ Li et al 2012
Atg51940	PPR containing protein	PLS-E	M		M ^a	M			¹ this report
Atg51630	PPR containing protein	PLS-E-DYW	m			pm			
Atg51710	PPR containing protein	P	M			pm			
Atg51790	PPR containing protein	P	M			pm			
Atg51820	PCB1	P	C	C (Zn ^a)		C		Processing yob4 transcript (C ^a)	¹ PPDB, ¹ Hao et al 2010
Atg51790	DOT4	PLS-E-DYW	C	C (Zn ^a)	C ^a	C			¹ PPDB, ¹ in house SUBA3
Atg51840	PPR containing protein	PLS-E	-		C ^a	c			¹ this report, ¹ in house SUBA3
Atg51875.1	PPR containing protein	P	C			pc			
Atg51911	PPR containing protein	PLS-E	M			pm			
Atg51920	PPR containing protein	PLS-E	M			pm			
Atg51940	PPR containing protein	P	-			-			
Atg51990	Glycosyl transferase-related	P-D	m			pm			
Atg52090	EMB1028	P	c		M ^a	M	confirmed ^a		¹ Lurin et al 2004, ¹ this report, ¹ SeedGenes
Atg52070	EMB1031	P	c	PM (At ^a)	C ^a	C	confirmed ^a		¹ Li et al 2012, ¹ this report, ¹ SeedGenes
Atg52070	PPR containing protein	PLS-E	M			pm			
Atg52106.1	PPR containing protein	PLS-E-DYW	-			-			
Atg521170	PPR containing protein	P	M		MC ^a	MC			¹ Nara et al 2011
Atg52190	EMB1017	P	M	C (Zn ^a)		c	confirmed ^a		¹ PPDB, ¹ SeedGenes

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation ¹	Domains ²	Localization				EMG ³	Molecular Function (localization) ⁴	References
			Predictions ⁵	Proteomics ⁶	Experimental ⁷	Conclusion ⁸			
Atg21300	PPR containing protein	PLS-E	M	C (Zn ²⁺)	C ^a	C			"PPDB, "in house SUBA3
Atg21708	PPR containing protein	P	M		M ^a	M			"this report
Atg21892	PPR containing protein	P	M		M ^a	M			"this report
Atg21902	PRORP3	P-D	-		N ^a	N		Processing tRNA and maturation of RNA (N ^a)	"Gibbert et al 2010, "Gutmann et al 2012
Atg22780	PPR containing protein	PLS-E	-		M ^a	re			"this report
Atg25270	OTF75	PLS-E	C		C ^{a,b}	C		Splicing rpoC1 intron (C ^a)	"Chaignon-Boutin et al 2011, "in house SUBA3
Atg25680	PPR containing protein	P	M			pM			
Atg25693	PPR containing protein	P	M			pM			
Atg26010	RPF5	P	M		M ^a	M		processing nascent apc9, 26S rRNA (M ^a)	"this report, "Häcker et al 2013
Atg30103	MEF29/ZnPPR2263	PLS-E-DYW	M		M/C ^a	M/C		Editing nas2, cob (M ^a)	"Sosso et al 2012
Atg30825	PPR containing protein	P	-	C (Zn ²⁺)		c			"PPDB
Atg31070	PPR containing protein	PLS-E	M			pM			
Atg31802	PGR3	P	-		C ^a	C		Translation stabilisation, peLL and rchA (C ^a)	"Lurin et al 2004, "Yamazaki et al 2004, "Cai et al 2011
Atg32430	PPR containing protein	PLS-E	M			pM			
Atg32450	MEF85	PLS-E-DYW	M		M ^a	M		Editing nas2, nas3 (M ^a)	"in house SUBA3, "Verdubelt et al 2012
Atg33170	PPR containing protein	PLS-E-DYW	M		M/C ^{a,b}	M/C			"this report, "in house SUBA3
Atg33390	EMB2756	PLS-E-DYW	M			pM	potential ⁹		"SeedGenes
Atg34600	MRL1	P	C	C (Al ³⁺) PM (Al ³⁺)		C		Processing stabilisation ribol (C ^a)	"PPDB, "AT_Chloro, "Li et al 2012, "Johnson et al 2010
Atg35130	PPR containing protein	PLS-E-DYW	C		C ^a	C			"in house SUBA3
Atg35650	PPR containing protein	P	M	M (Al ³⁺), Ols ²		M			"Miller et al 2001, "Headwood et al 2004, "Woodmann et al 2011, "Taylor et al 2011
Atg35690	PPR containing protein	P	M	M (Al ³⁺)		M			"Headwood et al 2004, "Woodmann et al 2011
Atg37170	PPR containing protein	PLS-E-DYW	-		M ^a	re			"this report
Atg37282	PPR containing protein	PLS-E-DYW	C			pC			
Atg38010	PPR containing protein	PLS-E	M			pM			
Atg38150.1	PPR containing protein	P	-	M (M ^a)		re			"Taylor et al 2011
Atg39530	PPR containing protein	PLS-E	M			pM			
Atg39620	EMB2453/ZnPPR5	P	C	C (Zn ²⁺)		C	confirmed ⁹	splicing hnd (C ^a)	"PPDB, "SeedGenes, "Beck et al 2008
Atg39852	PPR containing protein	PLS-E	M			pM			
Atg40110	PPR containing protein	P	C			pC			
Atg40280	PPR containing protein	P	C	C (Zn ²⁺)		C			"PPDB
Atg40290	PPR containing protein	P	M	C (Zn ²⁺)		c	potential ⁹		"PPDB, "Myoung et al 2010
Atg40360.2	PPR containing protein	P	M			pM			
Atg40800	EMB175	PLS-E-DYW	C	C (Zn ²⁺) PM (Al ³⁺)	M/C ^{a,b}	M/C	confirmed ⁹		"PPDB, "Karnath et al 2010, "in house SUBA3, "this report, "SeedGenes
Atg40780	PPR containing protein	PLS-E-DYW	-	C (Al ³⁺)		c			"Vierthum et al 2004
Atg40810	ZnPPR4	P-D	C	C (Al ³⁺)	C ^a	C		Splicing rps12 intron1 (C ^a)	"PPDB, "this report, "Schmidt-Innenweber et al 2006
Atg40840	PPR containing protein	PLS-E-DYW	M		C ^a , M/C ^a	re/c			"this report, "in house SUBA3
Atg40810	PPR containing protein	PLS-E	M			pM			"this report
Atg40490	SLD1	PLS-E	M		M ^a	M		Editing nas2 (M ^a)	"Yuan & Liu 2012
Atg40810	PPR containing protein	PLS-E	-		M ^a	re			"this report
Atg40450	PPR containing protein	P	M	M (M ^a)		M			"Woodmann et al 2011
Atg40690	MEF7	PLS-E-DYW	-		M ^a	M		Editing nas2, nas3L, cob, cob206 (M ^a)	"Lurin et al 2004, "Zehmann et al 2012
Atg41090	PPR containing protein	P-D	-	CL (Al ³⁺)	C ^a	C			"Bo et al 2011, "this report
Atg41310	PPR containing protein	P	M			pM			
Atg412100	PPR containing protein	P	M	M (M ^a)		M			"Tan et al 2008
Atg41320	PPR containing protein	PLS-E-DYW	M		M ^a	M			"Lurin et al 2004
Atg41320	- RARE1	PLS-E-DYW	-		C ^a	C		Editing nas2 (C ^a)	"Lurin et al 2004, "Roberts et al 2008
Atg41370	PPR containing protein	P	C	C (Zn ²⁺)		C			"PPDB
Atg41400	PPR containing protein	P	C	C (Al ³⁺)	mis ³	M/C			"PPDB, "this report
Atg41470	PPR containing protein	P	M	PM (Al ³⁺)	M ^a	M			"Li et al 2012, "Lurin et al 2004
Atg41482	PPR containing protein	P	M			pM			
Atg418010	PPR containing protein	P	C			pC			
Atg418280	PPR containing protein	P	M			pM			
Atg41930	PPR containing protein	PLS-E	-	PM (Al ³⁺)		-			"Nata et al 2009
Atg41940	PPR containing protein	PLS-E-DYW	M		M ^a	M			"Lurin et al 2004
Atg41990	PPR containing protein	P	M	M (Al ³⁺) PM (Al ³⁺)		M			"Woodmann et al 2011, "Zhang et al 2011
Atg416420	PPR containing protein	P	M			pM			
Atg416640	PPR containing protein	P	M			pM			
Atg416650	PPR containing protein	PLS-E-DYW	-			-			
Atg416390	PPR containing protein	P	M			pM			
Atg416478	PPR containing protein	P	M		M ^a	M			"this report
Atg418650	PPR containing protein	P	M		M ^a	M			"this report
Atg419020	MEF18	PLS-E	C			M		Editing nas2 (M ^a)	"Tahmaki et al 2010
Atg41722	AP14H1	P-D	-	S (Al ³⁺)	mis ³	re/c			"PPDB, "this report
Atg42430	PPR containing protein	P	-			-			
Atg42530	PPR containing protein	P	-	C (Zn ²⁺)		c			"PPDB
Atg42710	PPR containing protein	PLS-E	M		M/C ^a	re/c			"in house SUBA3
Atg42720	EMB206	P	C	C (Zn ²⁺)	C ^a	C	potential ⁹		"PPDB, "this report, "SeedGenes
Atg427490	PPR containing protein	P	M			pM			
Atg428400	PPR containing protein	P	M			pM			
Atg43030	pseudogene	P	-			-			
Atg43750	PPR containing protein	PLS-E	M			pM			
Atg438730	PPR containing protein	P	M			pM			
Atg43930	PPR containing protein	PLS-E	M		M ^a	M			"in house SUBA3
Atg43980	EMB2744	PLS-E-DYW	-		M ^a	re	potential ⁹		"Lurin et al 2004, "in house SUBA3, "SeedGenes
Atg439710	EMB2745	P	M		M ^a	M	potential ⁹		"Nata et al 2011, "SeedGenes
Atg43990	EMB2746	P	-	C (Zn ²⁺)		c	confirmed ⁹		"PPDB, "SeedGenes
Atg44000	PPR containing protein	P	-			-			
Atg44040	PPR containing protein	PLS-E-DYW	M		M ^a	re			"this report
Atg440410	PPR containing protein	PLS-E-DYW	re	C (Al ³⁺)		c			"AT_Chloro, "Kong et al 2011
Atg441170	PPR containing protein	P	M			pM			
Atg44210	Ortholog of Z. Mays CRP1	P	M	C (Al ³⁺)		C		Translation stabilisation peLL and pasC (C ^a)	"PPDB, "Nata et al 1998, "Schmidt-Innenweber et al 2006
Atg442450	PPR containing protein	PLS-E	-			-			
Atg443790	PPR containing protein	PLS-E	-		mis ³	re/c			"this report
Atg443820	PPR containing protein	P	M			pM			

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 103–110

(1) Functional annotations were obtained from TAIR web site using the *Arabidopsis* Genome Initiative (AGI) genome release ver10. ABO5, ABA OVERLAY-SENSITIVE; ATECB, EARLY CHLOROPLAST BIOGENESIS; BIR, BSO-INSENSITIVE-ROOTS; CLB, CHLOROPLAST BIOGENESIS; CREP, CHLOROPLAST RNA EDITING FACTOR; CHR, CHLOROPLAST RESPIRATORY REDUCTION; DG, DELAYED GREENING; DGT, DEFECTIVELY ORGANIZED TRIBUTARIES; EMB, EMBRYO DEFECTIVE; FAC, EMBRYONIC FACTOR; GRP, GLUTAMINE-RICH PROTEIN; GUN, GENOME UNCUPLED; HCF, HIGH CHLOROPHYLL FLUORESCENCE; LOI, LOVASTATINE INSENSITIVE; LOJ, LATERAL ORGAN JUNCTION; LPA, LOW PSII ACCUMULATION; MEF, MITOCHONDRIAL RNA EDITING FACTOR; MPR25, MITOCHONDRIAL PPR 25; MTSF, MITOCHONDRIAL STABILITY FACTOR; NFD, NUCLEAR FUSION DEFECTIVE; OGR1, OPAQUE AND GROWTH RETARDATION; OTP, ORGANELLE TRANSSCRIPT PROCESSING; PDE, PIGMENT DEFECTIVE; PDM, PIGMENT DEFICIENT MUTANT; PGN, PENTATRICOPEPTIDE GERMINATION ON NaCl; PGR, PROTON GRADIENT REGULATION; PNM, PROTEIN LOCALIZED TO THE NUCLEUS AND MITOCHONDRIA; PPR, PENTATRICOPEPTIDE REPEAT; PRORP, PROTEINACEOUS RNASE P; PTAC, PLASTID TRANSCRIPTIONALLY ACTIVE; REME, REQUIRED FOR EFFICIENCY OF MITOCHONDRIAL EDITING; RPF, RNA PROCESSING FACTOR; SLG, SLOW GROWTH; SVR, SUPPRESSOR OF VARIATION; VAC, VANILLA CREAM; YS, YELLOW SEEDLING; ZmempP, Z. mays EMPTY PERICARP, ZmPPR, Zea mays PPR. (2) PPR domains were recovered from FLAGdb¹¹ v5 (<http://urgv.evyri.nra.fr/projects/FLAGdb+/HTML/index.shtml>) and from manually curated published evidences. Domain identifiers are according to Lurin and co-workers:²¹ "P" for PPR P-type domains, "P-D" for PPR P-type with additional atypical domain, "PLS" for PPR PLS-type domains, "PLS-E" for PPR PLS-type with an E- or EE- type additional domain, and "PLS-E-DYW" for PPR PLS-type containing EE- and DYW additional domains. (3) Localization predictions were aggregated from the independent predictions provided by the following software: Predotar v1.03, TargetP server v1.1, iPSORT, Multi Loc, LocTree, and AtsubP server with the complete *Arabidopsis* proteome using default settings. The rules to propose a conclusive prediction were as follows: if four or more software give the same prediction, this prediction is proposed and noted in uppercase; if three software give the same prediction and the three others do not predict any localization, the prediction is proposed and noted in uppercase; if two software give the same prediction and the four others do not predict any localization, the prediction is proposed and noted in lowercase; if three software give the same prediction and another predict a different localization, the main prediction is proposed and noted in lowercase; in the other cases, no prediction is proposed (-). (4) Proteomic localizations were gathered from published studies and from organelle proteomic databases as indicated in corresponding references in the last column of the table. Additional information in brackets states in which species the proteomic investigation was (were) performed: "At" stands for *Arabidopsis thaliana*, "Zm" for *Zea mays*, and "Os" for *Oriza sativa*. (5) Experimental localizations of fluorescent proteins were collected from targeted published studies and systematic approaches,^{13,17} this report, unpublished data from SUBA3 either using targeting peptides or full-length proteins. (6) Conclusion column gives a probable subcellular localization by integrating prediction, proteomic, genetics, and fluorescent proteins data. The decision rule is as follows: reverse genetics is prevalent followed by fluorescent proteins, proteomic data, and prediction. The conclusion is indicated in uppercase if reverse genetics data is available, if two experimental results are identical, or if the experimental data fit with the prediction. If not, the conclusion is indicated in lowercase. If only predictions are available, the predicted localization is indicated with a preceding "p". (7) Data of PPR Embryo defective mutants (EMB) was obtained from SeedGenes database (<http://www.seedgenes.org/index.html>) and manually curated mutants from published studies. (8) Molecular function based on reverse genetics approaches were obtained from literature, the localization of the molecular function is indicated in brackets. Localization data is indicated as followed. M, mitochondria; C, chloroplasts; N, nucleus; V, vacuole; Ct, cytosol; PM, plasma membrane. N/Ct, nucleus and cytoplasm; M/C, mitochondria and chloroplasts; lower case, "probably"; "pX", predicted in compartment X (conclusion column).

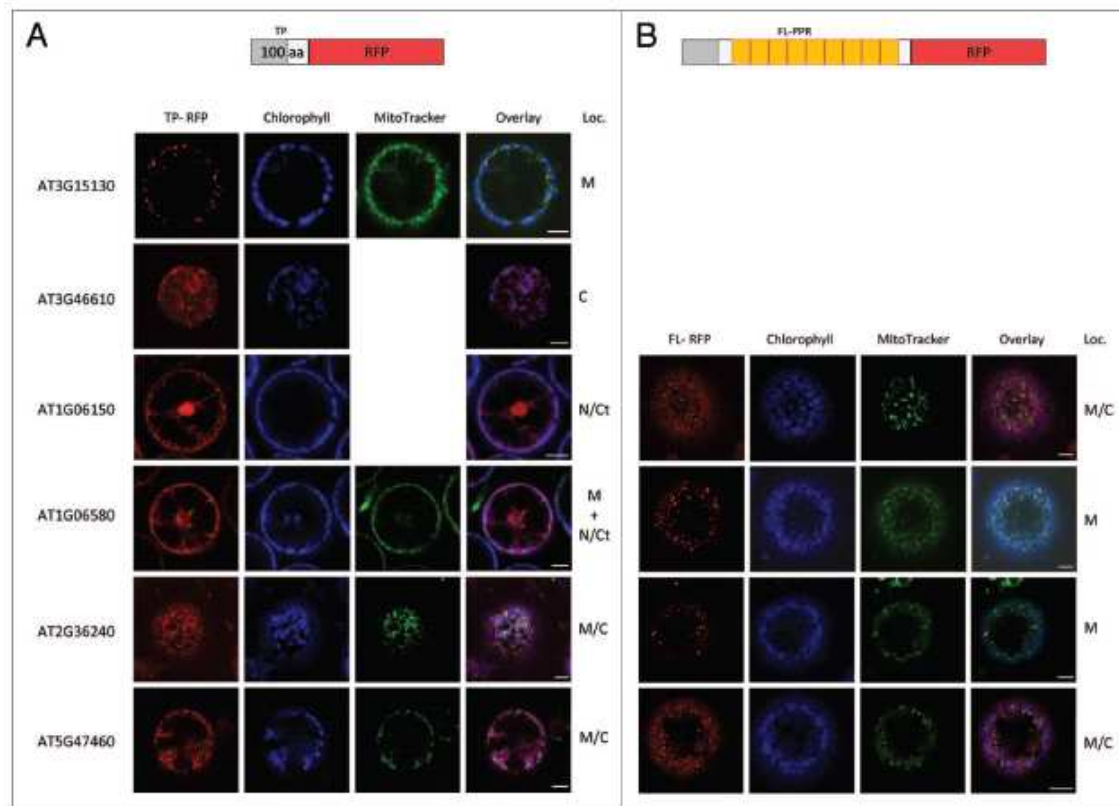


Figure 1. Examples of typical sub-cellular localizations observed using confocal microscopy. Confocal images of protoplasts obtained from *Nicotiana benthamiana* leaves infiltrated with constructs containing (A) the first 300 bp, coding for the first 100 amino acids, of six PPR ORFs, fused to the RFP coding sequence (TP-RFP) or (B) the full-length sequence of four PPR ORFs fused to the RFP coding sequence (FL-RFP). The RFP fluorophore (in red), the MitoTracker Green staining (in green), and the chlorophyll autofluorescence (in blue) were simultaneously visualized. Overlay panels show combined fluorescence from RFP, MitoTracker, and chlorophyll autofluorescence. Loc, deduced subcellular localization; M, mitochondria; C, chloroplasts; N/Ct, nucleus and cytosol; M/C, mitochondria and chloroplasts. Bars: 10 μ m.

data. The localization based on reverse genetics, when available, prevailed over any other approaches. Because we showed that the experimental localizations of fusion proteins were highly correlated with the localization of the molecular function when identified (Table 3), this data prevailed over the proteomics and bio-informatics ones. Additionally, PPR protein identification in organellar proteomes, though showing some discrepancies with functional data suggesting some errors of localization linked to this technique, was as far as we know more trustable than bio-informatics predictions. Finally, when no experimental data was available, we proposed a predicted localization in mitochondria or chloroplast (pM or pC). Figure 2 gives a graphical view of these results. The number of PPR proteins with a suspected or proved subcellular localization in at least one of the two organelles increased significantly with our study. For example, the experimental mitochondrial and chloroplast localization data increased by 50% (from 134 to 212) and with the addition 19 PPRs with experimental dual targeting to mitochondria and

chloroplast to the previously 10 known. Overall, 275 PPR proteins (60%) are expected to function in mitochondria, with 44% of them being validated in experimental studies. Additionally, 109 PPR proteins (24%) are expected to function in plastids, 82% being demonstrated experimentally. Forty-five PPR proteins (10%) are suspected to have a dual addressing to both plastids and mitochondria. Finally, five PPR proteins have been shown to have atypical localization: PROPR2 and PROPR3 were shown to be addressed to the nucleus,⁵⁸ PNM1 and GRP23 to both nucleus and mitochondria,^{27,28,57} and AT3G53170 was observed in both nuclear and chloroplastic extracts during proteomics studies.^{48,59} Only 24 PPR proteins (5%) do not have any clear localization based on experimental or bio-informatics reported investigations.

Discussion

RFP fusions with PPR-targeting peptides allowed us to study the subcellular localization of many members of the PPR

Table 3. Correlations between localization data sets

Data sets (number of PPR proteins with data in this set)	Fusion proteins			
	This study (126)	All data (208)	Proteomics (84)	Reverse genetics (68)
Predictions (377)	70/87 (80%)	135/159 (85%)	48/68 (71%)	44/55 (80%)
Reverse genetics (68)	17/18 (94%)	53/57 (93%)	12/15 (80%)	
Proteomics (84)	15/19 (79%)	30/36 (83%)		

In each cell of the table, the number and the percent of compatible localizations among the intersection of data available in both data sets are indicated. Two results are considered as compatible when their localizations are coherent: for example, experimental localization in both organelles and prediction or proteomics indicating only one of the two organelles.

family. Our aim in this study was to clarify the subcellular localization of 166 members of the large PPR family selected to have ambiguous localization predictions when we started the approach. In order to determine this, we used a strategy of high-throughput gateway cloning of the first 300 bp of PPR ORFs (corresponding to the N-terminal 100 amino acids of proteins) combined to a systematic microscopy investigation of the localization of transiently expressed RFP-tagged proteins. When it was determined that the first 100 amino acids displayed an interesting localization pattern, we performed in a second step a similar study using the whole ORF. Overall, with this work, we provided experimental information on the localization of 131 PPR proteins.

We have shown that 129 PPR proteins have functional targeting peptides able to address the RFP protein in one or both organelles. Seventeen have been previously published in dedicated studies and were shown to localize in agreement with our systematic results (Table 2 and 3).^{29,31,60-71} Additionally, 15 PPR proteins (HCF152 and OTP51 included) were identified in the same compartment using untargeted proteomic approaches (Table 2 and 3).^{44,59,72-76} These independent localization results largely validate our systematic strategy.

The strategy we used to study the localization of proteins can be performed at large scale to provide rapid functional information for organellar proteins. Nonetheless, some limitations have to be kept in mind when considering the results: first of all, the use of *Nicotiana benthamiana* is convenient as leaves are very comfortable to work with, but the evolution of addressing signals might be slightly different in distinct dicotyledonous species, explaining some discrepancies in the results. Second of all, the agro-infiltration to transform plant cells and generation of protoplasts to visualize expression are two steps known to generate stresses which, in some cases, may affect the conclusions. At least, the use of the very strong 2X35S promoter to trigger chimerical protein expression may overwhelm the translation and import machineries, leading to erroneous localization. However, the low number of discrepancy cases between our results and published information gained using a very large set of techniques largely validate our strategy and strengthen our results (Table 3).

Most discrepancies between our work and previous experimental localizations concern dual-localized proteins. Four of our dual-localized candidates (EMB175, AT5G14080, AT1G64100, AtC401) were previously shown in a single organelle using proteomic approaches⁷⁴ PPDB. Similarly, MEF11, and

AHG11 were functionally characterized in mitochondria editing,⁷⁷⁻⁷⁹ and AT3G62890-GFP fusion was previously observed in plastids in house SUBA3, whereas our results suggested a dual localization in both organelles for these three proteins. In contrast, three PPR proteins (AT2G37230, AT3G15130, AT5G06540) are suspected to have a dual localization because of proteomics results PPDB,^{42,44} or expression of fusion proteins (unpublished result from SUBA3), and were observed only in one of the two organelles in our study. Finally, five proteins previously observed in plastid extracts (AT1G09900, AT1G19720, AT2G28050, AT3G01580) or shown to be involved in plastid editing (AT3G14330) were observed in mitochondria in our study. Without any functional characterization, these differences cannot be definitively solved. Erroneous dual localization based on RFP-fusion localization could be explained by artifacts triggered by overexpression, whereas erroneous dual localization based on proteomics experiments could be due to sample contaminations. On the other hand, erroneous single localization might be common because of limitation in protein detection in one of the compartments during proteomics or microscopy experiments. The functional characterization of a protein in one of the two organelles does not refute the localization in the other one. Due to these experimental detection limitations, as well as the fact that we believe that dual-localized PPR proteins are mostly underestimated (see below), we have tentatively concluded that these 14 PPR proteins are localized in both organelles.

During this work, we did not observe the nuclear localization of GRP23 published by Ding and co-workers;²⁷ however, we did observe a mitochondrial localization of the TP fused to RFP, as described previously by Narsai and co-workers.⁵⁷ The GRP23 Nuclear Localization Signal, located at position 99–108, was not included in the 100 amino acid fragment used in our experiments.²⁷ Taken together, these results suggest that GRP23, as PNM1, may localize in both mitochondria and nucleus.

Addressing of PPR proteins to both organelles is underestimated. We identified 19 new PPR proteins that could have a role in both organelles. Integration of proteomic data and previous fluorescent subcellular localization studies suggest that overall at least 45 PPR proteins could be dual targeted. Recently, about 100 nuclear-encoded proteins were shown to be targeted to both mitochondria and plastids.⁸⁰ They are proposed to code for important cellular housekeeping activities. In addition, a study showed that in many cases, the dual targeting of proteins is conserved in three distant *Viridiplantae* species,⁸¹

allowing to assume that some PPR proteins could have the same dual localization in several species and probably with related functions.

Among the PPR family, five proteins were published to be dually addressed into mitochondria and plastids.^{57,58,69,82} The two orthologs, PPR2263 of maize and MITOCHONDRIAL EDITING FACTOR29 of Arabidopsis (included in our study), were shown to localize mainly in mitochondria, in which they edit *nad5* and *cob* transcripts, but also in plastids, in which their function remain to be elucidated.⁶⁹ Four other PPRs (PRORP1, OTP87, AT1G06270, AT4G21170) were not assayed in our investigation because their predicted localizations were not ambiguous according to our criteria. AT1G06270 and AT4G21170 are uncharacterized P-type PPR proteins shown as dually localized by Narsai and co-workers.⁵⁷ PROTEINACEOUS RNASE P 1 (PRORP1) was the first PPR protein shown to be dually addressed.⁵⁸ PRORP1 is an atypical PPR protein composed of 5.5 consecutive PPR repeats linked to a carboxyl-terminal (C-terminal) metallo-nuclease domain by a structural zinc-binding domain.⁸³ This protein is responsible for the nucleolitic maturation of tRNAs, an activity required in both organelles. By the use of targeting peptides fused to GFP protein, three proteins (OTP87, AT1G06270, AT4G21170) were also found in both organelles.^{57,82} OTP87 is an essential PPR protein required for RNA editing of mitochondrial *nad7* and *atp1* transcripts in *A. thaliana*. However, the depletion by an antisense strategy of OSPPR1, the ortholog of OTP87 in *O. sativa*, was described to affect the chloroplast biogenesis.⁸⁴ The predictions of localization corresponding to these five dual-localized proteins are either mitochondrial or plastidial (Table 2). Similarly, among 45 PPR proteins suspected to be localized in both organelles, eight are predicted in chloroplasts, 28 in mitochondria, and only nine do not have any predicted subcellular localization (Table 2). This suggests that many dual-targeted PPR proteins might be still unidentified. In particular, we suspect that many might be included in the 172 PPR proteins having a clear localization prediction in one of the two organelles. Moreover, although different mechanisms of dual targeting exist in the plant cell,⁸⁵ the current information does not help to hypothesize by which mechanism PPR proteins could be dual targeted, preventing the predictions of these dual localizations.

Dual targeting to mitochondria and chloroplast is an emerging class of localization in the plant cell and the PPR family seems to have an important contribution. Taking into account the functions of PPR proteins in RNA editing, RNA processing, and translation, this type of localization in the PPR family is not surprising and could be seen as a way to control or coordinate organelle RNA metabolism.^{86,87} However, this hypothesis requires testing because, until now, only one PPR protein has been shown to function in both organelles.⁵⁸ The analysis of domains in a PPR protein could help to infer its putative function. PPR proteins with dual localization seem to be present in all types of functional categories. However, among 45 dual-localized PPR proteins, 31 belong to the PPR-PLS subclass showing a probable overrepresentation of this subclass in the dual-targeted

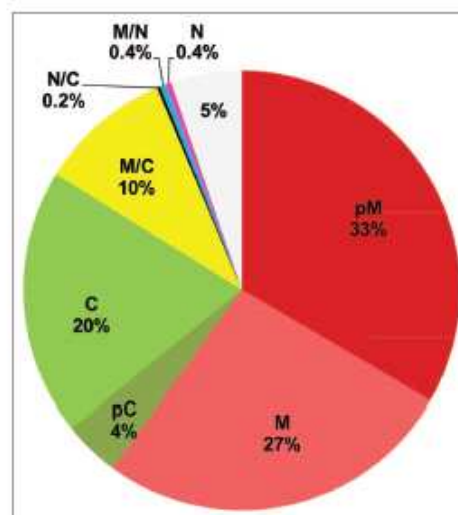


Figure 2. Distribution of the localization of *Arabidopsis thaliana* Pentatricopeptide Repeat (PPR) proteins. Classes of localization and percentage of each class in the PPR family are shown. pM, predicted mitochondria localization in dark red; M, mitochondria localization in light red; pC, predicted plastid localization in dark green; C, plastid localization in light green; M/C, mitochondria and plastid localization in yellow; N/C, nuclear and chloroplastic localization in black; M/N, mitochondria and nuclear localization in blue; N, nuclear localization in pink, unclear localization in light gray.

PPR proteins. Nevertheless, it is important to note that the localization of many PPR-P proteins (115) were not characterized yet, probably biasing this observation.

PPR proteins localized out of organelles seem to represent atypical examples in the family. Using the first 300 bp, we also identified nine PPR proteins potentially addressed out of the organelles, i.e. giving a nuclear and cytosolic localization. None were confirmed using the whole ORFs (Table 1, Fig. 1). This suggests that the number of PPR proteins being out of organelles is smaller than we thought when this work was initiated. In total, less than 1–2% of PPR proteins could function in the cytoplasm and/or the nucleus (Fig. 2). This value may be still overestimated as the model gene loci are sometimes miss-predicted, in particular, concerning the initiation codon. This may also suggest that the correct targeting sometime needs a peptide longer than the 100 amino acids we used for our work. Huang and co-workers showed that the length of mitochondrial pre-sequence varied greatly from 19–109 amino acids.⁸⁶ For GRP23, the beginning of the NLS signal has been located at the amino acid 99. Using the first 100 amino acids, we observed RFP signal into mitochondria (as previously described by Narsai and coworkers⁵⁷) whereas the full-length protein localizes in the nucleus.²⁷ This findings confirm that systematic localization using the whole proteins could give more accurate information on PPR localizations.

The case of PNM1 is even more complicated. The PNM1 nuclear localization is controlled by a NLS sequence in the C

terminus of the protein⁸² but the whole protein is addressed to mitochondria. The nuclear localization was only obtained with a truncated form of the protein without the predicted targeting peptide fused with the reporter fluorescent protein. This nuclear localization was confirmed using a specific antibody. The meaning of such a complex addressing system is still a matter of debate but suggests that a few very interesting PPR could be involved in signaling between organelles and nucleus.⁸⁶

Materials and Methods

Bioinformatic predictions and data collection. Subcellular localization prediction of the PPR proteins were performed using TargetP server (<http://www.cbs.dtu.dk/services/TargetP/>) (version 1.01 was used when we initiated this work to select the 166 PPRs and version 1.1 was used when we built **Tables 1 and 2**), Predotar v1.03 (<http://urgi.versailles.inra.fr/predotar/predotar.html>), iPSORT (<http://ipsort.hgc.jp/>), LocTree (<https://www.rostlab.org/wiki/index.php/LocTree>), Multiloc (<http://abi.inf.uni-tuebingen.de/Services/MultiLoc/>), and AtSubP (<http://bioinfo3.noble.org/AtSubP/?dowhat=About>) software using default setting. Proteomic data was recovered from published proteomic references and subcellular proteome databases: PPDB (Plant Proteome Database <http://ppdb.tc.cornell.edu/>),⁴³ SUBA3 (Subcellular location database for Arabidopsis proteins <http://suba.plantenergy.uwa.edu.au/>),⁵ and AT_CHLORO (http://www.grenoble.prabi.fr/at_chloro/).⁴²

Subcellular localization of proteins. The first 100 codons or the whole PPR ORFs were PCR amplified from *Arabidopsis thaliana* (ecotype Columbia-0) genomic DNA or cDNA using iProof DNA polymerase (Bio-Rad), specific primers (listed in **Table S1**) and a two-step amplification protocol as described previously.¹¹ PCR products were recombined into pDONR207 (Invitrogen) using Gateway® BP Clonase® II Enzyme mix (Invitrogen) as described.¹¹ For microscopic investigation, LR recombination reactions were performed using Gateway® LR Clonase® Enzyme Mix (Invitrogen) in order to transfer PPR sequences from Entry vectors to the pGREENII-derived destination vector p0229-RFP2¹¹ allowing C-terminal translational fusion with the RFP protein under the control of the 2X35S promoter. The proper ORF fusion was confirmed by sequencing using P35STL (5'-CGAATCTCAA GCAATCAAGC-3') and RFP2rev (5'-TGAACCTCGGT GATGACGTTTC-3') primers.

Binary vectors were introduced into thermo-competent *Agrobacterium tumefaciens* strain C58C1 harboring the helper plasmid pSOUP.⁸⁸ A single resistant colony was then used to inoculate 5 mL of Luria Bertani medium supplemented with 5 mg L⁻¹ Tetracycline, 50 mg L⁻¹ Kanamycine, and 2.5 mg L⁻¹ Rifampicine. This overnight pre-culture was then diluted 10 times and further grown overnight in similar conditions. After centrifugation, *Agrobacterium* cells were re-suspended in agro-infiltration buffer (10 mM MES/KOH pH 5.6, 10 mM MgCl₂, 150 µM 3',5'-Dimethoxy-4'-hydroxyacetophenone -Sigma-Aldrich-) with a final OD₆₀₀ between 0.2–0.3, and incubated at room temperature for 2 h. *Agrobacterium* suspensions were

infiltrated using 1 mL syringes without needle in leaves of *Nicotiana benthamiana*.

Protoplasts were prepared from leaf material (harvested 48–96 h after infiltration), cut into thin strips, and incubated in enzyme solution containing 4.3 g.L⁻¹ Murashige and Skoog Basal Salt Mixture (ICN Biomedicale), 0.5 g.L⁻¹ MES, 20 g.L⁻¹ sucrose, 80 g.L⁻¹ mannitol, KOH to pH 5.6, 0.4 g.L⁻¹ Pectinase from *Rhizopus* sp. (Sigma-Aldrich), 1 g.L⁻¹ Driselase® *Basidiomycetes* sp. (Sigma-Aldrich) and 2 g.L⁻¹ Cellulase Onozuka RS from *Trichoderma viride* (SERVA Electrophoresis GmbH) at 28 °C for 2–4 h.⁸⁹ Protoplasts were observed using an Eclipse TE2000S inverted microscope (Nikon) and RFP signal monitored using a custom filter block (exciter HQ546/12, emitter HQ605/75, beam-splitter Q560lp; Chroma Technology). For each construction, at least three independent agro-infiltrations were realized and each of them was observed independently by two of the authors. To confirm mitochondrial localizations, protoplasts were stained with 1 µM MitoTracker Green (Invitrogen) for 15–30 min. For confocal microscopy, proteins were visualized using a spectral Leica SP2 AOBs confocal microscope (Leica Microsystems) equipped with argon and HeNe lasers. Fluorescent signals were detected with a sequential configuration using a 488 nm laser line (MitoTracker Green: excitation/emission 488/510–530 nm) and a 543 nm laser line (RFP: excitation/emission 543/570–600 nm and chlorophyll autofluorescence: excitation/emission 543/600–700 nm). The images were coded red (RFP), green (MitoTracker Green), and blue (chlorophyll autofluorescence), giving yellow co-localization in mitochondria when green and red signals overlap in merged images and violet co-localization in plastid when blue and red signals overlap. Microscopic observations were performed using a Leica HCPL APO 633/1.20 Water Corr/0.17 Lbd.BL objective. Each image shown represents the projection of optical sections taken as a Z series.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Etienne Delannoy (Unité de Recherche en Génomique Végétale) for valuable discussions and comments on the article. We thank Olivier Grandjean (Institut Jean Pierre Bourgin), and Lionel Gissot (Institut Jean Pierre Bourgin) for their help with confocal microscopy. We thank Andeel Falcon de Longevialle (Genopole Plant Process Innovation) for providing us with beautiful healthy *Nicotiana benthamiana* plants. We thank Jennifer Yansouni for English editing. This research was supported a PhD fellowship to ML-O from the Université d'Evry Val-d'Essonne and the French Ministry of Research and Higher Education.

Supplemental Materials

Supplemental materials may be found here:
www.landesbioscience.com/journals/rnabiology/article/26128

References

1. Initiative AG; Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408:796-815; PMID:11130711; <http://dx.doi.org/10.1038/35048692>
2. Gray MW. Evolution of organellar genomes. Curr Opin Genet Dev 1999; 9:678-87; PMID:10607615; [http://dx.doi.org/10.1016/S0959-437X\(99\)00030-1](http://dx.doi.org/10.1016/S0959-437X(99)00030-1)
3. Unseld M, Maricfeld JR, Brandt P, Brennicke A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 1997; 15:57-61; PMID:8988169; <http://dx.doi.org/10.1038/ng0197-57>
4. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 1999; 6:283-90; PMID:10574454; <http://dx.doi.org/10.1093/dnares/6.5.283>
5. Tanzi SK, Castleden I, Hooper CM, Vacher M, Small I, Millar HA. SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. Nucleic Acids Res 2013; 41(Database issue):D1185-91; PMID:23406071; <http://dx.doi.org/10.1093/nar/gks1151>
6. Duncan O, Murcha MW, Whelan J. Unique components of the plant mitochondrial protein import apparatus. Biochim Biophys Acta 2013; 1833:304-13; PMID:22406071; <http://dx.doi.org/10.1016/j.bbamcr.2012.02.015>
7. Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. Uncovering the protein translocator at the chloroplast inner envelope membrane. Science 2013; 339:571-4; PMID:23372012; <http://dx.doi.org/10.1126/science.1229262>
8. Schleiff E, Becker T. Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat Rev Mol Cell Biol 2011; 12:48-59; PMID:21139638; <http://dx.doi.org/10.1038/nrm3027>
9. Binder S, Brennicke A. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. Philos Trans R Soc Lond B Biol Sci 2003; 358:181-8, discussion 188-9; PMID:12594926; <http://dx.doi.org/10.1098/rstb.2002.1179>
10. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 2008; 13:663-70; PMID:19004664; <http://dx.doi.org/10.1016/j.tplants.2008.10.001>
11. Lurin C, Andrés C, Aubourg S, Bellaoui M, Biton F, Bruyère C, Caboche M, Debaet C, Gualberto J, Hoffmann B, et al. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 2004; 16:2089-103; PMID:15269332; <http://dx.doi.org/10.1105/tpc.104.022236>
12. O'Toole N, Hattori M, Andres C, Tida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 2008; 25:1120-8; PMID:18343892; <http://dx.doi.org/10.1093/molbev/mrn057>
13. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, et al. The *Salicella* genome identifies genetic changes associated with the evolution of vascular plants. Science 2011; 332:960-3; PMID:21551031; <http://dx.doi.org/10.1126/science.1203810>
14. Fujii S, Small I. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 2011; 191:37-47; PMID:21557747; <http://dx.doi.org/10.1111/j.1469-8137.2011.03746.x>
15. de Longevialle AF, Small ID, Lurin C. Nucleary encoded splicing factors implicated in RNA splicing in higher plant organelles. Mol Plant 2010; 3:691-705; PMID:20603383; <http://dx.doi.org/10.1093/mp/psq025>
16. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 2012; 8:e1002910; PMID:22916040; <http://dx.doi.org/10.1371/journal.pgen.1002910>
17. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PLoS One 2013; 8:e57286; PMID:23472078; <http://dx.doi.org/10.1371/journal.pone.0057286>
18. Kotera E, Tasaka M, Shikama T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 2005; 435:326-30; PMID:15662426; <http://dx.doi.org/10.1038/nature03229>
19. Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. Plant Cell 2009; 21:558-67; PMID:19252080; <http://dx.doi.org/10.1105/tpc.108.064535>
20. de Longevialle AF, Meyer EH, Andrés C, Taylor NL, Lurin C, Millar AH, Small ID. The pentatricopeptide repeat gene OTTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in *Arabidopsis thaliana*. Plant Cell 2007; 19:3256-65; PMID:17965268; <http://dx.doi.org/10.1105/tpc.107.054841>
21. de Longevialle AF, Hendrickson L, Taylor NL, Delannoy E, Lurin C, Badger M, Millar AH, Small I. The pentatricopeptide repeat gene OTTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in *Arabidopsis thaliana*. Plant J 2008; 56:157-68; PMID:18557832; <http://dx.doi.org/10.1111/j.1365-3113X.2008.03581.x>
22. Schmitz-Linneweber C, Williams-Carter R, Barkan A. 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 2005; 17:2791-804; PMID:16141451; <http://dx.doi.org/10.1105/tpc.105.034454>
23. Uytendaele M, Mireau H, Rurek M, Hammani K, Arnal N, Quadrado M, Giegé P. PPR336 is associated with polyosomes in plant mitochondria. J Mol Biol 2008; 375:626-36; PMID:18054044; <http://dx.doi.org/10.1016/j.jmb.2007.11.011>
24. Lloyd J, Meinke D. A comprehensive dataset of genes with a loss-of-function mutant phenotype in *Arabidopsis*. Plant Physiol 2012; 158:1115-29; PMID:22247268; <http://dx.doi.org/10.1104/pp.111.192393>
25. Muralla R, Lloyd J, Meinke D. Molecular foundations of reproductive lethality in *Arabidopsis thaliana*. PLoS One 2011; 6:e28398; PMID:22164284; <http://dx.doi.org/10.1371/journal.pone.0028398>
26. Gutmann B, Gobert A, Giegé P. PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*. Genes Dev 2012; 26:1022-7; PMID:22549728; <http://dx.doi.org/10.1101/gad.189514.112>
27. Ding YH, Liu NY, Tang ZS, Liu J, Yang WC. *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 2006; 18:815-30; PMID:16489121; <http://dx.doi.org/10.1105/tpc.105.039495>
28. Hammani K, Gobert A, Hleibieh K, Choulier L, Small I, Giegé P. An *Arabidopsis* dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. Plant Cell 2011; 23:730-40; PMID:21297037; <http://dx.doi.org/10.1105/tpc.110.081638>
29. De Renzi S, Elemento O, Tavano S, Wieschaus EF. Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. PLoS Biol 2007; 5:e117; PMID:17456005; <http://dx.doi.org/10.1371/journal.pbio.0050117>
30. Bratic A, Wredenberg A, Grönke S, Stewart JB, Mourier A, Ruzzenente B, Kukat C, Wibom R, Habermann B, Partridge L, et al. The bicoid stability factor controls polyadenylation and expression of specific mitochondrial mRNAs in *Drosophila* melanogaster. PLoS Genet 2011; 7:e1002324; PMID:22022283; <http://dx.doi.org/10.1371/journal.pgen.1002324>
31. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2007; 2:953-71; PMID:17446895; <http://dx.doi.org/10.1038/nprot.2007.131>
32. Small I, Peeters N, Legasi F, Lurin C, Predotar A. A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 2004; 4:1581-90; PMID:15174128; <http://dx.doi.org/10.1002/pmic.200300776>
33. Gavel Y, Nilsson L, von Heijne G. Mitochondrial targeting sequences. Why 'non-amphiphilic' peptides may still be amphiphilic. FEBS Lett 1988; 235:173-7; PMID:3402595; [http://dx.doi.org/10.1016/0014-5793\(88\)81257-2](http://dx.doi.org/10.1016/0014-5793(88)81257-2)
34. Roise D. Recognition and binding of mitochondrial presequences during the import of proteins into mitochondria. J Bioenerg Biomembr 1997; 29:19-27; PMID:9067798; <http://dx.doi.org/10.1023/A:1022403604273>
35. Walther M, Hammen PK, Weiner H. Influence of the mature portion of a precursor protein on the mitochondrial signal sequence. J Biol Chem 1996; 271:21226-30; PMID:8702895; <http://dx.doi.org/10.1074/jbc.271.35.21226>
36. Huang S, Taylor NL, Whelan J, Millar AH. Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiol 2009; 150:1272-85; PMID:19474214; <http://dx.doi.org/10.1104/pp.109.137885>
37. Bionda T, Tillmann B, Simm S, Beilstein K, Ruprecht M, Schleiff E. Chloroplast import signals: the length requirement for translocation in vitro and in vivo. J Mol Biol 2010; 402:510-23; PMID:20688079; <http://dx.doi.org/10.1016/j.jmb.2010.07.052>
38. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 2002; 18:298-305; PMID:11847077; <http://dx.doi.org/10.1093/bioinformatics/18.2.298>
39. Nair R, Rost B. Mimicking cellular sorting improves prediction of subcellular localization. J Mol Biol 2005; 348:85-100; PMID:15808855; <http://dx.doi.org/10.1016/j.jmb.2005.02.025>
40. Höglund A, Dönnies P, Blum T, Adolph HW, Kohlhauser O. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. Bioinformatics 2006; 22:158-65; PMID:16428265; <http://dx.doi.org/10.1093/bioinformatics/btl002>
41. Kaundal R, Saini R, Zhao PX. Combining machine learning and homology-based approaches to accurately predict subcellular localization in *Arabidopsis*. Plant Physiol 2010; 154:36-54; PMID:20647376; <http://dx.doi.org/10.1104/pp.110.156851>
42. Ferro M, Brugière S, Salvi D, Seigneurin-Berny D, Court M, Moyer L, Ramus C, Miras S, Mellal M, Le Gall S, et al. AT_CHLORO, a comprehensive chloroplast proteome database with subplastid localization and curated information on envelope proteins. Mol Cell Proteomics 2010; 9:1063-84; PMID:20061580; <http://dx.doi.org/10.1074/mcp.M900325-MCP200>

43. Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD, van Wijk KJ. PPD8, the Plant Proteomics Database at Cornell. *Nucleic Acids Res* 2009; 37(Database issue):D969-74; PMID:18832363; <http://dx.doi.org/10.1093/nar/gkn654>
44. Huang S, Taylor NL, Narzai R, Eubel H, Whelan J, Millar AH. Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. *Plant Physiol* 2009; 149:719-34; PMID:19010998; <http://dx.doi.org/10.1104/pp.108.131300>
45. Huang M, Friso G, Nishimura K, Qu X, Olinares PD, Majeran W, Sun Q, van Wijk KJ. Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships: the concept of orthoproteomics. *J Proteome Res* 2013; 12:491-504; PMID:23198870; <http://dx.doi.org/10.1021/pr300952g>
46. Hummel M, Cordewener JH, de Groot JC, Smeekens S, America AH, Hanson J. Dynamic protein composition of Arabidopsis thaliana cytosolic ribosomes in response to sucrose feeding as revealed by label free MSE proteomics. *Proteomics* 2012; 12:1024-38; PMID:22522809; <http://dx.doi.org/10.1002/pmic.201100413>
47. Ito J, Bath TS, Petzold CJ, Redding-Johanson AM, Mukhopadhyay A, Verboom R, Meyer EH, Millar AH, Heazlewood JL. Analysis of the Arabidopsis cytosolic proteome highlights subcellular partitioning of central plant metabolism. *J Proteome Res* 2011; 10:1571-82; PMID:21166475; <http://dx.doi.org/10.1021/pr1009433>
48. Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JW, Shaw PJ. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol Biol Cell* 2005; 16:260-9; PMID:15496452; <http://dx.doi.org/10.1091/mbc.E04-09-0791>
49. Sakamoto Y, Takagi S. LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana. *Plant Cell Physiol* 2013; 54:622-33; PMID:23396599; <http://dx.doi.org/10.1093/pcp/pcr031>
50. Alexanderson E, Saalbach G, Larsson C, Kjellbom P. Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol* 2004; 45:1543-56; PMID:15574830; <http://dx.doi.org/10.1093/pcp/pcp209>
51. Jaquinod M, Villiers F, Kieffer-Jaquinod S, Hugouvieux V, Bruley C, Garin J, Bourguignon J. A proteomics dissection of Arabidopsis thaliana vacuoles isolated from cell culture. *Mol Cell Proteomics* 2007; 6:394-412; PMID:17151019; <http://dx.doi.org/10.1074/mcp.M600250-MCP200>
52. Keinath NF, Kierszniowska S, Lorek J, Bourdais G, Kessler SA, Shimosato-Asano H, Grossniklaus U, Schulze WX, Robatzek S, Panstruga R. PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. *J Biol Chem* 2010; 285:39140-9; PMID:20843791; <http://dx.doi.org/10.1074/jbc.M110.160531>
53. Li B, Takahashi D, Kawamura Y, Uemura M. Comparison of plasma membrane proteomic changes of Arabidopsis suspension-cultured cells (T87 Line) after cold and ABA treatment in association with freezing tolerance development. *Plant Cell Physiol* 2012; 53:543-54; PMID:22318864; <http://dx.doi.org/10.1093/pcp/pcr010>
54. Mitra SK, Walters BT, Clouse SD, Goshe MB. An efficient organic solvent based extraction method for the proteomic analysis of Arabidopsis plasma membranes. *J Proteome Res* 2009; 8:2752-67; PMID:19334764; <http://dx.doi.org/10.1021/pr801044y>
55. Szponarski W, Sommerer N, Boyer JC, Rossignol M, Gilbrat R. Large-scale characterization of integral proteins from Arabidopsis vacuolar membrane by two-dimensional liquid chromatography. *Proteomics* 2004; 4:397-406; PMID:14760709; <http://dx.doi.org/10.1002/pmic.200300607>
56. Zhang ZJ, Peck SC. Simplified enrichment of plasma membrane proteins for proteomic analyses in Arabidopsis thaliana. *Proteomics* 2011; 11:1780-8; PMID:21433285; <http://dx.doi.org/10.1002/pmic.201000648>
57. Narzai R, Law SR, Carrie C, Xu L, Whelan J. In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in Arabidopsis. *Plant Physiol* 2011; 157:1342-62; PMID:21908688; <http://dx.doi.org/10.1104/pp.111.183129>
58. Gobet A, Gutmann B, Taschner A, Gösinger M, Holzmann J, Hartmann RK, Rossmann W, Giegé P. A single Arabidopsis organelle protein has RNase P activity. *Nat Struct Mol Biol* 2010; 17:740-4; PMID:20473316; <http://dx.doi.org/10.1038/nmsb.1812>
59. Majeran W, Friso G, Anzura Y, Qu X, Huang M, Ponnala L, Watkins KP, Barkan A, van Wijk KJ. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. *Plant Physiol* 2012; 158:156-89; PMID:22065420; <http://dx.doi.org/10.1104/pp.111.188474>
60. Hashimoto M, Endo T, Peltier G, Tasaka M, Shikanai T. A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. *Plant J* 2003; 36:541-9; PMID:14617084; <http://dx.doi.org/10.1046/j.1365-3113.2003.01900.x>
61. Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-psbI-psbD* RNAs. *Plant Cell* 2003; 15:1480-95; PMID:12782738; <http://dx.doi.org/10.1105/tpc.010397>
62. Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast *psl2* pre-mRNA. *Plant Cell* 2006; 18:2650-63; PMID:17041147; <http://dx.doi.org/10.1105/tpc.106.046110>
63. Gutiérrez-Marcos JF, Dal Prà M, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Taitou C, Paul W, Perez P, et al. empty pericarp4 encodes a mitochondrial-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. *Plant Cell* 2007; 19:196-210; PMID:17259266; <http://dx.doi.org/10.1105/tpc.105.039594>
64. Cai W, Ji D, Peng L, Guo J, Ma J, Zou M, Lu C, Zhang L. LPA66 is required for editing *psbF* chloroplast transcripts in Arabidopsis. *Plant Physiol* 2009; 150:1260-71; PMID:19448041; <http://dx.doi.org/10.1104/pp.109.136812>
65. Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinzaki K, Small I, Shikanai T. Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* 2009; 21:146-56; PMID:19182104; <http://dx.doi.org/10.1105/tpc.106.046667>
66. Laluk K, Abuqamar S, Mengiste T. The Arabidopsis mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiol* 2011; 156:2053-68; PMID:21653783; <http://dx.doi.org/10.1104/pp.111.177501>
67. Takenaka M. MEP9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the *nad7* transcript in mitochondria of Arabidopsis. *Plant Physiol* 2010; 152:939-47; PMID:20018598; <http://dx.doi.org/10.1104/pp.109.151175>
68. Hammani K, Okuda K, Tanz SK, Chateigner-Boutin AL, Shikanai T, Small I. A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 2009; 21:3686-99; PMID:19934379; <http://dx.doi.org/10.1105/tpc.109.071472>
69. Soso D, Mbalo S, Vernoux V, Gendrot G, Dedieu A, Chambrier P, Dauzat M, Heurtevin L, Guyon V, Takenaka M, et al. PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial *nad5* and *cob* transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* 2012; 24:676-91; PMID:22319053; <http://dx.doi.org/10.1105/tpc.111.091074>
70. Zhu Q, Dugardeyn J, Zhang C, Takenaka M, Kihm K, Craddock C, Smalle J, Karampelis M, Denesek J, Peters J, et al. SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J* 2012; 71:836-49; PMID:22540321; <http://dx.doi.org/10.1111/j.1365-3113.2012.05036.x>
71. Chateigner-Boutin AL, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK, Small I. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J* 2013; 74:935-45; PMID:23251509; <http://dx.doi.org/10.1111/tpj.12180>
72. Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 2004; 16:241-56; PMID:14671022; <http://dx.doi.org/10.1105/tpc.016055>
73. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjölander K, Gruissem W, Baginsky S. The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Curr Biol* 2004; 14:354-62; PMID:15028209; <http://dx.doi.org/10.1016/j.cub.2004.02.039>
74. Klodmann J, Senkler M, Rode C, Braun HP. Defining the protein complex proteome of plant mitochondria. *Plant Physiol* 2011; 157:587-98; PMID:21841088; <http://dx.doi.org/10.1104/pp.111.182352>
75. Peltier JB, Ytterberg AJ, Sun Q, van Wijk KJ. New functions of the thylakoid membrane proteome of Arabidopsis thaliana revealed by a simple, fast, and versatile fractionation strategy. *J Biol Chem* 2004; 279:49367-83; PMID:15322131; <http://dx.doi.org/10.1074/jbc.M406763200>
76. Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* 2008; 3:e1994; PMID:18431481; <http://dx.doi.org/10.1371/journal.pone.0001994>
77. Murayama M, Hayashi S, Nishimura N, Ishide M, Kobayashi K, Yagi Y, Asami T, Nakamura T, Shinzaki K, Hirayama T. Isolation of Arabidopsis *ahp11*, a weak ABA hypersensitive mutant defective in *nad4* RNA editing. *J Exp Bot* 2012; 63:5301-10; PMID:22821940; <http://dx.doi.org/10.1093/jxb/err188>
78. Tang J, Kobayashi K, Suzuki M, Matsumoto S, Muranaka T. The mitochondrial PPR protein LOVASTATIN INSENSITIVE 1 plays regulatory roles in cytosolic and plastidial isoprenoid biosynthesis through RNA editing. *Plant J* 2010; 61:456-66; PMID:19929879; <http://dx.doi.org/10.1111/j.1365-3113.2009.04082.x>
79. Verbitakiy D, Zehrmann A, van der Merwe JA, Brennicke A, Takenaka M. 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* 2010; 61:446-55; PMID:19919573; <http://dx.doi.org/10.1111/j.1365-3113.2009.04076.x>
80. Carrie C, Small I. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim Biophys Acta* 2013; 1833:253-9; PMID:22683762; <http://dx.doi.org/10.1016/j.bbamcr.2012.05.029>
81. Xu L, Carrie C, Law SR, Murcha MW, Whelan J. Acquisition, conservation, and loss of dual-targeted proteins in land plants. *Plant Physiol* 2013; 161:644-62; PMID:23257241; <http://dx.doi.org/10.1104/pp.112.210997>

82. Hammani K, des Franco-Small CC, Takenaka M, Taniz SK, Okada K, Shikanai T, Brennicke A, Small I. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in *Arabidopsis* mitochondria. *J Biol Chem* 2011; 286:21361-71; PMID:21504904; <http://dx.doi.org/10.1074/jbc.M111.230516>
83. Howard MJ, Lim WH, Fierke CA, Koutmos M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A* 2012; 109:16149-54; PMID:22991464; <http://dx.doi.org/10.1073/pnas.1209062109>
84. Gothandam KM, Kim ES, Cho H, Chung YY. OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. *Plant Mol Biol* 2005; 58:421-33; PMID:16021404; <http://dx.doi.org/10.1007/s11103-005-5702-5>
85. Yogeve O, Pines O. Dual targeting of mitochondrial proteins: mechanism, regulation and function. *Biochim Biophys Acta* 2011; 1808:1012-20; PMID:20637721; <http://dx.doi.org/10.1016/j.bbame.2010.07.004>
86. Duchêne AM, Giegé P. Dual localized mitochondrial and nuclear proteins as gene expression regulators in plants? *Front Plant Sci* 2012; 3:221; PMID:23056004; <http://dx.doi.org/10.3389/fpls.2012.00221>
87. Krause K, Krupinska K. Nuclear regulators with a second home in organelles. *Trends Plant Sci* 2009; 14:194-9; PMID:19285907; <http://dx.doi.org/10.1016/j.tplants.2009.01.005>
88. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* 2000; 42:819-32; PMID:10890530; <http://dx.doi.org/10.1023/A:1006496308160>
89. Chupeau Y, Bourgin JP, Missonier C, Dorion N, Morel G. Preparation and culture of protoplasts from various *Nicotiana* species. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* 1974; D 278:1565-8

4.2. La localisation subcellulaire des protéines PPR : Actualisation de données

Les résultats globaux de l'approche d'intégration des données de localisation montrent qu'actuellement des données expérimentales permettant une assignation de localisation sont disponibles pour plus de 55% des protéines de la famille PPR (Figure 16). De plus, parmi les 206 protéines pour lesquelles aucune donnée expérimentale n'est disponible, il est possible de proposer une localisation pour 175 d'entre elles (37% de la famille) grâce à des logiciels de prédictions. Ainsi, l'ensemble de ces données montre qu'il est possible d'assigner une localisation à environ 93% des protéines PPR d'*A. thaliana*. Finalement, seulement 31 des 473 (environ 6,3%) protéines PPR ne sont associées à aucune localisation expérimentale ou prédictive (Figure 16). En ce qui concerne la double localisation de protéines PPR, l'intégration des données a permis de proposer que près de 10 % des protéines PPR auraient une localisation à la fois dans la mitochondrie et le chloroplaste, alors que 2 protéines ont été identifiées dans le noyau et les mitochondries et une pourrait être localisée dans le noyau et les chloroplastes (Figure 16).

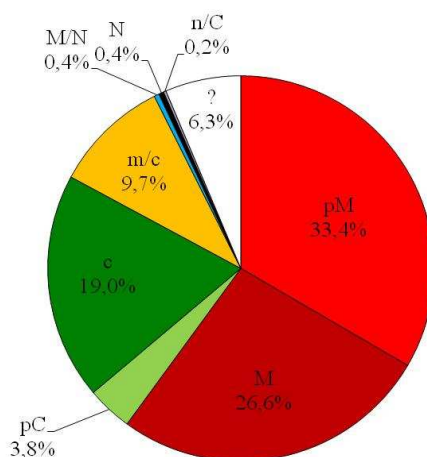


Figure 16. Distribution actualisé de la localisation des protéines de la famille PPR.

Les catégories de localisation et le pourcentage sont indiqués. pM: protéines prédites adressées aux mitochondries en rouge, M: protéines dont l'adressage aux mitochondries a été validé expérimentalement en rouge foncé, pC: protéines prédites adressées aux plastes en vert clair, C: protéines dont l'adressage aux plastes a été validé expérimentalement en vert foncé, m/c; protéines dont l'adressage aux mitochondries et aux plastes a été observé expérimentalement en orange, M/N: adressage des protéines aux mitochondrie et au noyau en bleu, N: protéine localisée dans le noyau en noir, n/C: protéines adressées au noyau et aux plastes en violet. La couleur blanche concerne la catégorie sans indication de localisation (Figure actualisée et adaptée de Colcombet *et al.*, 2013).

Il est important de noter que certaines protéines (6), incluant la protéine OTP71 récemment étudiée (Chateigner-Boutin *et al.*, 2013), présentent une double localisation lorsque l'on fusionne leur signal d'adressage N-terminal à une protéine fluorescente mais perdent cette double localisation lorsque la séquence pleine longueur est utilisée. Ces protéines ont été classées dans la catégorie correspondant à la localisation observée avec la fusion pleine longueur.

4.3. Analyse de la double localisation des protéines PPR.

Selon notre étude, 46 protéines PPR seraient potentiellement doublement localisées dans les mitochondries et les plastides. Parmi les 46 protéines PPR ayant une potentielle double localisation, seule une protéine a une activité caractérisée dans les deux organites. En effet, PRORP1 est nécessaire à la maturation des ARNt dans les mitochondries et les chloroplastes. Par ailleurs, 9 protéines ont été associées à l'édition chez l'un des deux organites (MEF11, MEF20, OTP87, PPR596, AHG11, CREF3, MPR25, LPA66 et MEF29) et 2 sont impliquées dans l'épissage mitochondrial (BIR6 et ABO5). Enfin, 34 des 46 protéines ayant une potentielle double localisation n'ont pas été caractérisées fonctionnellement. Parmi les 18 PPR doublement localisées étudiées en génétique inverse dans des publications, 4 (dont PRORP1) ont été montrées essentielles pour le développement de l'embryon chez *A. thaliana* (Cushing *et al.*, 2005, Gobert *et al.*, 2010, Savage *et al.*, 2013). La double localisation chez les protéines PPR est une caractéristique observée dans tous les types de sous-familles et sous groupes composant la famille (Figure 17). Ainsi, parmi les 46 protéines PPR ayant une double localisation dans les 2 organites, un tiers appartient à la sous-famille P et deux tiers à la sous-famille PLS. Cet enrichissement de la double localisation dans la sous-famille PLS doit être pris avec précaution. En effet, ces chiffres sont probablement biaisés du fait qu'il reste très difficile de prédire la double localisation d'une protéine actuellement et que la localisation de nombreuses protéines PPR de la sous-famille P (139) n'a pas encore été évaluée expérimentalement.

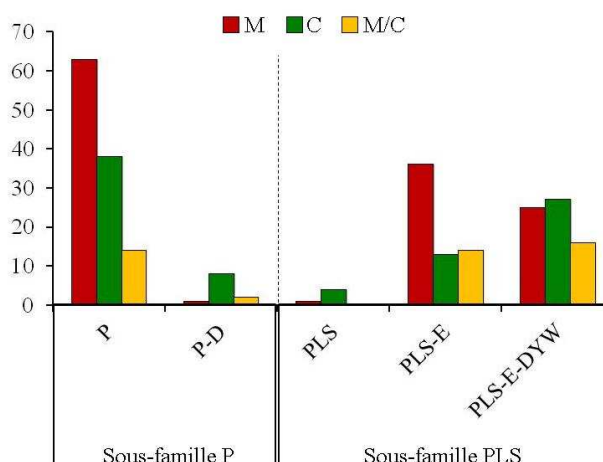


Figure 17. Distribution de la localisation des protéines PPR au sein des sous-familles (P et PLS) et des sous-groupes (P ; P-D ; PLS ; PLS-E ; PLS-E-DYW).

Seuls les résultats expérimentaux d'adressage des protéines PPR aux mitochondries (en rouge), aux plastes (en vert) et de la double localisation des protéines dans la mitochondrie et le plaste (en orange) ont été utilisés pour cette analyse.

Parmi les 46 protéines PPR ayant une potentielle double localisation, quatre groupes peuvent être distingués en fonction du type d'évidence expérimentale supportant la double localisation (Tableau 3). Le premier groupe contient 10 protéines pour lesquelles des données divergentes de localisation ont été obtenues selon les approches expérimentales utilisées. Un deuxième groupe rassemble 13 protéines de fonction inconnue pour lesquelles une fusion traductionnelle du signal d'adressage N-terminal avec une protéine fluorescente a permis d'observer une double localisation. Un troisième groupe comporte 7 protéines de fonction inconnue et 6 protéines, AHG11, BIR6, MEF11, MEF29, OTP87 et OTP90, ayant une fonction assignée dans la mitochondrie. Dans ce groupe, la double localisation est proposée sur la base de données expérimentales ayant permis d'observer la localisation de fusions traductionnelles composées des protéines pleine longueur et d'une protéine fluorescente (résultats complétés pour certaines protéines par les fusions avec les signaux d'adressage). Enfin, un quatrième groupe contient 10 protéines pour lesquelles différentes approches expérimentales (notamment microscopie en fluorescence et analyses protéomiques) apportent des résultats complémentaires suggérant une probable double localisation. Pour la plupart des protéines, les données fonctionnelles et de localisation disponibles concernent néanmoins majoritairement un seul organite (Tableau 3).

Comme détaillé précédemment, les protéines appartenant au premier groupe présentent de fortes divergences entre les différentes données de localisation les concernant.

Ainsi, lors d'expériences de microscopie en fluorescence utilisant la fusion du signal d'adressage avec des protéines fluorescentes, la plupart de ces protéines montrent une localisation mitochondriale. En revanche, des analyses protéomiques ont identifié ces mêmes protéines dans des échantillons chloroplastiques. Cinq protéines de ce groupe, ABO5, CREF3, MEF20, MPR25 et LPA66, ont par ailleurs fait l'objet d'études fonctionnelles les associant à un seul organite (Cai *et al.*, 2009, Takenaka *et al.*, 2010, Liu *et al.*, 2010, Toda *et al.*, 2012 Yagi *et al.*, 2013.). Néanmoins, il faut noter qu'une fonction n'a été recherchée dans le second organite que dans le cas de MPR25 (Toda *et al.*, 2012). Sans données complémentaires, il n'est pas possible de conclure quant à ces divergences de localisation. C'est d'ailleurs pour cette raison que ces protéines ont été retenues comme ayant une potentielle double localisation. Certaines d'entre elles sont probablement de faux positifs liés aux technologies utilisées, comme par exemple lors de la purification des organites utilisés pour l'analyse du protéome, technique pouvant facilement générer des contaminations.

Il est donc évident que la probabilité de double localisation des protéines du premier groupe et, dans une moindre mesure, de celles du deuxième groupe, doit être considérée comme relativement faible. Cette localisation devra être confirmée par des approches complémentaires de localisation ou par des études fonctionnelles. En revanche, les données de localisation concernant les protéines des troisième et quatrième groupes sont plus solides et la probabilité qu'elles soient doublement localisées est plus importante. En particulier, dans le groupe 4, en plus de la protéine PRORP1 qui est caractérisée fonctionnellement dans les deux organites (Gobert *et al.*, 2010), les données concernant les protéines AT2G15690 (DYW2) et AT3G4920 (OTP100) indiquent que ces deux protéines sont très certainement localisées dans la mitochondrie et le plaste bien que leurs fonctions n'y soient pas connues. Ces protéines ont été choisies pour des analyses fonctionnelles qui seront décrites dans une partie du chapitre V de cette thèse.

Tableau 3. Protéines de la famille PPR d'*Arabidopsis thaliana* ayant une potentielle double localisation dans la mitochondrie et dans le plaste

AGI ¹	Annotation ²	Fluorescence ³		Localisation			Références ⁷
		SAN	SPL	Protéomique ⁴	Fonction moléculaire ⁵	Conclusion ⁶	
Groupe 1							
<i>At1g09900</i>	PPR	M ^a		C (Zm ^b)		m/c	^a Colcombet et al 2013, ^b PPDB
<i>At1g19720</i>	PPR	M ^a		C (At ^{b,c} , Zm ^b)		m/C	^a Colcombet et al 2013, ^b PPDB, ^c Kong et al 2011,
<i>At1g51965</i>	ABO5	M ^a		C (At ^b)	Epissage (At-M ^a)	M/c	^a Liu et al 2010, ^b AT_Chloro,
<i>At2g19280</i>	PPR	M ^a		C (At ^b)		m/c	^a Lurin et al 2004, ^b PPDB,
<i>At2g28050</i>	PPR	M ^a		C (At ^b)		m/c	^a Colcombet et al 2013, ^b AT_Chloro,
<i>At3g01580</i>	PPR	M ^a		C (At ^b)		m/c	^a Colcombet et al 2013, ^b Kong et al 2011,
<i>At3g14330</i>	CREF3	M ^a			Edition (At-C ^b)	m/C	^a Colcombet et al 2013, ^b Yagi et al 2013
<i>At3g18970</i>	MEF20			C (At ^a)	Edition (At-M ^b)	M/c	^a AT_Chloro, ^b Takenaka et al 2010
<i>At3g22150</i>	MPR25	C ^a	M ^b	C (At ^c)	Edition (Os-M ^b)	M/c	^a SUBA3, ^b Toda et al 2012, ^c AT_Chloro
<i>At5g48910</i>	LPA66	M ^a , C ^{b,c}			Edition (At-C ^c)	m/C	^a Colcombet et al 2013, ^b SUBA3, ^c Cai et al 2009
Groupe 2							
<i>At1g02420</i>	PPR	M/C ^a				m/c	^a Colcombet et al 2013
<i>At2g34400</i>	PPR	M/C ^a				m/c	^a Colcombet et al 2013
<i>At2g41080</i>	PPR	M/C ^a				m/c	^a Colcombet et al 2013
<i>At3g02330</i>	PPR	M/C ^a				m/c	^a SUBA3
<i>At3g15130</i>	PPR	M/C ^{a,b}				m/c	^a Colcombet et al 2013, ^b SUBA3
<i>At3g21470</i>	PPR	M/C ^a				m/c	^a Colcombet et al 2013
<i>At3g23330</i>	PPR	M/C ^a				m/c	^a Colcombet et al 2013
<i>At4g04370</i>	PPR	M/C ^{a, b}				m/c	^a Colcombet et al 2013, ^b SUBA3
<i>At4g21170</i>	PPR	M/C ^a				m/c	^a Narsai et al 2011
<i>At4g33170</i>	PPR	M/C ^{a, b}				m/c	^a Colcombet et al 2013 ^b SUBA3
<i>At5g06540</i>	PPR	M/C ^{a, b}				m/c	^a Colcombet et al 2013, ^b SUBA3
<i>At5g27110</i>	PPR	M/C ^a				m/c	^a SUBA3
<i>At5g43790</i>	PPR	M/C ^a				m/c	^a Colcombet et al 2013

1. Code AGI (Arabidopsis genome initiative) du locus. 2. Annotation du gène. PPR: PENTATRICOPEPTIDE REPAT, ABO: ABA OVERLAY SENSITIVE, CREF: CHLOROPLAST RNA EDITING FACTOR, MEF: MITOCHONDRIAL RNA EDITING FACTOR, MPR: MITOCHONDRIAL PPR, LPA: LOW PSII ACCUMULATION, EMB: EMBRYO DEFECTIVE, OTP: ORGANELLAR TRANSCRIPT PROCESSING AHG: ABA HYPER-SENSITIVE AT GERMINATION, BIR: BSO-INSENSITIVE-ROOTS. Dans les colonnes concernant les données de localisation M: Mitochondrie C: Plaste, M/C; Mitochondrie et plaste. 3. Les expériences en fluorescence par fusions traductionnelles ont été divisées en deux catégories : celles utilisant le signal d'adressage N-terminal (SAN) et celles utilisant la séquence codant pour la protéine pleine longueur (SPL). Les indices correspondent aux références de chaque donnée. 4. Les données des analyses protéomiques ont été recueillies à partir des bases de données PPDB, SUBA3, AT-CHLORO et d'articles dont la référence en colonne 7 est précisée par l'indice. L'espèce chez laquelle la protéine a été identifiée est indiquée comme suit : At: *Arabidopsis thaliana*, Zm: *Zea mays*, Os: *Oriza sativa*. 5. La fonction moléculaire est indiquée selon les données de la littérature. La référence des travaux correspondant en colonne 7 est indiquée par l'indice ; L'espèce et l'organite chez lesquels la fonction a été déterminée sont indiqués entre parenthèses. 6. La conclusion de localisation correspond à une synthèse des différentes données. Une majuscule indique que la localisation a été observée dans au moins deux approches expérimentales différentes et une minuscule correspond à une localisation nécessitant d'être confirmée par une deuxième approche. 7. Les références bibliographiques correspondants à chaque donnée.*Protéines étudiées dans ce travail de thèse (analyses fonctionnelles)

Tableau 3. (suite) Protéines de la famille PPR d'*Arabidopsis thaliana* ayant une potentielle double localisation dans la mitochondrie et dans le plaste

		Localisation					
AGI ¹	Annotation ²	Fluorescence ³		Protéomique ⁴	Fonction moléculaire ⁵	Conclusion ⁶	Références ⁷
		SAN	SPL				
Groupe 3							
<i>At1g05670</i>	PPR		M/C ^a			m/c	^a Colcombet et al 2013
<i>At1g06145</i>	OTP91*		M/C ^a			m/c	^a Colcombet et al 2013
<i>At1g25360</i>	OTP90*	M/C ^a	M/C ^a		Edition (At-M ^b)	M/c	^a Colcombet et al 2013, ^b Cette thèse
<i>At1g74600</i>	OTP87	M/C ^a			Edition (At-M ^a)	M/c	^a Hammani et al 2011
<i>At2g44880</i>	AHG11		M/C ^a		Edition (At-M ^b)	M/c	^a Colcombet et al 2013, ^b Murayama et al 2012,
<i>At3g49710</i>	PPR		M/C ^a			m/c	^a Colcombet et al 2013
<i>At3g47530</i>	PPR	M/C ^a	M/C ^a			m/c	^a Colcombet et al 2013
<i>At3g48250</i>	BIR6	M ^a	M/C ^b		Epissage (At-M ^a)	M/c	^a Koprivova et al 2010, ^b Fuss et al 2013
<i>At3g50420</i>	PPR	M/C ^a	M/C ^a			m/c	^a Colcombet et al 2013
<i>At3g62890</i>	PPR	C ^a	M/C ^b			m/c	^a SUBA3, ^b Colcombet et al 2013
<i>At4g14850</i>	MEF11	M/C ^a			Edition (At-M ^b)	M/c	^a Colcombet et al 2013, ^b Verbitskiy et al 2010
<i>At4g30700</i>	MEF29	M/C ^a	M/C ^{a,b}		Edition (At-M ^b)	M/c	^a Colcombet et al 2013, ^b Sosso et al 2012
<i>At5g47460</i>	PPR	M/C ^a	M/C ^a			m/c	^a Colcombet et al 2013
Groupe 4							
<i>At1g64100</i>	PPR	M/C ^a		M (At ^a)		M/c	^a Colcombet et al 2013, ^b Klodmann et al 2011
<i>At1g80270</i>	PPR_596	M ^a		M(At ^b , Os ^c) C(At ^d)		M/c	^a Narsai et al 2011 ^b Klodmann et al 2011, ^c Huang et al 2009, ^d Froehlich et al 2003
<i>At2g15690</i>	DYW2*	M/C ^{a,b}	M/C ^a	M(Os ^c) C(Zm ^d)	Edition (At-M/Ca)	M/C	^a Cette thèse, ^b SUBA3, ^b Huang et al 2009, ^c PPDB,
<i>At2g32230</i>	PRORP1		M/C ^a	C (Zm ^b)	Clivage tRNA (At-M/C ^a)	M/C	^a Gobert et al 2010, ^b PPDB,
<i>At2g37230</i>	PPR*	M ^a	M ^b	C(At ^{c,d}) M(Os ^e)		M/c	^a Colcombet et al 2013, ^b Cette thèse, ^c AT_Chloro, ^d PPDB, ^e Huang et al 2009
<i>At3g42630</i>	PPR*	C ^a	M/C ^b	C(At ^c , Zm ^c)		m/C	^a Colcombet et al 2013, ^b Cette thèse ^c PPDB,
<i>At3g49240</i>	OTP100*	M/C ^a	M/C ^a	M(At ^{b,c}) C(At ^d , Zm ^d)		M/C	^a Colcombet et al 2013, ^b Ito et al 2006, ^c Klodmann 2011, ^d PPDB,
<i>At5g03800</i>	EMB175	M/C ^{a,b}		C(Zm ^c)		m/C	^a Colcombet et al 2013, ^b SUBA3, ^c PPDB,
<i>At5g14080</i>	PPR	M/C ^a		C(At ^b)		m/C	^a Colcombet et al 2013, ^b PPDB,
<i>At5g21222</i>	AtC401	M/C ^a		C(At ^b)		m/C	^a Colcombet et al 2013, ^b PPDB,

4.3.1. La double localisation et l'évolution des protéines PPR

Comme indiqué dans l'introduction (cf. section 3.4.2.), environ 45% (215) des gènes PPR ont un comportement évolutif caractéristique de gènes présents en copie unique dans le génome chez les angiospermes. Pour les autres, ils peuvent avoir un ou plusieurs paralogues dans une ou plusieurs espèces. D'autre part, les conservations d'orthologies sont moins importantes quand on compare la famille PPR des angiospermes avec celle d'autres clades comme par exemple *P. patens* (O'Toole *et al.*, 2008, Sugita *et al.*, 2013) ou *S. moellendorffii* (Banks *et al.*, 2012). Ces dernières disparités sont probablement associées à des différences dans le processus moléculaires impliqués dans le métabolisme des ARN des organites. Lorsque l'on regarde la répartition des différentes catégories de localisation des protéines PPR en fonction de leur comportement évolutif (relations d'orthologies simples ou existence de paralogues chez certaines espèces), aucun biais de répartition n'est observé (Figure 18). Ceci suggère que la localisation dans un seul organite ou dans deux organites n'influence pas le comportement évolutif du gène codant la PPR chez les angiospermes. D'autre part, le fait qu'une protéine PPR soit requise pour la viabilité d'une plante n'est pas corrélé avec la présence ou absence des paralogues chez les angiospermes. Par exemple, les gènes à copie unique *MEF11*, *MEF9*, *MEF12* ne sont pas essentiels pour la survie de la plante. La raison et le mécanisme pour la conservation de gènes à copie unique reste inconnue. Cependant il semblerait que pour certains de ces gènes, la génération des paralogues serait moins tolérée, conduisant ainsi à une perte des copies comme l'ont suggéré De Smet *et al.* (2013).

Bien que son origine soit encore discutée, la double localisation des protéines dans la mitochondrie et le plaste est un processus ancien, présent au moins depuis la divergence des plantes terrestres (Xu *et al.*, 2013). Ainsi, plusieurs protéines de la plante *P. patens* présentent une double localisation (Xu *et al.*, 2013, Fuss *et al.* 2013). D'autre part, cette catégorie de localisation est dynamique dans l'évolution des plantes car elle peut être perdue ou acquise dans une lignée spécifique. Malgré le très faible nombre de données concernant la localisation de protéines PPR orthologues chez plusieurs espèces, la double localisation des protéines PPR semble avoir une origine ancienne et une probable conservation entre espèces. En effet, trois protéines PPR de *P. patens* (PpPPR) ont été rapportées dans la littérature comme ayant une double localisation (Sugita *et al.*, 2013, Fuss *et al.* 2013). Dans ce même article, il est montré

RESULTATS

que la double localisation de la protéine PRORP1 d'Arabidopsis est conservée pour PpPPR_63 et PpPPR_104, ses deux orthologues chez *P. patens* (Sugita *et al.*, 2013).

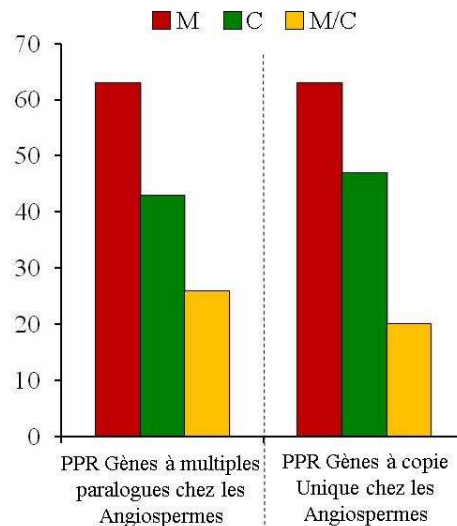


Figure 18. Distribution de la localisation des protéines PPR en fonction du comportement évolutif des gènes correspondants chez les angiospermes.

Seules les données expérimentales de localisation ont été utilisées pour cette figure : M : protéines adressées aux mitochondries (en rouge), C : protéines adressées aux plastides (en vert) et M/C : protéines localisées dans la mitochondrie et le plaste (en orange). Le nombre de gènes en copie unique codant des PPR a été obtenu à partir de données de De Smet *et al* (2012).

La double localisation de la troisième protéine de *P. patens*, PpPPR_43 du sous-groupe PLS-E-DYW, est plus controversée. En effet, cette protéine a initialement été associée à une localisation et à une fonction mitochondriale : l'épissage de l'intron 3 du gène *coxI* (Ichinose *et al.*, 2012). Cette localisation mitochondriale a été observée en utilisant le signal d'adressage N-terminal de la protéine fusionné à la GFP. Cependant, une étude plus récente suggère que PpPPR_43 aurait une double localisation. Cette double localisation est observée quelque soit le système d'expression utilisé : homologue chez *P. patens* ou hétérologue chez *N. tabacum* et *A. thaliana* (Fuss *et al.*, 2013). Il est important de noter que PpPPR_43 n'a pas d'orthologue clairement identifié chez les angiospermes, raison pour laquelle une conservation de localisation n'est pas caractérisable entre ces clades de plantes. De manière intéressante, il a été rapporté que deux transcrits alternatifs peuvent coder pour deux isoformes de la protéine PpPPR_43 par des codons d'initiation de la traduction différents (Ichinose *et al.*, 2012). Le premier transcrit, non-épissé, code la protéine plus petite et le second transcrit épissé code la protéine la plus longue. Malgré ces différences entre ces transcrits, les protéines prédites ne diffèrent que dans longueur de leur partie N-terminale

comprenant le signal d'adressage.(Ichinose *et al.*, 2012). Ichinose *et al.* (2012) ont observé une localisation mitochondriale pour le signal d'adressage N-terminal (84 aa) de l'isoforme la plus petite. Ces derniers auteurs ne donnent cependant aucune information sur la localisation de l'isoforme la plus longue présentant un signal d'adressage de 100 aa. Quant à Fuss *et al.* (2013), ils ont étudié la localisation de la plus petite isoforme de PpPPR_43 en utilisant un fragment d'ADNc incluant une partie de la région 5' non traduite et le signal d'adressage N-terminal. Aussi ces divergences pourraient simplement s'expliquer par la nature des séquences et des isoformes utilisées pour étudier la localisation subcellulaire de cette protéine.

Une situation similaire est observée chez *A. thaliana* pour le gène *AT3G42630*. Dans le cas de ce gène, j'ai pu cloner deux types de transcrits : un premier transcrit correspondant au modèle de gène TAIR et un second qui comprend 24 nucléotides de plus, cette séquence étant située en aval du site donneur d'épissage du seul intron prédit. Ce deuxième transcrit conduit à l'ajout de 8 acides aminés dans la partie N-terminale de la protéine (30 aa après le codon ATG). Cet ajout se traduit par le changement du signal d'adressage sans modifier la phase de lecture. Plusieurs transcrits (ADNc) correspondants à ce gène peuvent être identifiés dans la base de données du Genbank (BX824551, BX823210 AF378864) et permettent de renforcer mes résultats. Des études préliminaires de localisation des deux isoformes d'*AT3G42630* suggèrent qu'elles auraient une localisation différente, la forme la plus longue favorisant une localisation mitochondriale. Cependant, une étude de localisation utilisant des contrôles de co-localisation et des études fonctionnelles seront nécessaires pour déterminer la contribution de chacune des isoformes dans la localisation et l'activité de cette protéine. Récemment, dans le cadre d'une étude génétique portant sur un ensemble de protéines prédites avec localisation chloroplastique, il a été montré que la protéine d'*AT3G42630* est probablement essentielle car sa mutation entraîne un phénotype d'embryon létalité chez *A. thaliana* (Savage *et al.*, 2013). En effet, une lignée mutante, représentant la seule insertion disponible au sein du second exon, montre un phénotype de pigmentation lors du développement embryonnaire, ce stade photosynthétique étant nécessaire pour le développement embryonnaire chez *A. thaliana* (Savage *et al.*, 2013). Dans cette étude, les auteurs ont analysé une seconde lignée présentant une insertion localisée dans la région promotrice du gène et qui ne semble pas affecter le phénotype macroscopique de la plante. Cependant d'autres insertions proches du codon d'initiation sont disponibles et n'ont pas été analysées.

Il a également été observé récemment que deux couples de PPR orthologues chez les angiospermes (AtMEF29 - ZmPPR2263 et AtOTP87 - OsPPR1) présentent une conservation de leur double localisation (Sosso *et al.*, 2012, Xu *et al.*, 2013). Le gène codant pour les protéines AtMEF29 et ZmPPR2263 est présent en copie unique chez les angiospermes (Sosso *et al.*, 2012, De Smet *et al.*, 2012). Les deux protéines, chez *Arabidopsis* et chez le Maïs, sont observées dans les mitochondries et les plastes, avec toutefois une localisation principale et une activité identifiée dans mitochondries (Sosso *et al.*, 2012). Ainsi, chez *A. thaliana* et *Z. mays*, ces protéines sont nécessaires à l'édition mitochondriale des sites *nad5*-C1550 et *cob*-C908, le défaut d'édition à ces sites conduisant à une diminution de l'activité du complexe III et à un effet sur la biogenèse de la mitochondrie. Par ailleurs, aucune fonction dans l'édition des 34 sites majeurs du plaste n'a été observée.

Le couple d'orthologues AtOTP87 - OsPPR1 est sujet à plus de controverses en raison de la divergence des résultats obtenus pour chaque gène. En effet, AtOTP87 participe à l'édition des sites mitochondriaux *nad7*-C24 et *atp1*-C1178 mais, ni la biogénèse, ni les 34 sites majeurs d'édition chez le plaste, ne sont affectés chez le mutant *otp87* (Hammani *et al.*, 2011). La cause la plus probable du phénotype macroscopique de réduction de la croissance du mutant *otp87* est ainsi associée principalement au défaut d'édition du site *atp1*-C1178 qui modifie la nature d'un acide aminé (S393L) et causerait un défaut d'assemblage du complexe ATP synthétase de la mitochondrie (complexe V). Néanmoins, il a été proposé qu'OsPPR1, l'orthologue potentiel d'AtOTP87 chez le riz, aurait une fonction chloroplastique (Gothandam *et al.*, 2005). Dans cette étude, des plantes affectées dans leur expression d'OsPPR1 à l'aide d'une approche antisense, ont présenté un défaut de biogénèse des plastes. Néanmoins aucune donnée n'a permis d'identifier la cause moléculaire du phénotype cellulaire. Bien qu'incomplets, ces résultats suggèrent que des protéines PPR ayant une double localisation peuvent avoir des spécificités fonctionnelles dans un des organites dépendantes des espèces.

4.3.2. La double localisation et les réseaux d'interactions des protéines PPR

Les données concernant les réseaux d'interactions des protéines dans la cellule peuvent apporter des indices indirects concernant la localisation et la fonction moléculaire d'une protéine. Parmi les 46 protéines PPR présentant une potentielle double localisation, seulement 3 sont impliquées dans des interactions protéine – protéine connues (Figure 19). Ces données sont issues soit de l'interactome systématique d'Arabidopsis obtenu en double hybride chez la levure (Arabidopsis Interactome Mapping Consortium 2011) soit de l'analyse par immunopurification de complexes protéiques impliqués dans le cycle cellulaire (Van Leene *et al.*, 2007, Van Leene *et al.*, 2010). Le peu de données disponibles ne permet pas de conclure quant à l'éventuelle double localisation de ces protéines. Néanmoins, dans le cas de la protéine AT3G42630, une interaction avec la petite protéine chaperonne mitochondriale AtHSP23.6 supporte la localisation mitochondriale de cette protéine PPR.

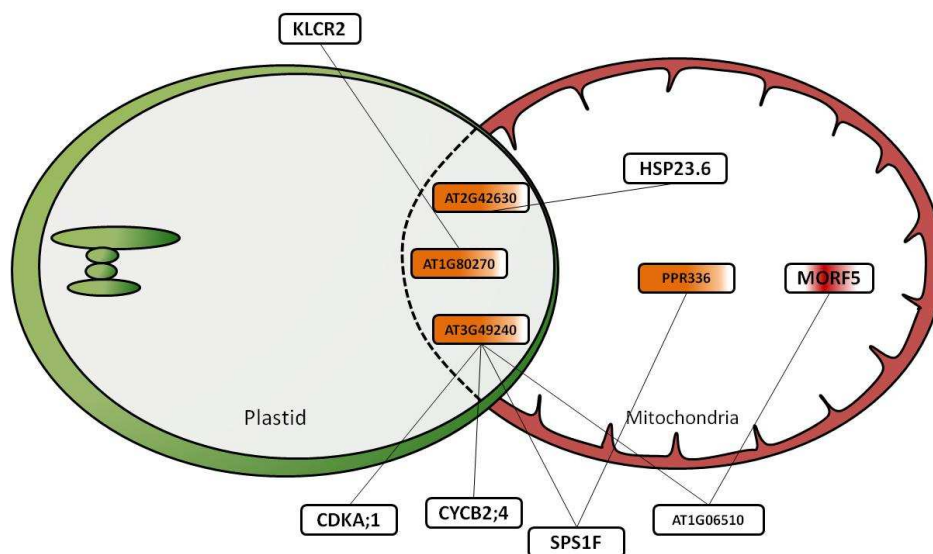


Figure 19. Réseau d'interactions protéine – protéine impliquant des protéines PPR ayant une potentielle ou double localisation.

RESULTATS

Chapitre V

Analyse fonctionnelle de protéines PPR ayant une double localisation.

Article II

Projet d'article III

Activités complémentaires du travail de thèse

Approches mises en place lors du travail de thèse pour contourner l'embryon létalité des mutants affectés dans DYW2 et OTP100 et déterminer leur fonction moléculaire.

Stratégie d'interférence artificielle inductible à l'aide de microARN.

Stratégie de complémentation spécifique du stade graine

Stratégie d'hémi-complémentation des mutants *dyw2* et *otp100*

Etude de l'interaction entre les protéines DYW2 et OTP100 par complémentation de fluorescence bimoléculaire ou BiFC.

Caractérisation fonctionnelle de la protéine OTP91.

5. Analyse fonctionnelle de protéines PPR ayant une double localisation.

Sur la base des résultats préliminaires de double localisation des protéines PPR ainsi que sur l'intérêt porté par l'équipe pour l'étude du mécanisme d'édition des ARN des organites, mon projet de thèse s'est focalisé initialement sur l'analyse fonctionnelle de 4 protéines PPR, AT1G06145, AT1G25360, AT2G15690 et AT3G49240, ces protéines ayant une potentielle double localisation et une potentielle fonction d'édition. Ainsi, les protéines AT1G25360 et AT1G06145, ci-après nommées OTP90 et OTP91, ont été choisies en raison de leur potentielle double localisation et de la présence dans leur séquence des domaines E et DYW associés principalement au phénomène d'édition. Les protéines AT2G15690 et AT3G49240, ci-après nommées DYW2 et OTP100, ont été choisies suite à leur identification dans un complexe protéique contenant le facteur d'édition CLB19 et en raison de leur probable double localisation suggérée par l'analyse des données protéomiques.

Dans le cadre du projet de thèse, l'analyse fonctionnelle de 3 des 4 protéines a pu être menée suffisamment à terme pour proposer les résultats les concernant sous forme d'articles (qui seront soumis prochainement). Cependant, l'ensemble des résultats partiels et des expériences additionnelles mises en œuvre au cours de la thèse sont présentés à la suite. Le premier projet d'article présente les résultats concernant la protéine OTP90, le deuxième article concerne les protéines DYW2 et OTP100. Les résultats complémentaires concernant, d'une part DYW2 et OTP100, et d'autre part, d'OTP91 sont présentés à la suite.

5.1. Article II

Running title: Interaction of organelle RNA metabolic processes

Title: The dual targeted PPR protein OTP90 and MORF1 share six target RNA editing sites in mitochondria and interact in vivo in *Arabidopsis thaliana*

Mauricio Lopez-Obando^{1,2}, Barbara Härtel³, Daniil Verbitskiy³, Anja Zehrmann³, Richard Berthomé^{1,a}, Mizuki Takenaka³, Claire Lurin^{1*}

1 INRA, UMR 1165, Unité de Recherche en Génomique Végétale (URGV), CNRS, ERL 8196, UEVE, 91057, Evry cedex, France.

2 Université d'Evry Val d'Essonne, Bd François Mitterrand, 91057, Evry cedex, France.

3 Molekulare Botanik, Universität Ulm, 89069 Ulm, Germany

Current addresses

a: Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR INRA 441/CNRS 2594, CS 52627, 31326 CASTANET TOLOSAN CEDEX, France

* Corresponding author: Claire Lurin^{1*} (lurin@evry.inra.fr), tel (33) 1.60.87.45.13.

Date of submission:

Number of tables: 0

Number of supplemental tables: 3

Number of figures: 7

Number of supplemental figures: 5

Total word count:

ABSTRACT

The Pentatricopeptide Repeat (PPR) protein family is an RNA binding protein family highly expanded in plant genomes. Previous reports have shown that they are mainly targeted to plastids, mitochondria or dually targeted to both organelles. To date, it is well known that members of this family are involved in organellar RNA metabolism at different levels. We recently showed through systematic targeting and data integration that dual targeting of PPR proteins to mitochondria and plastid is more common than previously thought in *Arabidopsis thaliana*. In order to understand the function of dual targeted PPR proteins, we selected and analyzed the function of ORGANELLAR TRANSCRIPT PROCESSING 90 PPR protein. Using a reverse genetic screen, we identified the *otp90* mutation which affects root growth at seedling stage. At the molecular level, we found that OTP90 protein targets at least six mitochondrial editing events in five different mitochondrial genes. Four of these sites belong to the first 10 predicted sites according to the refined code for PPR proteins. We also found that this protein interacts with four mitochondrial and one plastidial MORF editing factors. Moreover, we observed that this PPR protein is also directly or indirectly involved in splicing of the first intron of the *nad1* mRNA. In contrast, we did not find any evidence of the involvement of the OTP90 protein in editing of any of the 34 major sites reported in *Arabidopsis* plastids. In conclusion, our results confirm the model of organellar editing complex in which interactions between PPR and MORF proteins are necessary for total RNA editing efficiency. Secondly, we confirm that the refined PPR code can be used to accurately predict targets of PPR proteins acting in multiple sites. Finally, we propose that dually targeted PPR proteins could work in several organellar RNA metabolic processes as recently shown for PPR proteins with single localization.

Key Words: Arabidopsis, Dual targeting, Mitochondria, Plastid, PPR protein, RNA editing, Trans-splicing

INTRODUCTION

Organellar gene expression needs a number of transcriptional and post-transcriptional processing steps to produce mature transcripts. In particular, RNA splicing and RNA base conversions are important features in the genesis of functional organellar transcripts (Stern et al., 2010; Takenaka et al., 2013b). Organellar RNA base conversions, generally known as RNA editing, were discovered more than 20 years ago (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989; Hoch et al., 1991). To date, C to U, U to C and A to I types of RNA editing are known to proceed in plant organelles (Delannoy et al., 2009; Takenaka et al., 2013b). These nucleotide conversions have differences in their extent between plant species, but C to U is the most frequent reaching more than 2000 conversions in mitochondria of *Selaginella moellendorffii* and around 500 conversions in plastids of *Anthoceros formosae*. In flowering plants, 300 to 600 and 25 to 75 conversions have been identified in mitochondria and plastids, respectively (Huang et al., 2013; Takenaka et al., 2013b). For example, in *Arabidopsis thaliana* more than 400 RNA editing sites have been reported in mitochondria and more than 40 sites have been identified in plastids (Bentolila et al., 2013; Ruwe et al., 2013).

The implication of a specific trans-acting factor in organelle RNA editing was demonstrated when the protein named CHLORESPIRATORY REDUCTION4 (CRR4) was found to be involved in editing the *ndhD*-1 event of the *ndhD* plastid transcript (Kotera et al., 2005). In mitochondria the first trans-acting factor named MITOCHONDRAL EDITING FACTOR 1 (MEF1) affecting *rps4*-C956, *nad7*-C963 and *nad2*-C1160 editing sites was discovered some years later (Zehrmann et al., 2009). Many specific trans-acting factors involved in organelle RNA editing have been found since then, most belonging to the Pentatricopeptide Repeat (PPR) protein family. Members of this family, encoded by the nuclear genome, are mainly targeted to mitochondria, plastids or both organelles (Colcombet et al., 2013). The PPR domain, initially deduced from the observation of a degenerated sequence repeated along the protein (Aubourg et al., 2000; Small and Peeters, 2000), was used to characterize and classify the PPR family members. This classification divided the family into two major subfamilies, the PPR P and the PPR PLS subfamilies, the latter being further subdivided into two

RESULTATS

subgroups, the PLS-E and the PLS-E-DYW groups, which are characterized by specific C-terminal E or E and DYW domains (Lurin et al., 2004).

To date, at least 47 PPR proteins have been shown to be involved in mitochondrial and plastid RNA editing in *A. thaliana*. With the exception of PENTATRICOPEPTIDE REPEAT 596 (PPR596) protein (Doniwa et al., 2010), all PPR editing factors belong to the PLS subfamily, 20 belonging to the PLS-E and 26 to the PLS-E-DYW subgroups, respectively. The molecular functions of C-terminal E and DYW domains in the editing process are not clear yet. However, both domains are suspected to be required for editing. Although, no catalytic activity has been demonstrated for the DYW domain yet, based on convergent results, it has been hypothesized that it could contains the RNA editing enzymatic activity (Okuda et al., 2007; Salone et al., 2007; Boussardon et al., 2012; Rudinger et al., 2012).

Generally, PPR proteins are involved in all kinds of organellar RNA metabolic processes such as splicing, 5' and 3' processing, and transcript stabilization but also in translation initiation (Schmitz-Linneweber and Small, 2008). Most organellar introns are group II introns which require cis- or trans-splicing reactions mediated by domains of these group II elements and proteins factors. In flowering plants, to date, nine PPR proteins six in plastid and three in mitochondria were shown to play a role in group II either cis- or trans-splicing events (de Longevialle et al., 2010; Khrouchtchova et al., 2012; Zmudjak et al., 2013). Among them, only ORGANELLAR TRANSCRIPT PROCESSING 70 (OTP70), involved in the splicing of plastid *rpoC1* transcripts, belongs to the PLS subfamily which generally encompasses RNA editing factors (Chateigner-Boutin et al., 2011). Similarly, the *Physcomitrella patens* PPR 43 (PpPPR43) of the PLS-E-DYW subgroup is involved in intron 3 splicing of the mitochondrial *coxI* gene (Ichinose et al., 2012), suggesting that all types of PPR proteins could be implicated in RNA splicing.

The primary molecular property of PPR proteins is their ability to specifically interact with defined sequence motifs in RNA molecules. This was shown for PPR10 and OTP82, as well as many other PPR proteins (Prikryl et al., 2011; Okuda and Shikanai, 2012). Recently, using bioinformatics and statistics approaches, a code for PPR-RNA recognition was proposed

RESULTATS

(Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013). This code proposes that the combination of three amino acids, two located in the first PPR motif (3rd and 6th positions) and the third at the 1st position in the subsequent PPR motif is required for sequence specific recognition of the RNA target sequence. Recently, the structure of PPR-RNA complexes confirmed the implication of these amino acids and showed new features as interactions of amino acids in the PPR domain with the RNA backbone (Ke et al., 2013; Yin et al., 2013).

Besides PPR proteins, several other proteins have been shown to be implicated in organellar RNA metabolism. Members of a small family called Multiple Organellar RNA editing Factors (MORF) were found to be involved in RNA editing. These proteins function as general editing factors, can interact with many PPR proteins and, together with editing site specific PPR, are required for editing of many, if not all, organellar RNA editing sites, as shown for MORF1, MORF2, MORF3, MORF8 and MORF9 (Takenaka et al., 2012; Bentolila et al., 2013). Proteins containing classic RNA Recognition Motifs (RRM) were also found to be associated with RNA editing. For example, mutants in CHLOROPLAST RIBONUCLEOPROTEINS 31A or 31B (CP31A, CP31B) are affected in RNA editing of several chloroplast genes but also in organellar RNA stability (Tillich et al., 2009). Recently, the ORGANELLE RNA RECOGNITION MOTIF 1 (ORRM1) protein, containing both a RRM motif and domains related to MORF proteins, was shown to be implicated in Arabidopsis and maize chloroplast RNA editing (Sun et al., 2013). Similarly, proteins of Chloroplast RNA splicing and ribosome Maturation (CRM), Plant Organellar RNA Recognition (PORR) and some others families are involved in organellar intron splicing as detailed in recent reviews (de Longevialle et al., 2010; Stern et al., 2010). For example, five plastids and one mitochondrial CRM proteins have been implicated in many splicing events in plastid and mitochondria. At difference of PPR, each CRM protein could influence several intron splicing events of different genes, reason why they are considered as general factors of organelles intron splicing complexes.

Despite many factors were recently described as involved in organellar RNA metabolism, functional interactions between these factors, interrelationships of each RNA processing step on the others and functional regulations between organelles are still poorly understood. In this study, we characterized a PPR protein that is dually targeted to mitochondria and plastids, we

identified its involvement in mitochondrial RNA editing, and analyzed its interaction with editing factors belonging to the MORF family. Additionally, we present *in vivo* evidence showing that RNA editing and also RNA trans-splicing are interconnected probably by the function of the PPR protein in both processes.

RESULTS

AT1G25360 Encodes a PPR Protein Targeted to both Organelles which is Conserved in Angiosperm Species

A previous study suggested that dual targeting to mitochondria and plastid constitutes an emerging distinct class in the PPR family (Colcombet et al., 2013). To gather further information on such dual targeted PPR we selected potential candidates for functional analysis. Among these, the PPR protein encoded by AT1G25360 was shown to be potentially dually targeted, by the transit peptide alone and as the full length open reading frame (ORF) merged to the fluorescent DsRed2 protein coding sequence in transient expression assays (Colcombet et al., 2013). To extend our previous data, we monitored the subcellular localization of AT1G25360 in stable transgenic *Arabidopsis* lines expressing a AT1G25360:DsRed2 protein fusion under control of 35S promoter. Five transgenic lines were isolated and two of them were used for protein localization analysis. AT1G25360:DsRed2 fusion protein was observed in mitochondria of roots and cotyledon cells (Figure 1A and B) but also in chloroplasts of cotyledon cells (Figure 1B). We named this protein as ORGANELLAR TRANSCRIPT PROCESSING 90 (OTP90) due its putative function in organellar RNA metabolism.

The OTP90 protein belongs to the PPR PLS subfamily and more precisely to the PLS-E-DYW subgroup. In *Arabidopsis*, this PPR protein contains 15 predicted PPR domains upstream of E and DYW C-terminal domains (Figure 2A and supplemental Figure 2). Previous phylogenetic analysis had shown that this PPR protein has a potential ortholog in rice and that it is most closely related to the At2g22070 gene in *Arabidopsis* (O'Toole et al., 2008). In order to know if OTP90 is conserved in further plant species, we performed a phylogenetic analysis to search

RESULTATS

for putative orthologs in the genome data available on the phytozome website, in several angiosperms. We found a single homologous protein in each angiosperm species analyzed suggesting that despite multiple nuclear genome duplications in plants the gene coding OTP90 has kept or returned to a single-copy structure (Supplemental figure 1).

Orthologs of OTP90 appear to diverge mainly in the N-terminal sequences and to a lesser extent in PPR, E, and DYW domains. Even if the dual targeting prediction is not yet accurate, the integration of predictions suggests that the dual localization for OTP90 orthologs might be conserved. However, preferences of localization between species could be possible because the prediction in *gramineae* species seems to favor plastid localization (Supplemental Table 1).

***Arabidopsis otp90* Mutant is Impacted in its Plant Development**

To explore the molecular function of the OTP90 protein, we characterized two insertional mutant lines for this gene, SM_3_29668 and SM_3_29659. Since these two sister lines were supposed to have T-DNA insertions at similar positions, we sequenced PCR products corresponding to T-DNA junction region (see Figure 2). We confirmed the identical location of the T-DNA insertions in the two lines at position 806 bp after ATG codon. The line SM_3_29659 was used for the following experiments and it was named *otp90* (Figure 2A). The mutant line was backcrossed three times to *Arabidopsis thaliana* ecotype Columbia (Col-0) to segregate other mutations. Homozygous plants for the T-DNA insertion were identified after self-pollination of heterozygous plants obtained after back-crossing. In homozygous plants, we were not able to detect by RT-PCR any transcript corresponding to *OTP90* suggesting that the insertion induces *OTP90* gene knockout or strongly reduce its expression (Figure 2).

Backcrossed homozygous plants were characterized after germination on MS medium and growing under long-day conditions (16h light/8h dark). *otp90* mutant seeds did not show any difference in germination efficiency when compared to wild type Col-0 seeds. However, primary roots of 12 day old *otp90* seedlings (stages 1-1.02, Boyes et al 2001) showed delayed

RESULTATS

growth when compared to their wild type counterparts (Figure 2A and B). In soil and under long-day conditions, 3 week old *otp90* and wild type Col-0 plants did not show any growth or developmental difference (Figure 2C). Furthermore, plants reached reproductive stages and produced seeds at the same time as wild-type siblings. Since numerous PPR mutants show seeds phenotype, we compared seed sets obtained on four independent plants for each *otp90* and Col-0 lines. Seeds were separated in two categories according to their size using a 0.25mm sieve. We found a significant increased of the amount of small seeds (smaller than 0.25mm) in *otp90* sets (Figure 2D). All together, these data show that *otp90* mutation slightly but significantly affects root growth and seed size of plants.

The *otp90* mutant is impaired in Mitochondrial C to U Editing.

Previous reports suggested that most of PPR proteins with E and DYW domains at their C-terminal region are involved in RNA editing. Since OTP90 protein harbors such domains, we surveyed editing events in the *otp90* mutant. Direct sequencing of cDNA products or SNaPshot approaches were used to quantify editing rates in plastid and mitochondria, respectively. Among the 43 plastid editing sites reported in the literature, we characterized the editing level of 34 sites for which editing rate is higher than 5% in wild type plants. None of these sites was affected in the *otp90* mutant. In contrast, among the 415 mitochondrial sites investigated by SNaPshot in *otp90*, a total loss of editing was detected at 4 sites: *nad1*-C500 (288077), *ccmB*-C80 (30542), *ccmC*-C184 (240575) and *mttB*-C97 (157587) and a significant reduction was observed in 2 others: *ccmB*-C596 (31058) and *ccmFc*-C1246 (51406). All six impaired editing events change the amino acid sequence, three giving a substitution from Arginine to Tryptophan (R to W) and the three others leading to the substitution from Sérine to leucine (S to L) (Figure 3 and Supplemental table 2).

Complementation of the *otp90* mutant with the *Arabidopsis* *OTP90* genomic fragment containing the promoter region showed a recovered primary root growth phenotype now similar to Col-0 (Figure 2B and 2C). Moreover, all 6 RNA editing defects of the *otp90* mutant were reverted in the complemented line, some of them increasing their editing status compare to Col-0 (Figure 3).

OTP90 interacts with MORF proteins in the editing complex

Organellar RNA editing in plants requires several classes of proteins which may be active in a multi protein complex. Since several PPR editing factors share target editing sites with individual MORF proteins and indeed several interactions of such PPR and MORF proteins have been reported, it is hypothesized that the direct interaction of PPR and MORF proteins is indispensable for efficient RNA editing at respective target sites (Takenaka et al., 2012; Takenaka et al., 2013b). All seven target sites of OTP90 show reduced RNA editing also in the *morf1* mutant line and one of the six sites, *mttB*-C97, is also affected in a *morf3* mutant. Therefore, to evaluate the protein complex hypothesis, we analyzed the potential interactions between OTP90 and each of the MORF proteins in the Yeast-2-hybrid system (Figure 4). The results show that OTP90 interacts strongly with MORF1 and MORF3, but also with the plastid targeted MORF2 and the mitochondrial targeted MORF4, MORF6 proteins. Weak interaction is seen with MORF5. In contrast, no interactions between OTP90 and MORF7, MORF8 and MORF9, respectively, are detected (Figure 4).

To confirm the selective PPR - MORF interactions *in vivo*, YFP split analysis between either MORF1, MORF2, MORF3, MORF4, MORF5 or MORF9 and OTP90 was performed in the tobacco leaf system (Figure 5). YFP signals were detected only with the OTP90-MORF1 combination suggesting specific binding of OTP90-MORF1 *in vivo*.

nad1* Intron 1 Trans-Splicing is reduced in *otp90

Although OTP90 is localized in both mitochondria and chloroplasts, the *otp90* mutant does not show any defect in the 34 major RNA editing sites known in chloroplasts. To explore the status of the chloroplast transcripts we took the general approach of gene expression to pinpoint possible deregulated gene targets. Therefore, we quantified gene expression levels in chloroplast of mutant leaves compared to wild type leaves using qPCR as previously described (de Longevialle et al., 2008). These chloroplast gene expression analyses did not show a clearly target. Nevertheless a modest over-expression of the polycistronic operon

RESULTATS

rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-rpl36-rps11-rpoA was observed with a peak in the *rpl22* and *rps3* mRNAs (Supplemental figure 3). Because the macroscopic phenotype of *otp90* are not similar to PPR mutants with defects in chloroplast translation these over-expression seems to be consequence of an indirect effect in the mutant.

Whereas transcription in mitochondria is relaxed and subjected to little regulation post-transcriptional events are considered to play a main role. Taking in account the probably relationship between editing and splicing and the fact that in *otp90* two mitochondrial genes having introns are impacted in editing, we analyzed RNA splicing efficiency in mitochondria of *otp90* by qPCR as previously described (de Longevialle et al., 2007). In this analysis, most of splicing events did not seem affected in *otp90*. However, trans-splicing of *nad1* intron 1 was reduced in *otp90* (Figure 6). This observation suggests either a direct function of OTP90 in trans-splicing of *nad1* intron1 or an indirect functional link between editing and maturation of the *nad1* transcript (with putative involvement of *nad1*-C500 editing site).

Mitochondrial complex I is not impaired in the *otp90* Mutant

In plants, the mitochondrial NADH dehydrogenase complex I of the respiratory chain is a protein complex composed of 40 to 49 subunits (Klodmann et al., 2011). Plant *nad1* gene codes for a core subunit in the membrane arm of complex I and is located in the mitochondrial genome. One of the reasonable explanations of the slight root growth phenotype of the *otp90* mutant could be insufficient amounts and/or activity of complex I, since both RNA editing and splicing defects are observed in *nad1* transcripts in the mutant. Therefore, we compared mitochondrial complexes of wild type Col-0 and *otp90* 4 week old leaves by blue-native gel and their oxidative activity by staining of complex I (Figure 7). This analysis showed that both amount and activity of complex I in *otp90* mutant are similar to those in wild type.

DISCUSSION

Organellar genome expression is complex and relies on imported nuclear encoded proteins. In this study, we characterize the PPR protein OTP90 of the PLS-E-DYW subgroup and show its

RESULTATS

involvement in several mitochondrial editing events. Among the six mitochondrial editing sites targeted by OTP90, two, *nad1*-C500 and *ccmC*-C184, are always edited in all angiosperm species analyzed. Besides these, the *mttB*-C97 event also seems to be edited but the limited information for this gene in many species does not allow a definite conclusion about its general editing status in angiosperms. Furthermore the sites *ccmB*-C80 *ccmB*-C596 and *ccmFc*-C1246 are not edited in at least one specie, all being a C nucleotide in the mitochondrial genome of the respective specie(s) (Supplemental figure 4). All editing events targeted by OTP90 change the respective amino acid sequences in Arabidopsis towards the evolutionary conserved amino acids in these proteins (Supplemental figure 4). Whatever happened in evolution, the information suggest, at least, that *nad1*-C500 and *ccmC*-C184 are ancient angiosperm editing events targeted by the editing complex including OTP90 protein. We consequently hypothesize that absence of editing of one or a combination of these sites causes the specific developmental phenotype observed in *otp90*. We first suspected that the *nad1*-C500 editing defect and the observed reduction of *nad1* intron 1 trans-splicing efficiency was the main element explaining this phenotype but abundance and activity of mitochondrial complex I do not change significantly in the mutant at least in 4 weeks old leaves. It is still possible that complex I is significantly reduced in the early developmental stages or in roots. Otherwise, a combination of the several editing defects is likely to be the cause of the *otp90* phenotype but we cannot exclude the implication of other not detected molecular functions of the OTP90 protein.

PPR-MORF protein-protein interaction is a common feature of RNA editing complex.

The RNA editing complex selects specific target sites by binding cis-elements in the RNA sequences being a model of multi subunit complex recently proposed (Boussard et al., 2012; Takenaka et al., 2013b). This model is supported by our complete protein-protein interaction screening between OTP90 and MORF proteins. Nevertheless some particularities have been found in the OTP90-MORF interactions helping to show specific aspects of their interaction. Both, the yeast two hybrid and the *in vivo* split YFP (BiFC) analyses indicate a specific interaction of OTP90 and MORF1 (Figures 5 and 6). This is supported by the strong overlapping of the RNA editing profile between *otp90* and *morf1-1* confirming that both

RESULTATS

MORF1 and OTP90 are main factors implicated in the editing of these target sites. Analogous common target sites for MORF and MEFs have been identified for MORF1 and MEF21 (site *cox3*-C257), MORF1 and MEF19 (*ccmB*-C566), MORF3 and SLO2 (*mttB*-C144, *mttB*-C145, *nad7*-C739 and *nad4L*-C110) and MORF8 and MEF10 (*nad2*-C842) and also for MORF8 and RARE1, respectively (Takenaka et al., 2010; Bentolila et al., 2012; Takenaka et al., 2012; Zhu et al., 2012; Hartel et al., 2013).

The observations that OTP90 interacts also with MORF3 and other MORF in the yeast two hybrid system and that RNA editing of all OTP90 target sites is reduced but still detectable in the *morfl-1* mutant suggest a partial complementation of the MORF1 function by MORF3 (or other mitochondrial MORF) as interacting partners of OTP90 in the editing complex. Three MORF proteins have been reported as targeted to plastids, MORF2, MORF9 and MORF8 (Bentolila et al., 2013). 34 major editing sites are reported in plastids, none of them is affected in the *otp90* mutant. 9 additional sites with very low editing rates were recently described, they were not characterized in our study because they cannot be studied without deep sequencing. Potential interaction of OTP90 with MORF2 but not MORF8 or MORF9 may provide a clue to reveal a so far undetected function of the OTP90 protein in the plastid.

OTP90 RNA editing complex are able to recognize sequences with low similarity

Alignment of Arabidopsis nucleotide sequences located 5' and 3' of each OTP90 target editing site reveals little similarity in the primary target nucleotide sequences showing only three shared nucleotides at -11,-1 and +1 position related to impaired edited site (0) (Supplemental figure 5). In this way, OTP90 protein take part in the group of PPR proteins with multiple mitochondrial editing targets sites showing little sequence similarity in the cis-element that it target.

This does not contradict the code for RNA recognition of PPR proteins proposed previously (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013). Indeed, using the refined code (Takenaka et al., 2013a), four events, *ccmB*-C80, *nad1*-C500, *mttB*-C97 and *ccmFc*-C1246, of the six experimentally detected were found into the 10 best scores of predicted targets of

OTP90 protein, being located in the list at positions 1, 2, 5 and 10 respectively. For *ccmC*-C184 and *ccmB*-C596 the code does not give a good score. It suggests that more parameters are necessary for fit more accurately the prediction of target events. Explanations of the variability in the sequence recognition is the fact that the specificity could be helped by a slightly different composition in the complex for some sites or that PPR domains can be aligned in a variability position depending of the RNA sequence as has been suggested for CRR22 protein and its multiple target sites (Okuda and Shikanai, 2012). Alternatively, it is also possible that for one specific sequence recognition some PPR domains are not necessarily used as recently suggested in a mutagenesis study analyzing the implication of several amino acids in the function of PPR domains of PROTON GRADIENT REGULATION3 (PGR3) protein (Fujii et al., 2013). Thus, the editing complex targeting several sites that do not share upstream sequences could have a modulation in their composition or flexibility in their structure that can help it to bind in some extends unrelated nucleotide sequences as previously suggested for editing factors by Takenaka et al. (2013a).

What is the function of dual targeted PPR RNA editing factors in the two organelles?

All types of functional categories of PPR proteins are present inside of the dual targeted PPR protein class, including editing factors (Colcombet et al., 2013). To date, among the four dual targeted PPR proteins characterized, only the PROTEINACEOUS RNASE P 1 (PRORP1) has been shown to function in both mitochondria and plastid (Gobert et al., 2010). The three other dual targeted PPR proteins are editing factors, OTP87 (Hammani et al., 2011), maize PPR2263 and its *Arabidopsis* ortholog MEF29 (Sosso et al., 2012). However, their editing function have only been observed in mitochondria. Similarly, in this study, a molecular function for OTP90 was only shown in mitochondria. The reason of the dual targeting of these proteins was not discovered yet. Even if editing is a dynamic process, it was showed that in evolution of angiosperms loss of plastid editing is more frequent than editing gain (Takenaka et al., 2013b). A similar evolutionary scenario is also thought to occur in mitochondria based in the fact that angiosperm species have lost editing events in relation to gymnosperms and also during a more recent evolution of silene species (Richardson et al., 2013). In this context, it is possible to suggest a tentative explanation of dual targeted PPR editing factors showing

RESULTATS

functionality in a single organelle. Thus, it may be possible that during evolution the dual targeting of some PPR proteins were not modified even if their editing target(s) event(s) were lost in one organelle. That explanation implies that the retention of protein dual targeting was tolerated due to a minimum of perturbation in the metabolism. Nonetheless, these proteins could also have a not yet explored function in RNA processing or translation, for example. This last hypothesis needs also to be evaluated due to the recently found multiple functionality in RNA metabolism of maize PPR protein ATP4, the ortholog of the Arabidopsis SUPPRESSOR of VARIEGATION 7 (SVR7) (Zoschke et al., 2013). Zoschke and colleagues found that ATP4 in addition to be implicated in promoting *atpB* translation, it take part also in the stabilization of *rpl16-rpl14* dicistronic RNA. Thus, more studies will be necessary to verify these hypotheses. Many putative dual targeted PPR editing factors were identified in Arabidopsis (Colcombet et al., 2013) reason why we cannot exclude that, among them, some PPR involved in editing in both organelles could be discovered.

What is the link between RNA editing and splicing in *otp90*?

RNA editing is important to conserve the amino acid sequence of organelle proteins by correcting the mutational events in organelles genomes and also to conserve and regulate splicing events in plants through RNA structural stabilization (Farre et al., 2012; Kagami et al., 2012). These functions are not mutually exclusive and need to be integrated to better understand organelle RNA metabolism.

Although six proteins are found to be involved in *nad1* splicing, nMAT1, nMAT2, nMAT4, OTP43, PMH2 and mCSF1 (de Longevialle et al., 2007; Keren et al., 2009; Kohler et al., 2009; Keren et al., 2012; Zmudjak et al., 2013; Cohen et al., 2014) and seven trans acting factors, MEF11, MEF14, SLO2, MORF1, MORF3, MORF8 and MEF25, MEF32 are known to be involved in *nad1* editing events (Verbitskiy et al., 2011; Bentolila et al., 2012; Takenaka et al., 2012; Zhu et al., 2012; Arenas et al., 2013; Takenaka et al., 2013a), little is known about the mechanism of *nad1* transcripts maturation and less about the dynamic of the RNA-protein complexes acting in this process.

RESULTATS

Analysis of the putative connections between editing and splicing of *nad1* transcript in *otp90* might suggest a connection between these processes. RNA editing could affect the rate of splicing because of changes in RNA conformations affecting the fixation or recognition of sequences by RNA splicing complex. This hypothesis is supported by the observation that editing is frequently shown to proceed at early stages of RNA metabolism and probably before splicing events (de Longevialle et al., 2010). In yeast, it was showed that unedited intronic RNA molecules of *Oenothera* mitochondrial transcripts were not spliced (Borner et al., 1995). In addition, recent works have shown that RNA editing of non-coding sequences can affect RNA splicing of plant mitochondrial genes (Castandet et al., 2010; Farre et al., 2012). This influence could be easily explained if editing of specific nucleotides affects intron stability. In this particular case we think that this hypothesis is unlikely because the affected editing site in *otp90* is located in the third exon of the *nad1* primary sequence and more than 100 bases after intron 1 trans-splicing site if intron 2 cis-splicing is occurring at first place as suggested previously in a tentative model of *nad1* transcript maturation (de Longevialle et al., 2007; Farre et al., 2012). Furthermore, in a tentative of modeling the secondary RNA structure of *nad1* intron 1 using mFold software, we did not find any evidence supporting any change between structures of edited and non edited RNA.

An alternative explanation could be a direct function of OTP90 in trans-splicing of intron 1. This hypothesis is supported by the fact that PPR proteins could act in different RNA metabolic processes as showed for ATP4 (Zoschke et al., 2013). Moreover, even if there is no real evidence of splicing mitochondrial complexes, it was shown that several proteins can be involved in the splicing of the same intron, suggesting the existence of splicing complexes as was showed for *trnA* intron splicing in Maize and Arabisopsis plastids (Khrouchtchova et al., 2012) but also in *Chlamydomonas reinhardtii* plastid *psaA* intron trans-splicing (Jacobs et al., 2013). Finally, the involvement in splicing of PPR proteins from the PLS subfamily such as PpPPR43 and OTP70 suggests that this type of proteins can be part of splicing complex. Thus, a direct function of OTP90 in *nad1* splicing might be the simplest scenario to explain the observations in *otp90* mutant. Nevertheless, only structural studies with focus on RNA conformations and proteins involved in the *nad1* intron1 trans-splicing complex will help to understand the *otp90* splicing phenotype

Relationships of editing and splicing or editing and RNA processing are also found in genes disrupting the function of plastid localized OTP70 and SEEDLING LETHAL 1 (SEL1 also known as PMD1) PPR proteins (Chateigner-Boutin et al., 2011; Pyo et al., 2013). Between them, *otp70* mutant share with *otp90* mutant the aspects of be affected in RNA editing and splicing inside a gene. However the results suggest that OTP70 and OTP90 could influence these processes in different ways. For OTP70, Chateigner-Boutin and co-workers showed that the main function of the protein is on *rpoC1* intron 2 splicing and suggested that the increase of *rpoC1* edition in unspliced transcripts of *otp70* was probably attributed to changes in splicing kinetics. Here, we showed that OTP90 has its main function in promoting editing (*nad1-C500*) and suggested that it could be directly implicated in *nad1* intron1 trans-splicing. In this context, a more complex picture is emerging where editing could influence splicing but vice versa splicing could influence editing. It shows the intricate relationships in organelle RNA metabolism and the fact more analyses are necessary to better understand organelle RNA metabolism.

In summary, the work presented here supports the recent model of organellar C to U RNA editing complexes in plants requiring the presence of both PPR and MORF proteins for an efficient RNA editing (Takenaka et al., 2013b). In addition, even if there is no evidence of a direct function of OTP90 in splicing, its dual targeting, its interaction with the plastid editing factor MORF2, together with the molecular phenotype of *otp90* mutant in mitochondrial editing and splicing suggest a complex profile of functions for this editing factor in the plant cell.

MATERIALS AND METHODS

Subcellular protein localization

Sequences of primers used in this study are available in supplemental table 3. OTP90 full-length ORF without stop codon was amplified using OTP90-S and OTP90-E primers and cloned into pDNR207 as described previously (Lurin et al., 2004). P35S::OTP90:DsRed2 construct was obtained by LR recombination using p0229-DsRed2 destination vector (Lurin et

RESULTATS

al., 2004). C58C1 pSOUP *A. tumefaciens* strain containing the binary plasmid was used to transform *A. thaliana* Col-0 plants by floral dip procedure. Primary transformants were selected using BASTA in soil and PCR genotyping for T-DNA insertion was performed using P35S-B, DsRed2-Rev and OTP90-R1 primers. Subcellular localization of the OTP90:DsRed2 fusion protein was observed in root and cotyledon cells of 10-15 days old primary transformants grown on MS solid medium using a spectral Leica SP2 AOBS confocal microscope (Leica Microsystems) equipped with an argon laser and a HeNe laser. To confirm the mitochondrial localization, the seedlings were stained with 1 μ M of Mito-tracker green marker (Invitrogen) for 15-30 min. Signals were detected using laser lines at 488 nm (Mito-tracker green; excitation/emission 488/510-530nm) and 543 nm (DsRed2; excitation/emission 543/570-600nm / chlorophyll autofluorescence; excitation/emission 543/600-700nm). The images were coded green (Mitotraker green or chlorophyll autofluorescence) and red (DsRed2) giving yellow co-localization signals when green and red signals overlapped in merged images.

Plant material, phenotype characterization and complementation assay

Arabidopsis thaliana ecotype Columbia-0 (Col-0) was used as wild type plants. The two T-DNA insertional mutants, SM_3_29668 and the SM_3_29659 from John Innes Centre Enhancer/Suppressor-mutator lines (JIC SM) (Tissier et al., 1999), were obtained from the ABRC stock center. T-DNA homozygous plants were selected by PCR genotyping using the OTP90-RP, OTP90-LP and T-DNA Spm32 primers. Localization of the insertion was verified by sequencing of PCR products using the Spm32 primer. Homozygous *otp90* mutant plants were obtained after three backcrosses using SM_3_29659 as male donor and Col-0 as female donor. For *in vitro* phenotype characterization, seeds were surface sterilized, sown in Murashing and Skoog (MS) solid medium with 0.8% (W/V) of agar-agar type E (SIGMA ALDRICH) and grown in a growth chamber (16-h light/8-h dark cycle, 22°C, 50% hygrometry) after cold treatment for 48h at 4°C. Primary root length elongation was measured on 12 day old plantlets after cold treatment of seeds using EZ-Rhizo software (Armengaud, 2009). Plant used for phenotyping in soil were grown in chambers at 16h/8h light/dark cycles, 22°C at day and 17°C at night.

RESULTATS

For complementation of *otp90* mutant, a genomic 3595 bp fragment, containing the promoter, the Open Reading Frame (ORF) and the terminator region of the *OTP90* gene, was amplified by PCR in two steps as described previously (Lurin et al 2004) using OTP90_Fgene and OTP90_Rgene as primers and genomic DNA extracted from the Col-0 ecotype as template. This fragment was cloned in pDNR207 vector by BP recombination reaction (Invitrogen) to generate the entry vector containing the *OTP90* gene and then subcloned in pGWB1 binary vector (Nakagawa et al., 2007) by LR recombination reaction (Invitrogen). C58C1 pMP90 *Agrobacterium tumefaciens* strain containing *OTP90* in pGWB1 plasmid was used to transform *otp90* mutant plants by floral dip procedure as described previously (Clough and Bent, 1998). *Arabidopsis* transformed plants were selected for T-DNA insertion using Kanamycin antibiotic (50µg/mL) and PCR genotyped using pGWB4806 and OTP90-R1 primers.

Organelle RNA editing analysis

Total RNA from leaves of 20-22 day old plants was extracted using RNeasy plant mini kit (QIAGEN). RNA was treated with 1µl of DNase I (Ambion 2U/µL) two times for 30 min at 37°C. DNase I was inactivated after 10 min of incubation at 75°C in a solution containing 2.5mM EDTA. RNA was precipitated and diluted in RNase free water. cDNA was synthesized using random hexamer primers and superscript II (Invitrogen). To evaluate RNA editing in chloroplast transcripts, cDNA and specific primers surrounding known organelle editing events were used for PCR amplifications and sequencing. For mitochondrial genes, SNaPshot protocol (Takenaka and Brennicke, 2012) was used to evaluate editing. It was followed by sequencing when editing defects in *otp90* mutant was found by SNaPshot.

Chloroplast expression and mitochondrial splicing analysis

Total RNA from leaves of 3 week old Col-0 and *otp90* plants were extracted using RNeasy plant mini kit (QIAGEN). Three µg of total RNA was treated with 1µL of DNase I amplification grade (1U/µL, Invitrogen) two times for 20 min at 25°C. DNase I was inactivated by 10 min of incubation at 65°C in a solution containing 2.5mM EDTA. 1.5µg total RNA was used to synthesize cDNA using random hexamer primers and superscript II (Invitrogen). Synthesized cDNA was used as template for quantitative PCR amplification

RESULTATS

(qPCR). Ten μL of qPCR reactions were performed using MESA SYBR Green qPCR master mix (Eurogentec) and Roche LightCycler real time PCR system with the following thermal cycling program: 95°C for 10 min, followed by 40 cycles of 95°C for 10s, 60°C for 10s and 72°C for 20s. Chloroplast transcript expression analysis was performed with chloroplast specific primers as reported previously (de Longevialle et al., 2008). Mitochondrial splicing analyses were performed with mitochondrial specific primers as reported previously (de Longevialle et al., 2007).

Yeast two hybrid assays

Sequences coding for *OTP90* and the various MORF were cloned with the In-Fusion HD cloning system (Clontech). *OTP90* inserts were integrated into the binding domain containing vector pGBKT7, and MORF into the activator domain containing vector pGADT7 of the GAL4 Two Hybrid System 3 (Clontech). The vectors were co-transfected for expression into yeast cells (PJ69-4A) according to the protocol. Yeast cells with both bait and prey vectors were cultured in synthetic dropout medium without Leu and Trp. 5 μL of suspended cells with an OD_{600} of 0.3 were dropped onto the various selection media plates. To detect strong interactions, 2.5 mM of 3-Amino-1,2,4-triazole was added in the media plates.

In Vivo Protein–Protein Interaction Assay with BiFC analysis

For *in vivo* protein–protein interaction assays, *OTP90*, *MORF1*, *MORF2*, *MORF3*, *MORF4*, *MORF5* and *MORF9* were fused with either YFP-N(1-155) or YFP-C(156–239) at their C terminus, respectively and cloned under 35S promoter. *OTP90* was cloned in pGWB1 binary vector (Nakagawa et al 2007) by LR recombination reaction (Invitrogen) whereas MORF proteins were cloned in pMDC123 vector (Curtis and Grossniklaus, 2003) containing a multiple cloning site from pET41 (Merck Millipore Novagen®) with In-Fusion HD cloning system (Clontech). Each construction was transferred into *Agrobacteria*. Combinations of *agrobacteria* containing *OTP90*-YFP-N and each MORF-YFP-C, on one hand and *OTP90*-YFP-C and each MORF-YFPN, on the other hand, were used at equal concentration to agroinfiltrate 3-weeks old tobacco leaves. After 48 hours of incubation at 21°C, transformed tobacco cells were analyzed using a confocal microscope Leica, TCS SP5 II.

RESULTATS

Mitochondrial native protein complex and complex I activity assay

Isolation of Arabidopsis mitochondria was performed as previously (Kühn et al 2009) with slight modifications. Ten grams of green tissue from 3-4 week old Arabidopsis plants were crushed in 100 ml ice cold extraction buffer (0.3 M sucrose, 5 mM tetrasodium pyrophosphate decahydrate, 2 mM EDTA, 10 mM potassium dihydrogen phosphate, 1% polyvinylpyrrolidone K 25, 1% BSA, 1 mM DTT, 5 mM cysteine, pH 7.5) with following filtration through two layers of Miracloth (Calbiochem, USA). Enrichment of mitochondria was performed by following centrifugation steps; 10 min at 2000 g and collection of the supernatant, 10 min at 24000 g and collection of the pellet. The pellet was then resuspended in 5 ml of ice-cold wash buffer (0.3 M sucrose, 10 mM MOPS, 1 mM EGTA, pH 7.2) and homogenized with a potter. After centrifugation for 10 min at 3000 g, the supernatant was collected and centrifuged for 10 min at 24000 g. The final pellet was resuspended in 700 µl of wash buffer with a paintbrush, applied on a 18%-25%-50% Percoll (GE Healthcare, USA) step gradient (10 ml of 50%, 22 ml of 25%, 5 ml of 18%) containing 0.3 M sucrose and 10 mM MOPS (pH 7.2) and centrifuged 45 min at 40000 g. The mitochondria pellet was then collected, washed 3 times with wash buffer (3 pellet volumes) and stored at -80°C. Blue native electrophoresis was performed as previously indicated (Schagger et al., 1994). Fifteen mg of mitochondria was resuspended in 70 µl of ACA buffer (750 mM amino carpoic acid, 0.5 mM EDTA, 50 mM Bis-Tris, pH 7.0) and solubilized in presence of 1% (final concentration) dodecylmaltoside for 20 min on ice. Solubilized mitochondria was clarified by centrifugation for 5 min at 7,700 rpm in a bench centrifuge. Supernatant was mixed with loading buffer (750 mM amino carpoic acid, 5% coomassie G-250) and 7.5 mg of proteins were separated in a 4.5%-12% BN gradient gel at 12 mA. Activity of the Complex I was analysed by staining the gel in 1 mM nitroblue tetrazolium 0.14 mM NADH, 100 mM Tris, pH7.4 buffer. Staining was stopped by fixing the gel in 40% methanol, 10% acetic acid with following standard silver staining.

RESULTATS

FUNDING

This research was supported by the French Ph. D. fellowship associated to the Université d'Evry Val-d'Essonne and “Ministère de la Recherche et de l'Enseignement Supérieur” to M. L-O. M.T. is a Heisenberg fellow and his work is supported by grants from the DFG.

ACKNOWLEDGEMENTS

We thank Jose Caius (Unité de Recherche en Génomique Végétale, France) for his help with laboratory support, Olivier Grandjean (Institut Jean Pierre Bourgin, Versailles, France) for his help with confocal microscopy and Ian Small (Plant Energy Centre of Excellence ARC, University of Western Australia) for sharing predictions of OTP90 targeted editing sites. We appreciate the generous support of Stefan Britsch and Christoph Wiegrefe at Ulm University for letting us use their confocal microscopy equipment and for their kind technical instructions.

FIGURE LEGENDS

Figure 1. OTP90 is dually targeted to mitochondria and plastids in transgenic *Arabidopsis thaliana* plants. A. Mitochondrial localization of OTP90 in *Arabidopsis* roots. DsRed2 signals of OTP90:DsRed2 translational fusion are shown in red, Mito Tracker Green signals in root mitochondria are shown in green and yellow signals show the overlay of OTP90:DsRed2 and Mito Tracker Green signals. B. Chloroplast localization of OTP90 protein in *Arabidopsis* cotyledons. DsRed2 signals of OTP90:DsRed2 are shown in red, auto-fluorescence signals of chloroplast cotyledons are shown in green and yellow signals show the overlay of OTP90:DsRed2 and auto-fluorescence signals of chloroplast cotyledons.

Figure 2. Predicted domain structure of OTP90 and macroscopic phenotypes in *otp90* homozygous plants. A. Predicted PPR, E, and DYW domains of OTP90, location of T-DNA insertion and primers used for genotyping (LP: OTP90-LP, RP: OTP90-RP). B. Seedlings root length of *otp90* (Black column, n=72) is significantly shorter than that of Col-0 (white

RESULTATS

column, n=70) but *otp90* complemented line with *OTP90* locus (gray column n=72) is not different at 12 days after sowing on Murashine and Skoog medium with 16h/8h light/dark cycles. Error bars show confidence interval with $\alpha=0.001$ and asterisks show significant statistical differences between means using Student t-test with $\alpha=0.001$. Representative plants are shown in right panel with white line indicates scale bar of 1cm. C. Aerial fresh weight of *otp90* (Black column, n=33), Col-0 (white column, n=33) and *otp90* complemented plants (gray column n=35) are not statistical different at 22 days after sowing in soil at 22°C days and 16h/8h light/dark cycles. Error bars show confidence interval with $\alpha=0.05$. D. Percentage of seeds under 25mm size of the total seeds production of *otp90* (Black column, n=4) is higher than Col-0 (white column, n=4) plants. Error bars show confidence interval with $\alpha=0.05$. Asterisks show significant statistical differences between means using Student t-test with $\alpha=0.001$.

Figure 3. The *otp90* mutant shows RNA editing defect in 6 mitochondrial sites. Comparison of the cDNA sequence of 6 RNA editing sites in the *ccmFc*, *ccmB*, *nad1*, *ccmC*, *mttB* genes from wild type plant Col-0, *otp90* and *otp90* complemented by *OTP90* locus. Editing efficiency of each editing site are shown as percentage.

Fig. 4 Yeast two hybrid analysis of OTP90 interactions with MORF proteins. The open reading frame for OTP90 was cloned in frame with the binding domain (BD in vector pGBKT7) and was co-transformed into yeast cells with the various MORF proteins in frame with the activation domain (AD) in vector pGADT7. The interactions were evaluated on selective medium with SD-Ade-Leu-Trp-His after 7 days of incubation (7D) and also with 2.5 mM of 3-Amino-1,2,4-triazole (3-AT, +) after 13 days of incubation (13D). Positive control is a test of the known protein interaction using the vectors pGBKT7 p35 with pGADT7 T and the negative control using pGBKT7 Lam with pGADT7 T. The results show strong interaction of OTP90 with MORF1, MORF2, MORF3, MORF4 and MORF6.

Figure 5. BiFC analysis of OTP90 proteins with mitochondrial targeted MORF1, MORF3, and chloroplast targeted MORF2 proteins in tobacco leaves. OTP90 and all MORF are cloned both in YFP-N and YFP-C vectors. Labels in left sides of each set of panel indicate the

RESULTATS

various constructions co-transformed into the cells. Labels above each set of panels indicates the channels observed during the experimentation, protein-protein interacting (YFP) signal in yellow, mitochondrial localization control using AOX target sequence fusion with RFP signal in red (mt), chloroplasts control using auto-fluorescent signal in green (cp) and merged of all 3 fluorescent signals with the differential interference contrast image for each construct combination is also showed (merge). The white bars in each panel indicate the scale bar of 50 μ m. Except for the combination of OTP90 and MORF1, YFP signals were not detected (results of OTP90 with MORF4, MORF5, MORF6 or MORF9 are shown in Supplemental Figure).

Figure 6. Analysis of splicing of mitochondrial genes in *otp90* mutant using RT-qPCR approach. Splicing was measured in three biological replicates (black, white and gray) for all of 9 mitochondrial genes containing introns in *Arabidopsis thaliana*. Evaluation of splicing was performed using the relation showed in the vertical axis for each intron using specific primers as previously described (De Longevialle et al 2007). Bar errors show the standard deviation (SD) of qPCR replicates in each biological repetition.

Figure 7. Blue-native protein gel analysis of the mitochondrial extracts from wild type Col-0, *otp90* and *otp90* complemented by *OTP90*. 7.5 micrograms of mitochondrial extract from Col-0, *otp90* and *otp90* complemented plants were separated in the 4.5%-12% BN gradient gel. The gels were stained for complex I activity (A) prior to staining with silver staining (B). No significant difference in amounts and enzymatic activities of complex I in Col-0, *otp90* and complemented line were detectables.

AUTHOR CONTRIBUTIONS

Designed research: C. L., R. B. and M. L-O.

Performed research: M.L-O contributed to the production and characterization of mutant and transgenic lines, production of plasmids, localization of proteins and editing analyses. M. T.

RESULTATS

contributed to editing analyses in mitochondria, D. V. contributed the mitochondrial protein complexes analysis. B. H. and A. Z. contributed the Y2H and BIFC protein interaction.

Data analysis and write of the article: M.L-O, R. B., C. L., M. T.,

REFERENCES

- Arenas, M.A., Takenaka, M., Moreno, S., Gomez, I., and Jordana, X. (2013). Contiguous RNA editing sites in the mitochondrial nad1 transcript of *Arabidopsis thaliana* are recognized by different proteins. *FEBS Lett* 587:887-891.
- Armengaud, P. (2009). EZ-Rhizo software: the gateway to root architecture analysis. *Plant Signal Behav* 4:139-141.
- Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A. (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol* 42:603-613.
- Barkan, A., et al. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8:e1002910.
- Bentolila, S., et al. (2012). RIP1, a member of an *Arabidopsis* protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A* 109:E1453-1461.
- Bentolila, S., Oh, J., Hanson, M.R., and Bukowski, R. (2013). Comprehensive high-resolution analysis of the role of an *Arabidopsis* gene family in RNA editing. *PLoS Genet* 9:e1003584.
- Borner, G.V., Morl, M., Wissinger, B., Brennicke, A., and Schmelzer, C. (1995). RNA editing of a group II intron in *Oenothera* as a prerequisite for splicing. *Mol Gen Genet* 246:739-744.
- Boussardon, C., et al. (2012). Two Interacting Proteins Are Necessary for the Editing of the NdhD-1 Site in *Arabidopsis* Plastids. *Plant Cell* 24:3684-3694.
- Castandet, B., Choury, D., Begu, D., Jordana, X., and Araya, A. (2010). Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res* 38:7112-7121.
- Chateigner-Boutin, A.L., et al. (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1. *Plant J* 65:532-542.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735-743.
- Cohen, S., et al. (2014). nMAT4, a maturase factor required for nad1 pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in *Arabidopsis* mitochondria. *Plant J*.
- Colcombet, J., et al. (2013). Systematic study of subcellular localization of *Arabidopsis* PPR proteins confirms a massive targeting to organelles. *RNA Biol* 10.
- Covello, P.S., and Gray, M.W. (1989). RNA editing in plant mitochondria. *Nature* 341:662-666.
- Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* 133:462-469.

- 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.
- de Longevialle, A.F., Small, I.D., and Lurin, C. (2010). Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol Plant* 3:691-705.
- Delannoy, E., et al. (2009). *Arabidopsis* tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNA^{Arg}(ACG) and is essential for efficient chloroplast translation. *Plant Cell* 21:2058-2071.
- Doniwa, Y., Ueda, M., Ueta, M., Wada, A., Kadowaki, K., and Tsutsumi, N. (2010). The involvement of a PPR protein of the P subfamily in partial RNA editing of an *Arabidopsis* mitochondrial transcript. *Gene* 454:39-46.
- Farre, J.C., Akinin, C., Araya, A., and Castandet, B. (2012). RNA editing in mitochondrial trans-introns is required for splicing. *PLoS One* 7:e52644.
- Fujii, S., Sato, N., and Shikanai, T. (2013). Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of *Arabidopsis* PROTON gradient regulation3. *Plant Cell* 25:3079-3088.
- Gobert, A., et al. (2010). A single *Arabidopsis* organellar protein has RNase P activity. *Nat Struct Mol Biol* 17:740-744.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., and Grienemberger, J.M. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341:660-662.
- Hammani, K., et al. (2011). The Pentatricopeptide Repeat Protein OTP87 Is Essential for RNA Editing of nad7 and atp1 Transcripts in *Arabidopsis* Mitochondria. *J Biol Chem*. 286:21361-21371.
- Hartel, B., Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2013). MEF10 is required for RNA editing at nad2-842 in mitochondria of *Arabidopsis thaliana* and interacts with MORF8. *Plant Mol Biol* 81:337-346.
- Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A. (1989). RNA editing in plant mitochondria. *Science* 246:1632-1634.
- Hoch, B., Maier, R.M., Appel, K., Igloi, G.L., and Kossel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353:178-180.
- Huang, Y.Y., Matzke, A.J., and Matzke, M. (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* 8:e74736.
- Ichinose, M., Tasaki, E., Sugita, C., and Sugita, M. (2012). A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in *Physcomitrella patens*. *Plant J* 70:271-278.
- Jacobs, J., et al. (2013). Identification of a chloroplast ribonucleoprotein complex containing trans-splicing factors, intron RNA, and novel components. *Mol Cell Proteomics* 12:1912-1925.
- Kagami, H., Nagano, H., Takahashi, Y., Mikami, T., and Kubo, T. (2012). Is RNA editing implicated in group II intron survival in the angiosperm mitochondrial genome? *Genome* 55:75-79.

- Ke, J., et al. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struct Mol Biol* 20:1377-1382.
- Keren, I., et al. (2009). AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria. *RNA* 15:2299-2311.
- Keren, I., et al. (2012). nMAT1, a nuclear-encoded maturase involved in the trans-splicing of nad1 intron 1, is essential for mitochondrial complex I assembly and function. *Plant J* 71:413-426.
- Khrouchtchova, A., Monde, R.A., and Barkan, A. (2012). A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *Rna* 18:1197-1209.
- Klodmann, J., Senkler, M., Rode, C., and Braun, H.P. (2011). Defining the protein complex proteome of plant mitochondria. *Plant Physiol* 157:587-598.
- Kohler, D., Schmidt-Gattung, S., and Binder, S. (2009). The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of Arabidopsis thaliana. *Plant Mol Biol*.
- Kotera, E., Tasaka, M., and Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433:326-330.
- 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.
- Nakagawa, T., et al. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* 104:34-41.
- 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., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* 104:8178-8183.
- Okuda, K., and Shikanai, T. (2012). A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic Acids Res* 40:5052-5064.
- Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci U S A* 108:415-420.
- Pyo, Y.J., Kwon, K.C., Kim, A., and Cho, M.H. (2013). Seedling Lethal1, a pentatricopeptide repeat protein lacking an E/E+ or DYW domain in Arabidopsis, is involved in plastid gene expression and early chloroplast development. *Plant Physiol* 163:1844-1858.
- Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J., and Palmer, J.D. (2013). The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol* 11:29.
- Rudinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V. (2012). Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J Mol Evol* 74:37-51.
- Ruwe, H., Castandet, B., Schmitz-Linneweber, C., and Stern, D.B. (2013). Arabidopsis chloroplast quantitative editotype. *FEBS Lett* 587:1429-1433.
- Salone, V., et al. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581:4132-4138.

- Schagger, H., Cramer, W.A., and von Jagow, G. (1994). Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* 217:220-230.
- Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13:663-670.
- Small, I.D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25:46-47.
- Sosso, D., et al. (2012). PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* 24:676-691.
- Stern, D.B., Goldschmidt-Clermont, M., and Hanson, M.R. (2010). Chloroplast RNA metabolism. *Annu Rev Plant Biol* 61:125-155.
- Sun, T., et al. (2013). An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc Natl Acad Sci U S A* 110:E1169-1178.
- Takenaka, M., and Brennicke, A. (2012). Using multiplex single-base extension typing to screen for mutants defective in RNA editing. *Nat Protoc* 7:1931-1945.
- Takenaka, M., Verbitskiy, D., Zehrmann, A., and Brennicke, A. (2010). Reverse genetic screening identifies five E-class PPR-proteins involved in RNA editing in mitochondria of Arabidopsis Thaliana. *J Biol Chem*.
- Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013a). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* 8:e65343.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Hartel, B., and Brennicke, A. (2013b). RNA editing in plants and its evolution. *Annu Rev Genet* 47:335-352.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelman, M., Hartel, B., and Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A* 109:5104-5109.
- Tillich, M., et al. (2009). Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A* 106:6002-6007.
- Tissier, A.F., et al. (1999). Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. *Plant Cell* 11:1841-1852.
- Verbitskiy, D., Hartel, B., Zehrmann, A., Brennicke, A., and Takenaka, M. (2011). The DYW-E-PPR protein MEF14 is required for RNA editing at site matR-1895 in mitochondria of Arabidopsis thaliana. *FEBS Lett* 585:700-704.
- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8:e57286.
- Yin, P., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504:168-171.
- Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. *Plant Cell* 21:558-567.

RESULTATS

- Zhu, Q., et al. (2012). SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J* 71:836-849.
- Zmudjak, M., et al. (2013). mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in *Arabidopsis*. *New Phytol* 199:379-394.
- Zoschke, R., Watkins, K.P., and Barkan, A. (2013). A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* 25:2265-2275.

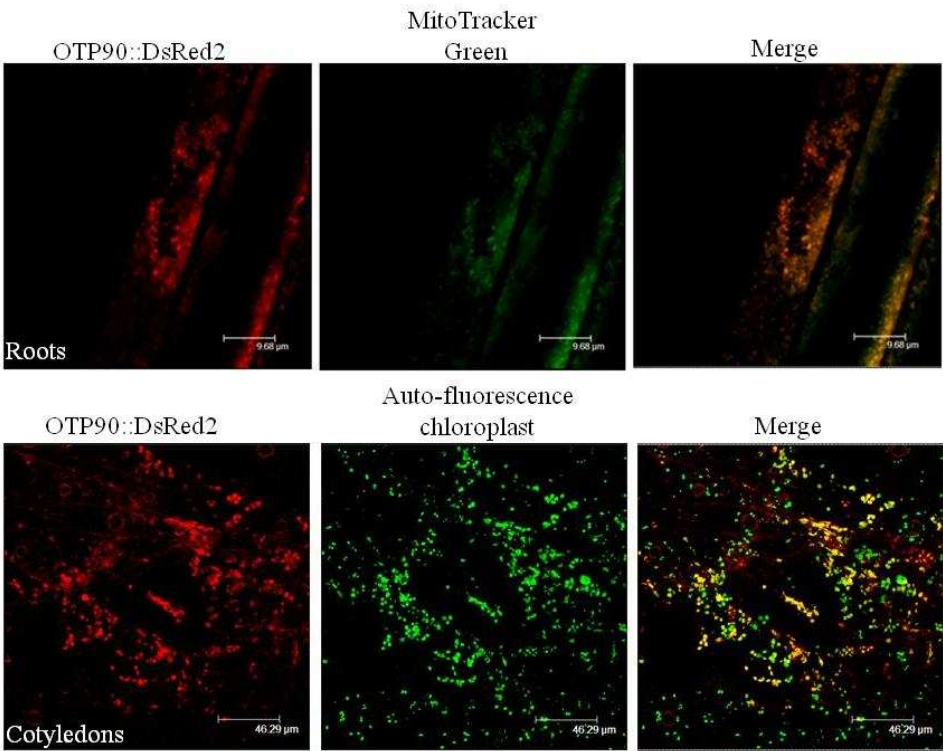


FIGURE 1

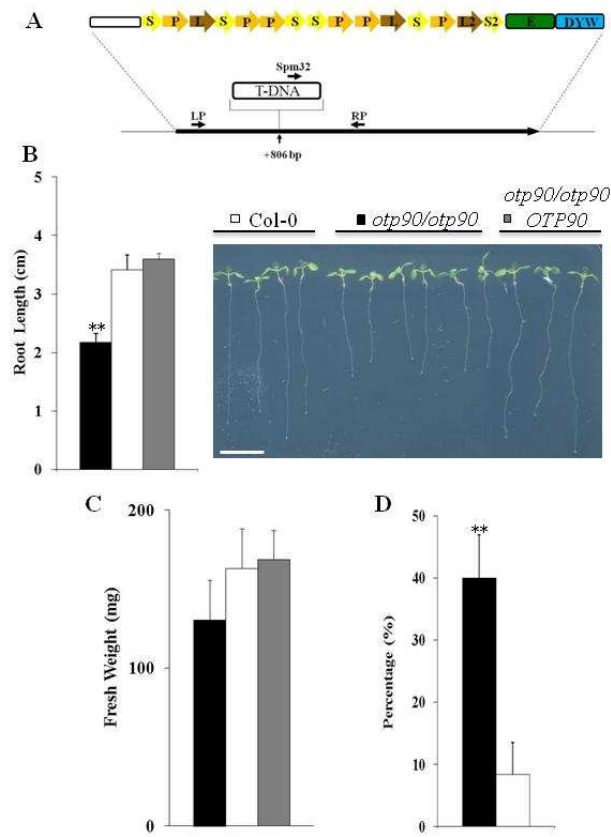


FIGURE 2.

RESULTATS

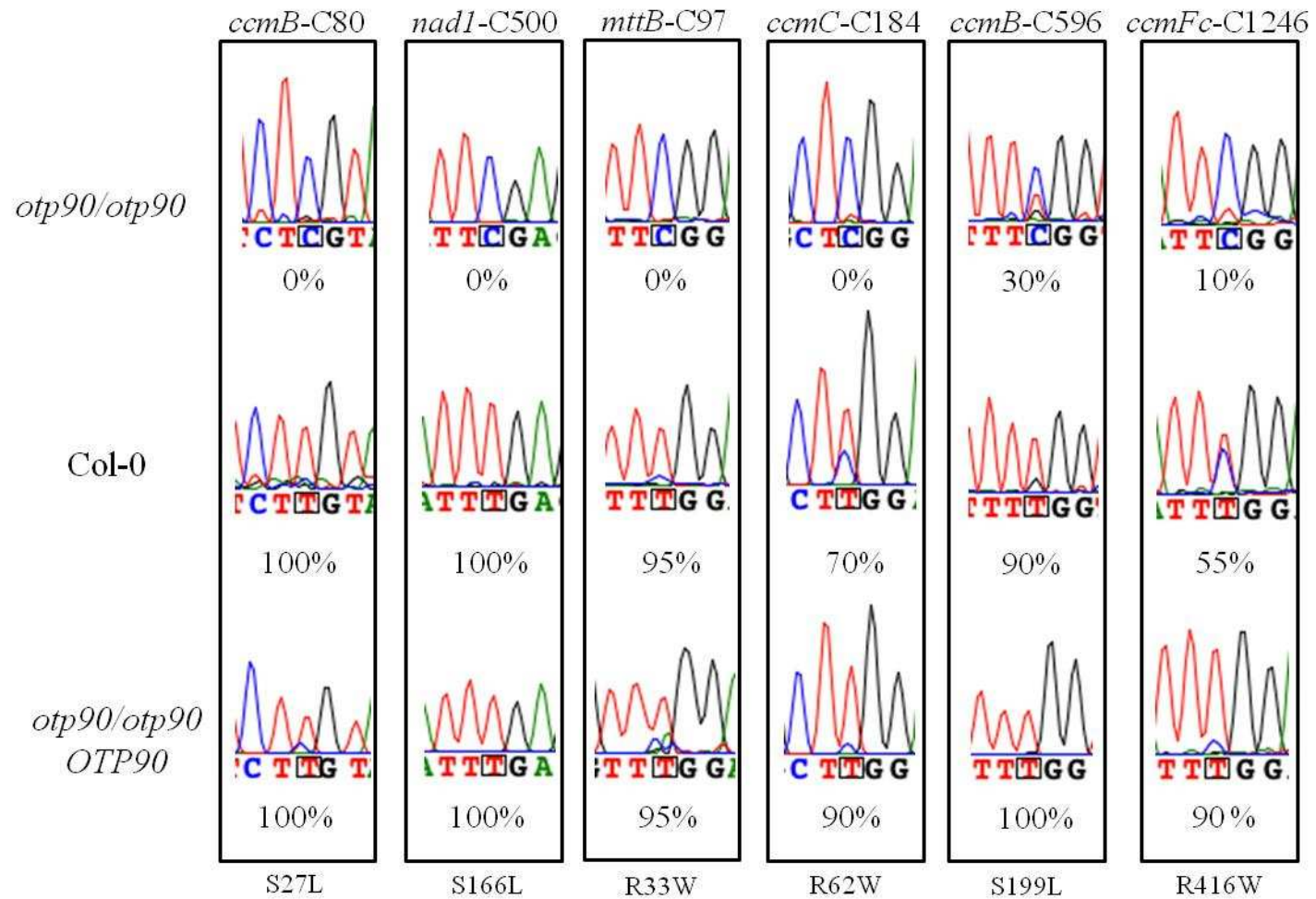


FIGURE 3.























	BD OTP90	
	7D	13D
3-AT	-	+
AD MORF1		
AD MORF2		
AD MORF3		
AD MORF4		
AD MORF5		
AD MORF6		
AD MORF7		
AD MORF8		
AD MORF9		
Positive control		
Negative control		

FIGURE 4.

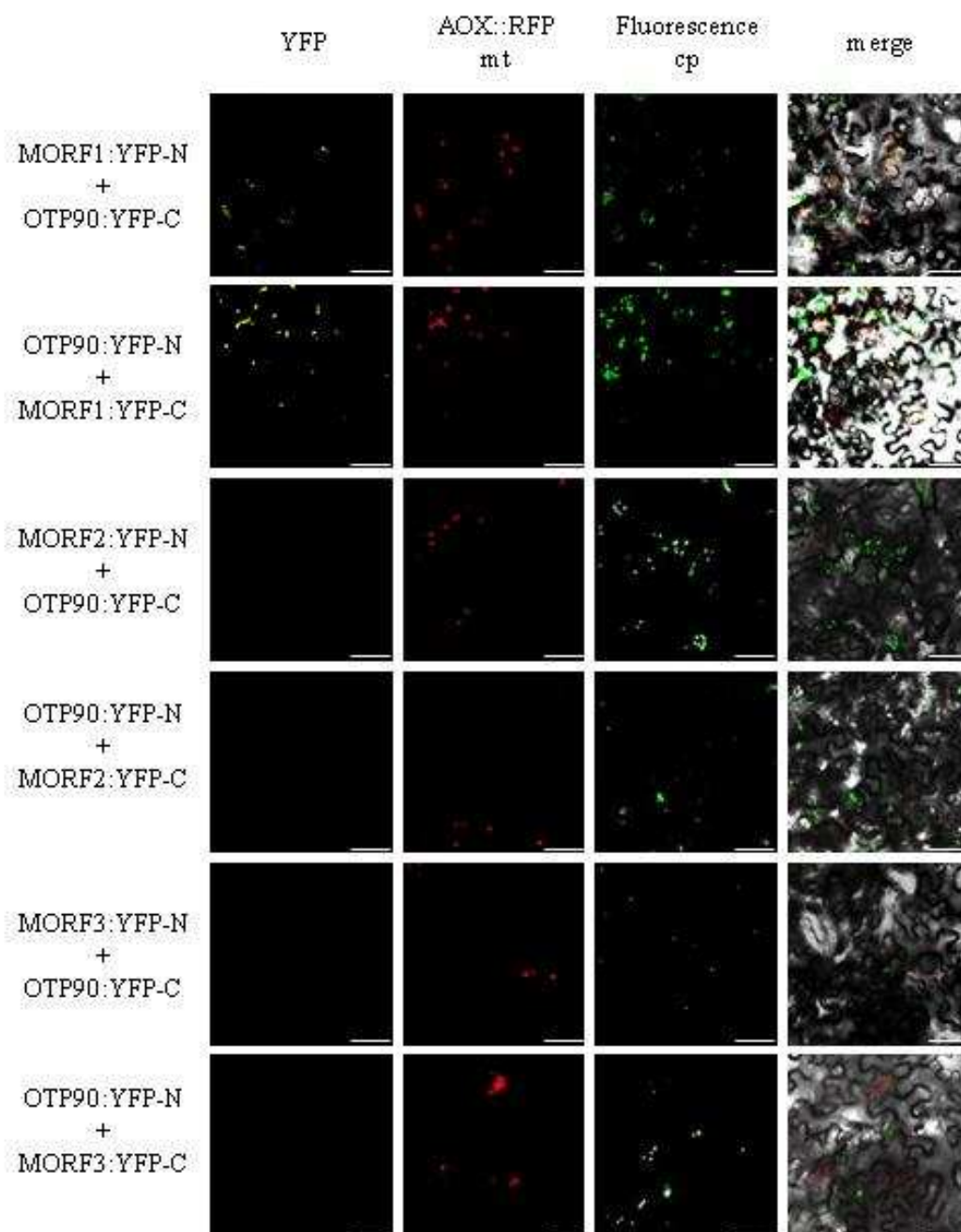


FIGURE 5.

RESULTATS

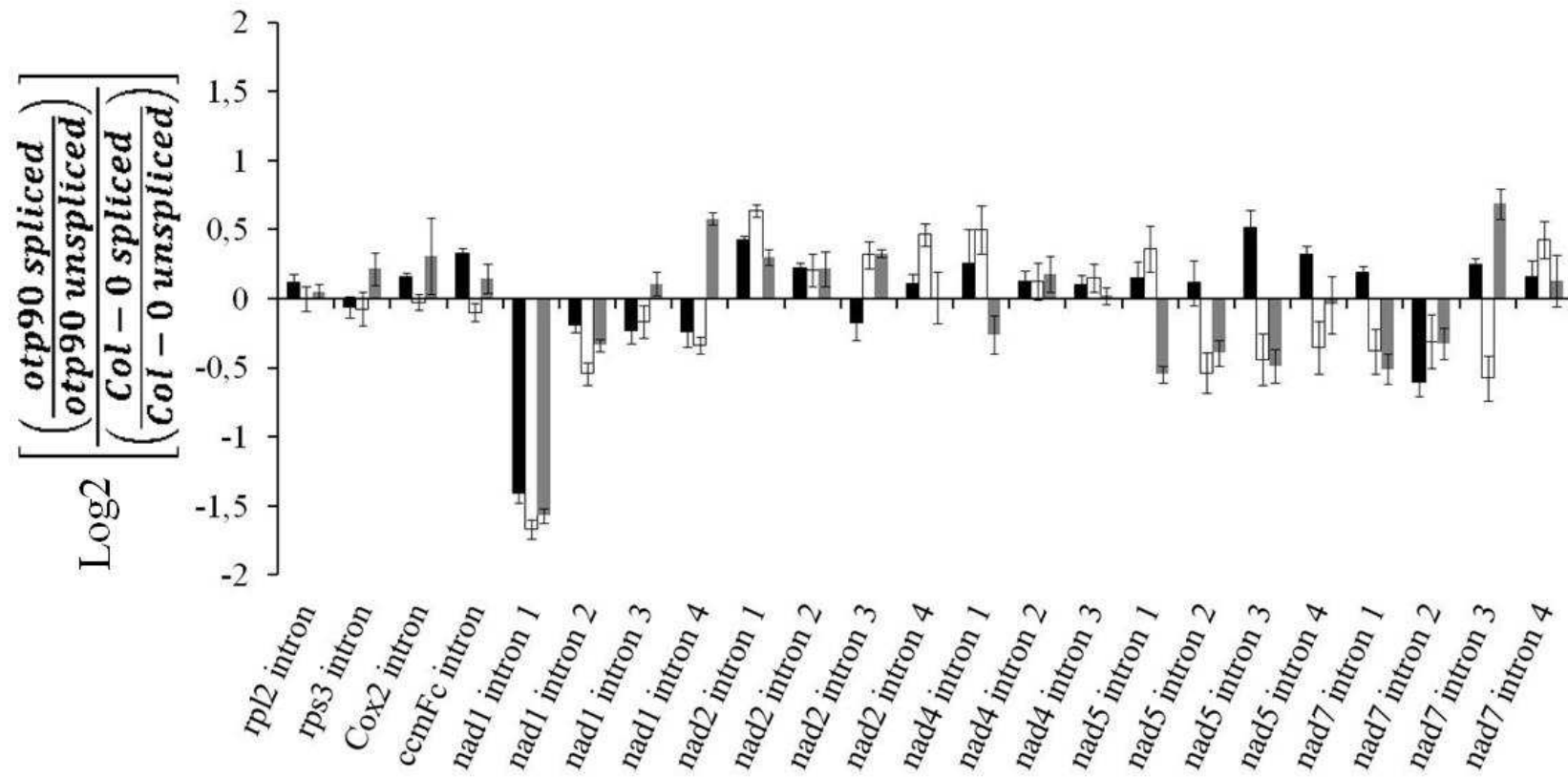


FIGURE 6.

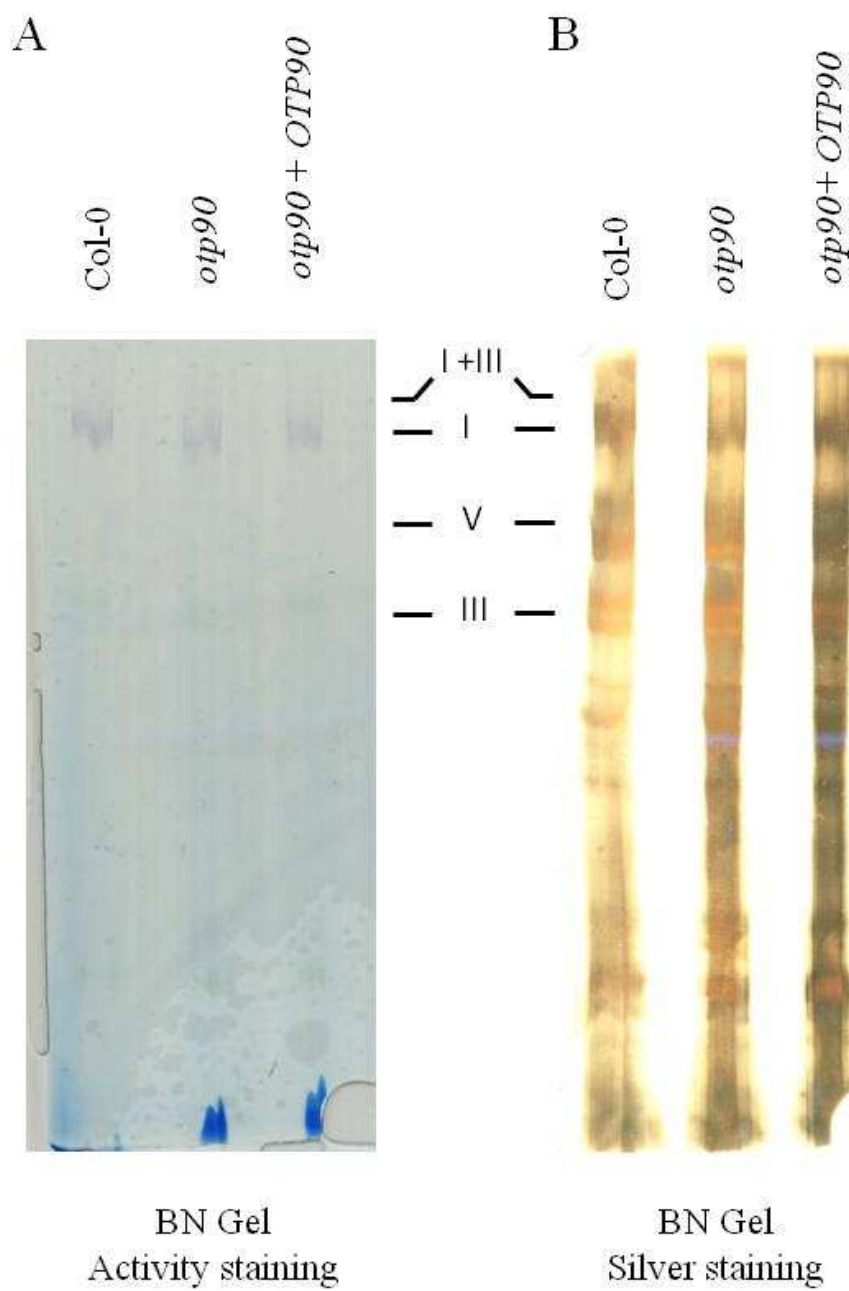
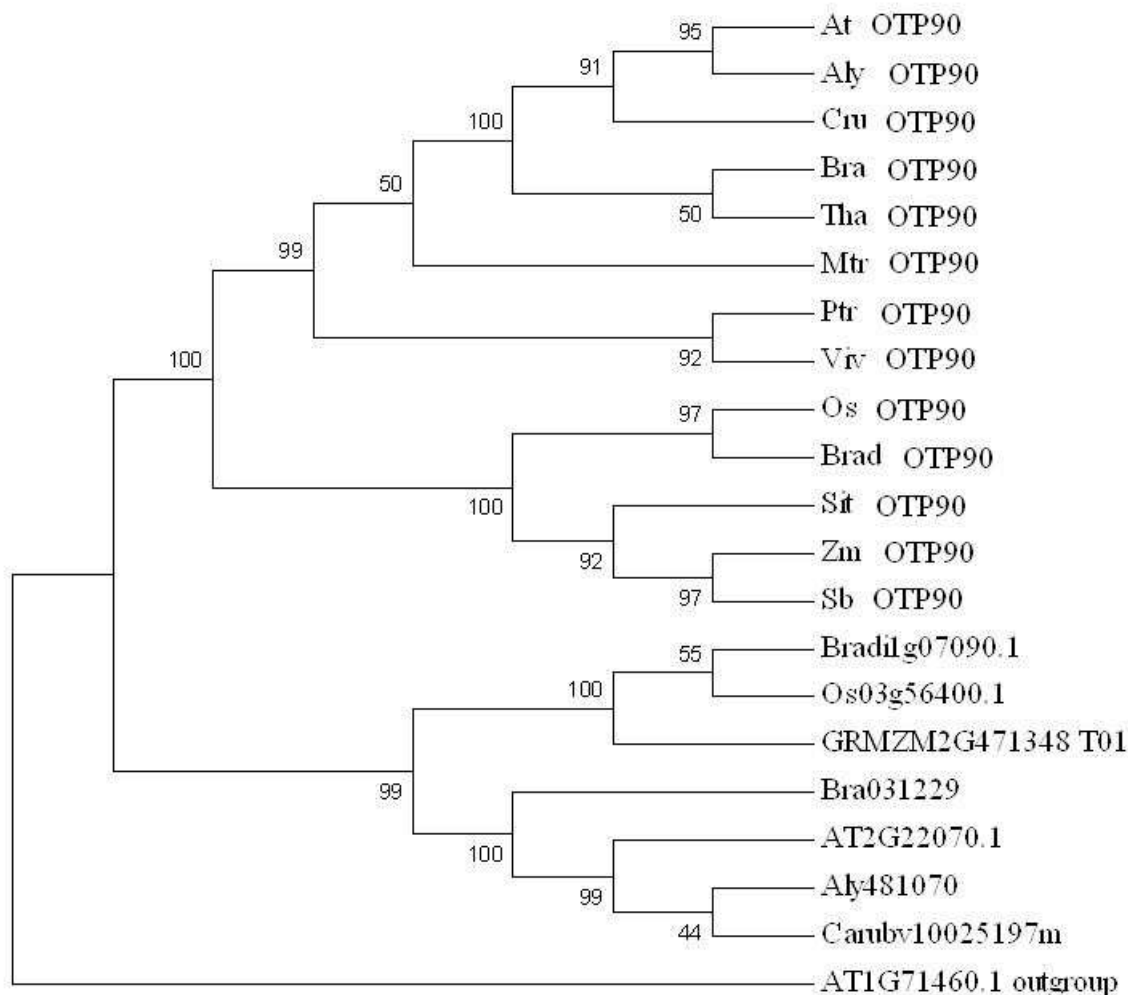


FIGURE 7.



Supplemental Figure 1. Phylogenetic relationship of OTP90 ortholog proteins in 13 angiosperms species. OTP90 protein sequences were obtained by blast from selected genomes in phytozome web page V9.1 Best matches in each species were blasted to others genomes and the most related sequences to OTP90 and AT2G22070 were recovered. Phylogenetics relationship was constructed in MEGA5 software (Tamura *et al.*, 2011) using sequence alignments obtained by MUSCLE. Maximum likelihood method using the best model fitting the data, the Jones-Taylor-Thornton (JTT) model and gamma distribution (G) with 1000 bootstraps was used to obtain the tree. *Arabidopsis thaliana* OTP90 (At OTP90: AT1G25360) orthologs in other species are: *Arabidopsis Lyrata* (Aly OTP90: 472839), *Capsella rubella* (Cru OTP90: Carubv10008344m), *Thellungiella halophila* (Tha OTP90: Thhalv10006846m), *Populus trichocarpa* (Ptr OTP90: Potri.002G027800.1), *Medicago truncatula* (Mtr OTP90: Medtrlg073300.1), *Brasica rapa* (Bra OTP90 : Bra010992), *Zea mays* (Zm OTP90: GRMZM2G057176_T01), *Sorghum bicolor* (Sb OTP90:Sb02g037960.1), *Setaria italica* (Sit OTP90: Si032758m.g), *Oriza sativa* (Os OTP90: Os07g39910.1), *Vitis vinifera* (Viv OTP90: GSVIVG01009866001), *Brachypodium distachyon* (Brad OTP90: Bradilg22990.1).

```

* * * * *
Mtr_OTP90 -----MDVRDLAYRYLTQLNHV-----STTQIIARAVHAHILTSGF---KPNTFILNRLINIYCKSSNITYARKLFDKIPK
Ptr_OTP90 ----MTTPDSIRTAANRYGSLQLICCLQS-----PISYSLARPVHAHMIASGF---QPRGHILNRLIDIYSKSSKLNRYARQLFDEIPQ
At_OTP90 ---MQPNPDLVRAIANRYAANLRLCLPLR-----RTSLQLARAVHGNIITFGF---QPRAHILNRLIDVYCKSSELNYARQLFDEISE
Aly_OTP90 ---MQPNPDLVRAIANRYAANLRLCLPLR-----RTSLQLARAVHGNIITFGF---QPHAHILNRLIDVYCKSSELDYARQLFDEISE
Bra_OTP90 -----
Cru_OTP90 ---MQPNPDLVRAIANRYAANLRLCLPIR-----RTSLQLARAVHANIITFGF---QPRAHILNRLIDVYCKSSELRYARELFDEISE
Tha_OTP90 ---MQPNPDLVRAIANRYAANLRLCLPLR-----RTSLQLARAVHGNIITFGF---QPRAHILNRLIDVYCKSSELHYARQLFDEISE
Zm_OTP90 -MPRPPPPPHGVTSLPHQCSVLLRRLAASRSLASS---SYFLRALRCLHARLLTSALLHAPPHPHLTLRLIHLTYTLSRDLPAAATLFCFCA--D
Sit_OTP90 -MLPPPPPPPAVTSLPYQCSVLLRRLVASRSLASS---TSFLRALRRLHARLLTAGLLHAQSHPHLTLRLIHLTYTLSDLPAAALLFRS--N
Sb_OTP90 MPPPPPARPQGVTSPLPHQCSVLLRRLAASRSLTS---SSFLRALRCLHARLLTSALLHAPPHPHLTLRLIHLTYTLSGDLPAAATLFRFRA--D
Os_OTP90 MRPWRPPPTAVASLAHQCSLLLRRLAERHSPAPTCPSSSFLRALRCLHARLLTADLLHAPSHPHLTLRLIHLTYTLSPDLATPAALFRSDPD
Brad_OTP90 --MPPPPLLAVASLPHQCSVLLRRLAARHSPVPSSP--SSFLRTLRLCLHARLLTAALLHDPSPHPLTLRLIHLTYTLSPDLATPAVLFRFRA--D
Viv_OTP90 ----MRNAANVRRLANQYAAQLQCCPHN-----PMSYSIARTVHAHMIASGF---KPRGYILNRLIDVYCKSSDLVSAHHFLFDEIRQ

```

```

* * * * *
Mtr_OTP90 PDIVARTTLLSAYSSSGNVKLAQQLFNATPLTIRDTVSYNAMITAYSHGNDGHAALNLFVQMKR--YGFLLPDPFTFSSVLSALSILIAD-EER
Ptr_OTP90 PDIVARTTLIAAYSAAAGDLKLSRKIFSDTPLGMRDSVFYNAMITAYSHNHDGHAAIELFCMDQR--DNFRPDNYTFTSVLGALALVAE-KEK
At_OTP90 PDKIARTTMVSGYCASGDITLARGVFEKAPVCMRDTVMYNAMITGFSHNNDGYSAINLFCKMKH--EGFKPDNFTFASVLAGLALVAD-DEK
Aly_OTP90 PDKIARTTMVSGYCASGDIALARSVFEETPVSMRDTVMYNAMITGFSHNNDGYSAINLFCKMKH--EGFKPDFTYASVLAGLALVVD-DEK
Bra_OTP90 -----
Cru_OTP90 PDKIARTTMVSGYCASGDITLARGVFEETSVSMRDTVMYNAMITGFSHNNDGYSAIKLFCKMKH--EGFQPDNFTFASVLAALALVVE-EEK
Tha_OTP90 PDKIARTTMVSGYCASGDVTLARCVEETPLSMRDTVMYNAMITGFTHNHDGYSAINMFCCEMKR--QGFKPDNFTFASVLAGLALVVE-EEK
Zm_OTP90 PCPVSAATSLVAAYAAADRLPAAVSFFDAVPPARRDTVLHNAVISAYARASHAAPAVAVFRSLLAS-GSLRPDDYSFTALLSAAHLPNLSVR
Sit_OTP90 PCPVAATSLVAHAAAGRLPAAVSFFDAVPPARRDTVLHNAVISAYARASHAAPAVAVFRSLLAS-GSLRPDDYSFTALLSAAHLPNLSVR
Sb_OTP90 PCPVAATSLVAAYAAAGRLPAAVSFFDAVPQARRDTVLHNAVISAYARASHAAPAVAVFRSLLAS-GSLRPDDYSFTALLSAAHLPNLSVR
Os_OTP90 PGPVAATSLVAHAAAGRLRDAAFFDAVPPARRDTVLHNAVISAFARASLAAPAVSVFHALLS-GSLRPDDYSFTALISAVGQMHNLAAP
Brad_OTP90 PGPVAATSLVSAYAVAGRLRDSAAFFDSVPVARRDTVLHNAVISAFARASLAAPAVSVFRSLLASDDSLRPDDYSFTSLLSAGVQMHDLAVS
Viv_OTP90 PDIVARTTLIAAHSSAGNSNLAREIFFATPLGIRDTCYNAMITGYSHNNDGFGAIELFRDLLR--NGFRPDNFTFTSVLGALALIVE-DEK

```

* * ** *** * *** * ***

Mtr_OTP90 HCQMLHCEVIKLGITLLIPSVTNALLSCYVCCASSPLVKSSQLMASARKVFDETPKNQIYEPSWTTMIAGYVRNDDLVAARELLDGLTYPID-

Ptr_OTP90 HCQQLHCAVVKSGTGFVTSVLNALISSYVKCAASPSAQSSSLMAEARKLFDEMPNRD--ELSWTTIITGYVKNNDLDAAKEFLNGTSKKLG-

At_OTP90 QCVQFHAAALKSGAGYITSVSNALVSVYSKCASSP-----SLLHSARKVFDEILEKD--ERSWTTMMTG YVKNGYFDLGEELLEGMDDNMKL

Aly_OTP90 QCVQFHAAALKSGAGYVTSVSNALVSVYSRCASSP-----SLLHSARKVFDDIPEKD--ERSWTTMMTG YVKNGCFDLGKELLKGMDENMKL

Bra_OTP90 -----MTGYVKNGFDSGLELLEGMDEGMKL

Cru_OTP90 QCLQFHAAALKSGAGSITSVSNALVSVYSKCASSP-----SLLHSARKVFDEMPEKD--ERSWTTMITGYVKNGCFDLGKELLEVM DENMKV

Tha_OTP90 DCVQFHAAAMKSGAGYITSVLNALVSVYSKCA-----SMLSSARKVFDEIPEKD--ERSWTTMITGYVKNGFDSGQDLLEGMDENMRL

Zm_OTP90 HCAQLHCSVLKSGAGGALSVCNALVALYMKCESP-----EATRDARKVLDEMPNKD--DLTWTTMVVGYVRRGDVGAARSVFEEVDGKFD-

Sit_OTP90 HCAQLHCSVLKSGAGGALSVCNALIALYMKCEAP-----VATREARKVLDEMPAKD--ELTWTTMVVGYVRRGDVGAARSVFEEVDGKFD-

Sb_OTP90 HCAQLQCSVLKSGAGGVLSVSNALVALYMKCEAL-----EATRDARKVLDEMPDKD--ALTWTTMVVGYVRRGDVGAARSVFEEVDVKFD-

Os_OTP90 HCTQLHCSVLKSGAAAVLSVSNALIALYMKCDTP-----EASWDARKVLDEMPDKD--DLTWTTMVVGYVRRGDVNAARSVFEEVDGKFD-

Brad_OTP90 HCTQLHCAVHKLGAVALSVSNALIALYMKCDAP-----GVTRDARKVLDEMPKD--ELTWTTIVVGHVRKGDVHAARSFAFEEIDGEFD-

Viv_OTP90 QCQQIHCAVVKSGSGFVTSVLNALLSVFVKCASSPLVSSSSSLMAAARKLFDEMTED--ELSWTTMIAGYVRNGELDAARQFLDGMTEKLV-

Mtr_OTP90 VAWNAMISGYVRRGLYEEAFDTRFRMHSMGIQEDEYTYTSLISACGSCNEKMGMFNCGRQVHG YILRTVVEPSHHFVLSVNNAITFYTKYD

Ptr_OTP90 VAWNAMISGYAHRGLYLEAFEMFRKMIMSKIQLDEFTFTSVISVCANA---GCFRLGKEMHAYFLKTVANPAPDVAMPVNNAITFYWKCG

At_OTP90 VAYNAMISGYVNRGFYQEALEMVRRMVSSGIELDEFTYPSVIRACATA---GLLQLGKQVHAYVLRRE----DFS FHF DNSLVS LYKCG

Aly_OTP90 VAYNAMISGYVNCGLYQEALEMVRRMVSSGIELDEFTYPSVIRACATA---RLQLGKQVHAYVLRRE----DFS FHF DNSLVS LYKCG

Bra_OTP90 VAYNALISGYVHRGLHREALMVRMVSSGIELDAFTYPSVIRACATS---GLLQFGKQVHAYALRRQ----DFS FHF DNSLVS LYKCN

Cru_OTP90 VAYNAMISGYVNRGLYQEALEMVRRMVSSGIELDEFTYPSVIRACATA---GLLQLGKQVHAYVLRRE----DFS FHF DNSLVS LYKCG

Tha_OTP90 VAYNAMISGYVHRGLYQEALEMVRRMVSSGIELDEFTYPSLIRACATA---RLQLGKQVHAYVLRRE----NFS FHF DNSLVS LYKCG

Zm_OTP90 VVWNAMISGYVHSGMAVEAFELFRRMVLERVPLDEFTFTSVLSACANV---GLFAHGKSVHGQ IIRLQPNFV EAAI PVNNALVTFY SKCG

Sit_OTP90 VVWNAMISGYVQSGMVVEAFELFRRMV LARALLDEFTFTSVLSACANA---GFFVLGKSVHGQ IIRLQPDFVPEAALPVNNALVTFY SKAG

Sb_OTP90 VVWNAMISGYVHSGMVVEAFELFRRMVLERVPLDEFTFTSVLSACANA---GFFAHGKSVHGQ IIRLQPNFVPEAALPVNNALVTLY SKCG

Os_OTP90 VVWNAMISGYVQSGMCADAFELFRRMVSEKVP LDEFTFTSVLSACANA---GFFVHGKSVHGQ IIRLQPNFVPEAALPVNNALVTLY SKGG

Brad_OTP90 VVWNAMISGYVQSGMCAEAFELFRRMVSKRIPPDEFTFTSVLSACANA---GFFLHGKSVHGQ IIRLQPDFVPEAALPVNNALVTLY SKSG

Viv_OTP90 VAWNAMISGYVHHGFFLEALEMFRKMYLLGIQWDEFTYTSVLSACANA---GFFLHGKQVHAYILRTEPRPSI DFSI SVNNALATLY WKCG

* * * * * * * * * * * * *

Mtr_OTP90 RMIEARRVEDKMPVRDIISWNAVLSGYVNAQRIEEANSIFSEMPERNVLTWTVMISGLAQNGFGEEGLKLFNQMKSEGLEPCDYAFAGAITA

Ptr_OTP90 KVDIAQEIFNKMPERDLVSWNIILSGYVNVRCMDEAKSFFNEMPEKNILSWIIMISGLAQIGFAEEALKFFNRMKLQGFEP CDYAFAGAIIS

At_OTP90 KFDEARAI FEKMPAKDLVSWNALLSGYVSSGHIGEAKLIFKEMKEKNILSWMIMISGLAENGFGEEGLKLFSCMKREGFEPCT YAFSGA IKS

Aly_OTP90 KFNEARAI FEKMPAKDLVSWNALLSGYVSSGHIGEAKLIFKEMKEKNILSWMIMISGLAENGFGEEGLKLFSCMKREGFEP CDYAFSGA IKS

Bra_OTP90 KFVEARAI FEKMPAKDLVSWNALLSGYVSSGHIGEAKLLFKEMEKEKNILTWMIMISGLADNGFGEEGLRLFS CMRKEGFEP CDYAFSGA IKS

Cru_OTP90 KFDEARAI FEKMPAKDLVSWNALLSGYVSSGHIGEAKLIFKEMKEKNILTWMIMISGLAENGFGEEGLKLFSCMKKEGFEP CDYAFSGA IKS

Tha_OTP90 KFGEARAI FEKMPAKDLVSWNALLSGYVSSGHIGEAKLLFKEMEKEKNILTWMIMISGLADNGFGEEGLRLFS CMRKEGFEP CDYAFSGA IKS

Zm_OTP90 NIAVARRIFDNMTLKD VVSWNTILSGYVSSCLDKAVEVFKEMPYKNELSWMVMVSGYVHGGFAEDALKLFNKMRS ENVPCTI YTYAGAI AA

Sit_OTP90 KIAVAKRIFDSMKS KD VVSWNTILSGYVSSCLDKAVEVFKEMPYKNELSWMVMVSGYVQGGRAEDALKLFNWMRADDVKPCDYTYAGAI AA

Sb_OTP90 NIAVARRIFDNMKS KD VVSWNTILSGYVSSCLDKAVEVFKEMPYKNELSWMVMVSGYVHGGFSEDALKLFNMRRAEDVKPCDYTYAGAI SA

Os_OTP90 KIVIAKRIFDTMNLKD VVSWNTILSGYIDSGCLDKAVEVFKEMPYKNELSWMVMVSGYVHGGFSEDALKLFNQMRRAEDVKPCDYTYAGAI AA

Brad_OTP90 KIAVATKIFDSMTLKD VVSWNTILSGYIESGCLDNAARIFKEMPYKSEL SWMVMVSGYVHGGFSEDALKLFNQMRSEDVKPCDYTYAGAVAA

Viv_OTP90 KVDEARQVFNQMPVKDLVSWNAILSGY-----

* ** * * * ** ** ** *

Mtr_OTP90 CSVLGSLDNGQQIHSQVIRLGHDSGLSAGNALITMYSRCGVVESAESVFLTMPYVDSVSWNAMIAALAQHGHGVKAIELFEQMMKEDILPDR

Ptr_OTP90 CSVLGSLKHGRQLHAQVVRGYESSLSAGNALITMYARCGVDAAHCLFINMPCVDAISWNAMIAALGOHQGTQAIELFEEMLKEGILPDR

At_OTP90 CAVLGAYCNGQQYHAQLLKIGFDSSLSAGNALITMYAKCGVVEEARQVFRITMPCLDSSVSWNALIAALGOHGHGAFAVDVYEEMLKKGIRPDR

Aly_OTP90 CAVLGAYCNGQQFHAQLVKIGFDSSLSAGNALITMYAKCGVVEEAQQVFRITMPCLDSSVSWNALIAALGOHGHGVEAVDVYEEMLKKGIRPDR

Bra_OTP90 CAVLGAYCNGQQFHAQLVKIGFDSSLSAANALITKYAKCGVVEEARVFRITMPCQDSVSWNALIAALGOHGHGAFAVGVYEEMLKEGIKPDR

Cru_OTP90 CAVLGAYCNGQQFHAQLVKIGFDSSLSAGNALITMYAKCGVDEARQVFRITMPCLDSSVSWNALIAALGOHGHGAFAVDVYEEMLKEGIRPDR

Tha_OTP90 CAVLGAYGNGQQFHAQLVKIGFDSTLSAGNALITMYAKCGVVEEARQVFLAMPQDSVSWNALIAALGOHGHGTEAVDAYEEMLKEGIRPDR

Zm_OTP90 CGELGALKKHGKQLHGHIVQLGFEGSNLSAGNALITMYARCGAVKEAHLMLVMPNIDSLSVSWNAMISALGOHGHGREALELFDQMVVAEGITPDR

Sit_OTP90 CGELGALKKHGKQLHGHVQLGFEGSNSAGNALITMYAKCGAVKEAHLVFLVMPNIDSLSVSWNAMISALGOHGHGREALDLFDQMVSEGITPDR

Sb_OTP90 CGELGSLKHGKQLHGHVQLGFEGSNSAGNALITMYARCGAVKEANMLVMPNIDSLSVSWNAMISALGOHGHGREALELFDQMVVAEGITPDR

Os_OTP90 CGELGALKKHGRQLHAHLVQCGFEASNSAGNALLTMYAKCGAVNDARLVFLVMPNIDSLSVSWNAMISALGOHGHGREALELFDQMVVAEGIDPDR

Brad_OTP90 CGELGALKKHGKQLHAHLVQCGFEASNSAGNALLTMYARCGAVKDARLVFLVMPNIDSLSVSWNAMISALGOHGHGREALELFDQMVVAQGITPDR

Viv_OTP90 -----AAHCLFLTMPYLDLSVSWNAMIAALGOHGHGAFALELFEMLKEDILPDR

* ** * * * ** ** ** *

Mtr_OTP90 ITFLTILTACNHAGLIIKEGRHYFDTMCTRYGITPGEDHYARLIDLLCRAGMFLKAQSVIKSMPFEAGAPIWEALLAGCRIHGNMELGIIQAAD

Ptr_OTP90 ISFLTIVISACSHAGLVKEGRKYFDSMHNVYGVNPDEEHYARIIDLLCRAGKFSEAKEVMESMPFEAGAPIWEALLAGCRIHGNIDLGIEAAE

At_OTP90 ITLLTVLTACSHAGLVQGRKYFDSMETVYRIPPQADHYARLIDLLCRSGKFSDAESVIESLPFKPTAEIWEALLSGCRVHGNMELGIIAAD

Aly_OTP90 ITFLTIVLTACSHAGLVQGRKYFNSMETVYRIPPQADHYARLIDLLCRSGKFSDAESVIESLPFKPTAEIWEALLSGCRVHGNMELGIIAAD

Bra_OTP90 ITFLTIVLTACSHAGLVQGREYFDSMETVYRIPPQADHYARLIDLLCRSGRFTAEESVIESLPFKPTAEIWEALLSGCRVYGNMELGIIAAE

Cru_OTP90 ITLLTVLTACSHAGLVQGRKYFDSMETIYRIPPQVDHYARLIDLLCRSGKFSDAESVIESLPFKPTAEIWEALLSGCRVHGNMELGIIAAD

Tha_OTP90 ITFLTIVLTACSHAGLVEQGRYFDSMETFYSIHPGPDHYARLIDLLCRSGRFTAEESVIESLPFKPTAEIWEALLSACRVYGNMELGIIAAE

Zm_OTP90 ISFLTIVLTACNHAGLVDEGFYFESMKRDFGIIPGEDHYARLIDLLGRAGRIGEARDLIKTMPPFEPTPSIWEALLSGCRTSGDMELGAHAAD

Sit_OTP90 ISFLTIVLTACNHAGLVDEGFYFESMKRDFGIIPGEDHYARLIDLLGRAGRIGEARDLIKTMPPFEPTPSIWEALLSGCRINGDMELGAYAAD

Sb_OTP90 ISFLTIVLTACNHAGLVDEGFYFESMKRDFGIIPGEDHYARLIDLLGRAGRIGEARDLIKTMPPFEPTPSIWEALLSGCRTSGDMELGAHAAD

Os_OTP90 ISFLTILTACNHAGLVDEGFHYFESMKRDFGISPGEDHYARLIDLLGRSGRIGEARDLIKTMPPFEPTPSIWEALLSGCRTNGDMELGAYAAD

Brad_OTP90 ISFLTILTACNHAGLVDDGFYFESMERDFGISPGEDHYARLIDLLGRAGRIGEARDLIKTMPPFEPTPAIWEALLSGCRINGDMELGAYAAD

Viv_OTP90 ITFLTIVLSTCASHAGLV-----

* * * * * * * * * * * * * * *

Mtr_OTP90 RLLELIPQGDGTYYIILSNMYAALGQWDEVARVRLLMRERGVKKEPGCSWVEVENMVHVFLVDDARHPEVQAVYTYLQQVLNEMKKLGYPVDT

Ptr_OTP90 RLFELKPQHDGTYYVLLSNMYAVAGQWDMKVRKLMRDRGVKKEPGCSWIEVENKVHSFLVGDANHPEVRQIYNYLEQLVLEMRKIGYPVDT

At_OTP90 KLFGLIPEHDGTYYMLLSNMHAATGQWEEVARVRKLMRDRGVKKEVACSWIEMETQVHTFLVDDTSHPEAEAVYIYLODLGKEMRRLGYPVDT

Aly_OTP90 KLFGLIPEHDGTYYMLLSNMYAATGQWEEVARVRKLMRDRGVKKEVACSWIEMETQVHTFLVDDTSHPEAEAVYKYLQDLGKEMRRLGYPVDT

Bra_OTP90 KLFELIPEHDGTYYMLLSNMYAATGKWEAARVRKLMRDRGVKKEVACSWIEVETQVHKFLVDDTSHPEAEAVYKYLQELGKEMRRFGYPVDT

Cru_OTP90 KLFHLIPEHDGTYYMLLSNMYAATGQWDEVARVRKLMRDRGVKKEVACSWIEVETQVHTFLVDDTSHPEAEAVYNYLQELGKEMRRLGYPVDT

Tha_OTP90 KLFELIPEHDGTYYMLLSNMYAATGQWEEMARVRKLMRDRGVKKEVACSWIEVETQVHTFLVDDTSHPEAEAVYNYLQELCEEMRRFGYPVDT

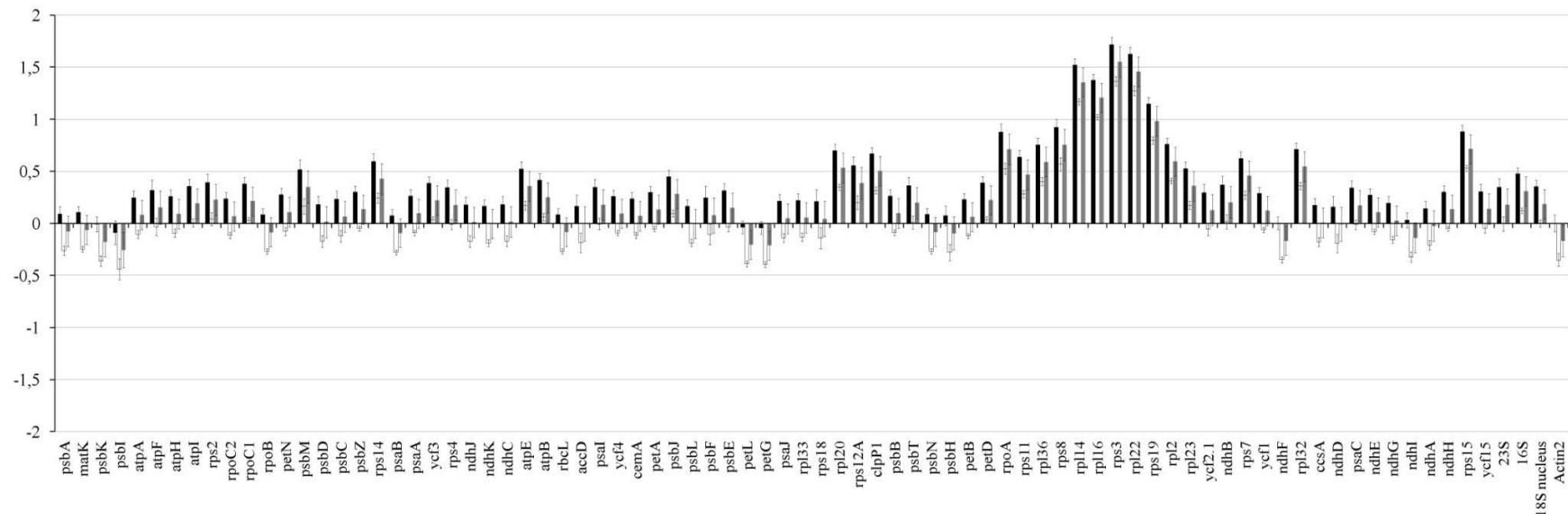
Zm_OTP90 QLFKMTPOHDGTYYIILSNTYSAAGCWVDAARVRKLMRDRGVKKEVACSWIEAGNKVHVFLVGDTHKHEAHEVYKFLQELGKEMRRLGYPVDT

Sit_OTP90 QLFKMTPOHDGTYYIILSNTYSAAGRWVDAARVRKLMRDRGVKKEVACSWIEVGNKVHVFLVGDTHKHEAHEVYHFLQELGKEMRRLGYPVDT

Sb_OTP90 QLFKMTPOHDGTYYIILSNTYSAAGRWVDAARVRKLMRDRGVKKEVACSWIEAGNKVHVFLVGDTHKHEAHEVYKFLQELGKEMRRLGYPVDT

Os_OTP90	QLFRMIPQHDGTYILLSENTYSAAGRWDAAARVRKLMRDRGVKKEPGCSWIEVGSKIHVFLVGD TKHPEAQEVYQFLIEVIGARMRKLGYVPDT
Brad_OTP90	QLFKMVP EHDGTYILLSENTYSAAGRWDAAARVRKLMRDRGVKKEPGCSWIEVG NKNVHVFLVGD TKHPDAHEVYRFLIEMVGAKMRKLGYVPDT
Viv_OTP90	-----QAVYNYLEELGLKMRKLGYPDT
Mtr_OTP90	KFVLHDMESE-HKEHSLSTHSEKLAVVYGIMKLP LGATIRVFKNLRI CGDCHNAFKYISKVVEREIVVRDRKRFHHFKNGECSCGNYW
Ptr_OTP90	KCVLHDVESD-LKEHELSTHSEKLAVAYGFMKLP HGATVRVFKNLRI CGDCHNAFKFMSKVVGREIVVRDGKRFHHFRD GKCSCGDYW
At_OTP90	SFVLHDVESDGHKEDMLTTHSEKIAVAFGLMKLP PGTTRIFKNLRTCGDCHNFFRFLSWVQORDIILRDRKRFHHFRNGECSCGNFW
Aly_OTP90	SFVLHDVESDGHKEDMLTTHSEKIAVAFGLMKLP PGTTRIFKNLRTCGDCHNFFRFLSKVVQORDIILRDRKRFHHFRNGECSCGNFW
Bra_OTP90	NFVLHDVESDGHKQDMLITHSEKIAVAFGLMKLP PGRAIRVFKNLRTCGDCHNFFIYLSRVVQREIVLRDRKRFHHFRNGECSCGNFW
Cru_OTP90	SFVLHDVESNGHKEDMLTTHSEKIAVAFGLMKLP PRTTIRVFKNLRTCGDCHNFFRFLSWVQORDIILRDRKRFHHFRNGECSCGNFW
Tha_OTP90	SFVLHDVESDGHKADMLTTHSEKLAVAYGLMKLP RGTAIRVFKNLRTCGDCHNFFRCLSRVQORDIILRDRKRFHHFRNGECSCGNFW
Zm_OTP90	KVVLHDMEPH-QKEHILFAHSERLAVGFGLLNLP PGATVTVLKNLRI CDDCHAAIMFMSKAVGREIVVRDVRRFHHFKDGECSGNYW
Sit_OTP90	KVVLHDMEPH-QKEHILFAHSERLAVGFGLLNLP PGATVTVLKNLRI CDDCHAVMMFMSKAVGREIVVRDVRRFHHFKDGECSGNYW
Sb_OTP90	KVVLHDMEPH-QKEHILFAHSERLAVGFGLLNLP PGATVTVLKNLRI CDDCHAVMMFMSKAVGREIVVRDVRRFHHFKDGECSGNYW
Os_OTP90	KFVLHDMEPH-EKEYILFAHSEKLAVGFGLLNLP PGATVTVLKNLRI CGDCHTAMMFMSKAVGREIVVRDVRRFHHFKDGECSGNYW
Brad_OTP90	KFALQDMAPH-EKEYVLF AHSERLAVSFGLLNLP AGATVTVLKNLRI CGDCHTAMMFMSRAVGREIVVRDVRRFHHFKDGECSGNYW
Viv_OTP90	KFVLHDMESE-QKEYVLSTHSEKLAVGFGLLNLP LGATVRVFKNLRI CGDCHNAFKFMSKVVEREIVVRDGKRFHHFKNGECSCGNYW

Supplemental Figure 2. Alignment of OTP90 ortholog protein sequences of 13 selected angiosperms. Alignment was performed using MUSCLE . PPR domains of *Zea mays* and *Vitis vinifera* were obtained using TPR prediction software (<http://tprpred.tuebingen.mpg.de/tprpred>) and are shown in boxes (continued boxes for PPR domains with $p\text{-value} < 10^{-5}$ and dashed boxes for PPR domains with $10^{-3} > p\text{-value} > 10^{-5}$). Arabidopsis E, and DYW domains are indicated with a continuous line, and a double line, respectively. Conserved amino acids in all species analyzed are highlighted and asterisks show conserved amino acids in all selected species other than one of the two species *Brasica rapa* or *Vitis vinifera*.



Supplemental Figure 3. Chloroplast transcripts levels in *otp90*. Chloroplast transcripts in *otp90* and Col-0 wild-type plants were quantified by qPCR using specific primers to each transcript as previously described (de longevialle et al 2008). Transcript levels in *otp90* was compared to the transcripts levels of wild-type and the ratios were normalized by nuclear genes ACTIN2 (white columns) and 18S (black columns) and by the mean of chloroplast transcripts *psbA*, *matK*, *psbK*, *psbI*, *atpA*, *atpF*, *atpH* and *atpI* (gray columns) seeming to do not change their level of expression between the mutant and the wild-type. Technical replicates were performed and standard deviation of the replicates is showed for each transcript.

RESULTATS

nad1 transcript sequences

C500



At_nE	aauauguguaggucccccguaaauucgagugagauugucau
At	aauauguguagguUcUguaauuUgagugagauugucau
Bna	aauauguguagguUcUguaauuUgagugagauugucau
Obe	aauauguguagguUcUguaauuUgagugagauugucau
Os	aauauguguagguUcUguaauuUgagugagauugucau
Tae	aauauguguagguUcUguaauuUgagugagauugucau
Bvu	aauauguguagguUcUguaauuUgagugagauugucau
Viv	aauauguguaggucccccguaaauUgagugagauugucau

ccmC transcript sequences

C184



At_nE	cucuauguacauguuccugcggcucggagagauuuauu
At	cucuauguacauguuccugUggcuUggagagauuuauu
Bna	cucuauguacauguuccugUggcuUggagagauuuauu
Obe	cUguauguacauguuccugUggcuUggagagauuucuu
Os	uUguauguacauguuccugUggcuUggagagauuaguu
Tae	uUguauguacauguuccugUggcuUggagagauuaguu
Bvu	cucuauguacauguuccuguggcUggagagauuucuu
Viv	cUguauguacauguuccugcggcUggagagauuucua

mttB transcript sequences

C97



At_nE	gagaaguucgaauccguuccguucggauauugaucggucu
At	gagaaguucgaauccguuccguuUggauauugaucggucu
Bna	gagaaguucgaauccguuccguuUggauauugaucggucu
Obe	aagaaguucgaauccguuccguuUggauauugaucggucu
Os	gagaaguucgaauccguuUcgauUggauauugauUggucu

ccmB transcript sequences

C80



At_nE	cacgaguuuuuucuaauccucucguauaucgucguaac
At	cacgaguuuuuucuuUauuUcuUuUguauaucgucguaac
Bna	cacgaguuuuuucucUauuccucUguauauUgucguaac
Obe	UacgaguuuuuucUuUguuccucUguauauUguUguaac
Os	cacgaguuuuuuuuuauuccucUguauauUgucgucac
Tae	cacgaguuuuuuuuuauuccucUguauauUgucgucac
Les	cacgaguuuuuucucuaauccucUguauauUgucguaac
Viv	cacgaguuuuuuuuccauuccucucguauaucgucgucac

NAD1 proteins (151-180aa)

S167L



IGLILITVLICVGP	NSSEIVMAQKQIWSG
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGS	CNLSEIVMAQKQIWF
IGLILITVLICVGP	RNLSEIVMAQKQIWF

CCMC proteins (47-76aa)

R62W



QGGNSRILYVHVP	AARMSIIVYIATAINTF
QGGNSRILYVHVP	AVMSIIVYIATAINTF
QGGNSRILYVHVP	AVMSIIVYIATAINTF
QGGNSRILYVHVP	AVMSILVYIATAINTF
QGGNSRILYVHVP	AVMSIIVYIATAINSF
QGGNSRILYVHVP	AVMSIIVYIATAINSF
QGGNSRILYVHVP	AVMSIIVYIATAINTF
QGGNSRILYVHVP	AAVMSILVYIATAINTL

MTTB proteins (18-47aa)

R33W



FASETILGEVRIR	SVRILIGLGLTWFTCYW
FAETILGEVRIR	SVWILIGLGLTWFTCYW
FAETILGEVRIR	SVWILIGLGLTWFTCYW
FAETILEEVRIR	SVWILIGLGLTWFTCYW
FAETFLGEVRIR	SVWILIGLGLTWFTCYW

CCMB proteins (12-42aa)

S27L



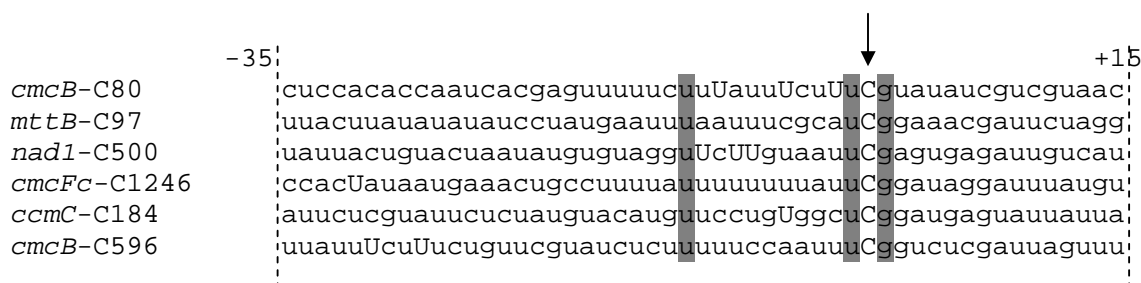
LIFSSTPITSFSS	FLSYIVVTPMLGFEKD
LIFSSTPITSFSL	FLLYIVVTPMLGFEKD
LIFSSTPITSFSL	FLLYIVVTPMLGFEKD
QIFSSTPITSFSL	FLLYIVVTPMLGFEKD
QIFSSTPITSFSL	FLLYIVVTPMLGFEKD
QIFSSTPITSFSL	FLLYIVVTPMLGFEKD
QIFSSTPITSFSL	FLLYIVVTPMLGFEKD
RIFLSTPITSF	FLSYIVVTPMLGFEKD

RESULTATS

<i>ccmB</i> transcript sequences		CCMB proteins (177-206aa)	
	C596 ↓		S199L ↓
At_nE	guucguaucucuuuuuuccaauuuCggucucgauuaguuu	TEWFHVLSSIGYSSLFVSLFPISVSISLQD	
At	guucguaucucuuuuuuccaauuuUggucucgauuaguuu	TEWFHVL LL IGY FF LFVSLFPILVSISLQD	
Bna	guuUguaucucuuuuuuccaauuuUggucucgauuaguuu	TEWFHVL LL IGY FF LFVSLFPILVSISLQD	
Obe	guuUguaucucuuuuuuccaauuuUggucucgauuaguuU	TEWFHVL LL IGY FF LFVSLFPILVSISLQD	
Os	uuuuguaucucuuuuauccaauuuUggucucgauuaguuU	TEGFHVL LL IGY FF LFVSLYPILVSISLQD	
Tae	uuuuguaucucuuuuauccaauuuUggucucgauuaguuU	TEGFHVL LL IGY FF LFVSLYPILVSISLQD	
Les	guuUguaucucuuuuuuccaauuuUggucucgauuaguuU	TEWFHVL LL IGY FF LFVSLFPILVSISLQD	
Viv	guucgucucuuuuuuuuccaauugcggucucgauuaguuU	TEWFHAPSSIGYSSSFVSFFSMSVSISLQD	
<i>ccmFc</i> transcript sequences		CCMFC proteins (177-206aa)	
	C1246 ↓		R416W ↓
At_nE	aacugccuuuuauuuuuuuuuauucggauaggauuuauugu	GWYTTIMKLPFIFFIRIGFMLASLGGSPSL	
At	aacugccuuuuauuuuuuuuuauuUggauaggauuuauugu	GWYTTIMKLPFIFFI W IGFMLASLGG I PSL	
Bna	aacugccuuuuauuuuuuuuuauuUggauaggauuuauugu	GWYTTIMKLPFIFFI W IGFMLASLGG L PSL	
Obe	aacugccuuuuauuuuuuuuuauuUggauaggauuuauugu	GWYTTIMKLPFIFFI W IGFMLASLGG L GSL	
Os	aaauugccuuuuucuuuuuuuaguauuUggauaggauuuucugu	GWYTTIMKLPFLFSI W IGFLLASLGGSRSL	
Tae	aaauugccuuuuauuuuuuuuaguauuUggauaggauuuucugu	GWYTTIMKLPFIFS I WIGFLLASLGGSRSL	
Bvu	aacugccuuuuucuuuuuuuuauuUggauaggauuuucugu	GWYTTIMKLPFLFFI W IGFLLASLGGSRSL	
Viv	aacugccuuuuauuuuuuuuuauucggauaggauuuauuga	GWYTTIMKLPFIFFIRIGFMMASSG L RSL	
Cpa	aacugccuuuuucuuuuuuuugauucggauaggauuuauugu	GWYTTIMKLPFLFLIRIGFMLASLGGSRSL	

Supplemental Figure 4. Surrounding RNA and protein sequences of each RNA editing site targeted by OTP90 in *Arabidopsis* are aligned with corresponding sequence of several other angiosperm species. REDIdb database (Picardi *et al* 2011) was used to obtain mitochondrial transcript sequences. Editing sites in the targeted by OTP90 are indicated by an arrow and all editing events at each species are shown in upper case U. Position of the amino acid (aa) in the protein sequence windows of *Arabidopsis* mitochondrial proteins are shown in brackets. Amino acid substitutions in protein sequences caused by editing are highlighted. Amino acid positions corresponding to the edited codons targeted by OTP90 in *Arabidopsis* are indicated by an arrow. At_nE: *Arabidopsis thaliana* no edited sequence, At: *Arabidopsis thaliana* edited sequence, Bna: *Brasica napus* edited sequence, Obe: *Oenothera berteriana* edited sequence, Os: *Oriza sativa* edited sequence, Tae: *Triticum aestivum* edited sequence, Bvu: *Beta vulgaris* edited sequence, Les: *Lycopersicon esculentum* edited sequence, Vvi: *Vitis vinifera* edited sequence, Cpa: *Carica papaya* edited sequence.

RESULTATS



Supplemental Figure 5. Alignment of transcripts sequences surrounding the all RNA editing sites targeted by OTP90. Nucleotide windows (-35 and +15) around each editing site (0, indicated by an arrow) are shown. Shared nucleotides in all aligned sequences are highlighted and other editing events in the surrounding sequence are showed in upper case U.

Supplemental Table 1. Prediction of localization of OTP90 protein orthologs using five predictors available in the web.

Protein name	Predotar	Target P	Target P*	WoLF PSORT	Plant-mPLOC	ATP
At OTP90	None	M	-	M, c	C	m/c (0.65653)
Aly OTP90	None	M	-	M, c	C, cyto, m	m/c (0.53062)
Cru OTP90	None	M	-	M, cyto, c	C	m/c (0.65653)
Bra OTP90	None	-	-	-	C	-
Tha OTP90	None	M	-	M, c	C	m/c (0.65653)
Mtr OTP90	None	M	-	cyto, c	C	m/c (0.52957)
Ptr OTP90	None	M	-	C	C	m/c (0.68813)
Viv OTP90	M	M	M	C	C	-
Os OTP90	M	C	C	C	C	-
Brad OTP90	C	C	C	C	C, m	-
Sit OTP90	C	C	C	C	C, m	m/c (0.53959)
Zm OTP90	C	C	-	C	C	m/c (0.57119)
Sb OTP90	C	C	C	C	C, m	-

Captions need to be considered as m: possibly mitochondria, M: Mitochondria, c: Possibly plastid, C: Plastid, cyto: Cytoplasm, m/c: possibly mitochondria and plastid. Predotar prediction was obtained using version 1.03 (<http://urgi.versailles.inra.fr/predotar/predotar.html>). Target P prediction was obtained selecting prediction for plant organism with no cutoffs and also with cutoffs using specificity>0.90, this last showed with an asterisk. Predictions using WoLF PSORT was obtained in its website (<http://wolfsort.org/>) and results were interpreted as followed: Uppercase was used when score was upper than 9 signaling a confidence in the prediction and lowercase for scores between 3 and 9 indicating a possible secondary location predicted. Plant-mPLOC prediction was obtained using version 2.0, in this predictor the first location was interpreted as the most accurate and the others as possible type of location (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>). Ambiguous Targeting Prediction (ATP) was obtained in its website considering score values rangin from 0.5 to 0.7 showed in brackets as possibly dual targeted protein by ambiguous targeting peptide sequence (<http://www.cosmoss.org/bm/ATP>).

atp/	1178	1392	1415	1404
Cell	100	100	100	100

Mitochondrial editing events levels are estimated for each gene in percentage related to the picks size relations. Arabidopsis mitochondrial nucleotide editing events analysed are indicated by their position in the transcript sequence. Symbol (-) means that no significant differences were found in *otp90* related to *col-0*.

Name	Sequence (5' to 3')
------	---------------------

5.2. Projet d'Article III^ψ

Running title: Organelle RNA editing protein complex

Title: A CLB19 interactor in the RNA editing complex is an essential Pentatricopeptide Repeat Protein influencing the editing of Chloroplast and Mitochondrial Transcripts of *Arabidopsis thaliana*.

Mauricio Lopez-Obando^{1,2}, Andeol Falcon De Longevialle^{1,2,b}, Mizuki Takenaka³, Alexandra Avon², Geert de Jaeger⁴, Richard Berthomé^{2,a}, Claire Lurin^{2,*}

1 INRA, UMR 1165, Unité de Recherche en Génomique Végétale (URGV), CNRS, ERL 8196, UEVE, 91057, Evry cedex, France.

2 Université d'Evry Val d'Essonne, Bd François Mitterrand, 91057, Evry cedex, France.

3 Molekulare Botanik, Universität Ulm, 89069 Ulm, Germany

4 Department of Plant Systems Biology - VIB, Gent University, Technologiepark 927 9052, Gent Belgium

Current addresses

a: Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR INRA 441/CNRS 2594, CS 52627, 31326 CASTANET TOLOSAN CEDEX, France

b: Laboratoire Plant Process Innovation, GENOPOLE, Evry, France

* Corresponding author: Claire Lurin (lurin@evry.inra.fr), tel (33) 1.60.87.45.13.

Date of submission

Number of tables: 1

Number of supplemental tables: 4

Number of figures: 4

Number of supplemental figures: 5

Total word count:

ψ: Ne sera publié dans l'état.

ABSTRACT

RNA metabolism in plant organelles involves several processes including RNA editing, splicing, polycistronic cleavage, 5' and 3' maturation and translation initiation. In particular, plant organelle RNA editing corresponds to nucleotide conversions involving multi-subunits protein complexes. In plant organelles, central players of this process are members of the Pentatricopeptide Repeat protein family. They act to specify the target editing site in interaction with the Multiple Organelle editing Factors. Nevertheless, the composition of organelle RNA editing complexes is still poorly known. Here, through a tandem affinity purification of tagged proteins approach, we explored a plastid RNA editing complex using CLB19 editing factor as a bait. This approach helped us to identify two essential PPR partners, dually targeted to both organelles. Genetic and molecular characterizations confirmed the implication of one partner of the complex in plastid editing at the same sites as CLB19. In addition, reverse genetics results showed that this dual targeted PPR protein is also involved in RNA editing in mitochondria. The results showed here strong support our previous results showing that in organelle RNA editing complexes, the PPR, E and DYW domains are always presents.

Key Words: RNA editing, Mitochondria, Chloroplast, PPR protein, Dual targeting, Embryo lethality, *Arabidopsis thaliana*.

INTRODUCTION

RNA modifications are wide spread mechanisms occurring in all living organisms. Several types of RNA modifications are known and could be divided in two major classes: i) structural changes in RNA sequences and ii) alterations of RNA sequence that change genetic information carried by RNA (Grosjean, 2005). This last class of modifications is also known as RNA editing. RNA editing was discovered more than 25 years ago when it was shown, in *Trypanosomma brucei*, that nucleotide insertions in the mitochondrial *cox2* transcripts were required to obtain functional transcripts encoding COX2 (Benne et al., 1986). After that, importance of the RNA editing mechanisms was clearly demonstrated in a wide range of organisms, including plants (Takenaka et al., 2013b). Two distinct mechanisms can lead to editing. Base deletions or insertions mediated by guide RNA molecules and conversions or substitutions. Conversion leads to the chemical alteration of an individual nucleotide by a protein complex recognizing a specific target sequence and is catalyzed by enzymes such as cytidine deaminases (C to U conversions) or adenosine deaminases (A to T conversions).

In plants, RNA editing is mainly associated to RNA nucleotide base conversions of organelle transcripts. Plant organelle RNA editing has been detected for the first time in mitochondrial transcripts independently by three teams (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989). These results showed that RNA editing process was involved in conversions of cytosine into uracil nucleotides (C to U). Conversions of uracil into cytosine (U to C) was described subsequently and occur in the mitochondrial transcripts less frequently than C to U conversions (Gualberto et al., 1990; Schuster et al., 1990). RNA editing events are now well documented in both plant organelles ; C to U conversions are common among clades of land plants whereas U to C changes are prevalent in only some clades (Malek et al., 1996; Freyer et al., 1997; Yoshinaga et al., 1997). To date, C to U, U to C and A to I conversions are known to proceed in land plant organelles but are not observed in green algae and in the special case of *Marchantia polymorpha* (Delannoy et al., 2009; Knoop, 2011; Takenaka et al., 2013b).

In land plants, C to U conversion is mostly studied. Current data shows that gain or loss of editing sites have been detected during different periods of plant evolution (Fujii and Small, 2011; Richardson et al., 2013). Except for *Marchantia polymorpha* in which

RESULTATS

organelle RNA are not edited, C to U RNA editing in land plant organelles range from 2 events in *Physcomitrella patens* plastids to more than 2000 in *Selaginella moellendorffii* mitochondria (Fujii and Small, 2011; Hecht et al., 2011; Takenaka et al., 2013b). In flowering plants, the number of editing events between species is more conserved ranging from 300 to 600 in mitochondria and from 25 to 75 in plastids (Fujii and Small, 2011; Huang et al., 2013; Takenaka et al., 2013b). In particular, more than 500 RNA editing sites were reported in Arabidopsis mitochondria whereas about 40 sites were identified in its plastids (Bentolila et al., 2013; Ruwe et al., 2013).

The molecular mechanism of editing in land plants was highlighted by the study of a site specific trans-acting factor named CHLORESPIRATORY REDUCTION4 (CRR4), involved in plastid RNA editing (Kotera et al., 2005). This striking contribution stimulates the subsequent finding of new site specific trans-acting factors required for plant organelle RNA editing, all of them belonging to the family of Pentatricopeptide Repeat (PPR) containing domain. The PPR domain is a degenerated polypeptide showing a conserved structural conformation of two antiparallel helixes joined by an unstructured sequence, this helix-loop-helix conformation being repeated in tandem in the protein (Small and Peeters, 2000; Delannoy et al., 2007; Howard et al., 2012). The PPR proteins are encoded by nuclear genomes of a wide range of eukaryotes. However, their distribution is not similar between all species and a clear expansion of the family has been observed in land plants (O'Toole et al., 2008; Schmitz-Linneweber and Small, 2008). Most members of this family are targeted to organelles. Indeed, in Arabidopsis, whereas 57% of the PPR proteins are localized or predicted to be localized in mitochondria, 24% in plastids, about 10 % in both organelles, very few of them have been convincingly detected outside the organelles (Colcombet et al., 2013). In land plants, the nature of PPR domain within proteins in conjunction with additional domains helped to classify the PPR family (Lurin et al., 2004; Rivals et al., 2006). This classification of PPR proteins divided the family into two subfamilies, the PPR P and the PPR PLS. This last subfamily was also subdivided in two subgroups, the PLS-E and the PLS-E-DYW, according to the presence of a C-terminal E domain alone or an E and DYW domains. Besides their involvement in RNA editing process, members of PPR family have been also implicated in all types of processes of organelle RNA metabolism such as 5' or 3' transcript stabilization and processing, splicing and translation (Schmitz-Linneweber and Small, 2008).

RESULTATS

Biochemical and structural evidence showed that PPR proteins could interact with RNA molecules through their PPR domains (Prikryl et al., 2011; Okuda and Shikanai, 2012). A code for RNA recognition by PPR proteins was recently suggested by biochemical and bioinformatics approaches (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013). In this code, and according to PROSITE PPR domain model, the nucleotide recognition is achieved by the combination of three amino acids, two located in a first PPR domain (3rd and 6th position) and the last one being at the 1st position (1') of the following domain. The structural determination of PPR-RNA complexes recently confirmed that the specificity of recognition is mainly associated to the 6th position as previously suggested (Ke et al., 2013; Yin et al., 2013). Furthermore, these structures showed that a PPR protein dimer could be involved in the interaction with a RNA molecule.

To date, in land plant, more than 60 PPR proteins were shown to be involved in RNA editing. Most of them belong to the PLS subfamily and particularly to the PLS-E and PLS-E-DYW subgroups. While a function in RNA-Protein interaction is well defined for PPR domains, the functions of E and DYW domains remain unclear and controversial. Molecular and phylogenetic evidences suggest that E and DYW C-terminal domains are required for editing activity. Although no catalytic activity has been demonstrated for the DYW domain yet, it has been hypothesized that it could contain the RNA editing enzymatic activity (Salone et al., 2007; Boussardon et al., 2012; Rudinger et al., 2012).

Besides PPR proteins, several proteins were also shown to be involved in C to U RNA editing process. Multiple Organellar editing Factors (MORF) were recently described as required for many editing sites of plant organelles. In *A. thaliana*, 9 MORF proteins are encoded by the genome and predicted to be localized in one or both organelles (Takenaka et al., 2012; Bentolila et al., 2013). These proteins function as general factors targeting several editing sites of organelle transcripts (Takenaka et al., 2012; Bentolila et al., 2013). Interactions with PPR proteins have been demonstrated in the case of MORF8-MEF10, MORF8-RARE1 and, MORF1-OTP90 using yeast two hybrid experiments or in planta BiFC experiments (Bentolila et al., 2012; Takenaka et al., 2012) (Lopez-Obando *et al.*, in preparation). Proteins containing RNA Recognition Motifs (RRM) were also found to influence RNA editing in plant organelles. For example, mutants in CHLOROPLAST RIBONUCLEOPROTEINS 31A or 31B (CP31A, CP31B) are affected in RNA editing of several chloroplast transcripts but also in organellar RNA stability (Tillich et al., 2009).

RESULTATS

Interestingly, the ORGANELLE RNA RECOGNITION MOTIF 1 (ORRM1) protein, containing both a RRM motif and domains related to MORF proteins, was shown to be implicated in several chloroplast RNA editing events in *A. thaliana* and *Z. mays* (Sun et al., 2013).

The RNA editing field has been subjected to intensive studies during last two decades. Different types of protein complex have been associated to several RNA editing systems, some of them being high molecular weight complexes, like the 20S editosome of *T. brucei* (Goringer, 2012). In land plant organelles, results showing that several proteins are required for the same editing events and that protein interactions occur between protein involved in RNA editing support the existence of an RNA editing protein complexes similar to editosomes of tripanosomes. However, plant editing complex protein composition remains unresolved and further studies are needed to identify members of these complexes and their functions.

Here, we implemented a Tap-tag approach to gain insight into the composition of a chloroplast editing complex. We use the known chloroplast editing factor CLB19 required for *rpoA* and *clpP* editing (Chateigner-Boutin et al., 2008) as a bait for Tap-tag experiments. We identified two additional PPR proteins, required for *A. thaliana* embryo development, that participate into the CLB19 editing complex. Two allelic mutations affecting one of these proteins altered the editing efficiency of CLB19 target sites together with mitochondria sites. The function of the third PPR protein of this complex was not characterized because of embryo lethality of its mutant lines. This work is the first demonstration that a PPR protein dually targeted to mitochondria and plastid can function in both organelles as an editing factor.

RESULTS

Exploring the CLB19 chloroplast editing complex

A TAP-Tag approach was developed using Arabidopsis cell culture to gain insight into the *in vivo* composition of an RNA editing complex of land plant chloroplasts. To this purpose, the previously characterized chloroplast editing factor CLB19, targeting *rpoA*-C200 (78691) and *clpP*-C559 (69942) editing sites (Chateigner-Boutin et al., 2008), was

RESULTS

use as bait. Three proteins were identified with high confidence as partners of CLB19 complex in four independent experiments (Table 1, supplemental figure 1). Even if others proteins have been identified in individual experiments, their scores and/or their absence in at least one of the 4 experiments lead us to exclude them from the CLB19 partner list (Supplemental Table 1). In the fourth TAP-Tag experiment performed, samples were subjected to RNase treatment before the purification procedure. This treatment did not affect the nature of the proteins identified in the complex suggesting that the proteins could either interact directly or form a complex in which the RNA sequences are protected from RNase treatment.

Four proteins were reproducibly purified during these TAP-tag experiments. The CHLOROPLAST CHAPERONIN 60 ALPHA (CNP60A, AT2G28000) was identified in all experiments, however it is also considered as a sticky protein often detected in this type of experiments, reason why it was excluded from the subsequent analysis. AT1G05750 protein is the CLB19 bait protein whereas AT3G49240 and AT2G15690 are two uncharacterized PPR proteins. According to the PPR classification, *AT3G49240* codes for a protein that belongs to the PPR P subfamily which was named here as ORGANELLE TRANSCRIPT PROCESSING 100 (OTP100) whereas *AT2G15690* encodes a PLS-E-DYW subgroup PPR protein. *In silico* prediction of domains showed that OTP100 harbors 11 predicted PPR domains, a predicted Nuclear Localisation Signal (NLS) located in its N-terminal sequence between amino acid positions 45 and 54, and a 100 amino acid long C-terminal region without any conserved domain. This C-terminal region has a low probability to form a coil-coiled structure, typical of transcription factors. In *Arabidopsis thaliana*, OTP100 is closely related to the PPR protein GLUTAMINE RICH PROTEIN 23 (GRP23) (Ding et al., 2006). The phylogenetic analysis performed using sequences of several available angiosperms genomes showed that 1 to 3 *OTP100* paralogs could be found in these genomes with an orthology link to GRP23 proteins (Supplemental figure 2).

The second PPR partner of CLB19, encoded by the *AT2G15690* gene, is an atypical PLS-DYW PPR protein containing five predicted PPR domains and a C-terminal DYW separated by an amino acid sequence that do not clearly correspond to an E domain. This unusual architecture is shared by only five others proteins in the *A. thaliana* genome among which is the recently described DYW1 chloroplast editing factor (Boussard et al., 2012). For this reason, this protein was named hereafter the DYW2 protein.

Arabidopsis DYW2 protein is related to DYW1 but also to the Mitochondrial Editing Factors MEF8, MEF8S and to two others PPR proteins with unknown function AT2G34370, named DYW3 and AT1G29710, named DYW4, these proteins forming a specific subfamily in *Arabidopsis* (O'Toole et al., 2008; Verbitskiy et al., 2012). To understand the relationships between these proteins, an expanded phylogenetic analysis was realized using the genomes of several angiosperms, available in phytozome web page version 9 (Supplemental figure 3). This analysis showed that DYW2 protein is present in all angiosperm species, mostly as a single copy gene with the exception of *Zea mays* and *Medicago truncatula* where several paralogs are found suggesting gene duplications in these species. Additionally, our results also suggest that the DYW2 orthologous group is the most closely related to the DYW1 group. Considering MEF8, MEF8S, DYW3 and DYW4 proteins, the analysis showed that these proteins have probably a more complex evolutionary history. In the *Brassicaceae* family, MEF8 protein has one ortholog in each species whereas MEF8S, DYW3 and DYW4 proteins were not identified in some species. For the *Gramineae* family, the situation is different since there is only one paralog of MEF8, MEF8S, DYW3 and DYW4 in each species suggesting evolutionary differences between monocotyledons and dicotyledons species.

OTP100 and DYW2 are dually targeted to chloroplast and mitochondria

In our previous study “Systematic study of subcellular localisation of *Arabidopsis* PPR proteins”, we showed that OTP100 is dually targeted to mitochondria and chloroplast and also proposed a putative dual targeting to both organelles for DYW2 (Colcombet *et al.*, 2013). To confirm the dual localisation of DYW2 deduced from available proteomic data, the localization of the full length DYW2 protein fused to DsRed2 expressed under the control of the Cauliflower mosaic virus 35S promoter was evaluated in stable *Arabidopsis* transgenic lines. The localization of this translational fusion was observed under confocal microscope using chlorophyll autofluorescence and mitotracker green as markers of chloroplast and mitochondria localization, respectively. A strong DsRed2 signal overlapping with mitotracker green signal was observed in root cells confirming that DYW2 was addressed to mitochondria. Similar strong punctuated signals were also recovered in cotyledons as well as faint signals co-localizing with chlorophyll autofluorescence. Even if mitotracker did not allow us to label cotyledon mitochondria in

our experiments, these last results strongly support previous proteomics data indicating that DYW2 is dually targeted to both organelles (Figure 1).

OTP100 and DYW2 proteins are crucial for *A. thaliana* embryo development.

A reverse genetics approach was used to identify the molecular functions of DYW2 and OTP100 proteins. Several T-DNA insertion lines were available in the *DYW2* gene (*AT2G15690*). We identified two mutant lines carrying insertions into the coding sequence: *dyw2-1* (GK_332A07) and *dyw2-2* (FLAG_435F11) in the Columbia (Col-0) and Wassilewskija (Ws) genotypes respectively. In addition, a third line, *dyw2-3* (SALK_002014.56), was selected with an insertion in the promoter in the Col-0 genotype. T-DNA insertions were confirmed by sequencing PCR products corresponding to T-DNA flanking sequences and their localizations are illustrated in Figure 2A. Because *dyw2-1* and *dyw2-2* T-DNA insertion were located in *DYW2* coding region, we initially performed genetic analysis using selfed progenies to identify homozygous plant for both mutations. However, plants homozygous for the *dyw2* mutations were never recovered for either allele. Overall, 198 and 107 seedlings, corresponding to *dyw2-1* and *dyw2-2* self progenies, respectively, were PCR genotyped (Supplemental Table 2). Results obtained showed that the T-DNA transmission rate of both allele mutants was lower than expected for a Mendelian trait. Indeed, among the viable seedlings, only two-third carried the T-DNA insertion (at heterozygous state), supporting that *dyw2-1* and *dyw2-2* mutations are embryo lethal mutations leading to the loss of homozygous mutants. To support these results we characterized the phenotype of these mutants in details. In soil and with long-day conditions, all heterozygous plants carrying *dyw2-1* or *dyw2-2* mutation were indistinguishable from their wild type counterparts. We stained and examined pollen by Alexander staining of mature anthers from both mutants. Results indicated that all pollen grains were viable in both heterozygous mutants and that mutations did not affect male gametophyte viability (Figure 2B). However, numerous white or small seeds corresponding to aborting embryos were observed in siliques of heterozygous plants. This phenotype is common in embryo lethal PPR mutants affected at late stages of embryo development (Colcombet et al., 2013)(Figure 2C). This phenotype seems to be dependent of genetic background since it was more severe in the *dyw2-2* mutant line in which several seeds were collapsed. *dyw2-3*, carries a T-DNA insertion located 94 nucleotides upstream of the translation initiation codon in the promoter region of the *DYW2* gene. Homozygous

RESULTATS

dyw2-3 plants were identified using PCR genotyping. They harbored a wild type phenotype in soil and under long-day conditions (Figure 3A). Finally, a genetic complementation assay between *dyw2-1* and *dyw2-2* mutant lines, using reciprocal crosses, confirmed that *dyw2-1* and *dyw2-2* are allelic mutations responsible for the observed embryo lethal phenotype (Supplemental Table 3).

Only two T-DNA insertion lines are available in the *OTP100* gene. They were named *otp100-1* (SALK_069042) and *otp100-2* (SAIL_784_A11) and were both identified in Col-0 genotype (Supplemental figure 4). These two mutant lines were previously characterized as embryo defective mutants *emb1796* during the seedgenes project (Bryant et al., 2010). The phenotype of homozygous mutants is characterized by an embryo development arrest before globular stage and seeds remain white whereas viable embryo became green due to photosynthesis activity involved in normal embryogenesis process of *A. thaliana* seeds (Cushing et al., 2005).

For both *dyw2* and *otp100* mutants, we tried to rescue the embryo development in order to obtain homozygous plants. We used sucrose enriched medium and specific growth conditions previously described to rescue other embryo lethal or PPR mutants (de Longevialle et al., 2008; Chi et al., 2010; Geerts et al., 2011). Unfortunately, we were not able to recover homozygous plants from any of the identified mutations.

DYW2 protein is involved in editing of chloroplast and mitochondria transcripts

RT-qPCR performed on *dyw2-3* homozygous mutant plants revealed a reduction in *DYW2* gene expression which did not affect *OTP100* transcript accumulation (Figure 3B). This result indicated that *dyw2-3* is a weak allele of *DYW2* mutant, partially disrupting its gene expression. We evaluated RNA editing in plastids and mitochondria of leaf tissue from *dyw2-3* homozygous mutant plants. The analysis of the editing behavior of the *dyw2-3* line at the 34 major chloroplast editing sites showed a low but clear and reproducible reduction of *rpoA* (78691) and *clpP* (69942) editing. Both sites are targets of the CLB19 protein. The analysis of 415 mitochondria editing sites showed that *dyw2-3* mutant line was also affected in 6 mitochondrial editing sites, *ccmC*-C421 (240338), *ccmC*-C673 (240086), *ccmFc*-C1280 (51372), *cox3*-C257 (218536), *cox3*-C413 (218692) and *mttB*-C144 (157634) (Figure 3C, supplemental Figure 5). To complete this study, we generated two

RESULTATS

heteroallele mutant lines by crossing the weak *dyw2-3* homozygous mutant with the knock-out *dyw2-1* and *dyw2-2* heterozygous lines. Plants were PCR genotyped in order to identify heteroallelic *dyw2-3/dyw2-1* or *dyw2-3/dyw2-2* plants. They were able to germinate and grow in soil without showing any macroscopic phenotype that could differentiate them from *dyw2-3* homozygous mutant or wild type plants (Figure 4A). Gene expression level of *DYW2* in *dyw2-1/dyw2-3* and *dyw2-2/dyw2-3* lines was characterized and a more pronounced reduction in *DYW2* gene expression was observed in comparison with *dyw2-3* heterozygous and *dyw2-3* homozygous plants (Figure 4B). Although reduction in *DYW2* gene expression did not affect the macroscopic phenotype of the plants, the chloroplast and mitochondria editing analysis of the heteroallele lines showed a more pronounced reduction of editing efficiency in the sites previously mentioned when compared to the *dyw2-3* homozygous mutant plants (Figure 4C). This positive correlation between *DYW2* gene expression level and editing efficiency at *DYW2* putative targets supports the hypothesis that *DYW2* may be probably required for editing of *rpoA-C200* and *clpP-C559* plastid sites, on one hand, and *ccmC-C421*, *ccmC-C673*, *ccmFc-C1280* *cox3-C257*, *cox3-C413* and *mttB-C144* mitochondrial sites, on the other hand. However, without null mutant, the requirement of *DYW2* could not be proven as well as we can not exclude that *DYW2* could be involved in some additional sites in both organelles.

DISCUSSION

After more than 25 years of study, the composition of plant RNA editing protein complexes is still misunderstood. Until now and mainly by genetics studies, two RNA binding protein families, the PPR and CRM families, as well as the potential RNA binding factors of MORF family were demonstrated to influence the editing status of organelle transcripts (Takenaka et al., 2013b). Among them, the PPR proteins are considered to be the specificity factors selecting the target sites to be edited. In contrast, CRM and MORF factors are thought to be general factors probably involved in the stabilization of the RNA-protein complexes. According to the current data, none of these factors have been demonstrated to carry the deamination activity supposed to be the main mechanism responsible for the plant organelle RNA editing. However, based on phylogenetic, molecular and genetic studies, a proposed hypothesis suggest that the DYW domain of PLS PPR proteins could be responsible for this enzymatic activity (Salone et al., 2007; Boussardon et al., 2012; Hayes et al., 2013). To challenge this hypothesis and gain insight

into the plant organelle C to U RNA editing complex composition, we used a tandem affinity purification proteomics approach using the CLB19 chloroplast editing factor as bait. The results of four independent experiments showed that two proteins are reproducibly eluted with CLB19 protein, both being PPR proteins. Surprisingly, general factors such as MORF2, MORF9 and the related ORRM1 factors, known to be involved in *rpoA*-C200 and *clpP*-C559 editing, were not identified in CLB19 complex. The absence of these proteins could be due to weak and transient interactions of these factors with the complex. Alternatively, their absence could be due to the stringency of statistical filters applied during TAP-tag result analysis. To exclude this possibility, we screened the raw data to find any of these factors. With the exception of two related proteins, none of these proteins was identified in the raw data. The two identified proteins were AT2G19380, a RRM containing protein, and AT1g62490, a member of the mTERF family. Both of them showed very low quality scores ranking these proteins as possible false-positive or background proteins.

Some additional proteins were identified in individual experiments but their absence in at least one of the 4 experiments lead us to exclude them from our analysis (Supplemental Table 1). Among these proteins, the LON PROTEASE 1 (LON1) known to be dually localized in mitochondria and plastid have been detected (Rigas et al., 2011). Interestingly, LON1 and the mitochondrial LON3 function as chaperon to regulate organelle protein homeostasis (Rigas et al., 2011; Solheim et al., 2012). Such proteins could play the role of regulatory proteins involved in the complex assembly or regulation.

Surprisingly, whereas *clb19* mutants are viable, T-DNA insertions in *DYW2* or *OTP100* coding regions lead to embryo abortion and lethality indicating that both *DYW2* and *OTP100* proteins are essential for embryo development. These PPR proteins are dually targeted to mitochondria and plastid, suggesting their putative involvement in common processes of RNA modification in organelles. Indeed, the function of *DYW2* protein in the CLB19 editing complex was confirmed by the molecular characterization of *dyw2-3* homozygous mutants. We showed that, in plastids, *DYW2* targets the same two editing sites as CLB19 protein, supporting its participation in this editing complex. Additionally, and in accordance with its dual localization in both organelles, we found that *DYW2* functions also in mitochondrial editing, targeting at least six sites. Two of these sites, *cox3*-C257 and *mttB*-C144, have been previously reported to be the targets of the PLS-E MEF21

RESULTATS

and SLO2 editing factors respectively. Strikingly, MEF21 and SLO2 are phylogenetically closely related to CLB19 in *A. thaliana* (O'Toole et al., 2008) suggesting that DYW2 protein may play a function in MEF21 and SLO2 editing complexes in mitochondria.

As previously showed, the refined code for PPR proteins can be used to identify the putative editing target sites of editing factors that belong to PLS PPR subfamily (Takenaka et al., 2013a). Using this code, CLB19, MEF21 and SLO2 proteins targets sites were accurately predicted, indicating a high specificity of recognition of the RNA sequences (Takenaka et al., 2013a). Thus, in the context of these editing complexes, these three PPR proteins are proposed to act as the main factors specifying the target sites. Interestingly, when used for DYW2 and OTP100, the code did not allow identifying any of these sites. This result suggests that the function of DWY2 and OTP100 in the editing complexes might not be associated to the RNA target specificity. Alternatively, they could be involved in catalyzing the edition reaction, in the stabilization of the complex, or in the interaction of the RNA editing complex with other factors involved in the process.

Unfortunately, the embryo lethal phenotype of *otp100-1* and *otp100-2* mutant lines and the lack of any weak mutation did not allow us to identify OTP100 molecular phenotype and function. However, OTP100 involvement in CLB19 editing complex and its dual targeting to organelles strongly suggest of its functional involvement in the editing complex. Moreover, OTP100 dual targeting to organelles as well as other structural characteristics such as the presence of conserved regions predicted to have a nuclear localization signals, the presence of a putative coil-coiled structure and the fact that in a previous study this protein was found to be part or interact with cyclins (Van Leene et al., 2007; Van Leene et al., 2010), suggest that this protein has a complex function probably linked to organelle RNA metabolism and the cell cycle. Indeed, some PPR proteins annotated in the full sequenced genomes of nine green algae possess cyclin domains (Tourasse et al., 2013). Furthermore, an example of the connection between these processes was demonstrated mainly at the level of transcriptional coordination in the red algae *Cyanidioschyzon merolae*, possessing a single plastid and single mitochondria. In this case, the implication of post-transcriptional regulators such as PPR proteins was not excluded (Kanesaki et al., 2012). Future studies on OTP100 might contribute to elucidate the putative involvement of this protein in both processes.

Composition of editing complexes: are DYW domains always needed for an efficient RNA editing?

During the last decade different RNA editing complexes have been described in several organisms, some of them having high molecular weight quaternary structures such as the 20S editosome of *T. brucei* (Goringer, 2012). However, plant organelle RNA editing complex is supposed to be more similar to the mammal ACF-APOBEC-1 editing complex (Shikanai, 2006). In this complex, ACF protein binds the RNA sequence around the target nucleotide to be edited, giving the specificity of recognition, and APOBEC-1 acts as the enzyme catalyzing the enzymatic reaction. It was suggested that plant editing complexes could have a similar constitution, involving at least one RNA binding factor giving the specificity, such as a PPR protein, and one catalytic factor contributing to the enzymatic reaction (Okuda et al., 2006; Chateigner-Boutin and Small, 2010; Boussardon et al., 2012). This model is based on genetic and molecular studies showing that PPR proteins act to specify the target nucleotide through their RNA binding properties. However, the enzymatic activity of editing remains elusive. Most of the PPR proteins involved in editing belong to the PLS subfamily, proteins which possess E and, in some cases, DYW domains. Genetics studies showed that E domain integrity is important to guarantee the editing efficiency (Okuda et al., 2007; Takenaka, 2010; Hartel et al., 2013). Two non-mutually excluding hypotheses are under discussion concerning the function of the E domain. Based on a predicted structural PPR-like conformation, the first hypothesis proposes that this domain stabilizes RNA-protein interaction through its binding to the nucleotides in the vicinity of the target site whereas the second hypothesis suggests a function in the recruitment of the editing enzyme (Chateigner-Boutin and Small, 2010; Boussardon et al., 2012; Yagi et al., 2013).

The DYW domain function is more controversial. It was suggested that it might confer the enzymatic activity for the editing reaction (Salone et al., 2007). One of the arguments against this hypothesis is based on the fact that the DYW domain is absent in the the PLS-E editing factors sequence (Okuda et al., 2007). Another argument is that in the case of CRR22 and CRR28 proteins, DYW domains can be disrupted without affecting the editing efficiency of the proteins (Okuda et al., 2009). We showed in a previous study that in the case of the DYW1 protein, its interaction with CRR4 editing factor contributes to the efficient editing of a same target site (Boussardon et al., 2012). In this complex, the CRR4

RESULTS

protein provides the RNA binding PPR domains and a N-terminal part of the E domain whereas the DYW1 protein contributes to the C-terminal part of the E domain, lost in CRR4, and to the DYW domain. The ability to complement the double *crr4/dyw1* mutant using a translational fusion protein harboring the PPR domains of CRR4, a reconstructed E domain composed of the junction of the CRR4 N-terminal part E domain with the DYW1 C-terminal part, and the DYW domain of DYW1, restoring a typical PLS-E-DYW protein, suggest that the interaction between these proteins could take place around the E domain. A similar PLS-E-DYW conformation might be needed and realized, in the CLB19 editing complex, by interaction between CLB19 and DYW2 proteins. In this complex, the CLB19 protein may contribute, with its complete E domain, to the interaction with the DYW2 protein providing the DYW domain.

The presence of the DYW2 protein in CLB19 editing complex supports our previous results showing that the DYW domain can be supplied in *trans*, reconstituting a PLS-E-DYW type protein. Additionally, the involvement of DYW2 protein in mitochondrial editing of *cox3*-C257 and *mttB*-C144 sites, which are also targets of the PLS-E MEF21 and SLO2 proteins, both harboring intact E domain but no DYW domain, strongly suggest a general model supporting the presence of DYW proteins in all editing complexes. Furthermore, the fact that all the 11 mitochondrial editing sites of *P. patens* are targeted by PLS proteins containing intact E and DYW domains is also in accordance with this model (Ichinose et al., 2013; Schallenberg-Rudinger et al., 2013). In conclusion, all these data strongly suggest that E and DYW domains, as well as PPR domains, are required for C to U editing in plant organelles.

The lack of deamination activity and the endoribonuclease activity of DYW domain containing proteins in *in-vitro* conditions (Nakamura and Sugita, 2008; Okuda et al., 2009) are still the main arguments against the hypothesis proposing the DYW domain as the enzymatic activity of editing. However, an alternative explanation is emerging in which the PPR proteins, including PLS-E-DYW, participate in several steps of organelle RNA metabolism. For example, Nakamura and Sugita (2008) demonstrated that the DYW domain of OTP85 PPR protein has an endoribonuclease activity *in-vitro* whereas Hammani *et al.* (2009) demonstrated through the genetic analysis of the *otp85* mutant that OTP85 is involved in the plastid edition in *A. thaliana* (Hammani et al., 2009). Additionally, an endoribonuclease activity was also suggested for the editing factor

CRR22 (Okuda et al., 2009). Moreover, the implication of the DYW containing CRR2 and *P. patens* PpPPR43 PPR proteins in the cleavage and the splicing of organelle transcripts, respectively, do not exclude the possibility that such proteins might be involved in editing or alternatively has lost its editing activity. Finally, the evidence that a single PPR protein can influence several RNA metabolic processes, for example RNA stabilization and translation for ATP4 (Zoschke et al., 2013), cleavage and editing for SEL1 (Pyo et al., 2013) or splicing and editing for OsPPR4 and OTP90 (Asano et al., 2013; Lopez-Obando *et al.*, in preparation) may explain the controversial results concerning PPR proteins harboring DYW domains.

Can dual localisation of a PPR editing factors be retained even though the protein lost its editing function in one organelle?

Actually more than 45 PPR proteins have a potential dual targeting to both organelles (Colcombet et al., 2013; Fuss et al., 2013). The dual localization of PPR proteins seems to be an ancient and conserved process because several *P. patens* PPR proteins were found to be dually localized as for example PpPPR_63 and PpPPR_104, the putative paralogs of the *A. thaliana* PRORP1 protein (Sugita et al., 2013). According to the literature, three editing factors, OTP87, OTP90 and MEF29, have been shown to be addressed in both organelles. However, they are exclusively involved in editing process in mitochondria, whereas their functions in plastids remain unknown (Hammani et al., 2011; Sosso et al., 2012) (Lopez-Obando *et al.*, in preparation). We provide in this study, the first evidence that a PPR editing factor, DYW2, is involved in the RNA editing process in both organelles. Based on the fact that angiosperm species have a tendency to lose C to U organelle editing sites by reversion to a T nucleotide in the organelles genomes (Huang et al., 2013; Richardson et al., 2013) and that this loss could influence the functional status of a PPR editing factor (Hayes et al., 2012), we previously suggested an explanation for the lack of editing factors activity in one of the two organelles. That particular situation would induce the loss of editing activity of a PPR in this organelle that has lost the editing site, this protein remaining dually localized if it does not alter the RNA metabolic process in that organelle (Lopez-Obando *et al.*, in preparation). Results obtained for DYW2 support partially this hypothesis. Indeed, a correlation between the presence of DYW2 protein and absence of editing of its plastid target sites *rpoA-C200* and *ClpP-C559* in several land plants, notably *O. sativa* and *N. tabacum*, was reported (Chateigner-Boutin and Hanson, 2003; Inada et al.,

RESULTS

2004; Jiang et al., 2012). Thus, considering that a single gene codes the DYW2 protein in *O. sativa*, it would be interesting to study if the localization and the function of OsDYW2 protein is conserved.

C to U editing protein complex composition may have been expanded during land plant evolution.

In terms of plant evolution, the protein content of editing complexes seems to increase in angiosperms compare to that of *P. patens*. Indeed, neither PLS-E nor MORF proteins have been identified in *P. patens* nuclear genome (Sugita et al., 2013; Takenaka et al., 2013b). Moreover taking in account the suggestion that the PLS-E proteins could originated from PLS-E-DYW proteins by deletions of DYW domains (Rivals et al., 2006), we can assume that during the evolution, the angiosperm C to U editing complex composition may have increased simply by decoupling the E and DYW domain from PLS proteins and by the addition of MORF proteins into the complex. Additionally, the presence of additional PPR proteins in the complex, such as the OTP100 protein identified in the present study, suggests an even more complex and unsuspected organization of the editing machinery in angiosperms. This complexity in the RNA editing protein complexes in angiosperms compare to *P. patens* is in the line of the fact that during evolution the molecular machines might increase its complexity by a neutral evolutionary process (Doolittle, 2012; Finnigan et al., 2012).

MATERIALS AND METHODS

Plant material and phenotype characterization of T-DNA insertion lines

A. thaliana ecotype Columbia-0 (Col-0) and Wassilewskija (Ws) were used as wild type plants. The T-DNA insertional mutants, GK_332A07 (*dwy2-1*), FLAG_435F11 (*dyw2-2*) and SALK_002014.56 (*dyw2-3*) for DYW2 and the mutants SALK_069042 (*otp100-1*) and SAIL_784_A11 (*otp100-2*) for OTP100 were obtained from the ABRC stock center. Mutants were genotyped by PCR using primers specific for each mutant line, flanking the T-DNA insertion and T-DNA specific primers in order to identify heterozygous or homozygous mutant plants. All sequences of primers used in this study are available in supplemental table 4. Localizations of the T-DNA insertions were verified by sequencing

RESULTATS

the PCR products corresponding to regions flanking the T-DNA. For phenotype characterization, seeds were sown in soil and cultivated in growth chambers under a 16h light/8h dark cycle at 22°C during the day and 17°C during night. For genetic studies, mutant seeds were surface sterilized for 10 min with a solution containing 10% (v/v) commercial bleach (BAYROL France SAS) diluted in 95° ethyl alcohol and 0.05% (v/v) Teepol 610 (SERVA Electrophoresis GmbH). Seeds were rinsed thoroughly with 95° ethyl alcohol and dried overnight. Seeds were sown on Murashige and Skoog solid medium (MS) containing 0.8% (W/V) agar-agar type E (SIGMA ALDRICH). Seeds were germinated in a growth chamber under long-day conditions (16h-light/8h-dark cycle, 22°C, 45% hygrometry) after cold treatment for 48 h at 4° C. Pollen viability was evaluated after Alexander staining (Alexander, 1969). For BiFC transient expression experiments, *A. thaliana* Landsberg erecta ecotype (Ler-0) plantlets were grown *in vitro* as described below. Seeds were sown on Arabidopsis medium (Estelle and Somerville, 1987) and germinated in a growth chamber (16h-light/8h-dark cycle, 22°C, 45% hygrometry) after a cold treatment for 48 h at 4°C.

TAP-Tag protein expression, purification and MS base detection of protein complex

The sequence of CLB19 without stop codon was cloned into the pKCTAP destination vector (Van Leene *et al.* 2007). *Agrobacterium tumefaciens* containing the TAP-Tag plasmid was used to transform Arabidopsis protoplasts. Protoplasts transformation, protein expression, tandem affinity purification and mass spectrometry protein based detection were performed as previously described by Van Leene *et al.* (2010).

Subcellular protein localisation

DYW2 full-length ORF without stop codon was amplified using DYW2-S and DYW2-E primers (supplemental table 4) and cloned into pDNR207 as described previously (Lurin *et al.*, 2004). P35S::DYW2:DsRed2 binary vector was obtained by LR recombination using p0229-DsRed2 destination vector (Lurin *et al.*, 2004). C58C1 pMP90 *A. tumefaciens* strain containing this binary plasmid was used to transform *A. thaliana* Col-0 plants by floral dip procedure (Clough and Bent, 1998). Primary transformants were identified in soil using BASTA selection and then PCR genotyped for T-DNA insertion verification using P35S-B and DsRed2-Rev primers (supplemental table 4). Subcellular localization of the DYW2:DsRed2 fusion protein was observed in root and cotyledon cells of 10-15 days old

RESULTATS

primary transformants grown on MS solid medium using a spectral Leica SP2 AOBs confocal microscope (Leica Microsystems) equipped with an argon laser and a HeNe laser. To confirm the mitochondrial localization, the seedlings were stained with 1 μ M of Mito-tracker green marker (Invitrogen) for 15-30 min. Signals were detected using laser lines at 488 nm (Mito-tracker green; excitation/emission 488/500-530nm) and 543 nm (DsRed2; excitation/emission 543/570-620nm and chlorophyll autofluorescence; excitation/emission 543/620-700nm). The images were coded green (Mitotraker green), blue (chlorophyll autofluorescence) and red (DsRed2) giving yellow or violet co-localization signals when green and red signals or blue and red signals overlapped in merged images.

Organelle RNA editing analysis

Total RNAs from leaves of 22 day old plants were extracted using RNeasy plant mini kit (QIAGEN). RNAs were treated with 1 μ L of DNase I (Ambion 2U/ μ L) two times for 30 min at 37°C. DNase I was inactivated after 10 min of incubation at 75°C in a solution containing 5.5mM EDTA. RNAs samples were precipitated by ethanol/acetate precipitation and diluted in RNase free water. cDNAs were then synthesized using random hexamer primers and superscript II as described by the manufacturer (Invitrogen). To evaluate RNA editing in chloroplast and mitochondrial transcripts, cDNA and specific primers surrounding known organelle editing events were used for PCR amplifications and sequencing (Boussardon et al., 2012; Takenaka and Brennicke, 2012).

Expression analysis by RT-qPCR

Total RNAs from leaves of 22 day old plants were extracted using RNeasy plant mini kit (QIAGEN). Total RNA integrity was evaluated using capillar Labchip Agilent 2100 bioanalyser station (Agilent). RNAs samples were then treated with 1 μ L of DNase I (Ambion 2U/ μ L) two times for 30 min at 37°C. DNase I was inactivated after 10 min of incubation at 75°C in a solution containing 5.5mM EDTA. RNAs were precipitated by ethanol/acetate precipitation and diluted in RNase free water. 1 μ g of DNase treated RNA samples were used for cDNAs preparation. cDNAs were synthesized using random hexamer primers and superscript II as described by the manufacturer (Invitrogen). Quantitative real time PCR experiments were performed in triplicates in a total volume of 10 μ L containing 5 μ L of Mesa Green qPCR mastermix Plus SYBR Assay (Eurogentec), 4 μ L

RESULTATS

of primer mix (1 μ M each one) and 1 μ l of matrix (cDNA dilutions) using the CFX384 TouchTM Real System (Biorad) with the following program: 95°C 30s ; 95°C 5s then 60°C 30s during 39 cycles ; 95°C 10s ; melting curve 65°C 5s, 95°C, +0,5°C/cycle for 60cycles.

ACKNOWLEDGEMENTS

We thank Pascal Audiger and Florie Vion for their help with plant facilities (Unité de Recherche en Génomique Végétale), José Caius for his technical assistance (Unité de Recherche en Génomique Végétale), Olivier Grandjean (Institut Jean Pierre Bourgin, Versailles, France) for his help with confocal microscopy.

AUTHOR CONTRIBUTIONS

Designed research: C. L., A. F., R. B., and, M. L- O.

Performed research: M.L-O contributed to the production and characterization of mutants and transgenic lines, production of plasmids, localization of proteins and editing analyses. A. F. initiated the projects and contributed to the first functional analyses. R. B. contributed to the characterization of mutants. M. T. contributed to editing analyses in mitochondria G. J. contributed with the Tap-tag experiments. A. A. contributed to the production of plasmids.

Coordination of the project: R. B. and C. L.

FUNDING

This research was supported by the French Ph. D. fellowship associated to the Université d'Evry Val-d'Essonne and "Ministère de la Recherche et de l'Enseignement Supérieur" to M. L-O. M.T. is a Heisenberg fellow and his work is supported by grants from the DFG.

RESULTATS

FIGURE LEGENDS

Figure 1. DYW2 is dually targeted to mitochondria and plastids in transgenic *A. thaliana* plants. Mitochondrial localization of DYW2 was observed in *Arabidopsis* roots and chloroplast localization in *Arabidopsis* cotyledons. DsRed2 signals of DYW2:DsRed2 translational fusion is shown in red (A, E, I), MitoTracker green signals of mitochondria are shown in green (B, F, J) and auto-fluorescence signals of chloroplast are shown in blue (C, G, K). In the merged images, yellow signals correspond to the overlay of fusion protein and Mitotracker green signals indicating mitochondrial localisation of DYW2:DsRed2 whereas purple signals correspond to the overlay of chloroplast autofluorescence and DYW2:DsRed2 signals in chloroplasts (D, H, L).

Figure 2. Predicted domain structure of DYW2 protein and embryo defective macroscopic phenotype in *dyw2* heterozygous plants. **A.** Glutamine rich region, predicted PPR and DYW domains of DYW2. Location of T-DNA insertions in each mutant line is indicated. **B.** Viability of pollen in mature anthers of heterozygous plants of *dyw2-1* and *dyw2-2* mutants examined by Alexander staining. **C.** White embryo phenotype observed in siliques of heterozygous *dyw2-1* and *dyw2-2* mutant plants.

Figure 3. *dyw2-3* homozygous mutant macroscopic phenotype and *DYW2* gene expression analysis. **A.** Macroscopic phenotype of *dyw2-3* homozygous mutant plants in long-day (16h-8h light-dark cycles) conditions. **B.** *DYW2* gene expression in *dyw2-3* homozygous mutant plants grown in long-day conditions. Gene expression was measured by RT-qPCR using total RNA extracted from leaves of two biological replicates. For each biological replicate, the mean expression level of at least three technical RT-qPCR replicates was normalized with the mean of nuclear 18S (black) and actin2 (white) expression, used as reference genes. *OTP100* expression was measured only in the first replicate. **C.** Plastid and mitochondrial editing sites affected in *dyw2-3* homozygous mutant plants. Editing efficiency was measured by cDNA sequencing in homozygous and wild type plants. Amino acid changes involving editing sites are showed below. Arrows point to the corresponding peaks.

Figure 4. *dyw2-3/dyw2-1* and *dyw2-3/dyw2-2* heteroalleles mutant lines macroscopic phenotype and *DYW2* gene expression analysis. **A.** Macroscopic phenotypes of *dyw2-*

RESULTATS

3/*dyw2-1* and *dyw2-3/dyw2-2* heteroalleles mutants in soil and grown in long-day (16h-8h light-dark cycles) conditions. The picture shows progenies of crosses performed using a *dyw2-1* or *dyw2-2* heterozygous mutant plant and a *dyw2-3* homozygous mutant plant. Red arrows indicate plants harboring *dyw2-3/dyw2-1* or *dyw2-3/dyw2-2* heteroalleles **B.** *DYW2* gene expression in heteroalleles lines. *DYW2* gene expression of plants grown in long-day conditions was measured by RT-qPCR using total RNA extracted from leaves of two biological replicates of *dyw2-3* heterozygous, *dyw2-3* homozygous, *dyw2-3/dyw2-1* and *dyw2-3/dyw2-2* mutant plants. For each biological replicate, the mean expression level of at least three technical RT-qPCR replicates was normalized with the mean of nuclear 18S (black) and actin2 (white) expression, used as reference genes. Wild type (Wt) refers to a sister seedling coming from the same self progeny of the heterozygous *dyw2-3/+* mutant line. **C.** Plastid and mitochondrial editing sites affected in *dyw2-3/dyw2-1* and *dyw2-3/dyw2-2* heteroallele mutant plants. Editing efficiency was measured by cDNA sequencing in heteroallele mutant and wild type plants. Amino acid changes involving the editing sites are showed below. Arrows point to the corresponding peaks.

Table 1. Proteins detected with high confidence in the CLB19 chloroplast complex using *A. thaliana* cell suspension cultures and tandem affinity purification mass spectrometric approach (TAP-Tag-MS).

REFERENCES

- Arenas, M.A., Takenaka, M., Moreno, S., Gomez, I., and Jordana, X. (2013). Contiguous RNA editing sites in the mitochondrial nad1 transcript of *Arabidopsis thaliana* are recognized by different proteins. *FEBS Lett* **587**, 887-891.
- Armengaud, P. (2009). EZ-Rhizo software: the gateway to root architecture analysis. *Plant Signal Behav* **4**, 139-141.
- Asano, T., Miyao, A., Hirochika, H., Kikuchi, S., and Kadowaki, K. (2013). A pentatricopeptide repeat gene of rice is required for splicing of chloroplast transcripts and RNA editing of *ndhA*. *Plant Biotechnology*. **30**, 57 - 64.
- Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A. (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol* **42**, 603-613.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* **8**, e1002910.
- Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted *coxII* gene from trypanosome

- mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819-826.
- Bentolila, S., Oh, J., Hanson, M.R., and Bukowski, R.** (2013). Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet* **9**, e1003584.
- Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., and Hanson, M.R.** (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A* **109**, E1453-1461.
- Borner, G.V., Morl, M., Wissinger, B., Brennicke, A., and Schmelzer, C.** (1995). RNA editing of a group II intron in *Oenothera* as a prerequisite for splicing. *Mol Gen Genet* **246**, 739-744.
- Boussardou, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C.** (2012). Two Interacting Proteins Are Necessary for the Editing of the NdhD-1 Site in Arabidopsis Plastids. *Plant Cell* **24**, 3684-3694.
- Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., and Meinke, D.** (2010). Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol* **155**, 1678-1689.
- Carrie, C., Giraud, E., and Whelan, J.** (2009). Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *Febs J* **276**, 1187-1195.
- Castandet, B., Choury, D., Begu, D., Jordana, X., and Araya, A.** (2010). Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res* **38**, 7112-7121.
- Chateigner-Boutin, A.L., and Hanson, M.R.** (2003). Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res* **31**, 2586-2594.
- Chateigner-Boutin, A.L., and Small, I.** (2010). Plant RNA editing. *RNA Biol* **7**, 213-219.
- Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S., and Small, I.** (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J* **65**, 532-542.
- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-Garcia, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., Jimenez, L.F., Lurin, C., Small, I., and Leon, P.** (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* **56**, 590-602.
- Chi, W., Mao, J., Li, Q., Ji, D., Zou, M., Lu, C., and Zhang, L.** (2010). Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in Arabidopsis cotyledons. *Plant J* **64**, 14-25.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**, 735-743.
- Cohen, S., Zmudjak, M., Colas des Francs-Small, C., Malik, S., Shaya, F., Keren, I., Belausov, E., Many, Y., Brown, G.G., Small, I., and Osterseizer-Biran, O.** (2014). nMAT4, a maturase factor required for nad1 pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in Arabidopsis mitochondria. *Plant J*.
- Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthome, R., and Lurin, C.** (2013). Systematic study of subcellular localization of

- Arabidopsis PPR proteins confirms a massive targeting to organelles. *RNA Biol* **10**.
- Covello, P.S., and Gray, M.W.** (1989). RNA editing in plant mitochondria. *Nature* **341**, 662-666.
- Curtis, M.D., and Grossniklaus, U.** (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**, 462-469.
- Cushing, D.A., Forsthoefel, N.R., Gestaut, D.R., and Vernon, D.M.** (2005). Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* **221**, 424-436.
- de Longevialle, A.F., Small, I.D., and Lurin, C.** (2010). Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol Plant* **3**, 691-705.
- de Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D.** (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana. *Plant Cell* **19**, 3256-3265.
- de Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H., and Small, I.** (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.
- 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.
- Delannoy, E., Le Ret, M., Faivre-Nitschke, E., Estavillo, G.M., Bergdoll, M., Taylor, N.L., Pogson, B.J., Small, I., Imbault, P., and Gualberto, J.M.** (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNA^{Arg}(ACG) and is essential for efficient chloroplast translation. *Plant Cell* **21**, 2058-2071.
- Ding, Y.H., Liu, N.Y., Tang, Z.S., Liu, J., and Yang, W.C.** (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**, 815-830.
- Doniwa, Y., Ueda, M., Ueta, M., Wada, A., Kadowaki, K., and Tsutsumi, N.** (2010). The involvement of a PPR protein of the P subfamily in partial RNA editing of an Arabidopsis mitochondrial transcript. *Gene* **454**, 39-46.
- Doolittle, W.F.** (2012). Evolutionary biology: A ratchet for protein complexity. *Nature* **481**, 270-271.
- Estelle, M.A., and Somerville, C.** (1987). Auxin-resistant mutants of Arabidopsis thaliana with an altered morphology. *Molecular and General Genetics MGG* **206**, 200-206.
- Farre, J.C., Aknin, C., Araya, A., and Castandet, B.** (2012). RNA editing in mitochondrial trans-introns is required for splicing. *PLoS One* **7**, e52644.
- Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., and Thornton, J.W.** (2012). Evolution of increased complexity in a molecular machine. *Nature* **481**, 360-364.
- Freyer, R., Kiefer-Meyer, M.C., and Kossel, H.** (1997). Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A* **94**, 6285-6290.
- Fujii, S., and Small, I.** (2011). The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol* **191**, 37-47.

- Fujii, S., Sato, N., and Shikanai, T.** (2013). Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON gradient regulation3. *Plant Cell* **25**, 3079-3088.
- Fuss, J., Liegmann, O., Krause, K., and Rensing, S.A.** (2013). Green Targeting Predictor and Ambiguous Targeting Predictor 2: the pitfalls of plant protein targeting prediction and of transient protein expression in heterologous systems. *New Phytol.*
- Geerts, P., Toussaint, A., Mergeai, G., and Baudoin, J.P.** (2011). Phaseolus immature embryo rescue technology. *Methods Mol Biol* **710**, 117-129.
- Gobert, A., Gutmann, B., Taschner, A., Gossringer, M., Holzmann, J., Hartmann, R.K., Rossmannith, W., and Giege, P.** (2010). A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol* **17**, 740-744.
- Goringer, H.U.** (2012). 'Gestalt,' composition and function of the Trypanosoma brucei editosome. *Annu Rev Microbiol* **66**, 65-82.
- Grosjean, H.** (2005). Fine-tuning of RNA functions by modification and editing (Berlin ; New York: Springer), pp. xxiv, 442 p.
- Gualberto, J.M., Weil, J.H., and Grienemberger, J.M.** (1990). Editing of the wheat coxIII transcript: evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. *Nucleic Acids Res* **18**, 3771-3776.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., and Grienemberger, J.M.** (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* **341**, 660-662.
- Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T., and Small, I.** (2009). A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* **21**, 3686-3699.
- Hammani, K., Colas des Francs-Small, C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A., and Small, I.** (2011). The Pentatricopeptide Repeat Protein OTP87 Is Essential for RNA Editing of nad7 and atp1 Transcripts in Arabidopsis Mitochondria. *J Biol Chem.* **286**, 21361-21371.
- Hartel, B., Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M.** (2013). MEF10 is required for RNA editing at nad2-842 in mitochondria of Arabidopsis thaliana and interacts with MORF8. *Plant Mol Biol* **81**, 337-346.
- Hayes, M.L., Giang, K., and Mulligan, R.M.** (2012). Molecular evolution of pentatricopeptide repeat genes reveals truncation in species lacking an editing target and structural domains under distinct selective pressures. *BMC Evol Biol* **12**, 66.
- Hayes, M.L., Giang, K., Berhane, B., and Mulligan, R.M.** (2013). Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem* **288**, 36519-36529.
- Hecht, J., Grewe, F., and Knoop, V.** (2011). Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol* **3**, 344-358.
- Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A.** (1989). RNA editing in plant mitochondria. *Science* **246**, 1632-1634.
- Hoch, B., Maier, R.M., Appel, K., Igloi, G.L., and Kossel, H.** (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* **353**, 178-180.

- Howard, M.J., Lim, W.H., Fierke, C.A., and Koutmos, M.** (2012). Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A* **109**, 16149-16154.
- Huang, Y.Y., Matzke, A.J., and Matzke, M.** (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* **8**, e74736.
- Ichinose, M., Tasaki, E., Sugita, C., and Sugita, M.** (2012). A PPR-DYW protein is required for splicing of a group II intron of *cox1* pre-mRNA in *Physcomitrella patens*. *Plant J* **70**, 271-278.
- Ichinose, M., Sugita, C., Yagi, Y., Nakamura, T., and Sugita, M.** (2013). Two DYW subclass PPR proteins are involved in RNA editing of *ccmFc* and *atp9* transcripts in the moss *Physcomitrella patens*: first complete set of PPR editing factors in plant mitochondria. *Plant Cell Physiol* **54**, 1907-1916.
- Inada, M., Sasaki, T., Yukawa, M., Tsudzuki, T., and Sugiura, M.** (2004). A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant Cell Physiol* **45**, 1615-1622.
- Ito, J., Heazlewood, J.L., and Millar, A.H.** (2006). Analysis of the soluble ATP-binding proteome of plant mitochondria identifies new proteins and nucleotide triphosphate interactions within the matrix. *J Proteome Res* **5**, 3459-3469.
- Jacobs, J., Marx, C., Kock, V., Reifschneider, O., Franzel, B., Krisp, C., Wolters, D., and Kuck, U.** (2013). Identification of a chloroplast ribonucleoprotein complex containing trans-splicing factors, intron RNA, and novel components. *Mol Cell Proteomics* **12**, 1912-1925.
- Jiang, Y., Fan, S.L., Song, M.Z., Yu, J.N., and Yu, S.X.** (2012). Identification of RNA editing sites in cotton (*Gossypium hirsutum*) chloroplasts and editing events that affect secondary and three-dimensional protein structures. *Genet Mol Res* **11**, 987-1001.
- Kagami, H., Nagano, H., Takahashi, Y., Mikami, T., and Kubo, T.** (2012). Is RNA editing implicated in group II intron survival in the angiosperm mitochondrial genome? *Genome* **55**, 75-79.
- Kanesaki, Y., Imamura, S., Minoda, A., and Tanaka, K.** (2012). External light conditions and internal cell cycle phases coordinate accumulation of chloroplast and mitochondrial transcripts in the red alga *Cyanidioschyzon merolae*. *DNA Res* **19**, 289-303.
- Ke, J., Chen, R.Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.K., Melcher, K., and Xu, H.E.** (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struct Mol Biol* **20**, 1377-1382.
- Keren, I., Tal, L., des Francs-Small, C.C., Araujo, W.L., Shevtsov, S., Shaya, F., Fernie, A.R., Small, I., and Ostersetzer-Biran, O.** (2012). nMAT1, a nuclear-encoded maturase involved in the trans-splicing of *nad1* intron 1, is essential for mitochondrial complex I assembly and function. *Plant J* **71**, 413-426.
- Keren, I., Bezawork-Geleta, A., Kolton, M., Maayan, I., Belausov, E., Levy, M., Mett, A., Gidoni, D., Shaya, F., and Ostersetzer-Biran, O.** (2009). AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in *Arabidopsis* mitochondria. *RNA* **15**, 2299-2311.
- Khrouchtchova, A., Monde, R.A., and Barkan, A.** (2012). A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *Rna* **18**, 1197-1209.

- Klodmann, J., Senkler, M., Rode, C., and Braun, H.P.** (2011). Defining the protein complex proteome of plant mitochondria. *Plant Physiol* **157**, 587-598.
- Knoop, V.** (2011). When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* **68**, 567-586.
- Kohler, D., Schmidt-Gattung, S., and Binder, S.** (2009). The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of *Arabidopsis thaliana*. *Plant Mol Biol*.
- Kotera, E., Tasaka, M., and Shikanai, T.** (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326-330.
- Lee, C.P., Eubel, H., O'Toole, N., and Millar, A.H.** (2008). Heterogeneity of the mitochondrial proteome for photosynthetic and non-photosynthetic *Arabidopsis* metabolism. *Mol Cell Proteomics* **7**, 1297-1316.
- Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I.** (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089-2103.
- Malek, O., Lattig, K., Hiesel, R., Brennicke, A., and Knoop, V.** (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *Embo J* **15**, 1403-1411.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T.** (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**, 34-41.
- Nakamura, T., and Sugita, M.** (2008). A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* **582**, 4163-4168.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I.** (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* **25**, 1120-1128.
- Okuda, K., and Shikanai, T.** (2012). A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic Acids Res* **40**, 5052-5064.
- Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T.** (2006). A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* **281**, 37661-37667.
- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T.** (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* **104**, 8178-8183.
- Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T.** (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* **21**, 146-156.
- Prikryl, J., Rojas, M., Schuster, G., and Barkan, A.** (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci U S A* **108**, 415-420.
- Pyo, Y.J., Kwon, K.C., Kim, A., and Cho, M.H.** (2013). Seedling Lethal1, a pentatricopeptide repeat protein lacking an E/E+ or DYW domain in *Arabidopsis*,

- is involved in plastid gene expression and early chloroplast development. *Plant Physiol* **163**, 1844-1858.
- Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J., and Palmer, J.D.** (2013). The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol* **11**, 29.
- Rigas, S., Daras, G., Tsitsekian, D., and Hatzopoulos, P.** (2011). The multifaceted role of Lon proteolysis in seedling establishment and maintenance of plant organelle function: living from protein destruction. *Physiol Plant* **145**, 215-223.
- Rivals, E., Bruyere, C., Toffano-Nioche, C., and Lecharny, A.** (2006). Formation of the *Arabidopsis* pentatricopeptide repeat family. *Plant Physiol* **141**, 825-839.
- Rudinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V.** (2012). Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J Mol Evol* **74**, 37-51.
- Ruwe, H., Castandet, B., Schmitz-Linneweber, C., and Stern, D.B.** (2013). *Arabidopsis* chloroplast quantitative editotype. *FEBS Lett* **587**, 1429-1433.
- Salone, V., Rudinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C.** (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* **581**, 4132-4138.
- Schagger, H., Cramer, W.A., and von Jagow, G.** (1994). Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* **217**, 220-230.
- Schallenberg-Rudinger, M., Kindgren, P., Zehrmann, A., Small, I., and Knoop, V.** (2013). A DYW-protein knockout in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J* **76**, 420-432.
- Schmitz-Linneweber, C., and Small, I.** (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* **13**, 663-670.
- Schuster, W., Hiesel, R., Wissinger, B., and Brennicke, A.** (1990). RNA editing in the cytochrome b locus of the higher plant *Oenothera berteriana* includes a U-to-C transition. *Mol Cell Biol* **10**, 2428-2431.
- Shikanai, T.** (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* **63**, 698-708.
- Small, I.D., and Peeters, N.** (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* **25**, 46-47.
- Solheim, C., Li, L., Hatzopoulos, P., and Millar, A.H.** (2012). Loss of Lon1 in *Arabidopsis* changes the mitochondrial proteome leading to altered metabolite profiles and growth retardation without an accumulation of oxidative damage. *Plant Physiol* **160**, 1187-1203.
- Sosso, D., Mbello, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., Dauzat, M., Heurtevin, L., Guyon, V., Takenaka, M., and Rogowsky, P.M.** (2012). PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* **24**, 676-691.
- Stern, D.B., Goldschmidt-Clermont, M., and Hanson, M.R.** (2010). Chloroplast RNA metabolism. *Annu Rev Plant Biol* **61**, 125-155.
- Sugita, M., Ichinose, M., Ide, M., and Sugita, C.** (2013). Architecture of the PPR gene family in the moss *Physcomitrella patens*. *RNA Biol* **10**.

- Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M.R., and Bentolila, S.** (2013). An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. *Proc Natl Acad Sci U S A* **110**, E1169-1178.
- Takenaka, M.** (2010). MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the nad7 transcript in mitochondria of Arabidopsis. *Plant Physiol* **152**, 939-947.
- Takenaka, M., and Brennicke, A.** (2012). Using multiplex single-base extension typing to screen for mutants defective in RNA editing. *Nat Protoc* **7**, 1931-1945.
- Takenaka, M., Verbitskiy, D., Zehrmann, A., and Brennicke, A.** (2010). Reverse genetic screening identifies five E-class PPR-proteins involved in RNA editing in mitochondria of Arabidopsis thaliana. *J Biol Chem*.
- Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K.** (2013a). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* **8**, e65343.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Hartel, B., and Brennicke, A.** (2013b). RNA editing in plants and its evolution. *Annu Rev Genet* **47**, 335-352.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelman, M., Hartel, B., and Brennicke, A.** (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A* **109**, 5104-5109.
- Tillich, M., Hardel, S.L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J.M., Lehwark, P., Leister, D., Small, I.D., and Schmitz-Linneweber, C.** (2009). Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A* **106**, 6002-6007.
- Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G., and Jones, J.D.** (1999). Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. *Plant Cell* **11**, 1841-1852.
- Tourasse, N.J., Choquet, Y., and Vallon, O.** (2013). PPR proteins of green algae. *RNA Biol* **10**.
- Van Leene, J., Stals, H., Eeckhout, D., Persiau, G., Van De Slijke, E., Van Isterdael, G., De Clercq, A., Bonnet, E., Laukens, K., Remmerie, N., Henderickx, K., De Vijlder, T., Abdelkrim, A., Pharazyn, A., Van Onckelen, H., Inze, D., Witters, E., and De Jaeger, G.** (2007). A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. *Mol Cell Proteomics* **6**, 1226-1238.
- Van Leene, J., Hollunder, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Stals, H., Van Isterdael, G., Verkest, A., Neiryneck, S., Buffel, Y., De Bodt, S., Maere, S., Laukens, K., Pharazyn, A., Ferreira, P.C., Eloy, N., Renne, C., Meyer, C., Faure, J.D., Steinbrenner, J., Beynon, J., Larkin, J.C., Van de Peer, Y., Hilson, P., Kuiper, M., De Veylder, L., Van Onckelen, H., Inze, D., Witters, E., and De Jaeger, G.** (2010). Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana. *Mol Syst Biol* **6**, 397.
- Verbitskiy, D., Hartel, B., Zehrmann, A., Brennicke, A., and Takenaka, M.** (2011). The DYW-E-PPR protein MEF14 is required for RNA editing at site matR-1895 in mitochondria of Arabidopsis thaliana. *FEBS Lett* **585**, 700-704.
- Verbitskiy, D., Zehrmann, A., Hartel, B., Brennicke, A., and Takenaka, M.** (2012). Two related RNA-editing proteins target the same sites in mitochondria of Arabidopsis thaliana. *J Biol Chem* **287**, 38064-38072.

- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T.** (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* **8**, e57286.
- Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.W., Wang, J., Zhu, J.K., Shi, Y., and Yan, N.** (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168-171.
- Yoshinaga, K., Kakehi, T., Shima, Y., Iinuma, H., Masuzawa, T., and Ueno, M.** (1997). Extensive RNA editing and possible double-stranded structures determining editing sites in the atpB transcripts of hornwort chloroplasts. *Nucleic Acids Res* **25**, 4830-4834.
- Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M.** (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* **21**, 558-567.
- Zhu, Q., Dugardeyn, J., Zhang, C., Takenaka, M., Kuhn, K., Craddock, C., Smalle, J., Karampelias, M., Denecke, J., Peters, J., Gerats, T., Brennicke, A., Eastmond, P., Meyer, E.H., and Van Der Straeten, D.** (2012). SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J* **71**, 836-849.
- Zmudjak, M., Colas des Francs-Small, C., Keren, I., Shaya, F., Belausov, E., Small, I., and Ostersetzer-Biran, O.** (2013). mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in *Arabidopsis*. *New Phytol* **199**, 379-394.
- Zoschke, R., Watkins, K.P., and Barkan, A.** (2013). A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* **25**, 2265-2275.

RESULTATS

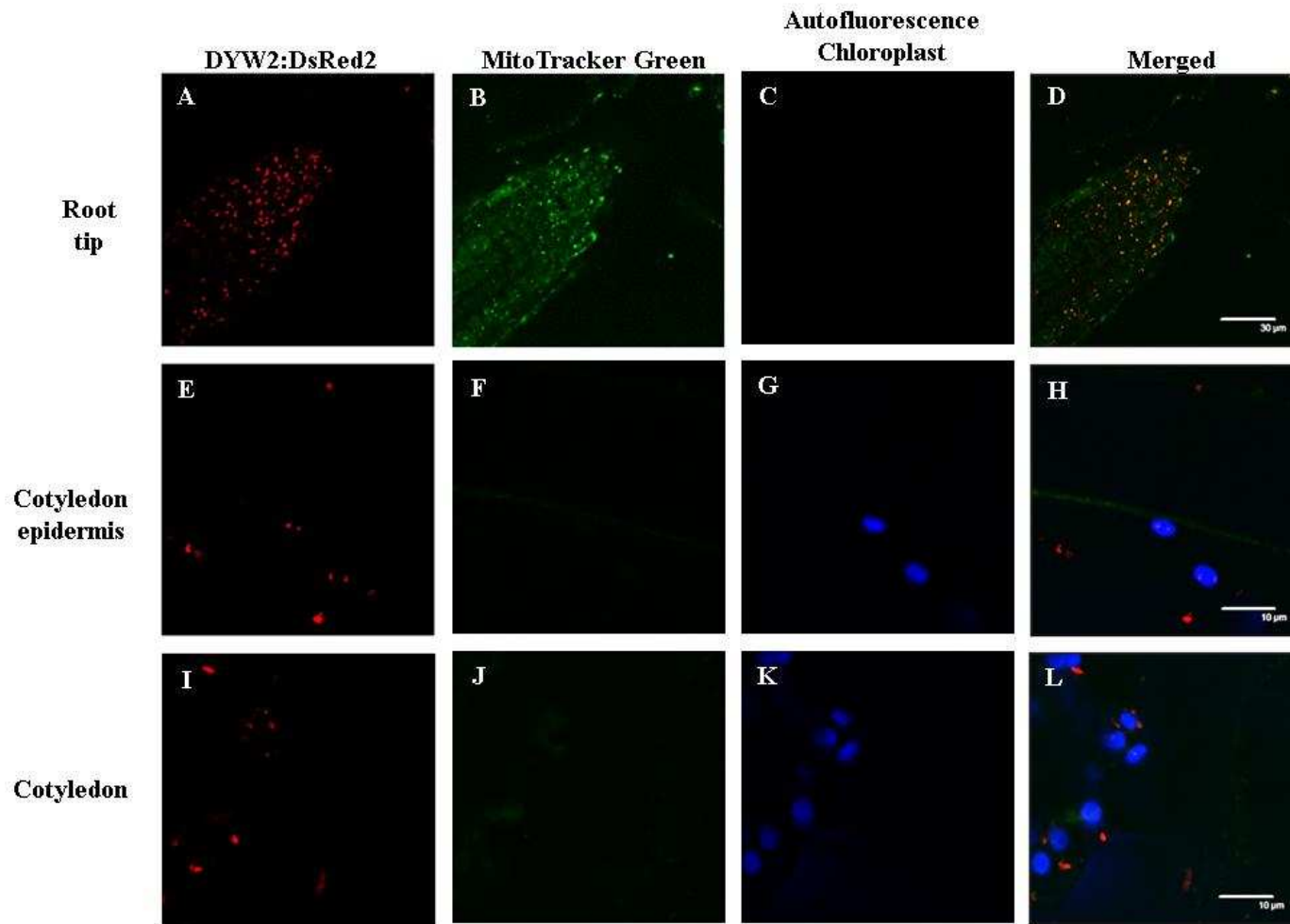


Figure 1.

RESULTATS

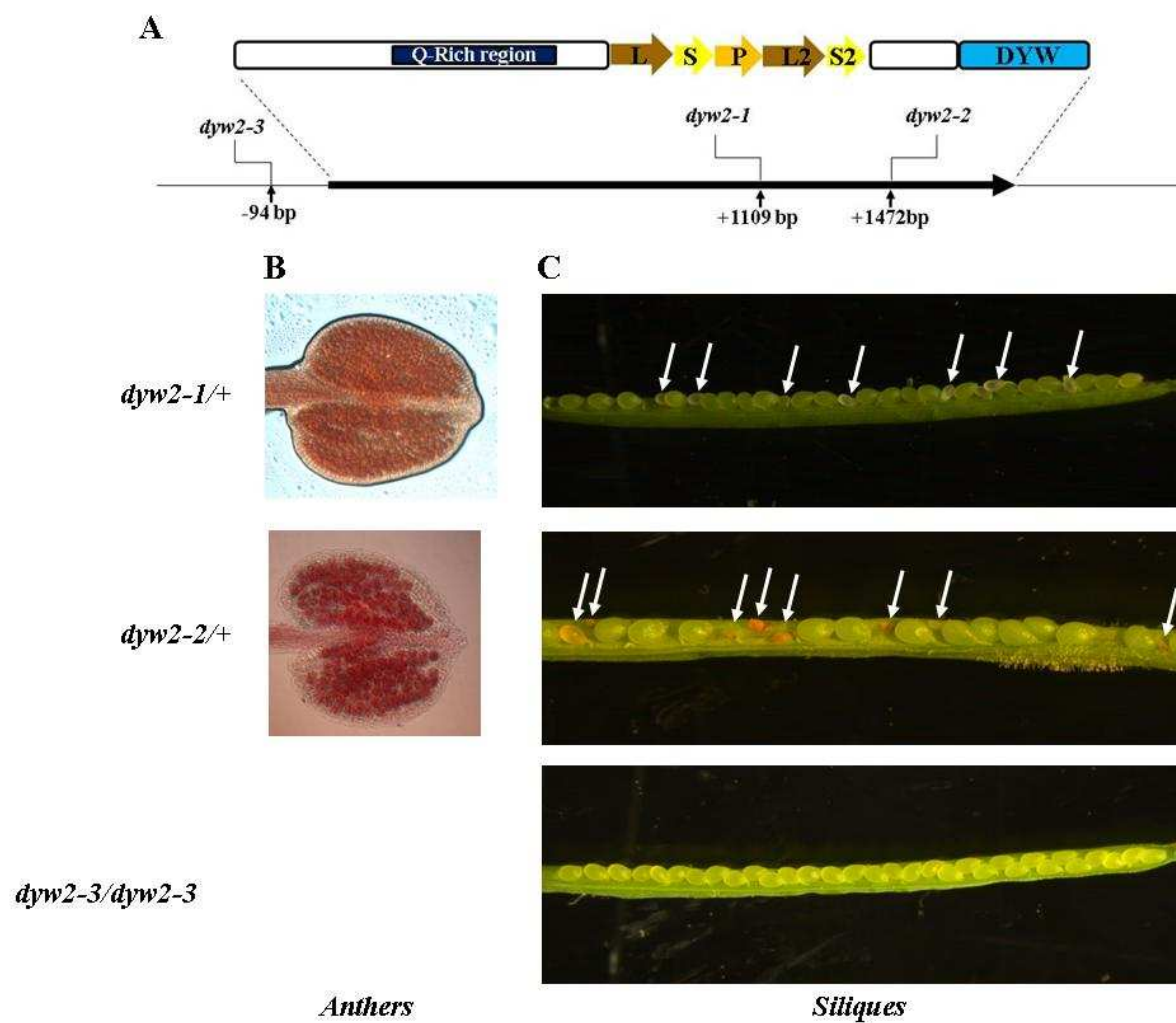


Figure 2.

RESULTATS

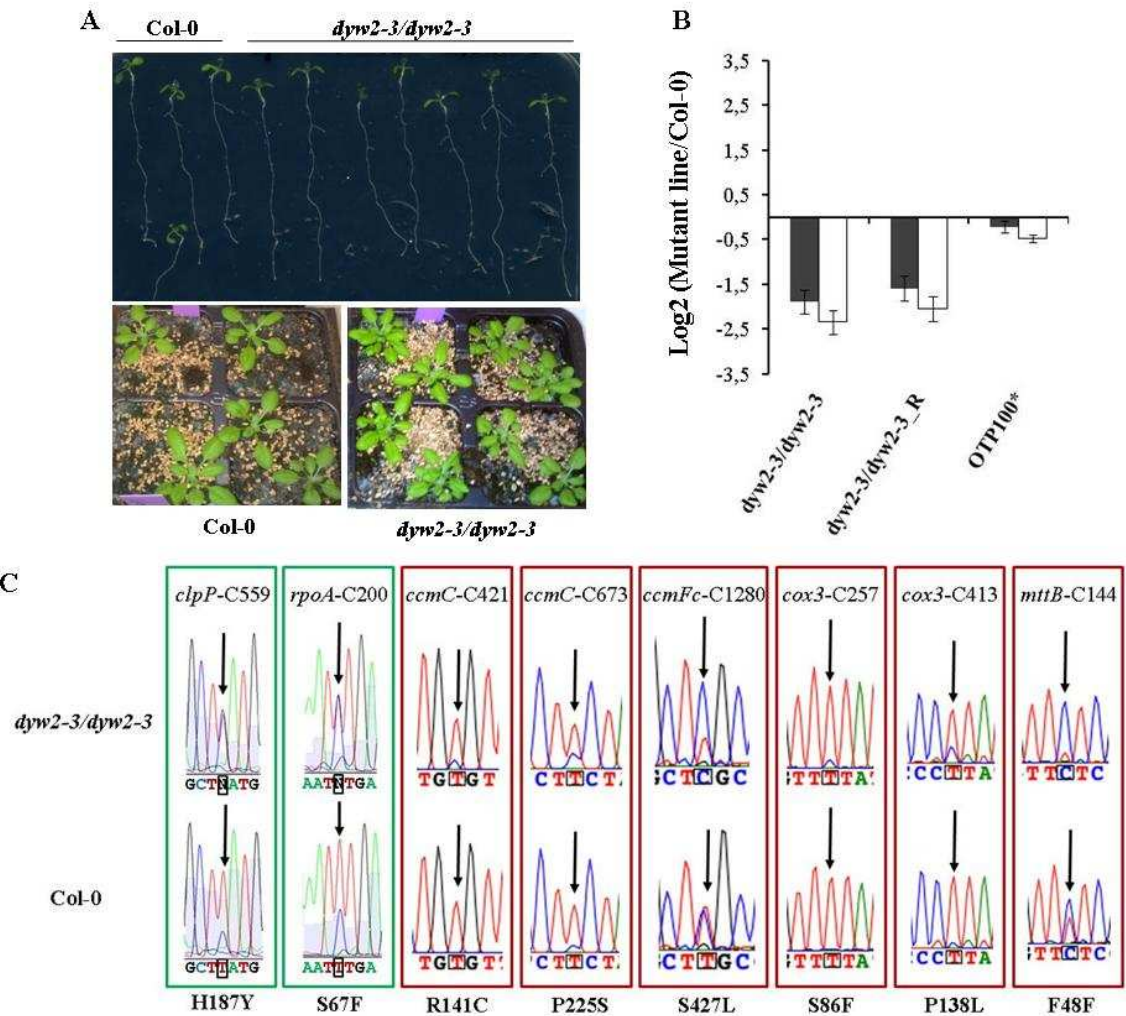


Figure 3.

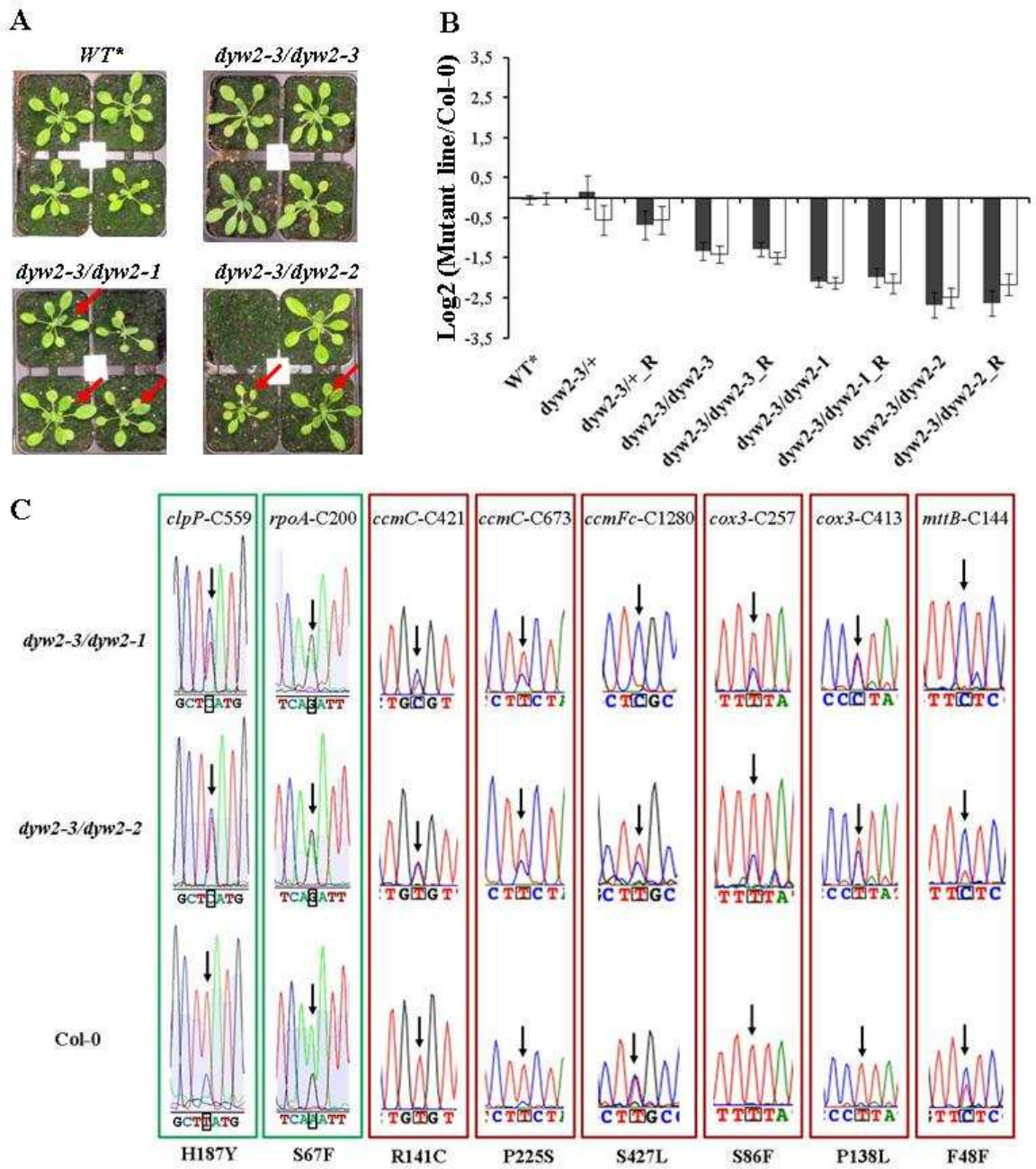


Figure 4.

RESULTATS

AGI ^(β)	Name ^(γ)	Prot. Weight (KDa)	Loc. ^(δ)	Experiment 1 ^(ε)			Experiment 2			Experiment 3			Experiment 4		
				PC	PS	IS	PC	PS	IS	PC	PS	IS	PC	PS	IS
AT1G05750*	CLB19	56.4	C	14	697	92.2	15	750	105.2	11	340	100.3	11	630	150.1
AT2G15690	DYW2	66.3	M/C**	8	167	60.7	10	243	103.9	2	66.1	60.4	6	123	80.4
AT2G28000	CPN60A	62.2	M/C	16	490	95.7	17	514	87.9	14	524	104.5	13	776	170.4
AT3G49240	OTP100	71.7	M/C	16	469	82.6	20	581	93.8	11	201	102.4	15	278	81.2

(β) Arabidopsis genome initiative annotation identifier in TAIR data base version 10

(γ) CLB19: ChLoroplast Biogenesis 19, OTP100: Organellar Transcript Processing 100, CPN60A: Chaperonin-60 α, LON1: Lon protease 1

(δ) Localisations of proteins were obtained form TAIR and Colcombet et al. (2013). C: plastid, M/C: Mitochondria and plastid

(ε) PC: Peptide Count, PS: Protein Score, IS: Best Ion Score

* Protein used as bait for all four experiments.

** This report.

Table 1.

RESULTATS

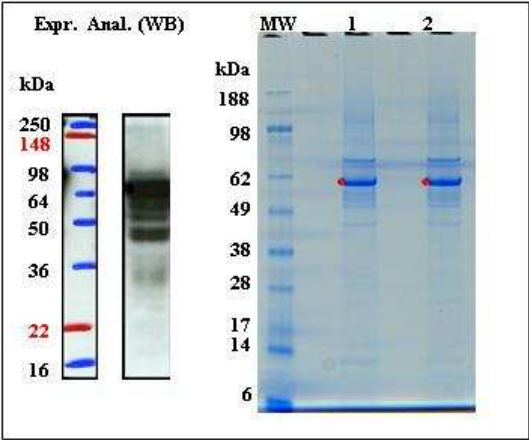
TAP Report 1

• Bait	CLB19 (in PSB-L !)
MW	55.8 kDa
Tag	CmTAP (22 kDa; after TEV cleavage: 7.5 kDa)
Protocol	standard
Status	MS results

TAP Report 2

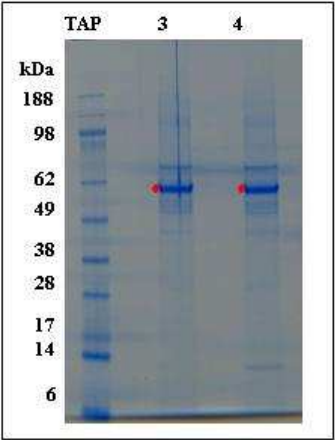
• Bait	CLB19 (in PSB-L !)
MW	55.8 kDa
Tag	CmTAP (22 kDa; after TEV cleavage: 7.5 kDa)
Protocol	standard vs. standard+Rnase treatment
Status	MS results

Expression analysis and TAP SDS-PAGE



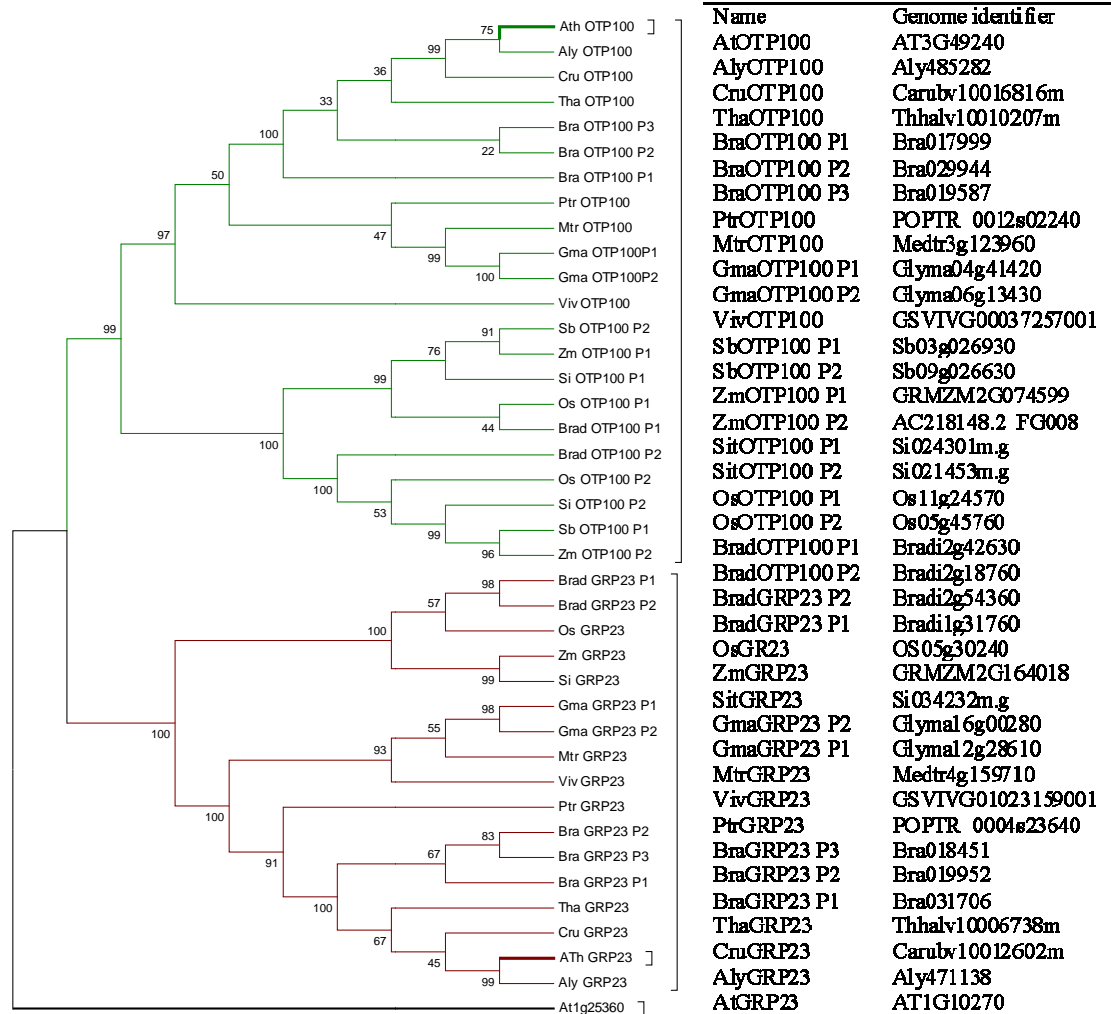
1: 35S-CLB19-CmTAP-1
2: 35S-CLB19-CmTAP-2

TAP SDS-PAGE

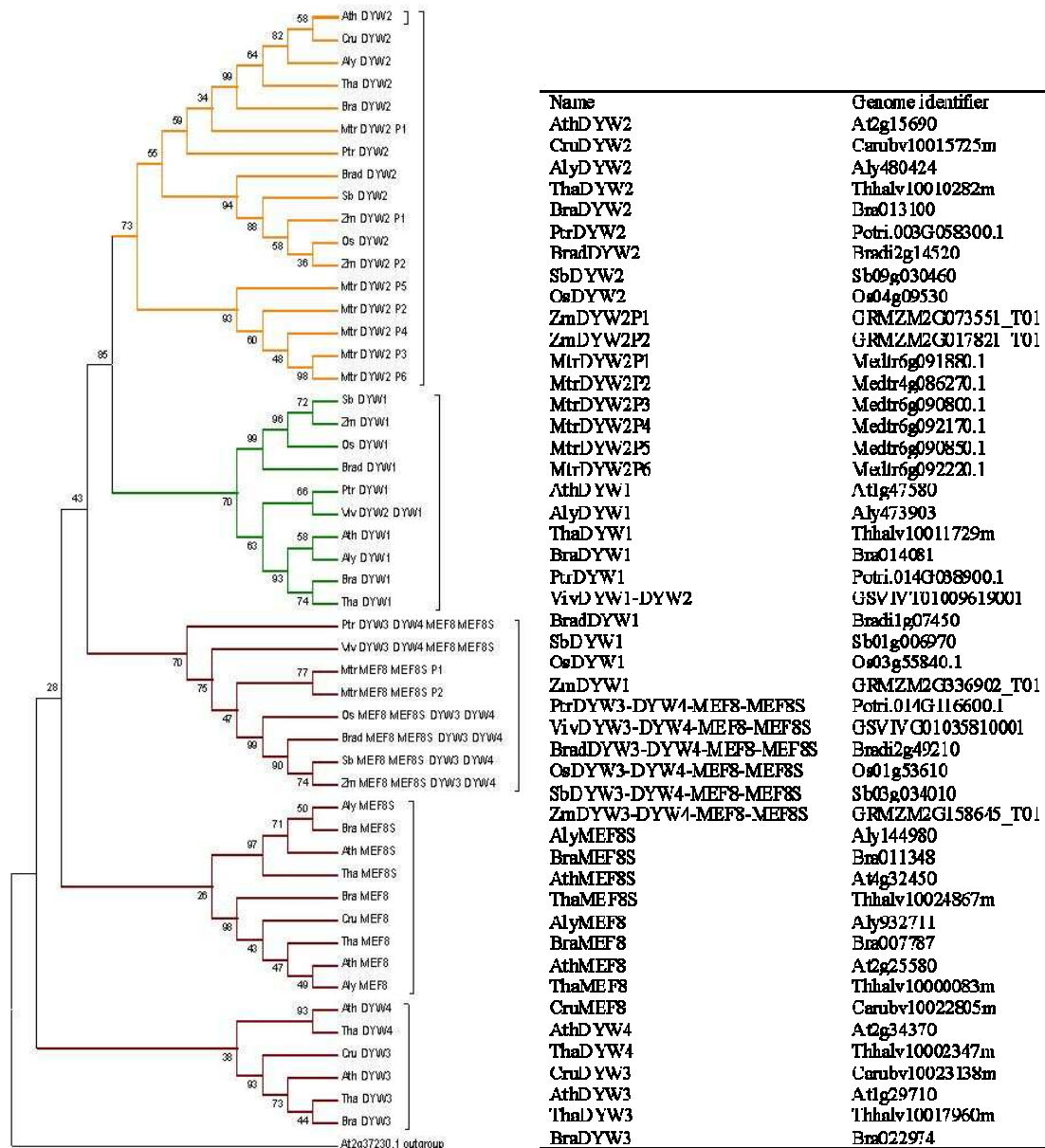


3: 35S-CLB19-CmTAP-3
4: 35S-CLB19-CmTAP-4 (+Rnase)

Supplemental Figure 1. Expression and SDS-PAGE gel analysis after tandem affinity Purification. Gels showing the migration of purified complexes were stained with coomassie blue and the protein use as bait is indicated with a red asterisk in each 4 independent experiments.

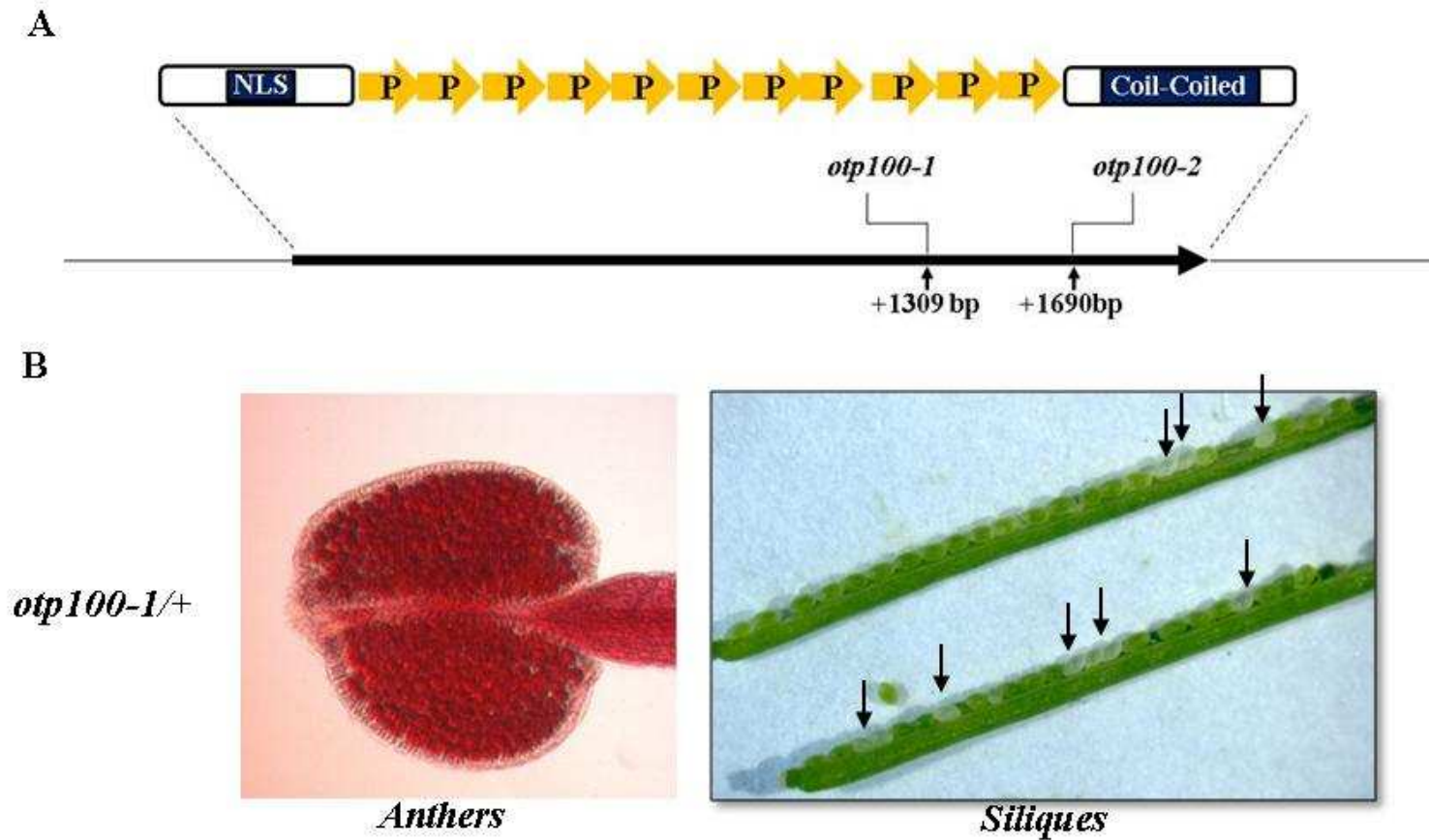


Supplemental Figure 2. Phylogenetic relationships of OTP100 orthologous proteins in 14 angiosperms species. OTP100 protein sequences were obtained by blast in selected genomes using phytozome web page V9.1. Best matches of each specy were blasted to others genomes and the most related sequences to OTP100 and GRP23 were recovered. Phylogenetic relationships were analyzed in MEGA5 software (Tamura *et al.*, 2011) using sequence alignments obtained by MUSCLE (Edgar R. 2004). Maximum likelihood method using the best model fitting the data, the Jones-Taylor-Thornton (JTT) model and gamma distribution (G) with 1000 bootstraps for the statistical representation of branch points were used to obtain the tree. Sequence of AT1G25360 was used as outgroup. AtOTP100: (*A. thaliana* OTP100) orthologs in other species are: AlyOTP100 (*Arabidopsis Lyrata* OTP100), CruOTP100 (*Capsella rubella* OTP100), ThaOTP100 (*Thellungiella halophila* OTP100), PtrOTP100 (*Populus trichocarpa* OTP100), MtrOTP100 (*Medicago truncatula* OTP100), BraOTP100 (*Brasica rapa* OTP100), GmaOTP100 (*Glycine max* OTP100) ZmOTP100 (*Zea mays* OTP100), SbOTP100 (*Sorghum bicolor* OTP100), SitOTP100 (*Setaria italica* OTP100), OsOTP100 (*Oriza sativa* OTP100), VivOTP100 (*Vitis vinifera* OTP100), BradOTP100 (*Brachypodium distachyon* OTP100)



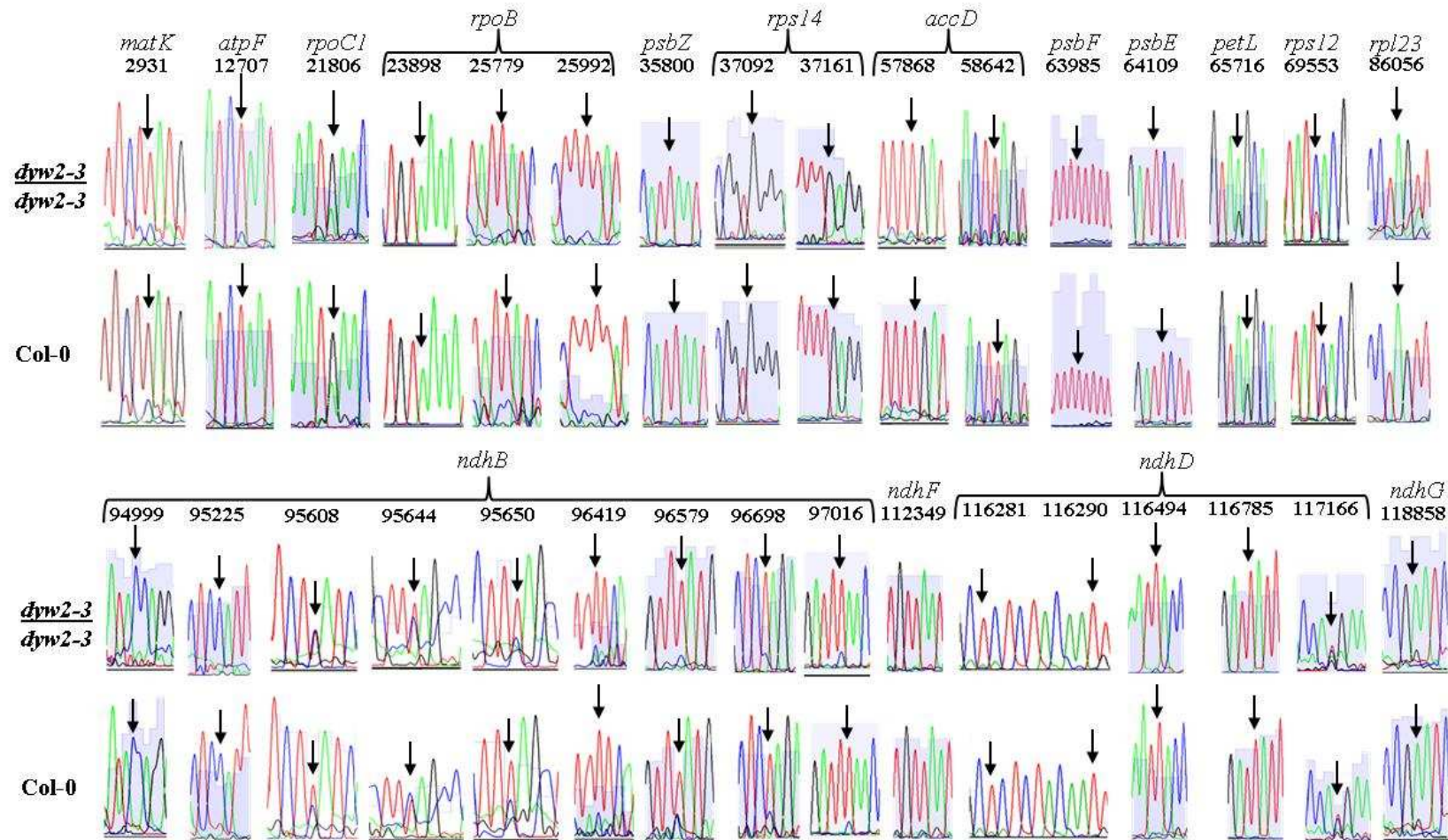
Supplemental Figure 3. Phylogenetic relationships of DYW2 orthologous proteins in 12 angiosperms species. DYW2 protein sequences were obtained by blast in selected genomes using phytozome web page V9.1. Best matches of each specy were blasted to others genomes and the most related sequences to DYW2 and GRP23 were recovered. Phylogenetic relationships were analysed in MEGA5 software (Tamura *et al.*, 2011) using sequence alignments obtained by MUSCLE (Edgar R. 2004). Maximum likelihood method using the best model fitting the data, the Jones-Taylor-Thornton (JTT) model and gamma distribution (G) with 1000 bootstraps for the statistical representation of branch points were used to obtain the tree. Sequence of AT2G37230 was used as outgroup AtDYW2: (*A. thaliana* DYW2) orthologs in other species are: AlyDYW2 (*Arabidopsis Lyrata* DYW2), CruDYW2 (*Capsella rubellal* DYW2), ThaDYW2 (*Thellungiella halophila* DYW2), PtrDYW2 (*Poplus trichocarpa* DYW2), MtrDYW2 (*Medicago truncatula* DYW2), BraDYW2 (*Brasica rapa* DYW2), ZmDYW2 (*Zea mays* DYW2), SbDYW2 (*Sorghum bicolor* DYW2), SitDYW2 (*Setaria italica*), OsDYW2 (*Oriza sativa* DYW2), VivDYW2 (*Vitis vinifera* DYW2), BradDYW2 (*Brachypodium distachyon* DYW2)

RESULTATS



Supplemental Figure 4. Predicted domain structure of OTP100 protein and embryo defective macroscopic phenotype in *otp100* heterozygous plants. **A.** Nuclear Localisation Signal (NLS), predicted PPR and coil-coiled domains of OTP100. Locations of T-DNA insertions in each mutant line are shown. **B.** Viability of pollen in mature anthers of heterozygous plants of *otp100-1* examined by Alexander staining (Alexander, 1969) and white embryo phenotype observed in siliques of heterozygous *otp100-1* mutants plants.

RESULTATS



Supplemental Figure 5. Chloroplast Editotype of *dyw2-3* homozygous mutant compared to its Col-0 counterpart. Sequencing chromatograms corresponding to nucleotide sequences of 32 RT-PCR products obtained using wild type Col-0 and *dyw2-3* mutant total RNA extracted from leaves are shown. Chloroplast genome positions of editing sites are indicated and arrows point to the corresponding peaks.

RESULTATS

Supplemental Table 1. Proteins with a low probability to participate in the CLB19 chloroplast RNA editing complex purified in at least one experiment of tandem affinity purification using mass spectrometry (TAP-Tag-MS)

AGI ^(β)	Name ^(γ)	Prot. Weight (KDa)	Loc. ^(δ)	Experiment 1 ^(ε)			Experiment 2			Experiment 3			Experiment 4		
				PC	PS	IS	PC	PS	IS	PC	PS	IS	PC	PS	IS
AT1G04810	Proteasome subunit	109.2	Ct/N	10	69.7	29.9									
AT1G69800	CBS protein	51.7	-	1	26.2	26.2	2	32	27.2				1	41.2	37.4
AT3G05780	LON3	103.9	M				2	38.1	38.1				1	34.7	34.7
AT3G24255	Polymerase protein*	85.3	pCt	3	49.2	39.4	5	58.5	41.7						
AT3G29310	Calmodulin binding	48.7	-	2	47.4	39.4	4	59.3	41.7						
AT3G32130	Unknown	25.3	pS				1	29.2	25.6				1	30	26.4
AT4G30990	ARM protein	240.1	G	7	36.8	35.4	6	36.3	35						
AT5G26860	LON1	109.7	M/C				10	114	44.2				8	57.3	34.7
AT5G51570	SPFH Band7	32.6	pN										5	78.9	45.9

(β) Arabidopsis genome initiative annotation identifier in TAIR data base version 10

(γ) CBS: Cystathionine beta-synthase, LON1: Lon protease 1, LON3: Lon protease 3 ARM: Armadillo like-fold protein

(δ) Localisation of proteins were obtained from TAIR, SUBA3 or PPDB databases. Ct/N: Cytosol and nucleus, pCt: Predicted cytosol, pS: Predicted secreted, G: Golgi, M/C: Mitochondria and plastid, pN: predicted nucleus

(ε) PC: Peptide Count, PS: Protein Score, IS: Best Ion Score

*RNA-direct DNA polymerase protein (Reverse transcriptase)

RESULTS

Supplemental Table 2. Genotype of the progenies of selfcross of *dyw2-1/+* and *dyw2-2/+* heterozygous mutants.

Selfcrosses		Genotype for the mutation*			
Female	Male	Wt	HT	HO	ungerminated
<i>dyw2-1</i>	<i>dyw2-1</i>	70	98	-	30
<i>dyw2-2</i>	<i>dyw2-2</i>	39	56	-	12

Supplemental Table 3. Genetic complementation test between heterozygous *dyw2-1/+* and *dyw2-2/+* mutant lines.

Reciprocal Crosses			Genotype for the mutation			
Female	Male	Wt	<i>dyw2-1/+</i>	<i>dyw2-2/+</i>	<i>dyw2-1/dyw2-2</i>	ungerminated
<i>dyw2-1</i>	<i>dyw2-2</i>	35	47	40	-	38
<i>dyw2-2</i>	<i>dyw2-1</i>	28	21	20	-	17

RESULTATS

Supplemental Table 4. Sequences of primers used in this study.

Name	Sequence (5' to 3')
Genotyping	
dyw2-1_LP	TCAAAATCCTCAACAAGGTGG
dyw2-1_RP	GCAATCGCTAACCTCTCACTG
dyw2-2_LP	GTTCTGTGGAAGCTTGGTGAC
dyw2-2_RP	ATGACGAAACATGGGTTGAAG
dyw2-3_LP	TCCAAAGAACATCGAAGGATG
dyw2-3_RP	TTGTTGAGGATTTTGACCACC
otp100-1_LP	TGGCTGTAAAGGTTTTGTGG
otp100-1_RP	TTTAGCCATCTCCACCACATC
otp100-2_LP	TGACTTGACCAGTTTGCTGTG
otp100-2_RP	TAATGGGTACTGTGCTGGAGG
SALK_LB1.3	ATTTTGCCGATTTTCGGAAC
SAIL_LB3	TAGCATCTGAATTCATAACCAATCTCGATACAC
Flag_Tag5	CTACAAATTGCCTTTTCTTATCGAC
Gabi_o8474	ATAATAACGCTGCGGACATCTACATTTT
Cloning	
CLB19_STap_Tag	AAAAAGCAGGCTCCACCATGGGTCTCCTTCCCGTCGTCG
CLB19_ETap_Tag	AGAAAGCTGGGTCAGCATTGAGGAGATCACCAGC
DYW2_Start	GGAGATAGAACCATGTCTTCTCTAATGGCCATTC
DYW2_End	TCCACCTCCGGATCCCCAGTAATCCCCGCAAGAAC
OTP100_Start	GGAGATAGAACCATGTCTGATTTCTAAAGCCGCCT
OTP100_End_Stop	TCCACCTCCGGATCMGCAGGACGGTGGATCCTG
P35S_C_F	GCCCAGCTATCTGTCACTTTA
DsRed2_Rev	TGAACTCGGTGATGACGTC
Otp100_R_center	TTCTCTCCAACAGCTTCCTCA
qPCR	
DYW2_qPCR_F	GTGTGTGTGGGGATTGTCAC
DYW2_qPCR_R	AAACGCCGCTCTACCAGTAA
OTP100_qPCR_F	GAGGGCGTTAAGTGAAGCTG
OTP100_qPCR_R	TCCCTCTCTTCCTCCTTTCC

5.3. Activités complémentaires du travail de thèse

5.3.1. *Approches mises en place lors du travail de thèse pour contourner l'embryon létalité des mutants affectés dans DYW2 et OTP100 et déterminer leur fonction moléculaire.*

Au début de ce travail de thèse, l'analyse des mutants d'insertion *dyw2-1*, *dyw2-2*, *otp100-1* et *otp100-2* a montré une absence de plantes homozygotes, suggérant un comportement de type embryon létal de ces mutations. Le fait que les protéines PPR DYW2 et OTP100 soient essentielles à la survie des plantes a été le défi majeur de ce travail de thèse. Différentes approches sont généralement proposées pour contourner l'embryon létalité de mutations, ils sont décrits dans Candela *et al.* (2011). Dans ce contexte, dans un premier temps, Richard Berthomé et Andéol F. de Longevialle ont tenté une approche de sauvetage d'embryons en utilisant des milieux enrichis similaires à ceux utilisés pour le sauvetage d'embryons de *Phaseolus vulgaris* (Geerts *et al.*, 1999). Après plusieurs tentatives, cette stratégie n'a pas permis l'obtention de matériel homozygote. De nouvelles approches ont été envisagées dans le cadre de ma thèse. J'ai ainsi choisi de mettre en œuvre deux stratégies différentes, d'une part, une approche d'ARN d'interférence inductible et, d'autre part, la complémentation dans les embryons à l'aide de promoteurs embryon spécifiques. Les résultats de ces approches sont décrits ci-dessous.

5.3.1.1. **Stratégie d'interférence artificielle inductible à l'aide de microARN.**

Dans la littérature, il a été montré que l'utilisation de systèmes d'ARN d'interférence est une méthodologie de génétique inverse envisageable pour étudier la fonction d'un gène donné (Eamens *et al.*, 2011, Li *et al.*, 2013). En particulier, j'ai choisi d'utiliser la stratégie de microRNA artificiels (amiR) mise au point par le laboratoire de D. Weigel (Max Plank Institut, Allemagne) et qui utilise des modifications du microARN MIR319a (Ossowski *et al.*, 2008). Pour cela, l'outil WMD3 (Web MicroRNA Designer) disponible sur le site wmd3.weigelworld.org a été utilisé pour proposer des séquences d'amiR spécifiques des gènes *DYW2* et *OTP100*. Parmi ceux-ci, nous avons sélectionné les « meilleures » séquences sur la

RESULTATS

base du score de qualité des amiR proposés et des régions ciblées dans les deux gènes. En effet, pour optimiser la spécificité de l'approche, les séquences ciblant les domaines protéiques conservés (domaines PPR et DYW) ont été évitées pour la sélection de l'amiR. Peu d'amiR ayant un bon score de qualité ont été proposés pour le gène *DYW2*. En revanche, plusieurs amiR ayant de bons scores de qualité ont été obtenus pour le gène *OTP100* (Annexe Tableau 13). Finalement, pour chacun des gènes *DYW2* et *OTP100*, deux amiR ont été sélectionnés, la figure 20 montrant leur localisation sur la séquence codante de chacun des gènes. Les amorces nécessaires à l'amplification des amiR sélectionnés ont ensuite été dessinées à l'aide du logiciel disponible sur le site web WMD3. Ces amorces ont été utilisées pour générer les amiR par PCR à partir de la séquence de MIR319a clonée dans le vecteur pRS300 selon la méthodologie proposée par le laboratoire de D. Weigel et détaillée dans la section « méthodes ». Les fragments contenant les amiR ont ensuite été clonés en deux étapes dans le vecteur pMLBART_AlcR. Ce vecteur, généré dans le laboratoire de P. Laufs (« Facteurs de transcription et architecture », IJBP Versailles), permet une expression inductible à l'éthanol (Roslan *et al.*, 2001).

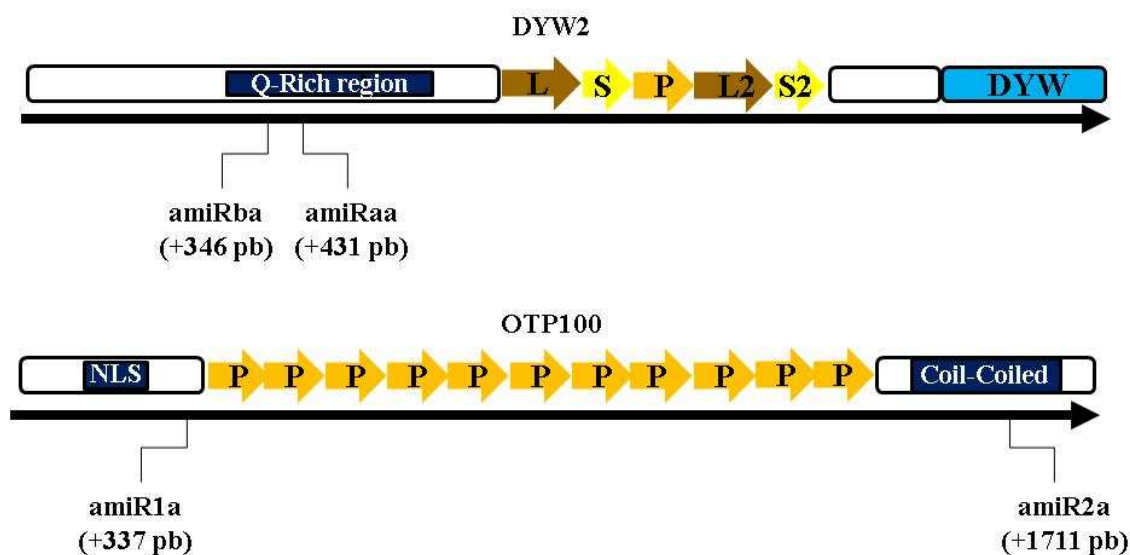


Figure 20. Positions des microARN d'interférence artificielle « amiR » sélectionnés dans les gènes *DYW2* et *OTP100*.

La position en pair de bases (pb) correspond au 5' des microARN, cette position étant définie à partir du codon d'initiation de la traduction ATG. Les domaines PPR (P, L, S), le domaine DYW et les régions prédites sont indiquées par des symboles au-dessus de la séquence codante de chaque protéine.

RESULTATS

Les vecteurs pMLBART_AlcR_amiR obtenus ont été utilisés pour la transformation stable de plantes de l'écotype Col-0 d'*A. thaliana*. Pour chacun des amiR, au moins 10 plantules T1 indépendantes ont été sélectionnées sur terreau à l'aide de l'herbicide BASTA. La descendance de chaque lignée (DYW2-amiRaa, DYW2-amiRba, OTP100-amiR1a, OTP100-amiR2a) a été utilisée pour des expériences d'induction en suivant le protocole détaillé dans la section « méthodes ». La descendance de transformants primaires a tout d'abord été utilisée pour chaque lignée amiR dans des expériences préliminaires en serre afin de déterminer s'il y avait un effet macroscopique de l'éthanol sur les plantes. Au cours de ces expériences préliminaires, et dans les conditions expérimentales testées, aucun effet macroscopique n'a été observé pendant les 4 premiers jours de traitement. En revanche, à partir du 5^{ème} jour, des effets du traitement ont été observés sur les plantes mais ils ont été associés à une toxicité de l'éthanol car toutes les plantes, y compris les plantes contrôle Col-0 de type sauvage, montraient un phénotype similaire (Figure 21).



Figure 21. Système d'induction des amiR sur terreau et phénotype macroscopique des plantes après 5 jours d'induction.

Les plantes sont cultivées sur terreau, en serre et en condition de jours longs (cycles 16h-8h lumière-obscurité).

Suite à ces résultats, j'ai décidé de quantifier l'expression des gènes *DYW2* et *OTP100* après trois jours d'induction (en chambre de culture) afin de minimiser l'impact des effets toxiques de l'éthanol.

Dans le cas du gène *DYW2*, deux types de constructions ont été générées (DYW2-amiRaa et DYW2-amiRba). Pour chacune d'elles, 8 lignées ont été caractérisées. La construction n'étant pas à l'état homozygote dans ces lignées, des feuilles de deux plantes différentes présentant une bonne résistance à l'herbicide ont été prélevées et poolées pour l'extraction d'ARN. Les résultats d'analyse d'expression par RT-qPCR n'ont pas montré

RESULTATS

d'effet majeur des constructions amiR sur l'expression de *DYW2* (Figure 22). De façon intéressante, et à l'inverse de ce qui était attendu, une faible augmentation de l'accumulation des transcrits a été observée dans toutes les plantes contenant la construction amiRba. Pour expliquer cette légère accumulation, on peut faire l'hypothèse que l'expression de cet amiR pourrait affecter la stabilité des ARNm en bloquant partiellement leur traduction (Fabian & Sonenberg 2012, Li *et al.*, 2013). Cependant, compte-tenu des résultats plus solides obtenus en parallèle avec les lignées d'insertion ADN-T homozygotes *dyw2-3* et les hétéroallèles *dyw2-3/dyw2-1* et *dyw2-3/dyw2-2*, j'ai décidé de ne pas continuer cette approche amiR dans le cas du gène *DYW2* et de ne pas tester cette hypothèse.

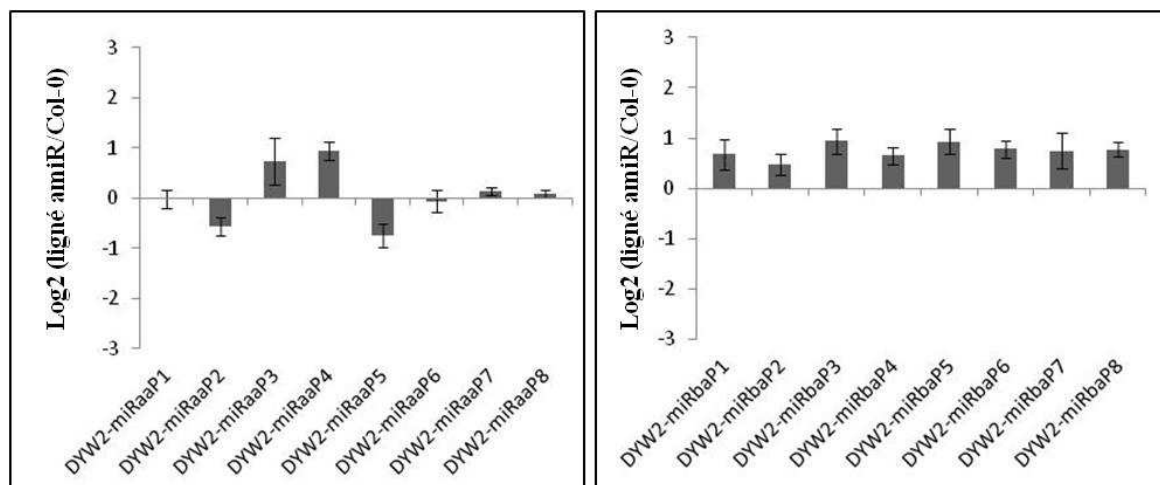


Figure 22. Expression du transcrit *DYW2* dans les lignées amiRaa et amiRba.

Pour chaque construction, l'expression du gène *DYW2* a été mesurée dans huit lignées indépendantes (P1 à P8) par RT-qPCR en utilisant le couple d'amorces *DYW2_qPCR_F* et *DYW2_qPCR_R*. L'expression du gène nucléaire codant l'ARN ribosomal 18S a été utilisée comme référence pour la normalisation.

Dans le cas des constructions amiR1a et amiR1b générées pour « éteindre » le gène *OTP100*, six lignées ont été testées dans des expériences d'induction. Les résultats d'expression par RT-qPCR ont montré que la plupart des lignées amiR n'étaient pas affectées dans l'expression du gène *OTP100* (Figure 23). Dans le cas de ces analyses, deux répétitions biologiques de l'expérience d'induction ont été réalisées afin de tester la reproductibilité du système. Les résultats ont montré une variabilité importante du niveau d'expression d'*OTP100* entre les deux expériences d'induction. Cette variabilité a rendu difficile l'interprétation des résultats en suggérant que des facteurs influençant l'expression n'étaient pas maîtrisés. Néanmoins, les lignées amiR1aP1, amiR1aP3 et amiR2aP2 ont montré dans les deux répétitions une légère diminution de l'expression d'*OTP100*. Elles sont donc les plus

RESULTATS

intéressantes pour continuer une analyse moléculaire. Cependant, leur faible nombre et le faible niveau d'extinction, rendent nécessaire l'identification de plus de lignées affectées afin de mener une analyse significative.

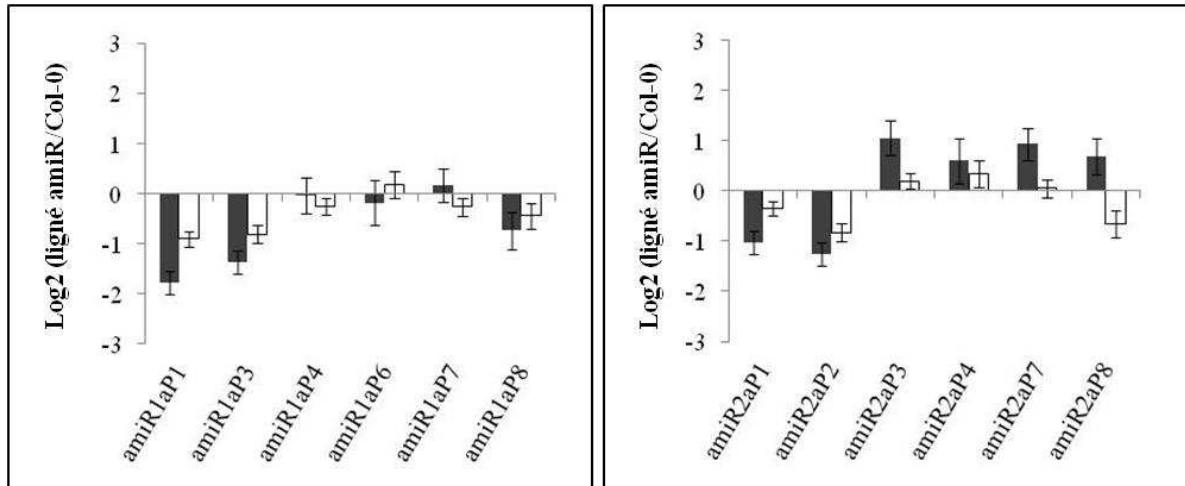


Figure 23. Expression du transcrit OTP100 dans les lignées amiR1a et amiR2a.

Les résultats de deux répétitions biologiques (histogrammes noirs et blancs) sont représentés. L'expression du gène *OTP100* a été mesurée dans six lignées indépendantes (P1 à P6) par RT-qPCR en utilisant le couple d'amorces *OTP100_qPCR_F* et *OTP100_qPCR_R*. L'expression du gène nucléaire codant l'ARN ribosomal18S a été utilisée comme référence pour la normalisation.

5.3.1.2. Stratégie de complémentation spécifique du stade graine

Une deuxième stratégie a été mise en place afin de contourner l'embryon létalité des mutants *dyw2* et *otp100* et de déterminer la fonction moléculaire des protéines DYW2 et OTP100. Notre idée était de compléter les mutants au cours du développement de l'embryon et de la graine en exprimant les protéines sous le contrôle de promoteurs embryon ou graine spécifiques. Le choix des promoteurs a été basé sur des données de la littérature. Ainsi, nous nous sommes intéressés aux 3 facteurs de transcription LEAFY COTYLEDONS (LEC1, LEC2 et LEC1-LIKE) qui sont exprimés de manière préférentielle dans la graine et particulièrement aux stades précoces du développement de l'embryon (Le *et al.*, 2010, Belmonte *et al.*, 2013, communication personnelle Bertrand Dubreucq, équipe « développement et qualité des graines », IJPB, Versailles). Cette expression spécifique de la graine est sous le contrôle de promoteurs régulés par des mécanismes épigénétiques réprimant l'expression de ces gènes en dehors du développement de la graine. Ceci a été en particulier

RESULTATS

observé dans le cas du promoteur du gène *LEC2* (Berger *et al.*, 2011). Ainsi, il a été observé qu'un fragment minimal de 500pb du promoteur de *LEC2* est suffisant pour obtenir une expression spécifique de la séquence du rapporteur GUS au cours du développement de la graine et particulièrement dans l'embryon. Pour les promoteurs de *LEC1* et *LEC1-LIKE* des résultats non publiés suggèrent un mécanisme similaire à celui de *LEC2* (communication personnelle Bertrand Dubreucq). Pour cela, trois vecteurs de destination de type Gateway ont été construits à partir du vecteur binaire pGWB13 et à l'aide d'enzymes de restrictions *HindIII* et *XbaI*. Le promoteur minimal de *LEC2* (-550 pb en amont du codon d'initiation de la traduction ATG utilisé comme référence 0) et les promoteurs de *LEC1* (-1012 pb) et *LEC1-LIKE* (-981 pb) ont été clonés en amont de la cassette de recombinaison Gateway. Les séquences codantes de *DYW2* et *OTP100* ont ensuite été clonées dans ces vecteurs et les constructions générées ont été utilisées pour la transformation de plantes hétérozygotes *dyw2-1/+*, *dyw2-2/+*, *otp100-1/+* et *otp100-2/+*. Ces travaux ont été menés dans le cadre du stage de Master 1 de Rachel Fourdin (co-encadrée par moi-même et Claire Lurin).

Après sélection, plusieurs transformants T1 hétérozygotes pour la mutation *dyw2-1*, *dyw2-2* et *otp100-2* et contenant l'insertion *PLEC2::DYW2* ou *PLEC2::OTP100* ont été obtenus, en revanche, le génotype des transformants T1 issus du mutant *otp100-1* reste à être confirmé (Tableau 4). La recherche de plantes homozygotes *dyw2-1*, *dyw2-2* et *otp100-2* par génotypage des plantes T2 est en cours au laboratoire. Ceci permettra d'évaluer si la stratégie peut être efficace pour l'obtention de matériel biologique permettant l'analyse de la fonction moléculaire de DYW2 et OTP100.

Tableau 4. Nombre de plantes et génotypes identifiés dans la première génération de transformants (T1) exprimant les protéines DYW2 ou OTP100 sous le contrôle du promoteur *PLEC2*.

Mutation	Construction	Nombre de plantes sélectionnées	Génotype pour la mutation*			
			S	HT	HO	à génotyper
<i>dyw2-1</i>	<i>PLEC2::DYW2</i>	>96	n.d.	10	n.d.	>86
<i>dyw2-2</i>	<i>PLEC2::DYW2</i>	>22	5	6	n.d.	>11
<i>otp100-1</i>	<i>PLEC2::OTP100</i>	6	n.d.	n.d.	n.d.	6
<i>otp100-2</i>	<i>PLEC2::OTP100</i>	59	2	11	n.d.	46

*Les allèles présents au locus endogène sont identifiés par PCR. La présence du transgène sous le contrôle de *PLEC2* est confirmée également par PCR. S: plante homozygote de type « sauvage », HT: plante hétérozygote, HO: plante homozygote, n.d. non déterminé.

Par ailleurs, des plantes hétérozygotes pour les mutations *dyw2-1*, *dyw2-2*, *otp100-1* et *otp100-2* ont été transformées avec les constructions *PLEC1::DYW2*, *PLEC1-LIKE::DYW2*, *PLEC1::OTP100*, *PLEC1-LIKE::OTP100* correspondantes. Bien que des plantes résistantes aient été obtenues lors de la sélection des T1, leur génotype pour les mutations *dyw2-1*, *dyw2-2*, *otp100-1* et *otp100-2* n'ont pas encore été caractérisés.

5.3.1.3. Stratégie d'hémi-complémentation des mutants *dyw2* et *otp100*

La recherche du phénotype moléculaire dans chacun des organites de mutants affectés dans des protéines PPR doublement localisées a été l'activité primaire de ce travail de thèse. Dans ce contexte, et en parallèle des approches décrites précédemment, j'ai développé une méthode, appelée « hémi-complémentation », spécifique aux protéines ayant une localisation subcellulaire complexe. L'hémi-complémentation est une complémentation fonctionnelle partielle qui a pour but d'essayer de comprendre la fonction d'une protéine dans une de ses localisations cellulaires de façon indépendante de son (ou ses) autre(s) localisation(s). Ainsi, dans cette approche, le transgène ne permet l'expression de la protéine que dans une de ses localisations et permet d'étudier la conséquence de la perte de fonction dans l'autre compartiment cellulaire. Par exemple, dans le cas d'une double localisation mitochondriale et chloroplastique, la protéine est adressée vers un seul organite à l'aide d'un peptide d'adressage spécifique de celui-ci, ce qui permet l'étude de la fonction dans l'autre organite. J'ai réalisé une série de clonages destinés à permettre l'hémi-complémentation fonctionnelle des mutants *dyw2* et *otp100* en sur-exprimant une fusion traductionnelle composée d'un peptide d'adressage mitochondrial ou chloroplastique et d'une protéine d'intérêt à laquelle on a supprimé son peptide d'adressage natif. Le peptide d'adressage de 56 acides aminés de la Formate déshydrogénase (FDH) a été utilisé comme peptide d'adressage pour la mitochondrie, et le peptide d'adressage de 51 acides aminés de la petite sous-unité nucléaire de la Rubisco (RBC1A) a été utilisé comme peptide d'adressage pour le chloroplaste. Pour cela, des plasmides binaires de type « Gateway » dérivés du plasmide pGWB15 ont été construits et nommés pGWB15_FDHTp (vecteur d'hémi-complémentation mitochondriale) et pGWB15_RBC1ATp (vecteur d'hémi-complémentation chloroplastique). Ces plasmides permettent de construire aisément des vecteurs d'hémi-complémentation pour toute protéine

RESULTATS

clonée dans un vecteur d'entrée Gateway. L'expression des protéines chimériques est placée sous le contrôle du promoteur P35S ; les protéines sont fusionnées au signal d'adressage mitochondrial ou chloroplastique (Tp) mais également à une étiquette HA. La construction obtenue après recombinaison est donc la suivante : P35S::Tp:HA:PROTEINE. Ci-dessous, je présenterai les résultats de cette approche. Ceux-ci restent malheureusement partiels et nécessitent d'être confirmés.

Les séquences codantes des protéines DYW2 et OTP100, tronquées dans leur région N-terminale correspondant à leurs signaux d'adressage (55 et 58 aa, respectivement), ont été clonées dans un vecteur d'entrée puis transférées dans chacun des vecteurs d'hémi-complémentation. Une fois construits, ces vecteurs ont été ensuite utilisés pour la transformation de plantes hétérozygotes pour les mutations embryon létales *dyw2-1*, *dyw2-2*, *otp100-1* et *otp100-2*. Les T1 ont été sélectionnés à l'aide d'antibiotiques et leur génotype confirmé par PCR (Tableau 5, Annexe Tableaux 17 et 18).

Tableau 5. Nombre de plantes et génotypes identifiés dans la première génération de T1 exprimant les protéines de fusion Tp:HA:DYW2 ou Tp:HA:OTP100 sous le contrôle du promoteur P35S.

Mutation	Insertion	Nombre de plantes sélectionnées	Génotype pour la mutation*		
			S	HT	HO
<i>dyw2-1</i>	P35S::FDHTp:HA:DYW2	17	9	8	
<i>dyw2-1</i>	P35S::RBC1ATp:HA:DYW2	26	17	9	
<i>dyw2-2</i>	P35S::FDHTp:HA:DYW2	1		1	
<i>dyw2-2</i>	P35S::RBC1ATp:HA:DYW2	2	1	1	
<i>otp100-1</i>	P35S::FDHTp:HA:OTP100	5	3	2	
<i>otp100-2</i>	P35S::FDHTp:HA:OTP100	9	1	5	3**
<i>otp100-2</i>	P35S::RBC1ATp:HA:OTP100	19		17	2**

*Les allèles présents au locus endogène sont identifiés par PCR. La présence du transgène d'hémi-complémentation est confirmée également par PCR. S: plante homozygote de type « sauvage », HT: plante hétérozygote, HO: plante homozygote **Les homozygotes nécessitent d'être confirmés.

Une fois obtenue, la descendance des T1 a été utilisée pour analyser la ségrégation de l'insertion d'hémi-complémentation grâce à la résistance aux antibiotiques portée par celle-ci. Ainsi, pour chacune des plantes T1 analysées (hétérozygotes pour la mutation *dyw2-1* et

RESULTATS

otp100-1), 100 graines de la génération T2 ont été semées sur milieu MS avec les antibiotiques correspondants, puis leur germination et la résistance des plantules a été évaluée 12 à 15 jours après semis. Des graines ont également été semées sur des milieux sans antibiotique afin de caractériser leur phénotype. L'étude de la ségrégation de l'ADN-T d'hémi-complémentation à la génération T2 a montré que ces lignées possèdent entre 1 et 2 insertions ADN-T (Tableau 6). Par ailleurs, j'ai observé un phénotype macroscopique chez certaines plantes (principalement du mutant *dyw2-1*) porteuses de l'ADN-T d'hémi-complémentation sur milieu MS avec ou sans ajout de l'antibiotique. En particulier, les feuilles de certaines plantes ont montré une coloration panachée avec des secteurs blancs et des secteurs verts (Figure 24 et Tableau 6). Une partie des plantes chlorotiques et toutes les plantes vertes ont survécu au repiquage en terre et sont restées viables et fertiles. Un semis contrôle du lot de graines utilisées pour la transformation a permis d'observer un phénotype similaire, indiquant que le phénotype observé chez les lignées transgéniques n'était probablement pas lié à l'introduction du transgène d'hémi-complémentation.

Tableau 6. Ségrégation de l'insertion ADN-T d'hémi-complémentation à la génération T2 issue de plantes T1 hétérozygotes pour les mutations *dyw2-1* et *otp100-1*.

Code ligné*	Insertion	Sensibles	Résistants		Non germées	N° Insertions estimées
			Vert pales	Verts		
P103	<i>P35S::RBC1ATp:HA:DYW2</i>	19	62	18	1	1
P104	<i>P35S::FDHTp:HA:DYW2</i>	13	70	15	2	1-2
P105	<i>P35S::RBC1ATp:HA:DYW2</i>	25	57	17	1	1
P107	<i>P35S::RBC1ATp:HA:DYW2</i>	28	61	10	1	1
P114	<i>P35S::FDHTp:HA:DYW2</i>	15	57	28	0	1-2
P118	<i>P35S::RBC1ATp:HA:DYW2</i>	20	71	6	3	1-2
P124	<i>P35S::RBC1ATp:HA:DYW2</i>	23	64	12	1	1
P125	<i>P35S::RBC1ATp:HA:DYW2</i>	30	46	20	4	1
P133	<i>P35S::RBC1ATp:HA:DYW2</i>	30	64	3	3	1
P138	<i>P35S::FDHTp:HA:DYW2</i>	26	61	13	0	1
P203	<i>P35S::FDHTp:HA:OTP100</i>	41	8	49	2	1
P205	<i>P35S::FDHTp:HA:OTP100</i>	45	17	36	2	1

Les résistances sont évaluées sur milieu Murashige-Skoog (MS) contenant de la kanamycine (*dyw2-1*) et de la kanamycine et de l'hygromycine (*otp100-1*). * Les plantes P100 à P199 correspondent à la mutation *dyw2-1* ou *dyw2-2*. Les plantes P200 à 299 correspondent à la mutation *otp100-1* ou *otp100-2*. Voir tableaux 17 et 18 de l'annexe.

RESULTATS

L'analyse préliminaire par génotypage n'a pas permis d'identifier des plantes homozygotes pour la mutation *dyw2-1* parmi les plantes de la génération T2 analysées. En revanche, pour une lignée portant l'ADN-T *P35S::FDHTp::HA::OTP100*, 5 plantes homozygotes pour la mutation *otp100-1* ont pu être identifiées parmi 96 plantes T2 issues de la même plante T1 (P203).

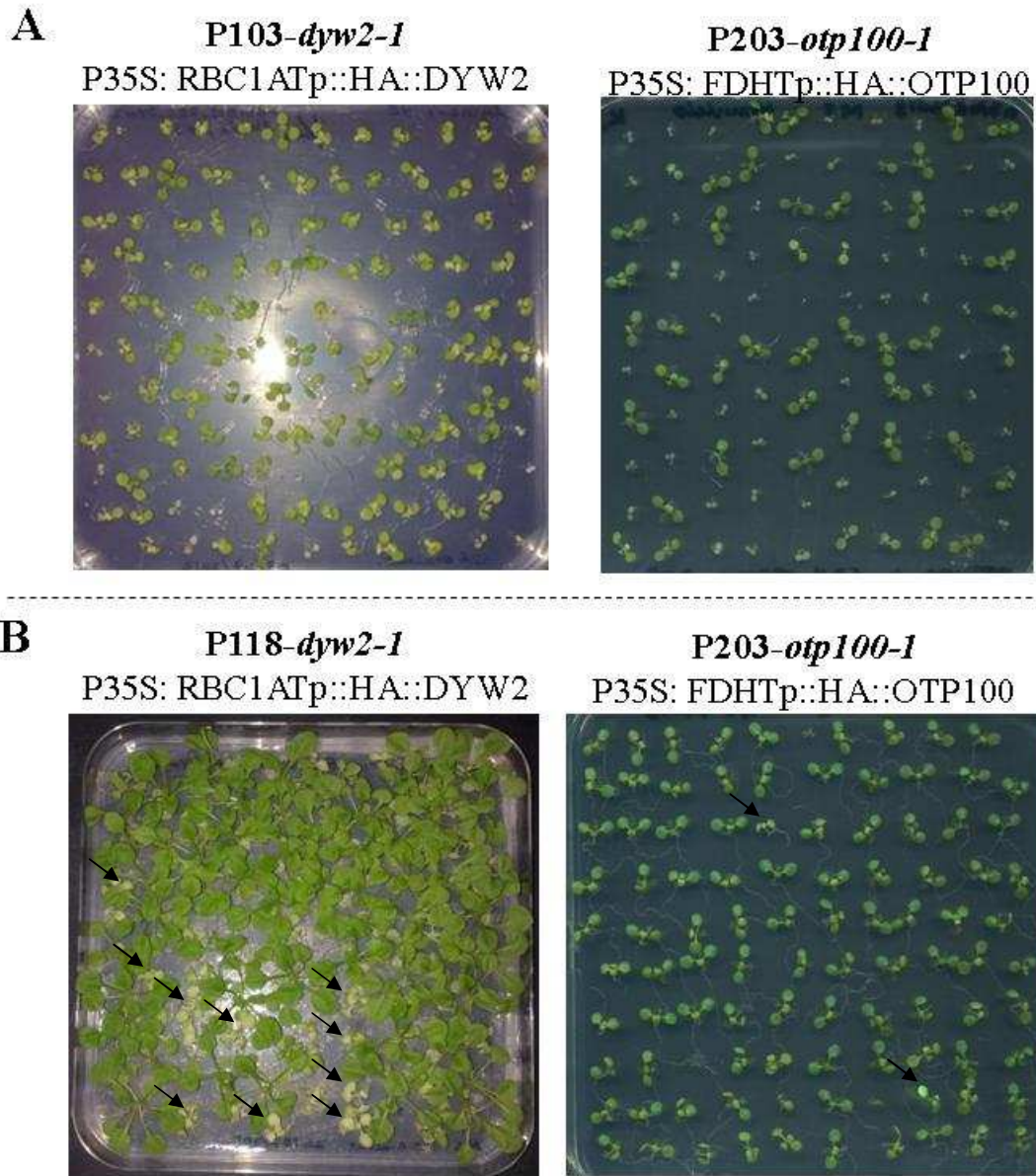


Figure 24. Phénotype macroscopique de descendants T2 de plantes hétérozygotes pour les mutations *dyw2-1* ou *otp100-1* et porteuses de l'ADN-T d'hémi-complémentation.

Les plantules sont observées *in-vitro* sur milieu Murashige-Skoog (MS) avec antibiotique (A) ou sans antibiotique (B). Les plantes montrant un phénotype chlorotique sont signalées par une flèche.

RESULTATS

Suite à ces résultats, j'ai cherché à vérifier l'expression des transgènes dans les plantes portant la mutation *dyw2-1*. Pour cela, j'ai testé l'expression par RT-PCR du transgène d'hémi-complémentation et ainsi que du gène endogène *DYW2* chez des plantes hétérozygotes issues des lignées P103 et P104. Les résultats montrent que le transgène et le gène endogène s'expriment dans ces lignées, indiquant que les deux populations d'ARN *DYW2* sont probablement présentes chez les plantes hémi-complémentés *dyw2-1* (Figure 25).

Dans le cas des mutants *dyw2-1*, aucune plante homozygote pour la mutation n'a pu être obtenue suite à l'utilisation de la stratégie d'hémi-complémentation. Ce résultat peut s'expliquer par un échec de notre stratégie (par exemple absence d'expression dans le compartiment ciblé, non fonctionnalité des protéines de fusion) ou par le rôle crucial de la protéine *DYW2* dans chacun des organites, son absence aboutissant à la létalité quelque soit l'organite complémenté. Avant de caractériser la présence et l'adressage des protéines chimériques dans les organites, nous avons testé la fonctionnalité du système d'hémi-complémentation par croisement des lignées hémi-complémentées dans le chloroplaste et dans la mitochondrie. En effet, si dans chacune des deux lignées, les protéines sont exprimées, bien adressés et fonctionnelles dans le compartiment ciblé, il doit être possible d'obtenir, par croisement des deux lignées hémi-complémentées, des plantes homozygotes pour la mutation *dyw2-1* complémentées dans chacun des deux organites grâce aux deux protéines chimériques. Un tel croisement a été effectué et sa descendance a été génotypée. Les résultats montrent que des plantes homozygotes pour la mutation *dyw2-1* ont été obtenues (Tableau 7). Ces plantes sont viables et fertiles sur terreau. Ces résultats suggèrent que les deux fonctions mitochondriale et chloroplastique de *DYW2* sont nécessaires à la viabilité de la plante. Cependant ces résultats restent préliminaires et nécessitent d'être confirmés par d'autres analyses, en particulier, l'étude de la perte des transgènes d'hémi-complémentation dans ces lignées (en cours), seront nécessaires pour confirmer ces résultats.

RESULTATS

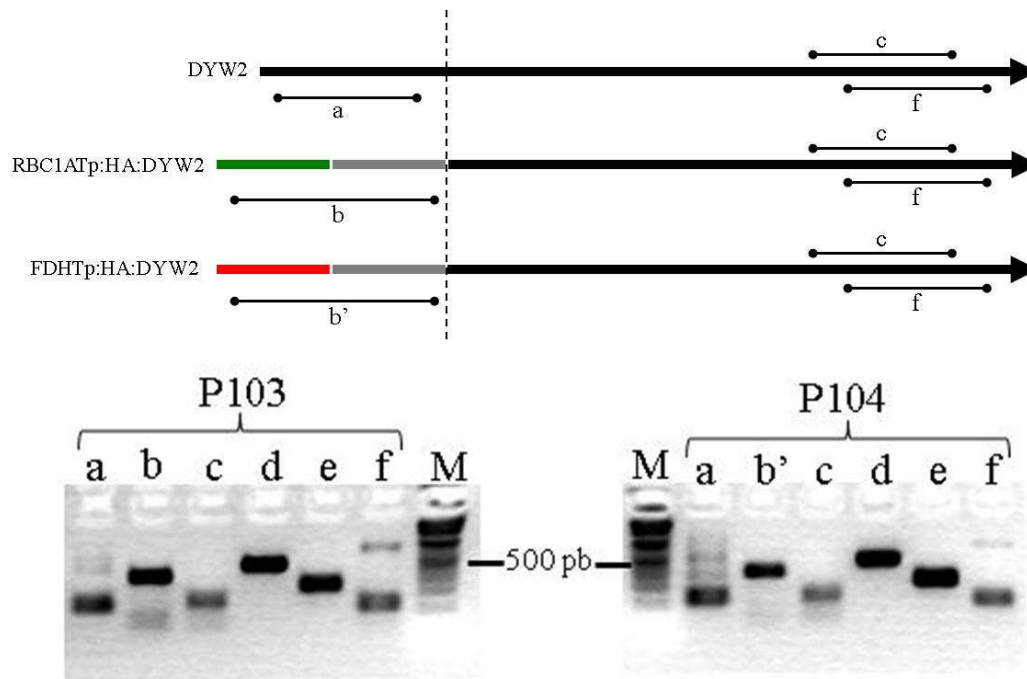


Figure 25. Analyse de l'expression du transgène et du gène endogène *DYW2* dans les plantes hémi-complémentées P103 et P104.

L'expression est évaluée par RT-PCR sur feuilles. Les positions des fragments amplifiés pour analyser l'expression de *DYW2* sont représentées dans le schéma au-dessus du gel de migration. **a** : fragment spécifique de l'expression du gène endogène *DYW2*. **b** : fragment spécifique de l'expression du transgène RBC1ATp:HA:*DYW2*. **b'** : fragment spécifique de l'expression du transgène FDHTp:HA:*DYW2*. Produits **c** et **f** : fragments correspondants à l'expression du transgène et du gène endogène. Les puits ou pistes **d** et **e** correspondent à l'amplification d'ADNc de contrôles chloroplastiques *rpoA* et *clpP*, respectivement. **M** : marqueur de taille 1Kb plus (Invitrogen).

Tableau 7. Génotypes identifiés lors du croisement de lignées hémi-complémentées *dyw2-1* mitochondriales et chloroplastiques.

Croisements		Génotype pour la mutation*		
Femelle	Male	S	HT	HO
P124 ^a	P114 ^b		14	5
P125 ^a	P114 ^b	1	6	7
P116 ^b	P125 ^a	1	3	5
P138 ^b	P118 ^a		23	7
P125 ^a	P138 ^b	2	31	8

a : plante portant la construction *P35S::RBC1ATp:HA:DYW2*. b : plante portant la construction *P35S::FDHTp:HA:DYW2*. *Les allèles présents au locus *DYW2* endogène sont identifiés par PCR. La présence du transgène d'hémi-complémentation est confirmée également par PCR. S: plante homozygote de type « sauvage », HT: plante hétérozygote, HO: plante homozygote.

5.3.1.4. Etude de l'interaction entre les protéines DYW2 et OTP100 par complémentation de fluorescence bimoléculaire ou BiFC.

L'approche de purification par TAP-tag d'un complexe contenant CLB19 a permis d'identifier les protéines DYW2 et OTP100. Cependant, l'information apportée par cette approche ne permet pas de conclure quant aux interactions protéine-protéine directes entre ces trois facteurs. Dans ce contexte, nous avons entrepris une étude des interactions entre protéines du complexe CLB19 *in vivo* grâce à la stratégie BiFC. Pour cela, nous avons utilisé une méthode d'expression transitoire des protéines dans les cotylédons de plantules d'*A. thaliana* (Marion *et al.*, 2008). Cette stratégie a été choisie car elle est couramment utilisée au laboratoire et qu'elle nous a permis de montrer l'interaction entre les deux protéines PPR CRR4 et DYW1 dans les chloroplastes (Boussardon *et al.*, 2012). Ainsi, les séquences codantes de DYW2 et OTP100 ont été transférées dans les vecteurs de destination BiFC disponibles dans l'équipe, permettant la fusion de la protéine d'intérêt avec les moitiés N-terminale et C-terminale de la protéine eYFP (fragment YFP en C-terminal de la protéine d'intérêt). J'ai ainsi cherché à montrer une interaction entre CLB19 et DYW2, d'une part, et CLB19 et OTP100, d'autre part. Malgré de nombreux essais, je n'ai jamais observé de fluorescence pour les deux couples testés. Dans les mêmes expériences, les signaux montrant l'interaction entre CRR4 et DYW1 ont pu être reproduits avec succès. L'absence de signal d'interaction entre deux protéines dans un système expérimental, ne permet pas de conclure à une absence d'interaction entre les protéines. De nombreux facteurs peuvent en effet expliquer ce résultat négatif (stabilité de l'interaction, encombrement stérique de l'YFP, stades de développement,...). Pour tenter de conclure sur l'interaction directe entre les protéines, il serait nécessaire de tester d'autres méthodologies ou conditions expérimentales que celles testées dans ce travail de thèse. Par exemple, une analyse de l'interaction de ces protéines dans des conditions *in vitro* par « Pull down » ou par double hybride chez la levure pourrait permettre une confirmation de l'interaction. D'autre part des analyses par double hybride avec des formes tronquées pourrait permettre d'identifier les domaines protéiques importants pour l'interaction.

5.3.2. Caractérisation fonctionnelle de la protéine OTP91.

Les données du projet systématique de localisation ont montré que la protéine OTP91, codée par le gène *AT1G06150*, a une localisation dans les deux organites lorsque sa séquence pleine longueur est fusionnée à la DsRed2 (Voir Figure 1 de Colcombet *et al.*, 2013). Par ailleurs, cette protéine appartient au sous-groupe PLS-E, principalement associé à l'édition. Ces caractéristiques ont motivé sa sélection pour l'étude de sa fonction moléculaire.

Néanmoins, avant de décrire mes travaux concernant la protéine PPR OTP91, je voudrais préciser quelques éléments bibliographiques concernant cette protéine. L'interruption du gène *OTP91* a été montrée comme probable cause d'un phénotype embryon-létal (mutant nommé de ce fait *embryon-defective* 1444 ; protéine EMB1444) (Tzafrir *et al.*, 2004). Cependant, ce résultat a été contredit par un deuxième travail, lequel a montré que l'insertion analysée n'était pas la cause de l'embryon-letalité (Cushing *et al.*, 2005). Ainsi, les résultats du test d'allélisme effectué par Cushing montrent que dans le croisement entre les mutants *salk_093892* et *salk_143087* il existe une complémentation de l'embryon-létalité qui suggère que ce caractère ne soit pas lié à la mutation dans OTP91. Il a été montré par la suite que l'insertion décrite par Tzafrir *et al.*, (2004) était localisée dans la région promotrice du gène et non dans la séquence codante comme proposé initialement. D'autre part, la structure du gène OTP91 portant les répétitions de motifs PPR n'est pas encore très claire à ce jour. En effet, la protéine OTP91 est codée par le locus *AT1G06145* dans la version 9 de l'annotation TAIR. Dans la version 10, l'exon codant pour les motifs PPR a été fusionné au gène *AT1G06150*. Cette « nouvelle » protéine correspond à une fusion des domaines PPR et E, avec une séquence contenant un domaine hélice-boucle-hélice de liaison à l'ADN. Cette modification de l'extrémité N-terminale induit un changement dans les prédictions de localisation subcellulaire de la protéine qui deviendrait, selon cette nouvelle annotation, une protéine nucléaire. Cette fusion de gènes n'est supportée par aucune évidence transcriptionnelle claire, est très atypique pour un gène PPR et est de plus contredite par le test d'allélisme effectué par Cushing *et al.*, (2005) (cf ci-dessus). Pour cet ensemble de raisons, dans cette thèse, le locus OTP91 a été assimilé au locus AT1G06145 (TAIR v9) ne présentant pas de fusion protéique.

Concernant les analyses fonctionnelles de la protéine OTP91, j'ai caractérisé dans un premier temps la lignée SALK_093892 (appelée *otp91-1*) et j'ai pu confirmer les données de

RESULTATS

Cushing *et al.* (2005) montrant que l'insertion de l'ADN-T dans la région promotrice n'entraîne pas de phénotype embryon-létal. Je n'ai cependant pas pu analyser le deuxième allèle décrit par Cushing *et al.* (2005) car la lignée correspondante (SALK_143087) a été retirée des « Stock Centers ». Pour contourner ce problème, j'ai analysé la lignée SAIL_549_E07 (appelée *otp91-2*) portant une insertion dans l'ORF, 1388 nucléotides après le codon d'initiation de la traduction d'AT1G06145 (Figure 26A). J'ai ensuite entrepris la recherche d'individus homozygotes pour l'insertion. 131 graines ont germé sur du milieu MS et les plantules ont été génotypées. Parmi celles-ci, aucun individu homozygote n'a pu être identifié (90 plantules étaient hétérozygotes pour l'insertion, 41 portaient deux allèles de type « sauvage »). Des observations sur le remplissage des graines dans des siliques des plantes hétérozygotes n'ont pas permis d'identifier de graines anormales (taille ou coloration des graines). Pour confirmer le résultat de génotypage, un second a été effectué. Les plantules étudiées étaient issues de graines provenant de quatre siliques prélevées sur deux plantes hétérozygotes (2 siliques par plante). Les graines ont été semées sur du milieu MS contenant 1% de saccharose. Dans ces conditions, j'ai pu identifier des individus homozygotes pour l'insertion dans la descendance d'une seule des deux plantes (Tableau 8). Ces résultats suggèrent que l'obtention d'individus homozygotes pourrait dépendre de facteurs que nous ne contrôlons pas bien comme peut-être des stress environnementaux affectant le processus de fécondation, la maturation des graines ou des problèmes de séchage des siliques au moment du prélèvement.

Tableau 8. Génotypage de plantules issues de deux siliques de deux plantes hétérozygotes pour l'insertion *otp91-2* après croissance sur milieu MS contenant 1% de saccharose.

Plante	Génotype			
	Sauvage	Hétérozygote	Homozygote	Non germées
1	20	52	0	26
2	24	46	7	14

Ces plantes homozygotes ont permis de caractériser un peu plus le phénotype induit par l'insertion. En particulier, j'ai pu montrer que l'insertion de l'ADN-T dans les individus *otp91-2* homozygotes conduisait à une très forte baisse voire une absence d'expression du gène *OTP91* (expression indétectable par RT-PCR, résultat non présenté). Par ailleurs, les analyses préliminaires du phénotype macroscopique ont montré que cette perte d'expression

du gène s'accompagne d'un retard de germination et affecte très fortement le développement et/ou la croissance des plantes (Figure 26B et 26C).

En collaboration avec le Pr. Ian Small et le Pr. Mizuki Takenaka, nous avons tenté de prédire les cibles ARN de la protéine OTP91 en utilisant le code de reconnaissance des PPR publié récemment (Barkan *et al* 2012, Takenaka *et al.* 2013). La protéine OTP91 portant un domaine E, typique de protéines impliquées dans l'édition, nous nous sommes penchés en particulier sur la prédiction de sites d'édition potentiels. Cette analyse a suggéré que plusieurs sites chloroplastiques et mitochondriaux pouvaient être ciblés par cette PPR (Annexe Tableau 19). En parallèle, j'ai caractérisé expérimentalement le niveau d'édition chez le mutant *otp91-2* de 31 des 34 sites chloroplastiques majeurs décrits chez *A. thaliana* (Figure 27). Les résultats n'ont montré aucun défaut majeur chez le mutant aux sites étudiés. Cependant, de façon très étonnante, j'ai observé chez le mutant *otp91-2* un niveau d'édition du site *ropC1-C509* (21806) significativement plus élevé que chez des plantes de type sauvage. Ce site n'étant pas prédit pour les analyses bioinformatiques. Ce résultat est similaire à celui observé pour le mutant *otp70* qui est affecté dans l'épissage de l'intron de *ropC1* et chez lequel l'édition au site *ropC1-C509* (21806) est augmentée (Chateigner-Boutin *et al.*, 2011). Afin de « nettoyer » le mutant *otp91-2* d'autres insertions d'ADN-T éventuelles dans le génome et de s'assurer ainsi du lien entre l'insertion dans le gène *OTP91* et le phénotype observé, j'ai fait deux rétrocroisements. L'obtention d'homozygotes issus de ces rétrocroisements devra permettre de confirmer les résultats préliminaires et d'approfondir la compréhension du phénotype moléculaire du mutant *otp91-2*.

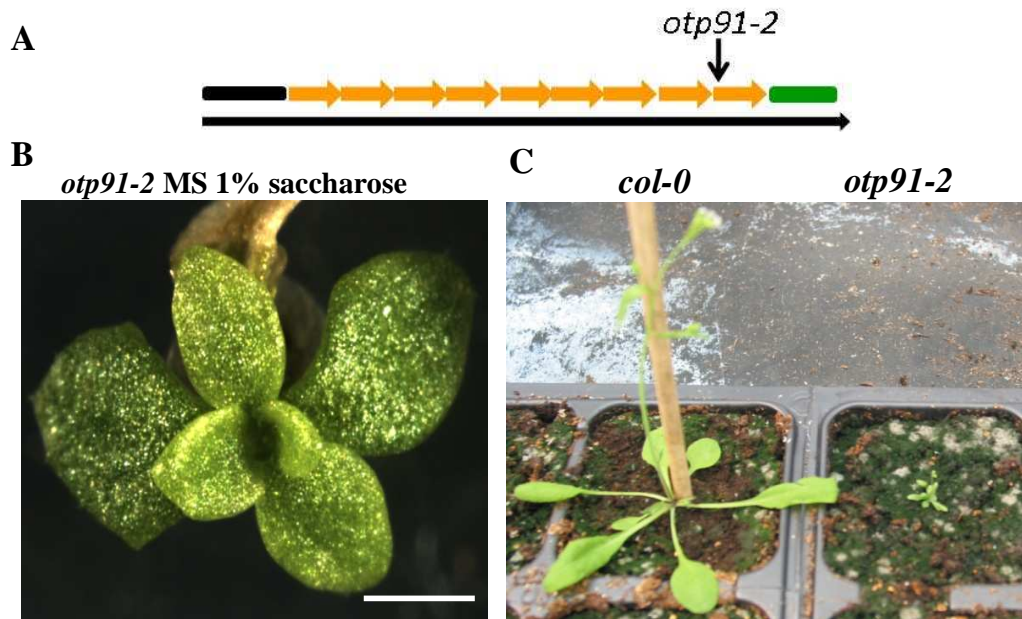


Figure 26. Structure de la protéine OTP91 et phénotype du mutant *otp91-2*.

A : Structure de la protéine OTP91 avec représentation de la séquence de la protéine (en noir), des domaines PPR (flèches jaunes), domaine E (en vert), et la position de l'insertion ADN-T dans le mutant *otp91-2*. B : Plantule mutante *otp91-2* homozygote observée entre les stades développementaux 3.0 et 3.5 (germination sur milieu MS contenant 1% de saccharose, placée sous une photopériode 16:8 lumière:obscurité). B : Plante mutante *otp91-2* observée au stade développemental 5 - 6 (après transfert du milieu MS/1% saccharose sur du terreau et sous une photopériode 16:8 lumière:obscurité).

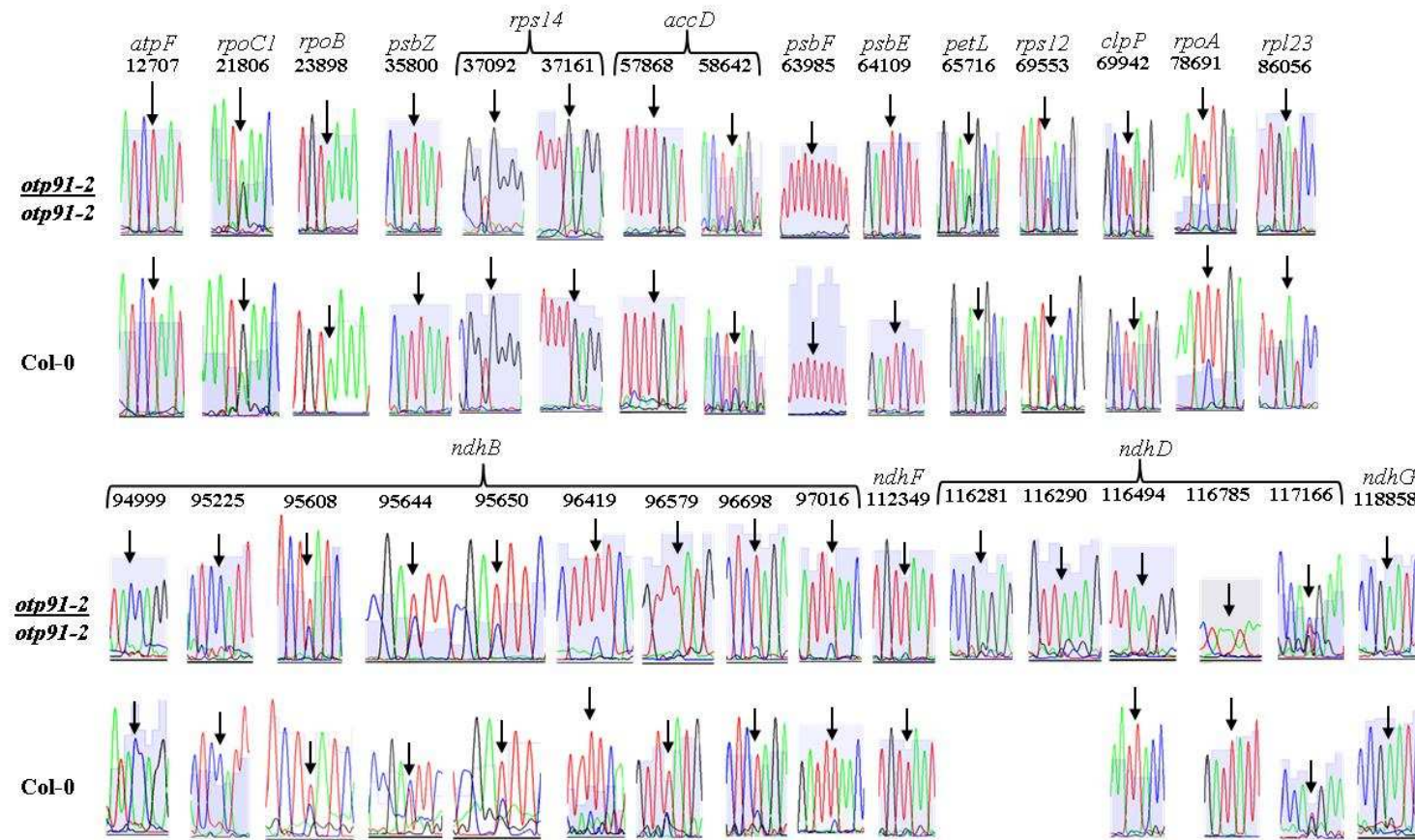


Figure 27. Éditeur chloroplastique du mutant homozygote *otp91-2*.

L'édition a été évaluée pour 31 des 34 sites majeurs d'édition chloroplastique par séquençage de produits de PCR à partir d'ADNc. Chaque site d'édition est indiqué par une flèche au dessus du chromatogramme de séquençage. Pour chaque site, la position dans le génome chloroplastique d'*A. thaliana* est indiquée. Les nucléotides sont représentés à l'aide du code couleur standard : rouge pour T, noir pour G, vert pour A et bleu pour C. Les séquences correspondant au transcrit *ndhD* ont été obtenues chez le mutant avec des amorces situées sur le brin inverse de celui utilisé pour séquençer les sites chez les plantes sauvages.

DISCUSSION

Chapitre VI

Bilan critique des résultats et éléments additionnels de discussion

Limites de l'approche d'intégration de données pour la double localisation des protéines de la famille PPR.

Les limites des approches pour contourner l'embryon létalité des gènes DYW2 et OTP100.

Quels facteurs régulent la localisation des protéines PPR ?

La double localisation des PPR est un phénomène ancien mais dynamique entre les espèces.

L'expression des génomes des organites est-elle coordonnée par des PPR doublement localisées ?

6. Bilan critique des résultats et éléments additionnels de discussion

En addition des discussions propres à chaque article, je présenterai ci-dessous un bilan critique des approches et des difficultés rencontrées lors de cette thèse. Par ailleurs, je présenterai des aspects supplémentaires de la discussion non traité dans les projets d'articles.

6.1. Limites de l'approche d'intégration de données pour la double localisation des protéines de la famille PPR.

Dans l'analyse de la localisation de la famille PPR, nous avons utilisé deux approximations pour l'obtention et la collecte des données de localisation. Tout d'abord, pour les données de localisation générées au sein de l'équipe, il a été considéré que l'utilisation d'un système hétérologue d'expression transitoire chez *N. benthamiana* pouvait rendre compte de la localisation des protéines PPR d'*A. thaliana*. D'autre part, pour les données de localisation subcellulaire recueillies à partir de données protéomiques, le concept d'ortho-protéomique a été utilisé (Huang *et al.*, 2013). Ce concept permet de faire l'hypothèse que la localisation subcellulaire de protéines orthologues serait transposable d'une espèce à une autre. Cependant, dans le cas de protéines ayant une double localisation, Fuss *et al.*, (2013) ont suggéré récemment que l'utilisation de systèmes hétérologues pourrait avoir un impact sur les résultats de localisation. Un tel résultat soulève également des questions liées au concept ortho-protéomique car il suggère que les systèmes de translocation peuvent varier entre espèces et modifier l'éventuelle double localisation d'une protéine dans une espèce donnée. Bien que l'approche d'intégration de données permette de minimiser les limitations que peut avoir une seule stratégie en la complétant par des données issues d'autres stratégies, il n'est donc pas à exclure qu'un certain nombre d'erreurs figurent dans notre analyse, en particulier en ce qui concerne la double localisation (présence de faux positifs et faux négatifs).

6.2. Les limites des approches pour contourner l'embryon létalité des gènes DYW2 et OTP100.

Les stratégies mises en œuvre dans ce travail de thèse pour contourner « l'essentialité » des protéines DYW2 et OTP100 n'ont pas été couronnées de succès. Tout d'abord, l'utilisation des amiR inductibles par l'éthanol s'est avérée une stratégie lourde et peu efficace pour l'obtention de matériel permettant l'analyse de la fonction moléculaire des protéines OTP100 et DYW2. Ceci peut s'expliquer par le faible impact que peut avoir une réduction de l'expression de gènes *DYW2* et *OTP100* sur le phénotype macroscopique des plantes, rendant difficile l'utilisation de cette stratégie. De plus, les effets toxiques de l'éthanol lorsqu'il est appliqué pendant plus de 5 jours limitent le temps d'induction des plantes avant analyse. Pour contourner ce désavantage du système d'induction et en raison du faible impact des amiR sur l'expression de nos gènes cibles, il aurait été probablement préférable de tester directement cette approche à l'aide de promoteurs constitutifs. Les résultats préliminaires concernant les amiR dirigés contre le gène OTP100 semblent néanmoins indiquer que cette stratégie pourrait être poursuivie dans ce cas.

La complémentation à l'aide de promoteurs spécifiques du stade embryonnaire n'a été mise en place que tardivement pendant la dernière année de thèse ce qui n'a pas permis de mener jusqu'au bout cette approche et de conclure quant à son efficacité. Cependant, les lignées produites pourraient être utiles pour une analyse de la fonction des protéines. Ces analyses sont actuellement en cours dans l'équipe.

Les résultats obtenus avec les lignées hémi-complémentées restent partiels et nécessitent d'être confirmés. Néanmoins, l'obtention de plantes homozygotes pour la mutation *dyw2-1* complémentées grâce aux deux insertions d'hémi-complémentation, alors qu'aucune survie de plante homozygote n'a pu être obtenue avec l'une ou l'autre des constructions, permet de proposer que les fonctions de DYW2 dans les deux organites sont nécessaires à la viabilité d'*A. thaliana*. Le mutant *clb19* est très affecté dans son développement mais viable sur des milieux riches. Ce résultat suggère donc que le mutant *dyw2* présenterait un phénotype chloroplastique additionnel correspondant à une fonction moléculaire de DYW2 encore inconnue. D'autre part, l'absence de survie de mutants *dyw2*

complémentés du côté mitochondrial, pourrait être expliquée par la perte d'édition de nombreux sites.

Par ailleurs, l'existence d'homozygotes dans les deux lignées hémi-complémentées d'*otp100* nécessite une vérification (en cours). Dans ce cas, les deux constructions permettant l'expression d'OTP100 soit dans les mitochondries soit dans les plastides conduiraient semble-t-il à une survie des plantes homozygotes. Ce résultat suggère que la suppression des fonctions de la protéine dans les deux organites est la cause de la létalité. Une autre explication pourrait être que le système d'hémi-complémentation pour OTP100 génère des « fuites de localisation » qui permettent la complémentation des mutants. Afin de confirmer ce résultat, il faudra vérifier l'homozygotie des lignées et analyser par des approches biochimiques la présence et l'absence des protéines chimériques dans les organelles des différentes lignées. Ces analyses pourront être effectuées par Western Blot en utilisant des fractions subcellulaires des mutants et en utilisant des anticorps dirigés contre le tag HA porté par les protéines dans nos constructions d'hémi-complémentation.

Malgré les nombreuses approches entreprises durant cette thèse, la fonction moléculaire de la protéine OTP100 reste inconnue jusqu'ici, aucun matériel viable permettant une approche de génétique inverse n'ayant pu être obtenu. Néanmoins, si les résultats préliminaires de la génération d'homozygotes hémi-complémentés sont confirmés, ils permettront de caractériser les fonctions moléculaires d'OTP100 dans les mitochondries et les chloroplastes. Une autre stratégie, non initiée à ce jour au laboratoire, pourrait être d'entreprendre une recherche de mutants EMS par « Tilling » afin d'identifier des mutations ponctuelles dans le gène *OTP100*. Cette approche pourrait permettre d'identifier des allèles moins forts que les insertions « perte de fonction » étudiées au cours de ma thèse.

6.3. Quels facteurs régulent la localisation des protéines PPR ?

En introduction de cette thèse, j'ai détaillé les mécanismes permettant la translocation des protéines dans les organites des cellules végétales. Malgré le grand nombre de données apportées par l'analyse systématique de la localisation des protéines PPR menée dans Colcombet *et al.*, (2013), ces données ne permettent pas de proposer

quels sont les mécanismes permettant l'import des protéines PPR dans les organites. Cependant, il semble que les deux principaux mécanismes de double localisation des protéines, impliquant un seul ou bien deux produits de traduction, soient utilisés pour l'import des protéines PPR. En effet, il a été montré que de nombreux événements d'épissage alternatif ont lieu dans des transcrits PPR chez *A. thaliana* (de Longevialle A. 2010). D'ailleurs les données de Ichinose *et al.* (2012) et de Fuss *et al.* (2014) montrent aussi que chez *P. patens* des transcrits alternatifs sont aussi possibles au moins pour la protéine PpPPR_43 et, par ailleurs, mes analyses suggèrent que chez la PPR AT3G42630 des transcrits alternatifs produisent des isoformes modifiées dans leur partie N-terminale ce qui peut probablement avoir une incidence sur la localisation subcellulaire de la protéine. Concernant le mécanisme de double adressage utilisant un seul produit de traduction, il n'existe pas à l'heure actuelle de preuve claire qu'un seul produit de traduction des PPR puisse générer une double localisation. Ceci est notamment dû au manque d'expérience d'import *in vitro* pour les protéines PPR. Cependant, si on considère que certains gènes codant pour des protéines PPR, telles que les protéines PRORP1, DYW2 et OTP100, ne produisent pas de transcrit alternatifs, et que les prédictions de double localisation du logiciel *Ambiguous Targeting Prediction* (ATP) donnent des scores importants pour certains membres de la famille PPR, il est fort probable qu'un mécanisme utilisant un seul produit de traduction existe et soit fonctionnel. Dans quelle mesure ces deux mécanismes contribuent à la localisation des protéines PPR au sein de la famille reste une question à résoudre.

Lors des expériences de localisation, nous avons observé pour plusieurs protéines PPR une localisation non homogène dans tous les organites, notamment dans le cas des mitochondries d'une même cellule. Par exemple, j'ai pu observer que certaines PPR produisaient un signal de localisation dans un nombre réduit de mitochondries. Une hypothèse pour expliquer ces résultats pourrait impliquer la diversité des complexes de translocation qui existent chez les organites de plantes. Ainsi, l'existence des différentes familles multigéniques codant pour les protéines des complexes TOM et TIM pourrait expliquer ces profils de localisation. Par ailleurs, l'hétérogénéité de la composition des génomes mitochondriaux pourrait également être un facteur influençant la localisation de certaines protéines PPR. En effet, il est connu qu'au sein de la population des mitochondries d'une cellule, toutes les mitochondries ne possèdent pas un génome entier, à tel point que certaines d'entre elles n'en ont pas du tout. Dans ce cas, et compte tenu des

fonctions des PPR, il ne serait donc pas surprenant qu'elles n'y soient pas adressées. Bien que les raisons de ces localisations hétérogènes dans les cellules soient encore inconnues, ces observations suggèrent qu'il existe un mécanisme de surveillance actif permettant une régulation fine de la localisation de ces protéines. Dans ce contexte, il serait intéressant de déterminer, dans des expériences futures, s'il existe une corrélation entre la présence (et l'abondance) de la cible ARN d'une protéine PPR et sa localisation subcellulaire.

Bien que cela soit encore inconnu dans le cas des protéines PPR, la biogénèse des organites au sein de la cellule pourrait également impacter la localisation des protéines. Il est important de considérer que les organites sont des compartiments dynamiques se transformant, leur nombre, forme et distribution évoluant au cours de la vie de la cellule. Ces changements sont plus évidents pour la mitochondrie (Welche *et al.*, 2014). Bien que nous ne disposions que de peu d'information concernant les changements qui s'opèrent dans la localisation des protéines lors des fissions et fusions des mitochondries, ces mécanismes peuvent jouer un rôle sur la distribution des protéines. Par ailleurs, si on considère le gradient de différenciation des cellules d'*Arabidopsis* depuis le méristème vers des organes différenciés tels que les feuilles, il n'est pas exclus que la biogénèse des organites, comme par exemple lors du passage du proplaste au chloroplaste, ait aussi un impact sur la localisation des protéines PPR, particulièrement pour les protéines doublement adressées.

6.4. La double localisation des PPR est un phénomène ancien mais dynamique entre les espèces.

Le nombre total de protéines doublement localisées dans la mitochondrie et le chloroplaste reste encore incertain (cf premier paragraphe de la discussion). Les données expérimentales recueillies dans cette thèse me permettent de suggérer que 123 protéines pourraient avoir une double localisation (Tableau 3 « considérant groupe 2 à 4 », Annexe Tableau 9). Au sein de ce groupe, 36 protéines (29%) sont des protéines PPR ce qui les place, avec les protéines de la famille des ARNt synthétases, comme une des familles majeures parmi les protéines ayant une double localisation mitochondriale et chloroplastique.

Par ailleurs, en intégrant les données expérimentales de localisation produites par des approches protéomiques et de fluorescence, issues des bases de données PPDB et SUBA3, il est possible d'estimer que 140 autres protéines d'*A. thaliana* pourraient avoir une double localisation (Voir Annexe I Tableau 10 et 11). Ainsi, au total, 273 protéines d'*Arabidopsis* pourraient appartenir à cette catégorie de localisation. Ce nombre est très similaire à celui de 277 protéines avancé par Carrie & Small (2013). De plus, le nombre de protéines de cette catégorie de localisation pourrait même être sous estimé car la localisation de très nombreuses protéines reste non caractérisée expérimentalement. Par exemple, plus de 40% des protéines de la famille des PPR n'ont jamais été localisées expérimentalement. Une étude récente estime que les protéines ayant une double localisation dans la mitochondrie et le plaste pourraient être plusieurs centaines (Baudisch *et al.*, 2014).

L'importance des protéines PPR dans cette catégorie de double localisation n'est pas surprenante si l'on considère le grand nombre de protéines PPR codées par le génome nucléaire des plantes terrestres et le fait qu'elles ont une fonction dans des processus du métabolisme des ARN très similaires chez les deux organites. Comme pour l'ensemble des protéines de cette catégorie de localisation, l'origine de la double localisation des protéines PPR n'est pas claire. Cependant, il est possible de proposer la même hypothèse que celle avancée pour la plupart des protéines ayant une double localisation : la double localisation des PPR a pu être mise en place pour diminuer le coût énergétique généré par les deux systèmes d'expression des génomes des organites et permettre en même temps au génome nucléaire de coordonner l'activité photosynthétique et la respiration oxydative au niveau post-transcriptionnel. Cette dernière hypothèse est supportée par le fait que les plantes terrestres ont principalement sélectionné un contrôle post-transcriptionnel (et non transcriptionnel) de l'expression des gènes des organites, ceci favorisant probablement l'émergence de la sous-famille de protéines PPR ayant une double localisation. Cependant, l'importance de la régulation post-transcriptionnelle n'explique pas à elle seule l'apparition de cette catégorie de localisation chez les protéines PPR. En effet, parmi les 105 protéines PPR décrites chez *P. patens*, très peu semblent avoir une double localisation. Ainsi, 29 localisations expérimentales ont été rapportées pour ces protéines et, parmi elles, seulement trois protéines ont été décrites avec une double localisation (Sugita *et al.*, 2013, Fuss *et al.*, 2013). D'autre part, parmi les 46 protéines PPR d'*A. thaliana* ayant une double localisation potentielle, seule la protéine PRORP1 a deux orthologues potentiels chez *P.*

patens. Cependant, même si Sugita *et al.*, (2013) proposent que ces deux orthologues, PpPPR_63 et PpPPR_104 sont adressés aux deux organites, aucune donnée expérimentale a été présentée pour le démontrer. Cette faible conservation ne permet pas, pour l'instant, d'évaluer la conservation de la double localisation entre ces espèces. Ces données suggèrent cependant que la double localisation serait un processus dynamique dont la fréquence aurait augmenté après la divergence des plantes vasculaires.

P. patens et *A. thaliana* possèdent uniquement 15 protéines PPR pour lesquelles il est possible de déterminer un lien d'orthologie. Par ailleurs, dans le cas de 10 protéines PPR d'*A. thaliana*, il est possible d'associer un couple de paralogues chez *P. patens*. Au total, il existe donc 35 protéines PPR similaires chez *P. patens* et *A. thaliana*. Il est intéressant de noter que ces 35 protéines font toutes partie de la sous-famille P, qu'il existe un enrichissement dans la localisation plastidiale et que la conservation de localisation subcellulaire semble importante entre ces deux espèces. Ainsi, parmi les 35 protéines, 7 ont une localisation chloroplastique prouvée chez les deux espèces et 15 protéines ayant une localisation plastidiale chez *A. thaliana* sont prédites dans le plaste chez *P. patens*. A l'inverse, la plupart des PPR mitochondriales de *P. patens* n'ont pas d'orthologue chez *A. thaliana*. Cette observation peut s'expliquer par le fait que le génome mitochondrial de *P. patens* est assez différent en structure et en taille de celui des angiospermes alors que les génomes chloroplastiques sont beaucoup plus similaires (Sugita *et al.*, 2003, Terasawa *et al.*, 2007). Ainsi, si l'on compare la famille PPR de *P. patens* à une famille PPR d'angiospermes, la proportion de protéines ayant une localisation mitochondriale passe de 45% à 60%, celle de protéines ayant une double localisation passe de 3 à 9%, alors que la localisation plastidiale diminue de 45% à 25%. Cette évolution de la répartition de la localisation subcellulaire des protéines PPR peut être reliée à l'augmentation du nombre d'événements post-transcriptionnels nécessaires à l'expression des génomes des organites des plantes vasculaires. En effet, l'expression du génome mitochondrial d'*A. thaliana*, et des angiospermes en général, nécessite un plus grand nombre d'événements post-transcriptionnels que celle du plaste, particulièrement dans l'édition (Takenaka *et al.*, 2013). Ceci pourrait indiquer que les protéines PPR ont contribué aux processus permettant une meilleure coordination de l'expression génomes mitochondriaux chez les plantes vasculaires.

En dehors d'*A. thaliana*, le nombre de protéines PPR dont la localisation subcellulaire a été caractérisée chez les angiospermes reste assez faible. Un seul cas de double localisation a été rapporté jusqu'ici, il s'agit de la protéine de Maïs ZmPPR2263 orthologue de la protéine MEF29 d'*A. thaliana* (Sosso *et al.*, 2012). Chez les deux espèces, les protéines ont été observées dans les deux organites et une même fonction moléculaire a également été décrite pour les deux protéines dans la mitochondrie. Ce cas unique ne permet bien évidemment pas de conclure à un éventuel comportement évolutif global de la famille. L'accumulation de données de localisation de membres de la famille PPR dans d'autres espèces, chez des angiospermes mais aussi chez des espèces appartenant à d'autres clades comme les lycophytes, les fougères ou les gymnospermes, sera nécessaire afin de comprendre la dynamique de la double localisation au sein de la famille.

6.5. L'expression des génomes des organites est-elle coordonnée par des PPR doublement localisées ?

Comme détaillé dans l'introduction de cette thèse, les organites sont des structures complexes nécessitant, pour l'expression de leur génome et leur biogénèse, l'activité coordonnée et régulée de facteurs nucléaires et de facteurs codés par les organites. En effet, par des analyses transcriptionnelles, il a été montré que l'expression des génomes des organites et du génome nucléaire sont interdépendants et complexes (Leister *et al.*, 2011, Schwarzländer *et al.*, 2012). D'une part, le noyau influence et contrôle les organites par l'activité de facteurs nucléaires adressés aux organites, ces processus sont connus sous le nom de « mécanismes de régulation antérograde ». Néanmoins, les organites ne sont pas des compartiments neutres sous le « joug dictatorial » du noyau et, en réponse à des stimuli, les organites influencent en retour l'activité des gènes nucléaires par des mécanismes décrits sous le nom de « signalisation rétrograde » (Woodson & Chory 2008, Leister 2012, Chi *et al.*, 2013). Cette interconnexion et le dynamisme de la biogénèse des organites rendent difficile l'interprétation du rôle que peut avoir chaque processus du métabolisme de l'ARN sur la coordination de l'expression des gènes. Cependant, compte-tenu de la régulation antérograde et de la localisation de certains facteurs nucléaires, deux types de mécanismes de coordination par le noyau de l'expression des gènes entre les deux organites sont imaginables. Soit le noyau produit et envoie deux facteurs, l'un dans la

mitochondrie et l'autre dans le plaste, de manière simultanée, soit un seul facteur est produit puis adressé aux deux organites. Il n'est pas possible pour l'instant d'évaluer la contribution de chacun de ces deux mécanismes sur la coordination de l'expression de gènes des organites. Néanmoins, leur co-existence semble probable compte-tenu du nombre croissant de protéines liées au métabolisme de l'ARN possédant une double localisation et du fait que les complexes protéiques des nucléoides, structures à laquelle est associé le métabolisme de l'ARN, semblent être composées d'une multitude de protéines présentant des localisations simples et doubles.

Chez les organites des plantes terrestres, la régulation de l'expression des gènes se fait principalement de façon post-transcriptionnelle (Binder & Brenicke 2003, Stern *et al.*, 2010, Small *et al.*, 2013, Hammani & Giegé 2014). Au sein de ces processus, les protéines PPR ont un rôle prépondérant. Dans ce contexte, les protéines PPR doublement localisées sont des bons candidats pour participer à la coordination de l'expression des gènes entre les deux organites. Nous avons essayé de tester cette hypothèse par l'analyse de protéines PPR impliquées dans l'édition. Malgré l'évidence, issue de cette thèse, de la fonction moléculaire de DYW2 dans les deux organites, les données actuelles ne nous permettent pas de valider l'hypothèse proposée au début du projet.

Il est important de considérer que le fait que le noyau envoie une machinerie protéique similaire vers les deux organites pour agir sur l'expression de leurs gènes ne prouve pas que l'expression des gènes mitochondriaux et plastidiaux soient coordonnées. Le lien de coordination doit être établi par une corrélation entre les cibles, c'est-à-dire sur la liaison entre les processus où agisse la protéine dans les deux organites. Dans ce contexte, il serait attendu que les protéines PPR contribuant à une coordination de l'expression entre les organites soient simultanément présentes dans les deux organites. Néanmoins, la complexité et l'interdépendance des différents processus post-transcriptionnels dans les organites ne permettent pas de bien distinguer la contribution de chaque processus. De plus, l'implication de processus comme l'édition C en U dans la régulation de l'expression des gènes n'est pas claire. En effet, deux hypothèses ne s'excluant pas mutuellement existent pour expliquer le rôle de l'édition C en U. La première hypothèse propose que l'édition serait un processus correctif pour contrecarrer les mutations T en C qui se produisent dans le génome des organites des plantes terrestres (Gray M. 2012). Cependant cette action correctrice de l'édition n'est pas très « pertinente » pour des changements

synonymes dans les séquences codantes. La seconde hypothèse, propose que l'édition contribue à l'existence des introns et à la régulation de l'expression des gènes des organites. Celle-ci est avancée par le fait que des événements d'édition dans des séquences non codantes, particulièrement dans les introns, ont une influence sur l'épissage (Borner *et al.*, 1995, Kagami *et al.*, 2012, Farre *et al.* 2013). Un des obstacles majeurs pour évaluer la coordination de l'expression de gènes des organites est le fait que les organites au sein de la cellule semblent proliférer de manière indépendante. En effet, cette différence importante rend difficile la caractérisation des relations de coordination entre organites.

Des analyses approfondies sur les PPR doublement localisées seront nécessaires pour mieux définir l'influence de ces facteurs nucléaires sur le contrôle et la coordination de l'expression de gènes entre les organites. Par exemple, il serait intéressant d'analyser de manière conjointe les transcriptomes nucléaires et des organites de plusieurs mutants de PPR doublement localisées en suivant une approche similaire à celle développée par Leister *et al.* (2011). Alternativement, des analyses permettant de travailler sur un matériel cellulaire végétal dont le développement des organites est synchronisé, comme par exemple les cellules BY-2 de tabac, adapté à des cultures cellulaires de mutants dans les protéines PPR pourrait éventuellement contribuer à la meilleure caractérisation du rôle de ces protéines doublement localisées.

Finalement, les choses sont évidemment beaucoup plus complexes que ça. En effet, en plus d'une coordination post-transcriptionnelle probable de l'expression des génomes, le contrôle du fonctionnement des complexes respiratoires, cordonné par des facteurs nucléaires et mitochondriaux, se fait principalement au niveau post-traductionnel (Giegé *et al.*, 2005). Ainsi, il n'est pas exclus que la coordination fonctionnelle des organites implique ces différents mécanismes.

7. METHODES ET MATERIELS

7.1. Matériels Biologiques

7.1.1. Espèces végétales

L'espèce modèle *A. thaliana* a été utilisée au cours de ce travail de thèse. Les écotypes Columbia (Col-0), Landsberg erecta (Ler) et Wassilewskija (Ws) et les lignées d'insertion ADN-T provenant des « Stock Centers » ont été utilisées dans le cadre des analyses fonctionnelles.

Des expériences de localisation par l'expression transitoire de protéines de fusions traductionnelles avec la protéine fluorescente DsRed2 ont été menées sur l'espèce *Nicotiana benthamiana*.

7.1.2. Mutants d'insertion chez *A. thaliana*

Les analyses fonctionnelles de cette thèse sont principalement basées sur une stratégie de génétique inverse utilisant des lignées d'insertion d'ADN-T. Pour cela les lignées suivantes ont été commandées à l'ABRC (*Arabidopsis Biological Resource Center*) au NASC (*Nottingham Arabidopsis Stock Center*) ou au CRB (Centre de Ressources Biologique, INRA de Versailles)

AGI	Nom de la lignée	Lignée	Code NASC
AT1G06145	<i>otp91-1</i>	SALK_093892	N593892
AT1G06145	<i>otp91-2</i>	SAIL_549_E07	N823289
AT1G25360	<i>otp90</i>	SM_3_29668	N116370
AT2G15690	<i>dyw2-1</i>	GABI_332A07	N431783
AT2G15690	<i>dyw2-2</i>	FLAG_435F11	
AT2G15690	<i>dyw2-3</i>	SALK_002014.56	N502014
AT3G49240	<i>otp100-1</i>	SALK_069042	N569042
AT3G49240	<i>otp100-2</i>	SAIL_784_A11	N835064

7.1.3. Souches bactériennes

La souche d'*Escherichia coli* DH5 α (génotype *fhuA2* Δ (*argF-lacZ*)*U169 phoA glnV44* Φ 80 Δ (*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) a été utilisée dans le cadre des expériences de biologie moléculaire pour les clonages et la multiplication d'ADN plasmidique.

La souche bactérienne d'*Escherichia coli* DB3.1 (génotype *F- gyrA462 endA1 glnV44* Δ (*srI-recA*) *mcrB mrr hsdS20*(*r_B⁻, m_B⁻*) *ara14 galK2 lacY1 proA2 rpsL20*(*Sm^r*) *xyl5 Aleu mtl1*), modifiée génétiquement, a été utilisée pour la génération et la multiplication d'ADN plasmidique contenant les cassettes Gateway dans lesquelles se trouve le gène *ccdB*. Cette souche est insensible à l'effet toxique du gène *ccdB* (Controlling Cell Death B).

La souche C58C1 d'*Agrobacterium tumefaciens* (génotype *pTi⁻, Ery^R, Chl^R*) a été utilisée pour la transformation génétique de plantes. Plus particulièrement, la souche contenant le plasmide pSOUP a été utilisée pour la transformation transitoire de plants de *N. benthamiana*. Pour la transformation stable d'*A. thaliana*, la souche C58C1 contenant le plasmide pMP90 a été utilisée.

7.2. Méthodes de biologie moléculaire

7.2.1. Clonage à l'aide d'enzymes de restriction

Avant de procéder au clonage utilisant des enzymes de restriction, il a d'abord fallu analyser et choisir le(s) site(s) de restriction afin d'éviter les sites se trouvant à l'intérieur des fragments d'ADN de l'insert à cloner et de garantir un clonage directionnel dans le site multiple de clonage (MCS) ou dans la région d'intérêt du plasmide. Par exemple, dans cette thèse, la modification de vecteurs pGWB a été nécessaire pour obtenir les vecteurs d'hémi-complémentation ou d'expression embryon spécifique.

Dans une stratégie basée sur le clonage de produits de PCR, des oligonucléotides ont été dessinés pour permettre la synthèse du fragment d'ADN d'intérêt portant des sites

de restriction aux extrémités 5' et 3' (quelques nucléotides sont ajoutés pour permettre une coupure efficace). La quantité de nucléotides dépend de l'enzyme de restriction. Les inserts sont synthétisés par PCR puis purifiés à l'aide du kit de purification « Qiaquick ADN purification kit » (QIAGEN).

7.2.1.1. Digestion de l'ADN par enzyme de restriction

Avant de procéder à la digestion, une quantification de la concentration d'ADN est réalisée à l'aide du spectrophotomètre NanoDrop ND-1000. L'ADN plasmidique et le produit de PCR purifié sont digérés suivant un protocole standard de digestion par des enzymes de restriction et les indications du fournisseur (New England Biolabs ou Fermentas).

Dans le cas d'une simple digestion (une seule enzyme) les conditions et réactifs pour un volume réactionnel de 30µL sont :

Réactifs	Quantité (volume total 30µL)
ADN (plasmide ou produit PCR)	XµL pour 2-5µg
Tampon spécifique à l'enzyme 10X	3µL
BSA 10X (si nécessaire)	1µL
Enzyme de restriction (10u/µL)	1µL
H ₂ O	Qsp 30µL

La réaction est effectuée à 37°C pendant le temps suggéré par le fournisseur et sans dépasser 16h. Ensuite l'enzyme est inactivée selon l'indication du fournisseur.

Dans le cas d'une double digestion (deux enzymes de restriction) les conditions et réactifs pour un volume réactionnel de 30µL sont

METHODES ET MATERIELS

Réactifs	Quantité (volume total 30µL)
ADN (plasmide ou produit PCR)	XµL pour 2-5µg de produit
Tampon 10X choisi pour optimiser la réaction	3µL
BSA 10X (si nécessaire)	1µL
Enzyme de restriction 1 (10u/µL)	1µL
Enzyme de restriction 2 (10u/µL)	1µL
Eau	Qsp 30µL

La réaction est effectuée à 37°C pendant le temps suggéré par le fournisseur, cette durée étant déterminée par l'enzyme ayant le plus court temps d'incubation. Ensuite les enzymes sont inactivées selon les indications du (des) fournisseur(s). Dans le cas d'une grande différence de temps d'incubation, une quantité d'enzyme a été ajoutée pour garantir une digestion totale des fragments tout en veillant à ce que le glycérol apporté avec les enzymes ne dépasse pas une proportion de plus de 1/10 du volume final. En effet, celui-ci peut inhiber la réaction.

7.2.1.2. Ligation de l'ADN

Avant de procéder à la ligation, les produits de digestion sont purifiés à l'aide du kit de purification « Qiaquick ADN purification kit » (QIAGEN). La ligation permet le clonage de l'insert dans le plasmide d'intérêt. Pour cela, la T4 DNA ligase (Promega) est utilisée dans les conditions préconisées par le fournisseur.

Réactifs	Quantités (volume total 10µL)
Vecteur digéré	100ng
Insert digéré	X ng*
Tampon ligase 10X	1µL
T4 DNA ligase	1µL
Eau	Qsp 10µL

* : X est la quantité en nano grammes d'insert pour obtenir un ratio de concentrations molaires vecteur : insert déterminé. En général, le ratio 3 :1 est recommandé mais ce ratio peut être optimisé pour chaque réaction. Il peut varier de 1 :1 à 10 :1. La quantité en nano grammes est calculée comme suit : $\left(\frac{\text{ng vecteur} \times \text{taille en kb d'insert}}{\text{taille en kb du vecteur}} \right) * \left(\frac{\text{mol de vecteur}}{\text{mol d'insert}} \right)$

La réaction est incubée à 4°C pendant une nuit ou durant 3h à température ambiante. Le produit de ligation est ensuite utilisé pour la transformation puis la sélection des bactéries (cf section 7.3.1.2.). Les colonies sont testées par PCR (cf section 7.2.8.) et l'ADN plasmidique issu des colonies positives est extrait et séquencé pour valider le clonage (cf section 7.2.11.).

7.2.2. Clonage GatewayTM (InvitrogenTM)

Principe

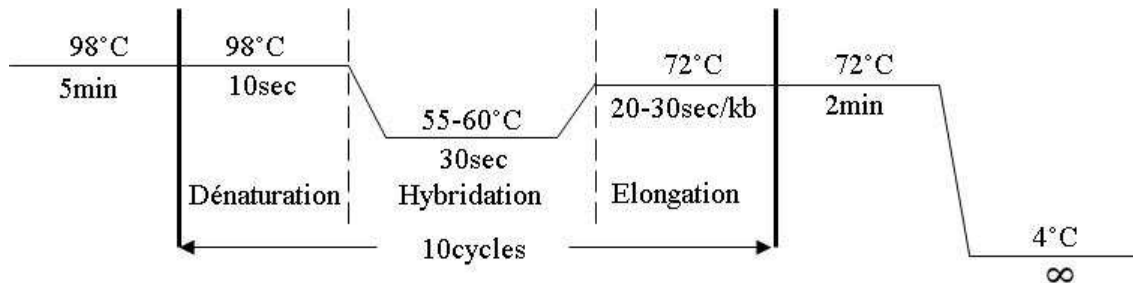
Il s'agit d'une technique de clonage par recombinaison permettant l'insertion d'un fragment d'ADN dans un vecteur. Pour ces clonages deux types de vecteurs sont utilisés, les vecteurs donneurs et les vecteurs de destination, pour générer respectivement les vecteurs d'entrée et les vecteurs d'expression. Le vecteur donneur porte les séquences *attP1* et *attP2*, lesquels par recombinaison avec les sites *attB1* et *attB2* (ajoutés aux extrémités du fragment à cloner, amplifié par PCR) vont permettre l'intégration du produit PCR dans le vecteur donneur. Cette réaction, appelée « réaction de BP » génère un nouveau vecteur (vecteur d'entrée) portant des sites *attL1* et *attL2*. Par la suite, un vecteur d'expression est construit par une seconde réaction de recombinaison (appelée « réaction LR ») entre le vecteur d'entrée possédant le fragment d'intérêt flanqué des sites *attL1* et *attL2* et un vecteur de destination portant les sites *attR1* et *attR2*. Tous ces vecteurs possèdent une origine de réplication pour leur multiplication dans *E. coli* et un gène conférant la résistance à un antibiotique permettant la sélection des bactéries transformées. Les vecteurs donneurs et de destination possèdent en plus la cassette suicide contenant le gène *ccdB*. Ils sont donc multipliés dans de bactéries DB3.1.

7.2.2.1. Amplification par PCR du fragment à cloner

Pour permettre le clonage d'un fragment d'ADN d'intérêt par recombinaison, la stratégie repose sur une amplification par PCR du fragment qui permet d'ajouter à ses extrémités les sites *attB1* et *attB2* nécessaires au clonage. Ainsi, deux PCR successives sont effectuées. La première (PCR1) permet d'amplifier le fragment à l'aide d'amorces spécifiques du fragment (amorces auxquelles une séquence étiquette de 12 bases est ajoutée). La deuxième réaction PCR (PCR2) a pour matrice le produit de PCR1 et permet d'apporter les séquences *attB1* et *attB2* grâce à des amorces universelles s'hybridant en 3' sur les étiquettes de 12 bases.

Réactifs PCR1	Quantités (volume total 10 μ L)
Amorce sens (0,5 μ M)	1 μ L
Amorce antisens (0,5 μ M)	1 μ L
dNTPs (10mM chacun)	0,2 μ L
Tampon polymérase iProof 5X	2 μ L
iProof polymérase (2u/ μ L)	0,1 μ L
Matrice (ADN g ou ADNc)	1 μ L
Eau	4,7 μ L

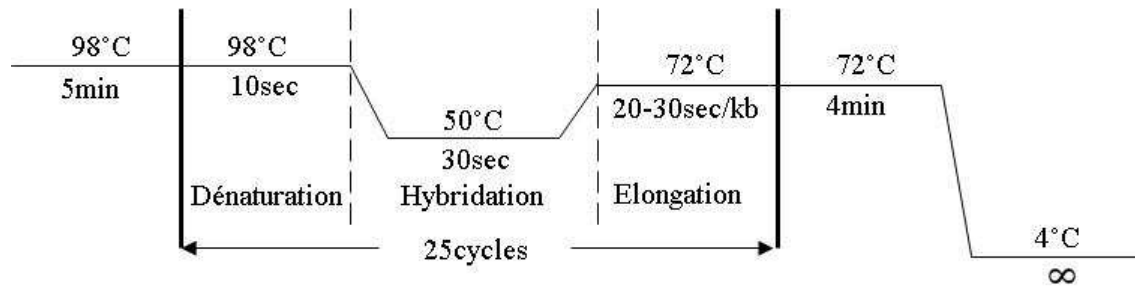
Conditions d'amplification PCR1:



Pour la réaction de PCR2 deux types d'amorces universelles sont utilisées selon la nature du fragment cloné (séquences codantes ou fragments génomiques). Les étiquettes portées par les amorces spécifiques sont modifiées en conséquence.

Réactifs PCR2	Quantité (volume total 50 μ L)
Amorce universelle sens (10 μ M)	1 μ L
Amorce universelle antisens (10 μ M)	1 μ L
dNTP (10mM chacun)	1 μ L
Tampon iProof 5X	8 μ L
iProof polymerase (2u/ μ L)	0,4 μ L
PCR1	10 μ L
Eau	28,6 μ L

Conditions d'amplification PCR2:



Un aliquot (5 μ L) de la réaction PCR2 est utilisé pour la vérification, par électrophorèse sur gel d'agarose, de la taille du fragment généré.

7.2.2.2. Purification du produit PCR

Deux protocoles ont été utilisés pour la purification des produits PCR avant clonage. Un premier protocole permet d'éliminer les oligonucléotides et les dimères d'amorces par simple précipitation au PEG. Pour cela, aux 45 μ L de produit PCR2 sont ajoutés 225 μ L de solution PEG/Mg/TE (2 volumes de PEG 30% 8000, 30mM MgCl₂ et 3 volumes de TE) et 1 μ L de glycogène azur (favorise la précipitation et colore en bleu les culots). Le mélange est ensuite soumis à une centrifugation à vitesse maximale (20000 g) pendant 15 min et à température ambiante. Le culot est resuspendu dans 20 μ L de TE, puis 2 à 5 μ L sont utilisés pour évaluer la taille et présence de l'ADN par électrophorèse sur gel d'agarose.

Un deuxième protocole de purification a été utilisé dans le cas où la spécificité de l'amplification n'était pas très bonne et des fragments « parasites » devaient être éliminés. Pour cela, le kit de purification « Qiaquick ADN purification kit » (QIAGEN) a été utilisé. Cette stratégie permet une purification sur gel du fragment PCR ayant la taille attendue.

7.2.2.3. Recombinaison BP (Invitrogen™)

Cette réaction permet le clonage du fragment flanqué par les sites *attB1* et *attB2* par recombinaison avec un vecteur donneur, dans notre cas le vecteur pDONR207, contenant les sites *attP1* et *attP2*. La réaction de recombinaison permet la formation d'un vecteur d'entrée similaire au vecteur donneur mais contenant, à la place de la cassette Gateway, le fragment d'intérêt flanqué des sites *attL1* et *attL2*.

Réactifs BP	Quantité (volume total 5µL)
Produit de PCR-attB	1-2µL
Vecteur d'entrée pDONR207 (150ng/µL)	1µL
Tampon BP clonase 5X	1µL
BP clonase	1µL
Tampon TE ou Eau	Qsp 5µL

La réaction est réalisée à 25°C durant 1h au minimum. Les 5µl de réaction BP sont utilisés pour la transformation puis la sélection des bactéries. Les colonies sont testées par PCR et l'ADN plasmidique des colonies positives est extrait et séquençé pour vérifier la séquence.

7.2.2.4. Recombinaison LR (Invitrogen™)

Cette réaction de recombinaison permet le transfert des inserts clonés dans le vecteur d'entrée dans un vecteur d'expression par recombinaison des sites *attL* du vecteur d'entrée et les sites *attR* du vecteur de destination. Les plasmides de destination utilisés dans cette thèse sont ceux de la série pGWB (Nakagawa *et al.*, 2007), en particulier les vecteurs pGWB1, pGWB2, pGWB13, pGWB15, et le vecteur p0229-DsRed2.

Réactifs LR	Quantité (volume total 5µL)
Vecteur de destination (100ng/µL)	1µL
Vecteur d'entrée (300ng/µL)	0,5-2µL
Tampon LR clonase 5X	1µL
LR clonase	1µL
Tampon TE ou Eau	Qsp 5µL

La réaction est incubée à 25°C durant 1h au minimum. 2 à 5µL de produit de réaction LR sont utilisés pour la transformation puis la sélection de bactéries. Les colonies sont testées par PCR et l'ADN plasmidique des colonies positives est extrait et séquençé si nécessaire.

7.2.3. Clonage pGEMT-easy

7.2.3.1. Ligation

Les produits PCR purifiés peuvent être directement clonés dans le vecteur pGEMT-Easy® (Promega) en suivant les instructions du fournisseur. Ce vecteur linéarisé possède à ses deux extrémités 3'-terminales une thymidine. Ces thymidines permettent le clonage de produits PCR générés par les polymérases classiques de type « Taq polymérase » qui insèrent une adénosine en position 3' terminale des fragments amplifiés. Dans le cas de Taq polymérases modifiées ayant une activité « proofreading » (par exemple la polymérase iProof® utilisée dans cette thèse), les adénosines sont éliminées par l'activité exonucléase de l'enzyme. Il est donc nécessaire d'ajouter après réaction PCR des adénosines en 3' des fragments par le biais de la technique du « A-tailing ». Pour cela, 7µL de produit de PCR purifié, 1µL de tampon polymérase 10X, 0,2 mM d'ATP, 5 unités de Taq polymérase dans un volume final de 10µL complété à l'eau déionisée sont mélangés. La réaction est incubée à 72°C pendant 10 à 15 minutes. La réaction de ligation se compose des éléments suivants :

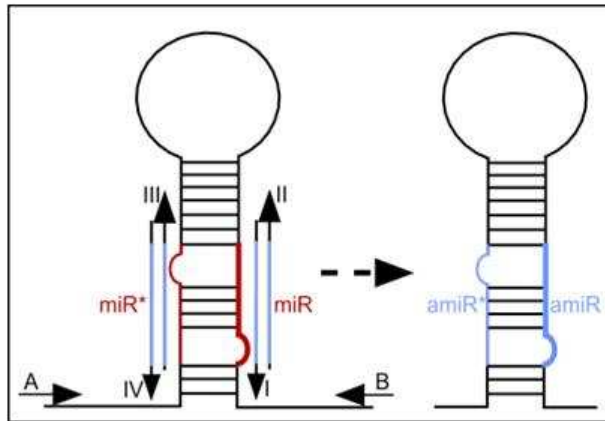
Réactifs	Quantité (volume total 10µL)
Vecteur pGEMT-easy®	50ng
Produit PCR	X ng*
Tampon ligase 10X	1µL
T4 DNA ligase	1µL
Eau	Qsp 10µL

* : X est la quantité en nano grammes d'insert pour une ratio de concentrations molaires vecteur : insert déterminé. En général, le ratio égal à 3 :1 est recommandé mais ce ratio peut être optimisé pour chaque réaction. Il peut varier de 1 :1 à 10 :1. La quantité en nano grammes est calculée comme suit :
$$\left(\frac{\text{ng vecteur} \times \text{taille en kb d'insert}}{\text{taille en kb du vecteur}} \right) * \left(\frac{\text{mol de vecteur}}{\text{mol d'insert}} \right)$$

La réaction est incubée à 4°C pendant une nuit. Le produit de ligation est ensuite utilisé pour la transformation et la sélection des bactéries blanches et bleues. Les colonies blanches, devant contenir un insert, sont testées par PCR et l'ADN plasmidique des colonies positives est extrait et séquençé pour sa vérification.

7.2.4. Clonage de microARN artificiels (*amiR*).

Les plasmides pRS300, pBJ36_AlcA et pMLBART_AlcR (voir cartes en annexe xx) proviennent du laboratoire du Pr. Detlef Weigel (Max Planck Institute For Developmental Biology, Tübingen, Allemagne) et du laboratoire de P. Laufs (« Facteurs de transcription et architecture », IJBP Versailles). Dans un premier temps, deux *amiR* ont été sélectionnés parmi la liste des *amiR* potentiels spécifiques des gènes *DYW2* (*amiRaa* et *amiRba*) et *OTP100* (*amiR1a* et *amiR2a*). Le choix des *amiR* a été réalisé en suivant les recommandations du site <http://wmd3.weigelworld.org>, en privilégiant des régions ne ciblant ni les domaines PPR ni le domaine *DYW* afin d'améliorer leur spécificité. La « création » de chaque *amiR* est réalisée par PCR ayant pour matrice le plasmide pRS300 contenant la séquence codante du *miR319a*. Pour chaque *amiR*, quatre amorces spécifiques (I, II, III, IV) sont définies grâce au logiciel en ligne Web MicroARN Designer <http://wmd3.weigelworld.org>. Ces amorces sont dessinées afin de modifier la séquence du *miR319a* et de la rendre complémentaire des transcrits *DYW2* ou *OTP100*. Les PCR (4) sont réalisées comme décrit dans la figure 28 en utilisant la polymérase iProof™ high fidelity (Biorad), les amorces universelles A et B et les amorces spécifiques à chaque *amiR*. La séquence de toutes les amorces et la carte des vecteurs utilisés sont présentés dans le tableau 22 de l'annexe. Le produit PCR (d) correspondant à chaque *amiR* a été cloné dans le vecteur pGEMT-Easy® selon le protocole décrit ci-dessus (cf section 7.2.3.). Les plasmides obtenus ont été vérifiés par séquençage en utilisant les amorces A et B. Le fragment a ensuite été libéré du vecteur pGEMT-Easy® par restriction avec les enzymes HindIII et XbaI puis il a été sous-cloné dans le vecteur pBJ36_AlcA. La séquence a été vérifiée par restriction et séquençage. Enfin, le construit chimérique *PALCA:amiR* a été transféré dans le vecteur binaire pMLBART_AlcR sous la forme d'un fragment NotI. Le nouveau plasmide a été vérifié par restriction. Ce dernier plasmide, contenant au sein de son ADN-T les constructions *P35S:ALCR* et *PALCA:amiR*, a été utilisé pour transformer la souche C58C1pMP90 d'*A. tumefaciens* (cf section 7.3.2.1.). La souche obtenue a été utilisée pour transformer des plantes d'*A. thaliana*, écotyle Col-0, selon la procédure par immersion de boutons floraux (cf section 7.4.4.).



PCR	Amorce sens	Amorce antisens	Matrice
(a)	A	IV	pRS300
(b)	III	II	pRS300
(c)	I	B	pRS300
(d)	A	B	(a)+(b)+(c)

Figure 28. Stratégie utilisée pour générer et amplifier des amiR.

7.2.5. Extraction d'ADN

7.2.5.1. Extraction rapide d'ADN génomique (ADNg) d'*A. thaliana*

Pour le génotypage des plantes, un protocole d'extraction rapide d'ADN génomique a été utilisé. Il est important de noter que ce protocole ne permet pas d'éliminer un grand nombre de contaminants, comme les protéines et des sucres par exemple. Le protocole est le suivant :

1 à 3 g de tissus sont collectés dans des tubes (eppendorf 2 ml) contenant deux billes métalliques. Les échantillons sont broyés dans de l'azote liquide à l'aide du TissueLyserII (QIAGEN) pendant 30 à 60 secondes à 20-30 Hz et avec les adaptateurs correspondants. Une fois les échantillons broyés, 400µL de tampon d'extraction [SDS 0,5 % ; NaCl 250 mM ; EDTA 25 mM; Tris-HCl (pH: 7,5) 200 mM] sont ajoutés à chaque tube. Après un léger mélange, les échantillons sont centrifugés 2 minutes à vitesse maximale (20000 g) dans une centrifugeuse de paillasse de type eppendorf et à température ambiante. 300 µL de surnageant sont collectés et placés dans un nouveau tube puis 300 µL d'isopropanol sont ajoutés à chaque tube pour précipiter les acides nucléiques. Après 2 minutes d'incubation à température ambiante, les échantillons sont centrifugés pendant 5 minutes à vitesse maximale (20000 g) dans une centrifugeuse de paillasse et à température ambiante. Le surnageant est éliminé et le culot est séché à température

ambiante. Finalement, les culots sont resuspendus avec 100 à 200 μ L de tampon TE pH 8,0 et les échantillons sont stockés à -20°C.

L'ADNg extrait est généralement assez propre pour effectuer une PCR par la suite. Pour obtenir des extraits d'ADNg contenant moins de contaminants, une étape d'élimination des protéines par « phénol/chloroforme » peut être ajoutée à la suite de l'extraction « rapide ». Le protocole est le suivant :

Un volume d'un mélange phénol/chloroforme est ajouté à un volume d'échantillon. Après avoir été mélangés pendant 30 secondes à l'aide d'un vortex, la solution est centrifugée durant 10 minutes à vitesse maximale (20000 g) dans une centrifugeuse de paillasse à température ambiante. La phase supérieure est ensuite prélevée et placée dans un tube propre. Un volume égal de chloroforme est ajouté, les échantillons mélangés au vortex et les tubes sont centrifugés pendant 10 minutes à vitesse maximale et à température ambiante. La phase supérieure est à nouveau collectée puis une précipitation des acides nucléiques est réalisée en ajoutant 1/10ème de volume d'acétate de sodium (5M) et 2 volumes d'ETOH 95°C froid. Cette solution est centrifugée 20 minutes à vitesse maximale et à 4°C. Le surnageant est éliminé et les culots rincés avec 200 μ L d'éthanol à 70%. Après centrifugation durant 15 minutes à vitesse maximale et à 4°C, le surnageant est éliminé et les culots sont séchés puis resuspendus dans 100 à 200 μ L de TE pH 8.0.

7.2.5.2. Extraction d'ADNg en plaque 96 puits

Une méthode d'extraction et purification d'ADNg de plantes consiste à utiliser des plaques 96 puits portant des membranes de purification. Cette méthode est couteuse en raison du cout des plaques (Whatman). Néanmoins, elle a pour avantage de permettre l'extraction d'ADNg « propre » et ceci de manière simple et simultanée pour 2 à 4 plaques (192 à 384 échantillons) en parallèle. Le protocole est le suivant :

Deux billes métalliques sont placées dans chaque puits de plaque de type *Mycrotube Rack System* (National Scientific Supply Co) puis 50-100 mg de tissu est prélevé et mis dans chaque tube de la plaque (il est conseillé de placer la plaque sur glace pendant le temps du prélèvement). Les tubes sont fermées et les planques sont ensuite

METHODES ET MATERIELS

congelées dans de l'azote liquide. Le tissu est broyé à l'aide du du TyssueLyserII (QIAGEN) pendant 60 secondes à 20 Hz. 300µL de tampon AP1 + RNase, préchauffé à 65°C, est ajouté dans chaque puits. Les plaques sont fermées et agitées énergétiquement pendant 15 secondes de haut en bas. Les plaques sont centrifugées à 3500g 5 minutes puis 100µL de tampon AP2 sont ajoutés dans chaque puit. Après avoir bouché chaque puit, une nouvelle agitation est réalisée pendant 15 secondes. Les plaques sont placés à -20°C pendant minimum 10 minutes (idéalement une nuit pour favoriser la précipitation des protéines). Les échantillons sont centrifugés pendant 5 minutes à 3500g, puis 300µL de surnageant sont transférés dans une nouvelle plaque. 450µL de tampon AP3/E sont ajoutés à chaque échantillon. Une nouvelle agitation est réalisée pendant 15 secondes. 600µL d'échantillon sont ensuite transférés sur la plaque de colonnes Whatman UNIFILTER® (GE Healthcare) sur le S-Block. Les échantillons sont centrifugés pendant 5 minutes à 3500g. Le liquide est jeté puis 600µL de tampon AW/E sont ajoutés à chaque colonne. La plaque est fermée avec une feuille spéciale perméable aux gaz puis les échantillons sont centrifugés pendant 15 minutes à 3500g.

Finalement la plaque Whatman est placé sur la plaque d'élution et 100µL de Tampon AE sont ajoutées à chaque colonne. Après 1 minute d'incubation à température ambiante les échantillons sont élués par centrifugation pendant 3 minutes à 3500g et les extraits d'ADNg sont stockés à -20°C

Composition des tampons :

Tampon AP1 (pour 500ml de solution)

Produits	Volume (mL)	Masse (g)
Tris 1 M pH 8,0	50	
EDTA 0,5M	50	
NaCl 2,5 M	100	
SDS 20 %	31,5	
PVP 40 000		5
Bisulfite de sodium		5
Eau	268	

METHODES ET MATERIELS

Tampon AP2 (K Ac. 3M – A. Ac. 5M, pour 200ml de solution)

Produits	Volume (mL)
Acétate de potassium 5M	120
Acide Acétique Glacial	23
Eau	57

Tampon AP3 (pour 150ml de solution)

Produits	Quantités
Chlorure de Guanidine	37,25 g
Ethanol 96%	100 ml
Eau	Qsp 150 ml

Dissoudre le Chlorure de Guanidine dans de l'eau pour atteindre un volume final de 50ml de solution puis ajouter l'éthanol.

Tampon AW (pour 1 L de solution)

Produits	Volume (mL)
Tris 1 M pH 8,0	22,5
EDTA 0,5M	0,2
Acétate de potassium 5M	32
Eau	945,3

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

7.2.6. *Extraction d'ADN plasmidique*

L'extraction de plasmide est effectuée à partir d'une culture liquide bactérienne ayant atteint son plateau de croissance. Celle-ci est culottée par centrifugation à 3000g pendant 15 minutes et à 4°C. Le surnageant est éliminé et le culot est conservé. De façon générale, l'extraction d'ADN plasmidique de bactéries consiste en une resuspension suivie d'une lyse, généralement avec un tampon basique, et finalement par la purification au sens strict de l'ADN plasmidique. Cette procédure a été standardisée et plusieurs kits utilisant des membranes à base de silice sont disponibles sur le marché. Dans cette thèse, le kit: « GeneJet plasmid miniprep » (Fermentas) a été utilisé. Le protocole rapide a été utilisé et est présenté ci-dessous.

Note : Toutes les étapes sont réalisées à température ambiante. Les centrifugations se font dans une centrifugeuse de paillasse (micro-centrifugeuse) avec une vitesse de rotation d'au moins 12000g.

1. Resuspension des cellules, lyse et neutralisation.

250 μ L de solution de resuspension contenant de la RNase sont ajoutés au culot bactérien, puis les bactéries sont resuspendues en les vortexant ou par pipetage. 250 μ L de solution de lyse sont ensuite ajoutés et mélangés grâce à 4 à 6 inversions douces. 350 μ L de solution de neutralisation sont ajoutés à cette solution visqueuse et mélangés par 4 à 6 inversions douces. Le mélange est centrifugé pendant 5 min.

2. Purification de l'ADN plasmidique.

Le surnageant est transféré dans une colonne de purification et centrifugé pendant 1 min. Après avoir jeté l'éluat, 500 μ L de solution de lavage sont ajoutés à la colonne de purification puis celle-ci est centrifugée pendant 1 min. Le lavage est répété une fois. Après avoir jeté l'éluat une nouvelle fois, la colonne est centrifugée à vide pendant 1 min pour éliminer toute trace d'alcool.

3. Elution de l'ADN

30 à 50 μ L de tampon d'élution sont ajoutés à la colonne (il est possible d'utiliser de l'eau mais le tampon garantit mieux l'intégrité de l'ADN pour un stockage long), après 2 min d'incubation, la colonne est centrifugée pendant 2 min et l'ADN plasmidique est élué dans un tube collecteur. Il est ensuite stocké à -20°C.

7.2.7. Quantification de l'ADNg ou l'ADNp

Au laboratoire, les acides nucléiques sont quantifiés grâce au spectrophotomètre NanoDropTM ND-1000. Ceci permet une estimation relativement précise de la concentration d'ADN dans l'échantillon. Il a l'avantage d'être simple et d'utiliser peu d'échantillon (1 μ L du produit d'extraction). La concentration en acides nucléiques (ADN et ARN) de l'échantillon est calculée en ng/ μ L. Deux ratios permettent d'estimer la qualité de l'échantillon par rapport aux contaminants. Le ratio 260/230 correspond au degré de

pureté de l'échantillon par rapport aux contaminants hydroxylés principalement. Les acides nucléiques ont un pic d'absorption autour de 260nm alors que les contaminants d'extraction comme les phénols et alcools ont un pic d'absorption aux alentours de 230nm. Plus le ratio 260/230 est proche de 2, moins l'échantillon est contaminé. Généralement, des ratios de plus de 1,6 sont acceptables pour des échantillons d'ADN. En revanche, pour l'ARN, les contaminants sont plus gênants et on privilégie des ratios supérieurs à 1,8. Le ratio 260/280 reflète la contamination de l'extrait par des protéines. Les protéines sont des contaminants pouvant gêner l'efficacité de futures manipulations, telles que des réactions PCR. Les protéines ont un pic d'absorption aux alentours de 280nm. On recherche des échantillons ayant un ratio 260/280 proche de 2, synonyme de faible contamination protéique.

7.2.8. *Amplification d'ADN par la réaction de polymérisation en chaîne (PCR).*

A partir d'une matrice ADN, il est possible d'amplifier, de façon très importante, une séquence d'intérêt par PCR. Le principe repose sur l'appariement d'oligonucléotides (ou amorces), spécifiques de la séquence à amplifier, qui sont « allongés » par des enzymes de la famille *Thermus aquaticus* polymérase (Taq) nécessitant le cofacteur Mg^{2+} et des désoxy-nucléotides. La plupart des amorces utilisées dans cette thèse ont été dessinées à l'aide du logiciel Primer3Plus en cochant la case « détection » et en indiquant les températures d'hybridation souhaitées. Ce logiciel est disponible sur internet (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>). Les amorces pour les lignées d'insertion ont été obtenues grâce à l'outil « *iSect Primer* » de la page web de « *T-DNA express: Arabidopsis gene mapping tool* » (<http://signal.salk.edu/cgi-bin/tdnaexpress>).

Plusieurs types de Taq DNA polymérase ont été utilisées pendant la thèse :

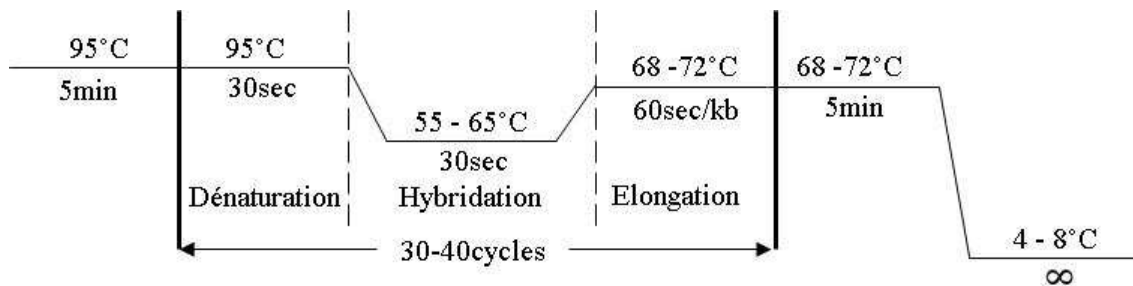
La Taq DNA polymérase du fournisseur Biolabs a été utilisée pour la plupart des manipulations : génotypages, sélection de colonies, analyses d'édition... Trois volumes réactionnels sont utilisés (10, 25 et 50µL).

METHODES ET MATERIELS

Réactifs	Volume (µL)		
	10µL	25µL	50µL
Tampon (10X)	1	2,5	5
Amorce sens (10µM)	0,2	0,5	1
Amorce antisens (10µM)	0,2	0,5	1
dNTPs (10mM chacun)	0,2	0,5	1
Taq polymérase (5u/µL)	0,1	0,15	0,25
ADNg*	0,5-1	1	1-2
Eau	Qsp 10	Qsp 25	Qsp 50

* Pour la PCR sur colonies, l'ADN est remplacé par de l'eau. La quantité d'ADN matrice peut varier de 10pg à 500 ng

Conditions d'amplification :

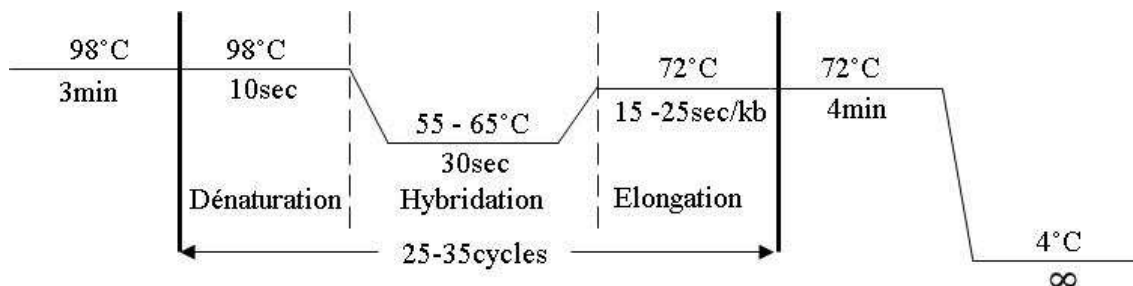


La Taq DNA polymérase iProof™ high fidelity (Biorad) a été utilisée lorsque les fragments générés ne devaient pas présenter d'erreur et devaient être clonés.

Réactifs	Quantités (µL)
	(volume total 50µL)
Tampon iProof (5X)	10
Amorce sens (10µM)	1
Amorce antisens (10µM)	1
dNTPs (10mM chacun)	1
Taq polymérase (2u/µL)	0,5
Matrice ADN*	1
Eau	35,5

* ADNg ou ADNc.

Conditions d'amplification :



7.2.9. PCR quantitative ou en temps réel (qPCR)

Cette technique est une modification de la PCR qui permet d'observer en temps réel la progression de la réaction. Ceci est possible grâce à l'utilisation de molécules fluorescentes qui sont intégrées dans les fragments d'ADN au cours de la réaction. Elle est utilisée dans cette thèse pour la quantification de l'expression des gènes. Selon la nature des molécules fluorescence utilisées, plusieurs types de qPCR peuvent être distingués. Dans cette thèse, j'ai utilisé le « *SYBR green* », une molécule s'intercalant dans l'ADN, pour évaluer le niveau d'expression des gènes nucléaires et des organites. Dans ce cas, la qPCR est réalisée en utilisant comme matrice de l'ADN complémentaire (ADNc) obtenu par reverse transcription (RT) (cf section 7.2.17.). Cette méthode est ainsi appelée RT-qPCR.

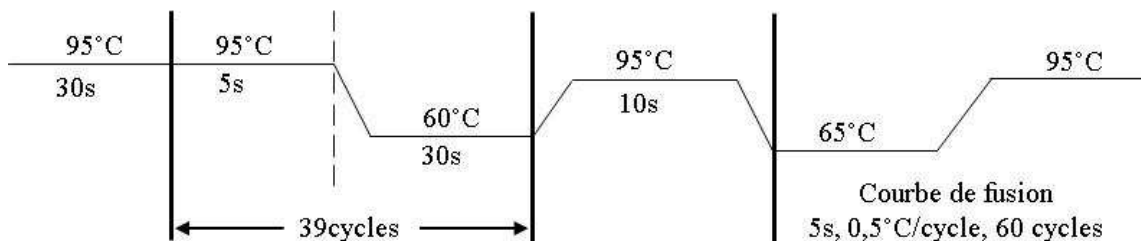
Plusieurs variables sont importantes pour la quantification par RT-qPCR de l'expression des gènes. Parmi elles, il faut souligner la génération et la validation des amorces spécifiques du gène d'intérêt, la qualité des échantillons d'ARN, les enzymes ou kits utilisés, l'utilisation de gènes de référence et la méthode de quantification. Dans cette thèse, j'ai suivi les recommandations de l'article « *Minimum Information for publication of Quantitative real time PRC Experiments* » (MIQE) (Bustin *et al.*, 2009). Ce document précise l'ensemble des paramètres à prendre en compte lors d'expériences de RT-qPCR et qui nécessitent d'être décrits dans les publications. La méthode de quantification choisie dans cette thèse est la quantification relative utilisant des courbes standards. Les gènes de référence choisis lors des expériences sont : le gène ribosomal 18S et le gène *ACTIN2*.

Les réactions de qPCR ont été réalisées en utilisant la machine CFX384 Touch™ Real System (Biorad) ou le système « *Roche LightCycler real time PCR system* » dans un volume final de 10µL contenant les réactifs suivants :

Réactifs	Volume (µL)
Mesa Green qPCR mastermix Plus SYBR Assay*	5
Amorces sens et antisens (1µM chacun)**	4
Matrice ADNc (1/10°, 1/100°, 1/1000°, 1/5000°)	1

*Fournisseur EUROGENTEC, **Pour les réactions de qPCR pour l'étude du transcriptome chloroplastique ou mitochondrial, la concentration de chaque amorce est de 0,75µM.

Conditions générales d'amplification



Les conditions d'amplification dans le système « *Roche LightCycler real time PCR system* » sont légèrement différentes : 95°C pendant 10 min, suivi de 40 cycles à 95°C pendant 10s, puis 60°C pendant 10s et 72°C durant 20s.

7.2.9.1. Dessin des Amorces de qPCR

Les amorces de qPCR ont été dessinées grâce au logiciel Primer3Plus cochant la case « General settings, Load and Save, qPCR » et avec une température d'hybridation optimale de 60°C. La taille du fragment a été choisie entre 80 et 200 pb et avec un contenu optimal en GC des amorces de 45 à 60%. Les amorces ont été validées sur une matrice d'ADNc obtenue à partir de feuilles en vérifiant la spécificité d'amplification, l'absence de formation de dimères et une efficacité d'amplification (E) comprise entre 90% et 105%. Ces tests sont réalisés à partir d'une large gamme de dilutions de l'ADNc (1/10^e, 1/100^e, 1/1000^e, 1/5000^e).

7.2.10. Electrophorèse sur gel agarose

L'électrophorèse sur gel d'agarose permet la séparation d'échantillons d'acides nucléiques linéaires, par exemple des produits de PCR, en fonction de leurs masses moléculaires et grâce à un champ électrique permettant la migration des acides nucléiques chargés négativement vers le pôle positif du champ induit.

Le gel d'agarose est préparé en présence de bromure d'éthidium (BET), un intercalant des acides nucléiques qui fluoresce sous les UV. Un autre intercalant, le SYBR Gold, peut aussi être utilisé pour colorer les acides nucléiques. La concentration en agarose

détermine la capacité de séparation des molécules d'acides nucléiques en fonction de leur taille. Elle est généralement comprise entre 0,8 et 2% (masse d'agarose/volume de tampon). Une concentration élevée est préconisée pour séparer les fragments de petite taille. Dans cette thèse, la concentration d'agarose la plus souvent utilisée est de 1% dans du tampon Tris-Acetate-EDTA (TAE ; 40 mM TRIS-Acetate, 1 mM EDTA, pH=8). Une solution de glycérol contenant un colorant chargé négativement (par exemple rouge de crésol dans cette thèse) est ajoutée à chaque échantillon avant dépôt sur gel afin d'alourdir la solution et de suivre la migration. La migration a lieu dans une cuve d'électrophorèse remplie de tampon TAE sous une tension constante généralement comprise entre 50 et 300V (le temps et le voltage variant en fonction de la taille du gel et des échantillons testés).

7.2.11. Séquençage de produits de PCR ou de plasmides

Le séquençage a été effectué par une société privée spécialisée (GATC BIOTECH, <http://www.gatc-biotech.com/en/index.html>). Les conditions d'envoi sont définies par le prestataire. Les résultats sont reçus 2 à 3 jours après envoi et les séquences sont analysées en utilisant le logiciel A plasmid Editor (ApE) version 2.0.45.

7.2.12. Extraction d'ARN totaux

L'ARN est une molécule fragile. Tout d'abord, en raison de la présence de nombreuses RNases dans les tissus vivants et de la stabilité de ces enzymes. D'autre part, les molécules d'ARN sont sensibles à la dégradation physique comme sous l'effet de fluctuations thermiques prolongées ou d'un pH alcalin. Lorsque l'on prépare des échantillons d'ARN, il est donc impératif de travailler avec des gants, des tubes, des cônes et des solutions « RNase free ». Il est aussi important de manipuler à 4°C pour préserver les échantillons. Deux protocoles d'extraction d'ARN ont été utilisés dans cette thèse.

7.2.12.1. Extraction à l'aide du kit RNeasy plant Minikit (QIAGEN)

L'utilisation du kit RNeasy plant Minikit permet l'extraction d'ARN totaux grâce à des membranes à base de silice qui permettent une bonne reproductibilité de la qualité de l'ARN. Cependant, ce kit est onéreux, limité dans la quantité de matériel de départ utilisé et ne permet pas d'extraire la totalité des molécules d'ARN. En effet, les ARN de taille inférieure à 200b ne sont pas retenus par la colonne. Cette extraction a été néanmoins choisie pour la plupart des extractions d'ARN réalisées pendant ma thèse. Le protocole suivi est celui préconisé par QIAGEN et disponible en ligne sur leur site internet.

7.2.12.2. Extraction au Trizol® «RNA Now »™ (Ozyme)

L'extraction des ARN peut être réalisée également à l'aide de Trizol®. Ceci permet d'extraire tous les ARN sans exclusion de taille mais nécessite l'utilisation de substances toxiques comme le phénol et le chloroforme. De plus, certaines étapes sont critiques et peuvent influencer dans la qualité de l'ARN extrait. Le protocole est le suivant (en utilisant des gants et en travaillant sous sorbonne) :

Le matériel végétal est prélevé et placé immédiatement dans de l'azote liquide. Il est ensuite broyé à l'aide d'un TissueLyser II ou de pilons et de mortiers propres et autoclavés (tout en maintenant la chaîne du froid dans l'azote liquide). Après broyage, les échantillons sont homogénéisés, sous sorbonne chimique, dans 1ml de Trizol® pour 50 à 100mg de tissu. Les échantillons sont ensuite incubés 5 minutes à température ambiante et 0,2mL de chloroforme est ajouté pour 1ml de Trizol. Les tubes sont inversés 15 secondes et incubés 2 à 3 minutes à température ambiante. Les échantillons sont centrifugés à 12 000 g maximum durant 15 minutes à 4°C.

Après centrifugation, le mélange est séparé en une phase inférieure rouge, la phase de phénol-chloroforme, une interphase, et une phase aqueuse supérieure incolore. Les ARN se trouvent la phase aqueuse qui est prélevée et transférée dans un nouveau tube. Les ARN sont précipités avec 0,5ml d'isopropanol pour 1ml de Trizol. Les échantillons sont incubés 10 minutes à température ambiante et centrifugés à 12 000 g maximum durant 15

minutes à 4°C. Le surnageant est éliminé et le culot est rincé avec 1ml d'éthanol à 75% pour 1ml de Trizol. Une dernière centrifugation à 7 500g maximum durant 5 minutes à 4°C est enfin nécessaire. Les culots sont resuspendus dans de l'eau « RNase free ».

7.2.13. *Traitement de l'ARN total à la DNase*

L'extraction des ARN ne permet pas d'éliminer toutes les molécules d'ADN (nucléaire, mitochondrial et chloroplastique). Dans certaines expériences, il est impératif d'éliminer toute trace d'ADN, notamment lors de la caractérisation du niveau d'expression de gènes ou de l'étude de l'édition des l'ARN. Pour cela, un traitement à la DNase doit être réalisé. Dans cette thèse, deux types de DNase ont été utilisées, la « DNase I amplification grade » (Invitrogen) et la Turbo™ DNase (Ambion). Les protocoles sont les suivants :

Réactifs	Composition de la réaction (DNase I Ampl. grade 1U/μL)	Composition de la réaction (Turbo™ DNase 2U/μL)
ARN	1-5 μg	1-10μg
Tampon 10X	1μL	5μL
DNase	1μL	1μL
Eau	Qsp10μL	Qsp50μL

Dans le cas de la DNase I amplification grade :

Le mélange réactionnel est placé à 25°C pendant 20 minutes puis 1μL de DNase I est à nouveau ajouté et l'échantillon est de nouveau incubé à 25°C pendant 20 minutes. La réaction est arrêtée par ajout de 1μL d'EDTA 25mM et incubation à 65°C pendant 10 min. Les ARN sont ensuite précipités et stockés à -80°C.

Dans le cas de la Turbo™ DNase :

Le mélange réactionnel est placé à 37°C pendant 30 minutes puis 1μL de Turbo™ DNase est à nouveau ajouté et l'échantillon est incubé à 37°C pendant 30 minutes. La réaction est arrêtée par ajout de 14μL d'EDTA 25mM et incubation à 75°C pendant 10 min. Les ARN sont ensuite précipités et stockés à -80°C.

Une étape de vérification de la contamination ADN par PCR est nécessaire afin de vérifier l'absence totale d'amplification sur les extraits d'ARN. Dans le cas d'une amplification positive, une nouvelle étape de traitement à la DNase est réalisée.

7.2.14. Précipitation des ARN

Plusieurs méthodes sont possibles pour la précipitation des acides nucléiques tels que les ARN. Dans cette thèse, les ARN sont précipités par la méthode « acétate de sodium/éthanol ». Ainsi, il est ajouté 1/10^e de volume acétate de sodium 3M pH 5,2 et 2,5 volumes d'éthanol 96° froid (0°C) sont ajoutés à un volume d'échantillon ARN. Les échantillons sont placés à -20°C durant une heure au minimum (une nuit est conseillée) puis centrifugés à 20 000 g pendant 15 minutes à 4°C. Le culot est rincé avec de l'éthanol 70% puis l'échantillon est centrifugé durant 10 à 15 minutes à 10000g et à 4°C. Le culot est repris dans de l'eau « RNase free ».

7.2.15. Quantification des l'ARN totaux

De la même façon que pour l'ADN, l'ARN peut être quantifié au Nanodrop (cf section 7.2.7.). Les mesures effectuées au Nanodrop sont néanmoins peu précises ce qui peut poser problème pour certaines expériences.

Pour plus de précision, les ARN peuvent être dosés au Ribogreen (Molecular Probes). Le Ribogreen est une molécule qui fluoresce peu à l'état libre mais qui, lorsqu'il est complexé aux acides nucléiques, fluoresce de manière linéaire dans une gamme de concentrations. De cette manière les ARN peuvent être quantifiés de façon très précise. Le protocole donné ici est celui utilisé par la plateforme transcriptomique de l'URGV.

7.2.15.1. Préparation de la gamme

76µL de tampon TE sont ajoutées à un échantillon de 4µL de Ladder ARN 0,5-10kb (100ng/ µL). Une gamme d'étalonnage est préparée directement dans une plaque à fond noir selon le tableau suivant :

METHODES ET MATERIELS

Puits*	Volume de TE (μL)	Volume d'ARN gamme (μL)	Quantité (ng)
1	30	20	100
2	34	16	80
3	38	12	60
4	42	8	40
5	46	4	20
6	48	2	10
7	49	1	5
8	49.5	0.5	2.5

*D'une plaque à fond noir.

7.2.15.2. Préparation des échantillons

Les ARN à quantifier sont dilués avec du TE en se plaçant à une concentration approximative de 50ng/μL. Dans la plaque à fond noir, 49μL de TE sont déposés dans chaque puit, puis 1μL d'échantillon dilué est ajouté par puit.

7.2.15.3. Préparation du Ribogreen

Une solution de Ribogreen dilué est préparée dans un tube noir de 1,5mL en ajoutant 50μL de TE et 0,25μL de Ribogreen par échantillon à doser. 50μL de cette solution de Ribogreen diluée sont ensuite ajoutés à chaque puits de la plaque (gamme + échantillons). La plaque est ensuite couverte et les réactions sont agitées pendant 5 minutes à température ambiante. La lecture du signal fluorescent est réalisée dans le fluorimètre à plaques FLUOstar *Galaxy* et à l'aide du logiciel BIOLISE.

7.2.16. Test de la qualité et de l'intégrité des ARN issus d'une extraction

Le test d'intégrité des ARN se base sur la comparaison du profil de l'échantillon étudié avec un profil type connu pour des ARN de bonne qualité. Ce test peut être réalisé par électrophorèse sur gel d'agarose ou bien, pour des résultats plus précis, par analyse de l'échantillon sur puce « Agilent » (Agilent, RNA 6000 Nano Kit). Le protocole est celui utilisé par la plateforme transcriptomique de l'URGV.

7.2.16.1. Préparation des échantillons

20µL de RNA 6000 ladder sont dénaturés durant 10 min à 65°C. 1µL de chaque échantillon est prélevé et placé dans un tube de 1,5mL. Les échantillons sont ensuite dénaturés pendant 2 minutes à 65°C puis placés sur glace. Le gel est ensuite préparé et les échantillons sont déposés selon les instructions d'Agilent.

7.2.17. Réaction de transcription inverse

Il s'agit de la synthèse du brin d'ADN complémentaire (ADNc) à partir d'une matrice ARN. Le but est d'obtenir un équivalent ADN d'un transcrit. A partir de ce brin d'ADNc, il est possible d'effectuer les manipulations comme des PCR qualitatives ou quantitatives. La synthèse d'ADN complémentaire est effectuée en utilisant le kit de reverse transcriptase SuperscriptII™ (Invitrogen™).

La synthèse se fait à partir de 150ng à 5µg d'ARN total traité à la DNase (7.2.13.) auxquels sont ajoutés 1µL de dNTP (10mM), 150ng d'héxanucléotides aléatoires ou 2µL oligonucléotide dT (10µM), le tout dans un volume final de 12µL. Après 5 minutes à 65°C puis 1 minute sur glace, 4µL de tampon de réaction 5X, 2µL de DTT, 1µL de RNase OUT™ sont ajoutés. Dans le cas d'héxanucléotides aléatoires, le mélange est incubé à 25°C pendant 2 minutes ; dans le cas d'un oligonucléotide dT, il est incubé à 42°C pendant 2min. Ensuite, 1µL de transcriptase inverse est ajouté et la réaction est incubée à 25°C pendant 10 minutes puis 50 min à 42°C (héxanucléotides aléatoires) ou directement 50 minutes à 42°C (oligonucléotide dT). Finalement, l'enzyme est inactivée par 15 minutes d'incubation à 70°C.

En général, un témoin négatif est généré en parallèle : réaction sans transcriptase inverse qui est alors remplacée par de l'eau. Ceci permet par la suite de vérifier qu'il ne restait pas de contamination d'ADN génomique dans les ARN.

7.3. Méthodes relatives aux bactéries

7.3.1. *Souche d'E. coli*

7.3.1.1. Préparation de bactéries thermo-compétentes d'E. coli.

Une préculture de 10mL est inoculée par une souche bactérienne dans du milieu liquide (LB pour Luria-Bertani) et est incubée une nuit à 37°C et à 200-250rpm. La préculture est utilisée le lendemain pour ensemer 500mL de milieu (SOB pour Super Optimal Broth) qui est incubé à 19°C sous agitation jusqu'à l'obtention d'une densité optique DO (600nm) comprise entre 0,5 et 0,6. La culture est refroidie 10 minutes sur la glace puis centrifugée à 5000g pendant 10 minutes à 4°C. Le culot bactérien est repris dans 160ml de solution TB froide puis incubé pendant 10 min. Les bactéries sont à nouveau centrifugées à 5000g pendant 10 minutes et à 4°C puis reprises dans 40ml de TB froid. Après ajout de 5,6ml de DMSO, elles sont incubées 10 minutes sur la glace, aliquotées et finalement congelées dans de l'azote liquide avant stockage à -80°C.

Solution TB : 10mM Pipes, 15mM CaCl₂, 250mM KCl, ajusté à pH 6,7 avec du KOH, 55mM MnCl₂, stérilisé par filtration.

Milieu LB liquide : 10g/L bacto-tryptone, 5g/l extrait de levure, 10g/L NaCl, pH ajusté à 7 avec du NaOH.

Milieu LB solide : 10g/L bacto-tryptone, 5g/L extrait de levure, 10g/L NaCl, Agarose 15g/L pH ajusté à 7 avec du NaOH.

Milieu SOB : 20g/l bacto-tryptone, 5g/L extrait de levure, 10mM NaCl, 2,5mM KCl, 10mM MgCl₂, 10mM MgSO₄, pH ajusté à 7 avec du NaOH.

7.3.1.2. Transformation de bactéries thermo-compétentes et sélection

Le principe repose sur l'application de deux chocs thermiques successifs permettant l'entrée du vecteur dans la bactérie. Le protocole est le suivant :

Les bactéries thermo-compétentes sont décongelée sur glace puis 100µL de bactéries sont ajoutées à 10 à 100ng (1-2 µL) de vecteur ou de réaction de clonage dans un

tube froid. Le mélange est incubé sur glace durant 5 à 20 minutes. Un choc thermique est ensuite réalisé par 45 secondes d'incubation à 42°C (bain Marie) puis 5 minutes d'incubation à 0°C (bain eau-glace). Après choc thermique, les bactéries sont mises en culture dans 900µL de milieu SOC (pour « Super Optimal broth with Catabolic repressor ») à 37°C pendant une heure sous agitation (150rpm). 50 à 150µL de la culture sont ensuite étalés sur milieu LB solide contenant l'antibiotique nécessaire pour la sélection des souches contenant le vecteur et incubés à 37°C pendant une nuit.

Une PCR sur colonies est réalisée le lendemain pour sélectionner des clones positifs. Pour cela, chaque colonie est piquée avec une pointe de cône stérile de 10µL, inoculée sur une boîte de LB solide et la pointe est ensuite immergée dans 10 ou 25µL de mélange réactionnel PCR. Les conditions de la PCR sont adaptées en fonction des amorces utilisées. Le produit PCR est ensuite observé par migration sur gel d'agarose.

Les colonies positives suite au test PCR sont inoculées dans 5 à 10 mL de milieu LB liquide à 37°C pendant la nuit. Le lendemain, 700µL de culture sont mélangés avec 300µL de glycérol 50% stérile puis stockés à -80°C. Le reste de la culture liquide est utilisé pour l'extraction de l'ADN plasmidique (cf section 7.2.6.) puis le plasmide est envoyé au séquençage avec les amorces correspondantes.

Milieu SOC: 20g/l bacto-tryptone, 5g/L extrait de levure, 10mM NaCl, 2,5mM KCl, 10mM MgCl₂, 10mM MgSO₄, 20mM glucose, pH ajusté à 7 avec du NaOH.

Milieu LB solide : 10g/L bacto-tryptone, 5g/L extrait de levure, 10g/L NaCl, Agarose 15g/L pH ajusté à 7 avec du NaOH.

7.3.2. Souche d'*A. tumefaciens*

7.3.2.1. Préparation de bactéries C58C1 thermo-compétentes

Une préculture est effectuée en ensemençant 8mL de LB agro, contenant les antibiotiques nécessaires, avec une colonie fraîche d'*A. tumefaciens*. La préculture est incubée une nuit à 28°C puis ajoutée à 192mL de milieu LB agro contenant les antibiotiques de sélection. La culture est incubée à 28°C sous agitation jusqu'à une DO

(600 nm) comprise entre 0,4 et 0,6 puis centrifugée pendant 15 min à 3400g et à 4°C. Le culot est repris dans 10ml de Tris-HCl 10 mM pH 7,5 froid, centrifugé 15 min à 3400g et à 4°C et resuspendu dans 4mL de LB agro froid. Des aliquotes de 200µL sont préparés et congelés dans de l'azote liquide avant stockage à -80°C.

Milieu LB agro liquide : 10g/L bacto-tryptone, 5g/l extrait de levure, 5g/L NaCl, pH ajusté à 7,5 avec du NaOH

Milieu LB agro solide : 10g/L bacto-tryptone, 5g/L extrait de levure, 5g/L NaCl, Agarose 15g/L pH ajusté à 7,5 avec du NaOH

7.3.2.2. Transformation de bactéries C58C1 thermo-compétentes

200µL de bactéries compétentes sont décongelées sur glace puis ajoutées à 1 µg de plasmide avant incubation pendant 5 min sur un mélange eau-glace. Deux chocs thermiques sont effectués par transfert des tubes 5 min dans de l'azote liquide puis en les incubant à 37°C pendant 5 min. Après choc thermique, les agrobactéries sont mises en culture dans 800µL de LB agro pendant 2h à 28°C sous agitation douce. 200µL de cette culture sont ensuite étalés sur milieu LB agro solide contenant les antibiotiques de sélection nécessaires. Après 48h à 28°C, les clones sont testés par PCR.

7.4. Méthodes relatives aux plantes

7.4.1. Culture d'*A. thaliana* en serre et chambre de culture

Deux compartiments (serre et chambre de culture) ont été utilisés pour le semis d'*Arabidopsis* en terre. Des plantes cultivées en serre ont été généralement utilisées pour la transformation, la sélection et la multiplication de plantes. Les conditions de culture dans ce compartiment sont de 21°C en moyenne et avec une périodicité de lumière de type « jours longs » (16h de lumière – 8h d'obscurité).

La culture des plantes dans de chambres de culture a été utilisée principalement lorsque des conditions plus contrôlées étaient nécessaires : phénotypage et expériences de

caractérisation de l'expression de gènes. Les conditions de culture dans la chambre « *Arabidopsis* » de l'URGV sont de 21°C pendant le jour et de 17°C la nuit, de 65% d'humidité et avec une périodicité de lumière de type « jours longs » (16h de lumière – 8h d'obscurité).

7.4.2. Stérilisation des graines

Deux protocoles de stérilisation des graines ont été utilisés lors de cette thèse : un traitement « acide » dans le cas de graines fortement contaminées et un second dans le cas d'une contamination plus faible.

Pour le premier protocole, une pastille d'eau de javel (CHLOR'O, Chimie Plus) est dissoute dans 50mL d'eau puis 50 µL de tween 20 sont ajoutés. 1 mL de cette solution est ensuite ajouté à 9 ml d'éthanol 96 % (v/v) et cette solution est utilisée pour stériliser les graines par incubation pendant 10 min maximum. Deux lavages à l'éthanol 96% sont ensuite effectués sous hotte stérile et les graines sont séchées sur papier stérile.

Dans le deuxième protocole, les graines sont stérilisées dans une solution d'éthanol 70% contenant 0,05% de SDS pendant 10 minutes. Après élimination de la solution, les graines sont incubées 10 minutes dans de l'éthanol 96% sous la hotte. Les graines sont finalement séchées sur papier stérile.

7.4.3. Culture *in vitro*

Deux types de milieux pour la culture de matériel végétal *in vitro* ont été utilisés lors de cette thèse.

Le milieu Murashige et Skoog (MS : MES 0,5g/L, 4,4g/L mix Murashige et Skoog salts SIGMA, pH 5,7), liquide ou solide, avec et sans addition de saccharose, a été utilisé pour la sélection de transformants et pour les expériences de phénotypages et de génotypages des lignées d'insertion. Pour la sélection de transformants ou pour l'évaluation de la ségrégation des ADN-T, des antibiotiques, Kanamycine (50 mg/L) et l'Hygromycine (25 mg/L), sont également ajoutés au milieu.

Un deuxième milieu utilisé pour la culture d'*Arabidopsis* est le milieu Duchefa (Ester & Somerville 1987) qui a été utilisé pour les expériences de BiFC et pour certains phénotypes.

Les graines stériles sont semées sur boîtes de milieu solide à l'aide de cure-dents stérilisés ou « à la volée ». Après traitement au froid (de 48h à 4°C) pour la synchroniser la germination des graines, les boîtes sont placées dans une chambre rotative à 21°C, avec un pourcentage d'humidité de 45% et une luminosité de type « jours longs » (16h de lumière – 8h d'obscurité).

7.4.4. Transformation stable de plantes d'*A. thaliana*

Une colonie d'*A. tumefaciens* portant le plasmide binaire d'intérêt est utilisée pour inoculer 5mL de milieu LB agro contenant de la rifampicine (100 µg/mL) et le(s) antibiotique(s) associés au plasmide. Les bactéries sont incubées durant 48h à 28°C en agitation à 150rpm. 500 µL de cette préculture sont utilisés pour inoculer 500mL de LB agro contenant les antibiotiques appropriés. La culture est incubée à 28°C jusqu'à l'obtention d'une DO (600nm) de 0,8 (environ 12 à 16 heures) puis centrifugée 30 min à 5000g et 4°C. Le surnageant est éliminé et le culot est resuspendu dans 500mL de milieu de transformation. Les inflorescences des plantes à transformer sont trempées dans la solution d'*Agrobacterium* pendant 2 min, puis les plantes sont laissées sous cloches pendant 24h.

Milieu de transformation : 5% Saccharose, 200 µL/L Silwet L-77

7.4.5. Croisement des plantes d'*A. thaliana*

A. thaliana est une plante essentiellement autogame mais qui tolère une fécondation croisée (allofécondation) entre génotypes. La fécondation a lieu naturellement avant ouverture des bourgeons floraux grâce à une synchronisation du développement de la partie male (étamines générant le pollen) et de la partie femelle (pistil).

Afin de croiser manuellement différents génotypes au laboratoire, les plantes sont cultivées dans la serre jusqu'au stade de croissance 6.10 – 6.50 (Boyce *et al.*, 2001). Une

hampe florale de la plante utilisée comme femelle est identifiée et les siliques et bourgeons floraux ouverts sont éliminés. A l'extrémité de cette hampe, deux bourgeons floraux bien développés mais non ouverts sont sélectionnés, les autres étant éliminés (les boutons sélectionnés sont généralement au stade 10-11 du développement de la fleur selon l'échelle du site <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>). Les pétales, les sépales et les étamines sont ensuite délicatement enlevés sous loupe binoculaire à l'aide de pinces fines et seul le pistil est conservé. Des étamines portant du pollen mature sont ensuite prélevées sur la plante considérée comme donneur mâle et le pollen est déposé sur le stigmate jusqu'à saturation. 2 à 3 jours après croisement, un début de croissance de la silique est normalement observé. Les siliques matures sont prélevées 2 à 3 semaines après croisement et laissées sécher dans un tube ouvert.

7.4.6. Induction de l'expression des amiR sous chambre de culture

Cette méthode a été adaptée du protocole d'induction du laboratoire de P. Laufs (« Facteurs de transcription et architecture », IJBP Versailles) http://www-ijpb.versailles.inra.fr/fr/bc/equipes/Meristeme2/systeme_ethanol_ind/sei_protocoles.html. Il a été choisi en raison que, sous les conditions du laboratoire « Facteurs de transcription et architecture », il semble induire moins d'effets secondaires dus à la toxicité de l'alcool. Après 48h à 4°C pour la synchronisation de la germination, des graines de chaque ligné amiR et le contrôle Col-0 ont été semées (2-4 graines par pot, 4 pots par ligné dans une barquette de 24 pots) et mis sous conditions de chambre de culture à 21°C pendant le jour et de 17°C la nuit, de 65% d'humidité et avec une périodicité de lumière de type « jours longs » (16h de lumière – 8h d'obscurité). Après 18 jours de croissance une feuille de chaque plante a été traitée avec une goutte de l'herbicide Basta pour déterminer les plantes résistantes et sensibles. 20 jours après semis l'induction de l'amiR par l'éthanol a été initiée en distribuant 5 tubes eppendorf de 1.5mL sur la barquette, puis 500 µL d'éthanol 80% ont été mis dans chaque tube. Les plantes résistantes ont été signalées et chaque barquette de 24 pots a été ensuite mise sous cloche. L'induction a été entretenue pendant 3 jours et pour cela l'éthanol a été rechangé chaque jour en retirant celle du jour précédent et ajoutant 500 µL d'éthanol 80% nouvelle. Finalement, les échantillons d'un pool de plantes résistantes et des contrôles ont été prélevés au troisième jour d'induction et les ARN ont été extraits.

7.5. Etude de la localisation subcellulaire des protéines

Dans ce travail de thèse, la localisation des protéines a été évaluée en utilisant une approche de fusion traductionnelle entre la protéine d'intérêt et une protéine fluorescente. Pour cela deux systèmes ont été utilisés : un système d'expression transitoire chez *N. benthamiana* et un système utilisant des lignées stables chez *A. thaliana*. Le signal de localisation est observé par microscopie confocale en utilisant le microscope confocal Leica SP2 AOBS (Leica Microsystems).

7.5.1. Expression transitoire chez *N. benthamiana*

Une colonie d'*Agrobacterium* contenant le plasmide d'intérêt est utilisée pour ensemer 5mL de LB agro liquide contenant les antibiotiques nécessaires. Cette préculture est incubée à 28°C pendant 48h sous agitation (150rpm) puis 500µL sont utilisées pour ensemer 5mL de milieu LB agro liquide contenant les antibiotiques nécessaires. La culture de bactéries est incubée à 28°C pendant 12 à 16h et sous agitation (150rpm). Les bactéries sont ensuite « culotées » par centrifugation pendant 10 min à 4°C et resuspendues dans un milieu d'agro-infiltration (10mM MES pH5,6, 10mM MgCl₂, 150µM d'acétosyringone) afin d'obtenir une densité optique (DO) à 600nm comprise entre 0,2 et 0,3. Cette solution est incubée une heure à température ambiante, puis utilisée pour infiltrer des feuilles de *N. benthamiana*. Pour cela, les agrobactéries sont injectées à l'aide d'une seringue de 1 mL sans aiguille sur la face inférieure des feuilles dans l'espace intercellulaire.

L'expression des protéines est observée après 72h d'infiltration et jusqu'à 120h sur des cellules protoplastisées. Pour obtenir des protoplastes, l'épiderme de la face inférieure de feuilles est retiré dans les régions agro-infiltrées puis les fragments de feuilles sont incubés dans un milieu de digestion enzymatique de la paroi cellulaire (Maceroenzyme R10 4g/L, Driselase 1g/L, Cellulase « Onusuka » 2g/L, Murashige et Skoog Salt Mix 4,4g/L MES 0,5g/L, Sucrose 20g/L, Mannitol 80g/L) pendant 2h à 28°C et à l'obscurité.

Pour confirmer la localisation mitochondriale, les protoplastes sont incubés avec 0,5 à 1µM de Mito-tracker green (marqueur de potentiel membranaire) pendant 10 à 15

min. Les signaux de fluorescence sont observés au microscope confocal de manière séquentielle : excitation à 488nm du Mito-tracker green suivie par une excitation à 543nm de la DsRed2. Les signaux d'émission sont détectés avec des fenêtres de 500 à 530nm, 570 à 620nm et 620 à 700nm pour le Mito-tracker, la protéine de fusion et l'auto-fluorescence de la chlorophylle respectivement.

7.5.2. Localisation des protéines dans les lignées stables d'*A. thaliana*.

Les descendants de transformants primaires exprimant des fusions PPR::DsRed2 sont semés *in vitro* sur milieu MS contenant ou non de l'herbicide Basta (en cas d'absence de sélection, les plantes transformées sont sélectionnées par PCR). 10 à 15 jours après semis, les plantules sont observées en microscopie confocale de la même façon que les protoplastes.

7.6. Test d'interaction protéine-protéine par complémentation biomoléculaire de fluorescence (BiFC) chez *A. thaliana*

Après clonage dans un vecteur d'entrée, les séquences codantes des PPR ont été introduites par recombinaison LR dans les vecteurs de la série pBiFC disponibles au laboratoire. La sélection des bactéries *E. coli* transformées est réalisée sur du milieu contenant 50mg/L de Spectinomycine, puis les bactéries sont testées par PCR et les plasmides de clones positifs sont purifiés. Finalement ces plasmides sont utilisés pour la transformation d'agrobacteries, lesquelles sont utilisées pour la transformation transitoire de plantules d'*Arabidopsis*.

Le protocole ci-dessous est basé sur celui décrit dans la publication de Marion *et al.* (2008). 4mL de milieu solide Duchefa sont ajoutés à chacun des puits d'une plaque à six puits (Saatec, PA500/38). Des graines de l'écotype Landsberg sont semées sur le milieu recouvert d'une grille en plastique de maille 1 mm. Les boîtes sont placées à 4°C pendant 48h puis incubées dans la chambre de culture *in vitro* d'*Arabidopsis*.

24h après mise en culture des semis d'Arabidopsis, les colonies d'agrobactéries contenant les plasmides pBiFC (pBiFC1 et pBiFC4) portant les inserts d'intérêt sont utilisées pour ensemer 5mL de milieu LB agro liquide contenant les antibiotiques nécessaires (spectinomycine/rifampycine). Les agrobactéries sont incubées pendant 16 à 20h et 1mL de cette préculture est utilisé pour inoculer 30mL de milieu LB agro liquide contenant les antibiotiques nécessaires. Les cultures sont incubées pendant 48h puis centrifugées 15min à 4000g et resuspendues dans 10mL de milieu Duchefa liquide contenant 200 μ M d'acétosyringone.

Les plantules sont transformées avec un mélange de 4 mL d'agrobactéries combinant les deux protéines du couple testé (2mL de culture contenant pBiFC1 et 2 mL contenant pBiFC4). La transformation est effectuée à l'aide d'une pompe à vide, la solution est éliminée et les plantules sont placées dans la chambre de culture pour une nouvelle phase de croissance. Trois jours après agro-infiltration, les signaux de fluorescence sont observés dans les cotylédons grâce à un microscope confocal. Le microscope est configuré en utilisant le laser 488nm comme source d'excitation. Les signaux d'émission sont détectés avec des fenêtres de 500 à 550nm et 600 à 700nm pour la protéine YFP reconstituée et l'auto-fluorescence de la chlorophylle, respectivement.

ANNEXE

Tableau 9. Protéines d'*Arabidopsis thaliana* connues pour être localisées à la fois dans les mitochondries et les plastes disponibles dans de publications.

AGI V10	Annotation	Localisation			Compartiment plastidial	Références
		Fluorescence	Proteomique	Autre		
At1g08130.1	DNA ligase	M; C; N	Ct			Xu <i>et al.</i> , 2013
At1g11870.1	Seryl-tRNA synthetase (SerRS)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g13900.1	Purple acid phosphatase 2 (AtPAP2)	M; C	C		membrane externe	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g21400.1	Thiamin diphosphate binding fold protein	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g22660.1	Polynucleotide adenylyltransferase family protein	M; C	M			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g30680.1	Toprim domain-containing protein, DNA helicase (Hel)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g31170.1	Cysteine-sulfinic acid reductase (Sulfiredoxin, AtSrx)	M; C		M; C		Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g48520.1	Glutamyl-tRNA(Gln) amidotransferase subunit B	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g49630.1	Presequence protease 2 Zinc metalloprotease (AtPrep2)	M; C	C; MP; n	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g50200.1	Alanyl-tRNA synthetase (AlaRS)	M; C; Ct	Ct	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g50840.1	DNA Polymerase gamma 2	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g63940.1	Monodehydroascorbate reductase 6 (MDHAR)	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g74260.1	Purine biosynthesis 4	M; C	M; C		stroma	Berthomé <i>et al.</i> , 2008
At1g74350.1	Intron Maturase ATnmAT4	M; C				Xu <i>et al.</i> , 2013
At1g74600.1	Organelle transcription processing 87 (OTP87)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At1g79230.1	Mercaptopyruvate sulfurtransferase 1 (MST1)	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At2g03050.1	mTERF, Soldat10, EMB93	M; C	C			Fuss <i>et al.</i> , 2013
At2g04842.1	Threonyl-tRNA synthetase (ThrRS-2)		C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At2g25840.1	Tryptophanyl-tRNA synthetase (TrpRS)	M; C	m; C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At2g30320.1	Pseudouridine synthase	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At2g31170.1	Cysteinyl-tRNA synthetase (CysRS)	M; C			stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At2g32230.1	Proteinaceous RNase P I		C	M; C	nucléotide	Gobert <i>et al.</i> , 2010
At2g35370.1	Glycine decarboxylase complex H	M; C	M; C; n	M; C		Baudisch <i>et al.</i> , 2013
At2g37250.1	Adenylate kinase	M; C	M; C			Xu <i>et al.</i> , 2013
At2g46110.1	Ketopentolate hydroxymethyltransferase (KPHMT1)	M; C	n	M; C		Baudisch <i>et al.</i> , 2013
At2g47510.1	Fumarase 1	M; C	M	M; C		Baudisch <i>et al.</i> , 2013
At3g02660.1	Tyrosyl-tRNA synthetase (TyrRS)	M; C		M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g05790.1	Lon protease 4	M; C	M; C	M		Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g10690.1	DNA GYRASE A	M; C	M; C			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g12370.1	Ribosomal protein L10 (RPL10)	M; C	M			Xu <i>et al.</i> , 2013
At3g13490.1	Lysyl-tRNA synthetase (LysRS), class II	M; C	C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g15000.1	Multiple organellar RNA editing factor 8 (MORF8); RIP1	M; C	M; C		nucléotide	Bentolila <i>et al.</i> , 2012
At3g15660.1	Glutaredoxin 4	M; C	M	M; C		Baudisch <i>et al.</i> , 2013
At3g16480.1	Mitochondrial processing peptidase α 2 (MPP α 2)	M; C	M	M; C		Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013

Les données de la colonne « Fluorescence » et « Autre » ont été collectés à partir des références citées dans la colonne correspondante. Les données de la colonne « Protéomique » ont été collectés des bases de données SUBA3 et PPDB. M: mitochondrie, C: Plastes, MP: Membrane Plasmique, Ct: Cytosol N: Noyau. Dans la colonne « Protéomique », une localisation en minuscule indique une localisation potentielle car une seule observation supporte cette localisation.

Tableau 9. (suite) Protéines d'*Arabidopsis thaliana* connues pour être localisées à la fois dans les mitochondries et les plastides disponibles dans de publications.

AGI V10	Annotation	Localisation			Compartiment plastidial	Références
		Fluorescence	Protéomique	Autre		
At3g19170.1	Presequence protease 1 Zinc metalloprotease (AtPrep1)	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g20540.1	DNA Polymerase gamma 1	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g23830.1	Glycine-rich RNA binding protein (GRP)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g24320.1	MutS Homolog1 (MSH1)	M; C			nucléotide	Xu <i>et al.</i> , 2011, Xu <i>et al.</i> , 2013
At3g25660.1	Glutamyl-tRNA amidotransferase subunit A (GATA)		M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g25740.1	Methionine aminopeptidase 1C	M; C	C			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g46100.1	Histidyl-tRNA synthetase 1 (HisRS)	M; C	M; C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g48110.1	Glycyl-tRNA synthetase (GlyRS)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g48250.1	PPR protein BIR6	M; C				Fuss <i>et al.</i> , 2013
At3g54660.1	Glutathione reductase 2	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g55400.1	Methionyl-tRNA synthetase MetRS	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At3g58140.1	Phenylalanyl-tRNA synthetase PheRS		M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g02930	Elongation factor Tu family	M; C	M; MP	M; C		Baudisch <i>et al.</i> , 2013
At4g03320.1	TIC 20-IV	M; C	C		membrane interne	Xu <i>et al.</i> , 2013
At4g08390.1	Stromal ascorbate peroxidase	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g10320.1	Isoleucine-tRNA synthetase (IleRS)	M; C	Ct			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g17300.1	Asparaginyl-tRNA synthetase (AsnRS)	M; C	M; C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g25200.1	Small heat shock protein 23.6 (AtHsp23.6)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g26500.1	Chloroplast sulfur acceptor E1 (AtSFE1)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g30490.1	AFG1-like ATPase family protein	M; C	M			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g30700.1	Mitochondrial Editing Factor 29	M; C				Sosso <i>et al.</i> , 2012
At4g31210.1	DNA topoisomerase (Top)	M; C	M; C		nucléotide	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g32400.1	Brittle 1	M; C			membrane interne	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g32915.1	Glutamyl-tRNA amidotransferase subunit C (GATC)		C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g33760.1	Aspartyl-tRNA synthetase (AspRS)		M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g36580.1	AAA-type ATPase family protein	M; C	M			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At4g37000.1	Accelerated cell death 2 (ACD2)	M; C	C		thylakoides	Pattanayak <i>et al.</i> , 2012
At4g37040.1	methionine aminopeptidase 1D (MAP1D)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g04140.1	Ferredoxin-dependent glutamate synthase (GLS1)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g05730.1	Anthranilate synthase α subunit 1 (ASA1)	M; C	m; C			Baudisch <i>et al.</i> , 2013
At5g06810.1	Mitochondrial transcription termination factor (MTERF)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g08710.1	Regulator of chromosome condensation (RCC1)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g08740.1	NAD(P)H dehydrogenase C1	M; C	C	M; C	plastoglobules	Carrie <i>et al.</i> , 2013
At5g14660.1	Peptide deformylase 1B (PDF1B)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g15700.1	RNA polymerase RPOTmp	M; C	M			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g16200.1	50 S ribosomal protein-related	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g16715.1	Valine-tRNA synthetase	M; C	C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g22800.1	Alanyl-tRNA synthetase (AlaRS-2), class IIc		C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013

Tableau 9. (suite) Protéines d'*Arabidopsis thaliana* connues pour être localisées à la fois dans les mitochondries et les plastes.

AGI V10	Annotation	Localisation				Références
		Fluorescence	Protéomique	Autre	Compartiment plastidial	
At5g23060.1	Calcium sensing receptor	M; C	m; C; mp		thylakoides	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g23300.1	Dihydroorotate dehydrogenase	M; C	M; mp	M; C		Baudisch <i>et al.</i> , 2013
At5g24850.1	Flavin adenine dinucleotide Cryptochrome 3 (Cry3)	M; C	C	M; C	nucléotide	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g26860.1	Lon protease 1 (Lon1)	M; C	M			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g26940.1	Defective in pollen organelle DNA degradation1 (DPD1)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g35630.1	Glutamine synthetase 2 (GLN2)	M; C	M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g38710.1	Methylenetetrahydrofolate reductase (MTHFR)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g40810.1	Cytochrome C1 protein	M; C	M	M; C		Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g44785.1	Organellar single-stranded DNA binding protein (OSB3)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g51440.1	Small heat shock protein 23.5 (AtHsp23.5)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g52520.1	Prolyl-tRNA synthetase (ProRS)		M; C	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g54770.1	Thiazole biosynthetic enzyme, chloroplast (ARA6)	M; C	C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g55200.1	Co-chaperone GrpE protein (GrpE)	M; C	M			Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g56940.1	Ribosomal protein S16 (RPS16)	M; C				Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g63980.1	SAL1	M; C	C; Ct	M; C	stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013
At5g64050.1	Glutamate tRNA synthetase (GluRS)	M; C	M; C		stroma	Carrie <i>et al.</i> , 2013, Xu <i>et al.</i> , 2013

Tableau 10. Protéines d'*Arabidopsis thaliana* probablement doublement localisées dans les mitochondries et les plastes.

AGI V10	Annotation	Localisation		
		Fluorescence	Protéomique	Compartiment plastidial
At1g11860.1	Glycine cleavage T-protein		M; C; mp	
At1g15390.1	Peptide deformylase 1A	M; C	M; C	
At1g48420.1	D-cysteine desulfhydrase	M; C; Ct	C; ct	
At1g52080.1	Actin binding protein	M; C		
At1g65630.1	DegP protease 3	M; C		
At1g74130.1	Rhomboid protease	M; C		
At1g79440.1	Aldehyde dehydrogenase 5F1		M; C	
At2g01140.1	Aldolase superfamily protein	C	M; C; mp	
At2g02510.1	NADH dehydrogenase (ubiquinone)	M; C; Prx	M	
At2g04030.1	Hsp90-5	C	M; C; mp	
At2g07698.1	ATPase, F1 complex, α subunit protein		M; C	
At2g19490.1	RecA DNA recombination family protein recA2	M; C	M	
At2g21170.1	triosephosphate isomerase-1 (TPI-1)	C; Ct	M; C; ct; mp	
At2g25140.1	Casein lytic proteinase B4		M; C	
At2g26080.1	Glycine decarboxylase P-protein 2		M; C	
At2g28000.1	Encodes chaperonin-60 α -1	C	M; C	stroma
At2g34630.1	Geranyl diphosphate synthase 1	M; C	M	
At2g35370.1	Glycine decarboxylase complex H		M; C; n	
At2g35605.1	SWIB/MDM2 domain superfamily protein	M; C	m	
At2g40300.1	Ferritin 4	M; C	C	
At3g01280.1	Voltage dependent anion channel 1	M	M; C; MP	
At3g02780.1	Isopentenyl pyrophosphate	M; C	M; C; ct	
At3g10270.1	DNA GYRASE B1		M; C	nucléoide
At3g13930.1	Dihydrolipoamide acetyltransferase		M; C; MP	
At3g47520.1	Malate dehydrogenase [NAD],	C	M; C	stroma
At3g57090.1	Tetratricopeptide repeat (TPR)-like protein	M; C; Prx	C	
At3g58610.1	Ketol-acid reductoisomerase		M; C; mp	
At3g61440.1	Cysteine synthase C1		M; C	
At3g62530.1	ARM repeat superfamily protein		M; C	
At4g00290.1	Mechanosensitive ion channel protein		M; C	
At4g02990.1	RUGOSA2	M; C	C	
At4g25130.1	Peptide met sulfoxide reductase 4		M; C	
At4g26970.1	Aconitase 2 (ACO2)	m; c	M; c; MP	
At4g33010.1	Glycine decarboxylase P-protein 1		M; C	
At4g34200.1	D-3-phosphoglycerate dehydrogenase (3-PGDH)	C	M; C; MP	
At4g34620.1	Small subunit ribosomal protein 16		M; C	
At4g39460.1	S-adenosylmethionine carrier 1(SAMC1)	M; C	m; C; MP	
At5g04130.1	DNA GYRASE B2		M; C	nucléoide
At5g13420.1	Aldolase-type TIM barrel family protein		M; C; mp	stroma
At5g13490.1	ADP/ATP carrier 2		M; C; MP; n	
At5g14590.1	Isocitrate dehydrogenase family protein		M; C	
At5g14780.1	Formate dehydrogenase		M; C; mp	
At5g20720.1	Chaperonin 20 (Cpn20)	C; Prx	M; C	
At5g39410.1	Saccharopine dehydrogenase		M; C; MP	
At5g36950.1	DegP protease 10	M; C		
At5g53170.1	FTSH protease 11		M; C	
At5g58270.1	ABC transporter family protein		M; C; MP	
At5g67500.1	Voltage dependent ion channel VDAC2		M; C; mp	

Ces données sont issues de l'analyse de protéomes mitochondriaux et plastidiaux des plantes. Dans ces protéomes une recherche des données expérimentales de double localisation a été mise en œuvre sur les bases de données SUBA3 et PPDB en omettant les protéines de la famille PPR. M: Mitochondrie, C: Plastes, MP: Membrane Plasmique, Ct: Cytosol N: Noyau, Prx: Peroxisome. Dans la colonne "protéomique", une localisation en minuscule indique une localisation potentielle car une seule observation supporte cette localisation.

Tableau 11. Protéines d'*Arabidopsis thaliana* pouvant être doublement localisées dans les mitochondries et les plastes.

AGI V10	Annotation	Localisation		
		Fluorescence	Protéomique	Compartiment plastidial
At1g03860.1	Prohibitin 2	M	M; c; mp	
At1g07040.1	Unknown protein		m; C	
At1g07510.1	FtsH protease 10		M; c; MP	
At1g08640.1	Chloroplast J-like domain 1 CJD1 (ARC6 interactor)		m; C	
At1g09130.1	ATP-dependent caseinolytic (Clp) protease	C	m; C	
At1g10760.1	Pyruvate phosphate dikinase		m; C	
At1g13280.1	Allene oxide cyclase 4		m; C	stroma
At1g14810.1	Aspartate-semialdehyde dehydrogenase		m; c	stroma
At1g15060.1	Unknown protein α /beta hydrolase		m; c	stroma
At1g25375.1	Metallo-hydrolase/oxidoreductase superfamily protein		m; C	
At1g26340.1	Cytochrome B5 isoform A		m; c; prx; mp	
At1g29910.1	Chlorophyll A/B binding protein 3	C	m; C	
At1g32470.1	Glycine cleavage system H protein		M; c	
At1g50940.1	Electron transfer flavoprotein α (ETF- α)		M; c	
At1g56190.1	Phosphoglycerate kinase family protein	C	m; C	
At1g62750.1	Translation elongation factor EFG/EF2 protein	C	m; C	
At1g64190.1	6-phosphogluconate dehydrogenase family protein		m; C; mp; ct	
At1g68680.1	Unknown protein		M; c	
At1g73990.1	Signal peptide peptidase (SppA).		m; C; mp; n	
At2g13560.1	NAD-dependent malic enzyme 1		M; c	
At2g15620.1	Nitrite reductase 1		m; C; mp	
At2g24120.1	RpoT-3 -Nuclear-encoded RNA polymerase (NEP)	C	m; C	
At2g25830.1	YebC-related		m; C	
At2g28900.1	Amino acid transporter (OEP16:HP15)	C	m; C; mp; n	outer membrane
At2g35010.1	Thioredoxin O1	C	M	
At2g35720.1	DNAJ heat shock N-terminal domain-containing protein		m; c; mp	
At2g37500.1	Arginine biosynthesis protein ArgJ family		m; C	
At2g38550.1	Transmembrane proteins 14C		m; C; mp	inner membrane
At2g43750.1	O-acetylSérine (thiol) lyase B	C	m; C	
At2g44640.1	Unknown protein		m; C; mp	
At3g01480.1	Tlp-40; CYCLOPHILIN 38 - CYP38		m; C	
At3g03590.1	SWIB/MDM2 domain superfamily protein	C	M	
At3g06310.1	Cox19-like CHCH family protein		M; c	
At3g10920.1	Manganese superoxide dismutase 1 MnSOD1		M; c	
At3g11670.1	Digalactosyl diacylglycerol (DGDG) synthase 1 (DGD1)		M; c	
At3g15020.1	Malate dehydrogenase family		M; c	
At3g15640.1	Cytochrome c oxidase subunit Vb		M; c	
At3g16540.1	DegP protease 11	M	c	
At3g19480.1	D-3-phosphoglycerate dehydrogenase		m; C	
At3g20390.1	Endoribonuclease L-PSP		m; C	
At3g27740.1	Carbamoyl phosphate synthetase small subunit A		m; C	stroma
At3g47450.1	Nucleoside triphosphate hydrolases protein	M	C	stroma
At3g48870.1	ClpC2; HSP93-III; similar to ClpC1	C	m; C	stroma
At3g52230.1	Unknown protein		m; C	outer membrane
At3g55010.1	PUR5	C	m; C	
At3g55250.1	Unknown protein		m; C	
At3g57050.1	Cystathionine beta-lyase		m; C	stroma
At3g59760.1	O-acetylSérine (thiol) lyase isoform C		M; c	
At4g00570.1	NAD-dependent malic enzyme 2		M; c; mp	

Ces données sont issues de l'analyse de protéomes mitochondriaux et plastidiaux des plantes. Dans ces protéomes une recherche des données expérimentales de double localisation a été mise en œuvre sur les bases de données SUBA3 et PPDB en omettant les protéines de la famille PPR. M: Mitochondrie, C: Plastes, MP: Membrane Plasmique, Ct: Cytosol N: Noyau, Prx: Peroxisome. Dans la colonne "protéomique", une localisation en minuscule indique une localisation potentielle car une seule observation supporte cette localisation.

Tableau 11. (suite) Protéines d'*Arabidopsis thaliana* pouvant être doublement localisées dans les mitochondries et les plastes.

AGI V10	Annotation	Localisation		
		Fluorescence	Protéomique	Compartment plastidial
At4g08870.1	Arginase/deacetylase superfamily protein		M; c	
At4g10750.1	Phosphoenolpyruvate carboxylase family protein		m; C; n	
At4g17740.1	Peptidase S41 family protein	C	m; C	
At4g23620.1	Ribosomal protein L25/Gln-tRNA synthetase		m; c	
At4g24280.1	Chloroplast heat shock protein 70-1 (cpHSP70-1)		M; C; MP; N	
At4g24830.1	Arginosuccinate synthase family		m; C	
At4g25100.1	Fe superoxide dismutase 1 (FSD1)	C	M; C; MP; n	
At4g27070.1	Tryptophan synthase beta-subunit 2 (TBS2)		m; c	
At4g28390.1	ADP/ATP carrier 3		M; c; mp	
At4g29130.1	Hexokinase 1	M	M; c; mp	
At4g29840.1	Pyridoxal-5'-phosphate-dependent enzyme family protein		m; C; mp; ct	
At4g31990.1	Aspartate aminotransferase 5	C	m; C; mp; ct	
At4g33030.1	Sulfoquinovosyldiacylglycerol 1		m; C	
At4g39980.1	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1		m; C; mp	
At5g01600.1	Ferretin 1		m; C; mp	stroma
At5g03455.1	Rhodanese protein	M	c	
At5g05570.1	Transducin family protein / WD-40 repeat family protein	M	c; mp	
At5g09650.1	Pyrophosphorylase 6		m; C; mp	stroma
At5g12040.1	Apolipoprotein N-acyltransferase family protein		m; C; ct	stroma
At5g13630.1	GUN5	C	m; C	
At5g14220.1	PPO2; MEE61		m; C; mp	
At5g15450.1	Casein lytic proteinase B3 (ClpB3) Hsp101 homologue		m; C	stroma
At5g16290.1	Valine-tolerant 1		m; c; ct	
At5g16390.1	Biotin carboxyl carrier protein BCCP-1		m; C	
At5g42390.1	Insulinase (Peptidase family M16) family protein		m; C	stroma
At5g42650.1	Allene oxide synthase	C	m; C; mp	thylakoid
At5g43780.1	Pseudouridine synthase; ATPsulfurylase-4 (APS4)		m; C	
At5g46800.1	Mitochondrial substrate carrier family protein		M; c; mp	
At5g50920.1	ClpC1; HSP93-V		m; C	
At5g55610.1	Unknown protein		M; c	
At5g56500.1	TCP-1/cpn60 chaperonin family protein		m; C	
At5g58330.1	Malate dehydrogenase [NADP]		m; C	
At5g63890.1	Histidinol dehydrogenase (ATHDH)		m; C	
At5g65780.1	Branched-chain amino acid aminotransferase 5 (BCAT5)	C	m; C	

Tableau 12. Liste des protéines PPR impliquées dans l'édition de l'ARN des organites chez *A. thaliana*

AGI V10	Annotation	Localisation	Références
<i>A. thaliana</i>			
At1g05750	CLB19	C	Chateigner-Boutin et al. 2008
At1g08070	OTP82	C	Okuda et al. 2010, Okuda et al. 2012
At1g11290	CRR22	C	Okuda et al. 2009
At1g15510	AtECB2	C	Yu et al. 2009, Tseng et al. 2010
At1g47580	DYW1	C	Boussardon et al. 2012
At1g59720	CRR28	C	Okuda et al. 2009
At2g02980	OTP85	C	Hammani et al. 2009
At2g29760	OTP81	C	Hammani et al. 2009
At2g45350	CRR4	C	Kotera et al. 2005
At3g22690	YS1	C	Zhou et al. 2008
At3g57430	OTP84	C	Hammani et al. 2009
At3g63370	OTP86	C	Hammani et al. 2009
At4g18750	DOT4	C	Petricka et al. 2010, Hayes et al 2013
At4g37380	ELI	C	Hayes et al. 2013
At5g13270	RARE1	C	Robbins et al. 2009
At5g55740	CRR21	C	Okuda et al. 2007
At5g59200	OTP80	C	Hammani et al. 2009
At5g66520	CREF7	C	Yagi et al. 2013
At1g06140	MEF3	M	Verbistkiy et al. 2012
At1g34160	OGR1	M	Kim et al. 2009
At1g62260	MEF9	M	Takenaka et al. 2010
At1g64310	OTP71	M	Chateigner-Boutin et al. 2013
At2g03880	REME1	M	Bentolila et al. 2010
At2g13600	SLO2	M	Zhu et al. 2012
At2g20540	MEF21	M	Takenaka et al. 2010
At2g22410	SLO1	M	Sung et al. 2010
At2g25580	MEF8	M	Vervitskiy et al. 2012
At3g05240	MEF19	M	Takenaka et al. 2010
At3g09040	MEF12	M	Hartel et al. 2013
At3g11460	MEF10	M	Hartel et al. 2013
At3g12770	MEF22	M	Takenaka et al. 2010
At3g13880	OTP72	M	Chateigner-Boutin et al. 2013
At3g25060	MEF25	M	Arenas-M et al. 2013
At3g26782	MEF14	M	Verbitskiy et al. 2011
At4g15720	REME2	M	Bentolila et al. 2013
At4g32450	MEF8S	M	Vervitskiy et al. 2012
At5g08490	SLG1	M	Yuan & Liu 2012
At5g09950	MEF7	M	Zehrmann et al. 2012
At5g19020	MEF18	M	Takenaka et al. 2010
At1g25360	OTP90	M/c	Cette Thèse
At1g74600	OTP87	M/c	Hammani et al 2011
At1g80270	PPR 596	M/c	Doniwa et al. 200
At2g44880	AHG11	M/c	Murayama et al. 2012
At3g18970	MEF20	M/c	Takenaka et al. 2010
At4g14850	MEF11	M/c	Verbitskiy et al. 2010
At4g30700	MEF29	M/c	Sosso et al. 2012
At3g14330	CREF3	m/C	Yagi et al. 2013
At5g48910	LPA66	m/C	Cai et al. 2009
At2g15690	DYW2	M/C	Cette Thèse

* Cette thèse

Tableau 12. (Suite) liste des protéines PPR impliquées dans l'édition de l'ARN des organites chez d'autres plantes terrestres.

AGI V10	Annotation	Localisation	Références
<i>Z. mays</i>			
	EMP5	M	Liu et al. 2013
	PPR2263	M/c	Sosso et al. 2012
<i>O. sativa</i>			
	OGR1	M	Kim et al. 2009
	MPR25	M/c	Toda et al. 2012
	PPR4	C	Asano et al. 2013
<i>P. patens</i>			
	PpPPR_56	M	Ohtani et al. 2010
	PpPPR_65	M	Schallenberg-Rudinger et al. 2013
	PpPPR_71	M	Tasaki et al. 2010
	PpPPR_77	M	Ohtani et al. 2010
	PpPPR_78	M	Rudinger et al. 2011
	PpPPR_79	M	Uchida et al. 2011
	PpPPR_91	M	Ohtani et al. 2010
	PpPPR_98	M	Schallenberg-Rudinger et al. 2013

Tableau 13. Liste de microRNA artificiels obtenus pour les gènes *DYW2* et *OTP100*

Séquence amiR	Hybridization	Gène ciblé
TACTGAGGCCTGTGACCACCG	-52,28	At2g15690
TCGTATTGAGGCCTCTGACCA	-48,73	At2g15690
TCGTACTGAGGCCTGTGACCA	-50,94	At2g15690
TGTATTGAGGCCTCTGACCAC	-48,69	At2g15690
TGTACTGAGGCCTGTGACCAC	-50,9	At2g15690
TGACCACCGTACTGAGGCCTG	-52,34	At2g15690
TTGACCACCGTACTGAGGCCT	-50,5	At2g15690
TGTCCACCGTATTGAGGCCTC	-50,27	At2g15690
TTGACCACCGTATGGAGGCCT	-50,64	At2g15690
TTGGAGGCCTTTGACCCCGT	-52,33	At2g15690
TGGTCCACCGTATTGAGGCCT	-50,67	At2g15690

Les amiR surlignées correspondent aux amiR sélectionnées.

Tableau 13. (suite) microRNA artificiels obtenus pour les gènes *DYW2* et *OTP100*.

Séquence amiR	Hybridization	Gène ciblé
TAAACGGTTATACACAGCCAG	-42,25	AT3G49240
TAAGGTGGTTTAAAAAGGCGC	-41,34	AT3G49240
TATGAAATACCCCTTCAGCAA	-41,68	AT3G49240
TAATTTACAGCCAAATGGCTC	-40,62	AT3G49240
TTATGTTAATACTTTGAGCCC	-38,46	AT3G49240
TATGTTAATACTTTGCGCCTT	-38,01	AT3G49240
TTTTACAGCCAAATGCCGCGG	-45,48	AT3G49240
TGTGGTTTAAAAAGGCGGCTC	-43,98	AT3G49240
TTATGTTAATACTTTGCGCCT	-38,04	AT3G49240
TCAAACGGTTATACACGGCCA	-44,66	AT3G49240
TTGGTCTTAGATTAGACTCAG	-41,39	AT3G49240
TATTGCCCATAAACGACACCG	-43,96	AT3G49240
TAAATAAGATCAGCCGGACGG	-44,44	AT3G49240
TTATGTGTACTCATCACGCAT	-41,49	AT3G49240
TAACAACCTCGTTGTCACGCAG	-43,67	AT3G49240
TAAACGGTTATACACACCCAA	-40,45	AT3G49240
TGATATTAGGTGCACTACCAG	-43,87	AT3G49240
TAATCATAGCTTATTGCCCCA	-42,05	AT3G49240
TTTTTGTAATACGCAGCGCCT	-41,67	AT3G49240
TAATTATAAGCCATCGCACTC	-40,75	AT3G49240
TAATTATAAGCCATACCACTC	-39,16	AT3G49240
TTTAGCTTGTCGACGCTGCGA	-46,34	AT3G49240
TGTGACGAGTGTACAGCGCAG	-49,1	AT3G49240
TGTGACGAGTGTACAGACCAG	-47,72	AT3G49240
TAACAACCTCGTTGTCACACAG	-42,07	AT3G49240
TATCATAGCTTATTGTGCGAA	-39	AT3G49240
TAATCATAGCTTATTGTCCCG	-40,59	AT3G49240
TTATGTTAATACTTTGAGCCT	-36,68	AT3G49240
TAATCGTTGTCACGCAGCTG	-46,28	AT3G49240
TAAATAAGATCAGCAGGGCGG	-45,06	AT3G49240
TCAAACGGTTATACACAGCCT	-43,02	AT3G49240
TGTGGTTTAAAAAGGCGCCTT	-41,79	AT3G49240
TCCTACCGTGTCATCATGCAG	-45,37	AT3G49240
TAAACGGTTATACACAGCCAA	-40,47	AT3G49240
TTGGTCTTAGATTAGACCCAA	-41,79	AT3G49240
TATGTTAATACTTTGACCCTT	-36,63	AT3G49240
TTGACATGTATCGGACCGCGA	-47,16	AT3G49240
TACGAGTGTACAGAGCACCTT	-45,88	AT3G49240
TTTTACAGCCAAATGTCTCGG	-41,92	AT3G49240
TATGAAATACCCCTTCATCCA	-42,24	AT3G49240

Tableau 14. Séquences des amorces utilisées pour les approches d'hémi-complémentation, complémentation embryon spécifique et des amiR.

Nom	Séquence (5' à 3')
At1g25360_49aa_start	GGAGATAGAACCATGCATATCCTGAACCGTTTGAT
At2g15690_55aa_start	GGAGATAGAACCATGGCAAACGATTACCACCAG
At3g49240_57aa_start	GGAGATAGAACCCCGGTTAATTCATTCAACAGAT
FDH_F_XbaI	GACTCTAGAATGGCGATGAGACAAGC
FDH_R_XbaI	TGTTCTAGAAAGGAAGTTAGGGTTCTTGG
FDH_R_center	TTGCTATCACCAGAAGATGC
PLec2_Fw	CAGAAGCTTTGAATGGCTATTAATGGTGTTTA
PLec2_Rv	TGTTCTAGATTTTCCCGGAGAGAGAGAG
PLec1_Fw	CAGAAGCTTACTTTTAGCAGAGTTCCATGAGATTTT
PLec1_Rv	TGTTCTAGATGTTTCTCTGCCGCTTTTTTTTTTTT
PLec1Like_Fw	CAGAAGCTTGAAGTGGAGAGAATACGCCG
PLec1Like_Rv	TGTTCTAGAATCTTTCTTGTCTAAATACACGAC
Tag_HA_R	AGCGTAATCTGGAACGTCATA
pBARTAlcR_seqF	CTGCAGGCGTTCAAACATT
pBARTAlcR_NosF	ACAAAATATAGCGCGCAAAC
pBJ36AlcA_OCSRv	TTTAGGTTTGACCGGTTCTG
pBJ36AlcA_AlcAF	TAGAGCAGAGACGGAGCACT
pRS300_mirRv	ACCGCATCATTCAATTCATTT
mir_Aprimer	CTGCAAGGCGATTAAAGTTGGGTAAC
mir_Bprimer	GCGGATAACAATTCACACAGGAAACAG
I_miRs_otp100_1a	GATACGAGTGTACAGAGCACCTTCTCTCTTTTGTATTCCA
II_miRa_otp100_1a	AGAAGGTGCTCTGTACACTCGTATCAAAGAGAATCAATGA
III_miRs_otp100_1a	AGAAAGTGCTCTGTAGACTCGTTTCACAGGTCGTGATATG
IV_miRa_otp100_1a	GAAACGAGTCTACAGAGCACTTTCTACATATATATTCCTA
I_miRs_otp100_2a	GATTATGTTAATACTTTGAGCCTCTCTCTTTTGTATTCCA
II_miRa_otp100_2a	AGAGGCTCAAAGTATTAACATAATCAAAGAGAATCAATGA
III_miRs_otp100_2a	AGAGACTCAAAGTATAAACATATTACAGGTCGTGATATG
IV_miRa_otp100_2a	GAATATGTTTATACTTTGAGTCTCTACATATATATTCCTA

Tableau 15. Séquences des amorces pour clonage BP des séquences codantes de gènes utilisées dans cette thèse.

Nom	Séquence (5' à 3')
At1g06145_start	GGAGATAGAACCATGAATGCTTTCGCCAATGTGC
At1g25360_start	GGAGATAGAACCATGCAACCGAATCCAGACCTTG
At2g37230_start	GGAGATAGAACCATGGCTTTCATTTCAGATCTAAAC
At3g42630_start	GGAGATAGAACCATGTTGTCGCTCAATCTATCTTTG
At2g15690_start	GGAGATAGAACCATGTCTTCTCTAATGGCCATTC
At1g29710_start	GGAGATAGAACCATGGTGCGATTGTGGTGT
At4g32450_start	GGAGATAGAACCATGATATACACGTTGACGAGAGG
At2g25580_start	GGAGATAGAACCATGTATACGAAGTTATCATTGTTCCC
At1g06145_end	TCCACCTCCGGATCCATATACATTCTCAGTTTCCTGCAC
At1g25360_end	TCCACCTCCGGATCCCCAGAAGTTACCACAAGAACAC
At2g37230_end	TCCACCTCCGGATCCTAGAGAAACATTGTTTTGTTTCTTGATTC
At3g42630_end	TCCACCTCCGGATCCGTAATTCCAGAAAATCTGGTCTCTC
At2g15690_end	TCCACCTCCGGATCCCCAGTAATCCCCGCAAGAAC
At1g29710_end	TCCACCTCCGGATCCCCAAAGATTGTTGCACCGG
At4g32450_end	TCCACCTCCGGATCCCCAATACTCTCTACAGGAGCAAA
At2g25580_end	TCCACCTCCGGATCCCCAATAATCTTTACAGGTGCAAGC
At3g49240_end	TCCACCTCCGGATCCGCAGGACGGTGGATCCTGA
At1g06145_stop	TCCACCTCCGGATCAATATACATTCTCAGTTTCCTGCAC
At1g25360_stop	TCCACCTCCGGATCACCAGAAGTTACCACAAGAACAC
At2g37230_stop	TCCACCTCCGGATCATAGAGAAACATTGTTTTGTTTCTTGATTC
At3g42630_stop	TCCACCTCCGGATCAGTAATTCCAGAAAATCTGGTCTCTC
At2g15690_stop	TCCACCTCCGGATCACCAGTAATCCCCGCAAGAAC
At1g29710_stop	TCCACCTCCGGATCACCAAGATTGTTGCACCGG
At4g32450_stop	TCCACCTCCGGATCACCAATACTCTCTACAGGAGCAAA
At2g25580_stop	TCCACCTCCGGATCACCAATAATCTTTACAGGTGCAAGC
At3g49240_stop	TCCACCTCCGGATCAGCAGGACGGTGGATCCTGA
At1g25360_gene_F	AAAGCAGGCTTCAATTGGGCTGGAGTTTAAAT
At1g25360_gene_R	GAAAGCTGGGTCAAAAAGCCTTACAATTGTTGC
At1g25360_Start_Prom	GGAGATAGAACCCTGGCAAAAAAGGCAAAATGCAGC
At1g25360_End_Prom	TCCACCTCCGGATCCTTCATGTTTATGGTTTCGGTTCG
At2g15690_Start_Prom	GGAGATAGAACCCTGGAGAAAGCCATATCAGTT
At2g15690_End_Prom	TCCACCTCCGGATCCTGGCCATTAGAGAAGACATT
At3g49240_Start_Prom	GGAGATAGAACCCTTAGTCACAGAAGCCGAAGT
At3g49240_End_Prom	TCCACCTCCGGATCCGAAATCGACATGGTTGAAAT

Tableau 16. Séquences des amorces pour évaluer l'édition de 34 sites majeurs du chloroplaste

Nom	Séquence (5' à 3')
matK_F	AGCGGCGTATCCTTTGTTGC
matK_R	CGTTACCGGGTAAAAGATGC
atpF_F	TTAATAGTACGTAAATGTAACTC
atpF_R	AGAAGAGGAGATTATATGAAA
rpoC1_F	GTCCACGGATTCCATTATCAA
rpoC1_R	TTGGACCTATAAAGAGTGGAATTTG
rpoB_A_F	CAAGTACTTCTTGGCGAGCTCTA
rpoB_A_R	CCTTTATACTGACACTGAGAAGATCG
rpoB_B_F	GTCTCCAATTAATATTTTCGGCG
rpoB_B_R	GAAAACCAGTAGGAATATGC
psbZ_F	AGAACATAGCCCTATGAGTTAATACGA
psbZ_R	GATAAGAGAATTAAGGATACCCACCA
rps14_F	AAAGTTGCGGAGACAGGATTT
rps14_R	TTATAGGGAGAAGAAGAGGC
accD_A_F	TTCATTTGTAGTGAAAGCGG
accD_A_R	AGAGGTAAACATTGATTGGTAGCAT
accD_B_F	TATGGGATCCGTAGTAGGCG
accD_B_R	AGAATCTGATCTAACAACAGGGAA
psbEF_F	CGTTGGATGAACTGCATTGC
psbEF_R	ACAGGAGAACGTTCTTTTGC
petL_F	GCCAAAGACGATGAATTGAGTT
petL_R	GAGAGGTACTATCTATTCCTAGCTCTTAATC
rps12_F	CTTGTAACAATTCACATTCTTTGGC
rps12_R	CAAGACAGCCAATCCGAAAC
clpP_F	TGAACCGCTACAAGATCAAC
clpP_R	GTAATGATCCATCAACCCGC
rpoA_F	GGAGTTAAACTTCCATTTGTCCA
rpoA_R	GCGATGCGAAGAGCTTTACT
rpl23_F	TCCAATTACGTAAATCAAATAGTTCAA
rpl23_R	AATTCCTACTGGATGCACGC
ndhB_A_F	AGGTCAGATTTTCGCCTATTCC
ndhB_A_R	TCATTGCTATTACTCAAACAAGCA
ndhB_B_F	TCCGATTTGACCTATGGACG
ndhB_B_R	GTGCTCCTACGGAACCAAGG
ndhB_C_F	TCTCAGAGATAGATAGAGAGGTAGGAATTT
ndhB_C_R	CATTGAATGTACAGAAATGGCTATAA
ndhB_D_F	ATTTCTTGAAGCTCAATCTCTCCCCGGAT
ndhB_D_R	TCATGATCTGGCATGTACAG
ndhF_F	GCATTCGCTGCAATAGGTGC
ndhF_R	AAAACCTTCGCCGCATGTGG
ndhD_A_F	CAAGCCTAATTCTATCATAACTCG
ndhD_A_R	ACCTGGTGTATCTTGTC
ndhD_B_F	TCGATCCATTTATAATCTTCGGA
ndhD_B_R	GACTGTGTTGGTTGTAAGAGATGTG
ndhG_F	AGAATTATTGAAATGAGTTC
ndhG_R	TATAATGGATTTGCCTGGAC

Tableau 17. Lignées T1 d'hémi-complémentation des mutants *dyw2-1* et *dyw2-2*.

Code	Génotype		Insertion
	<i>dyw2-1</i>	<i>dyw2-2</i>	
P100	Wt		P35S::FDHTp:HA:DYW2
P101	Wt		P35S::RBC1ATp:HA:DYW2
P102	Wt		P35S::RBC1ATp:HA:DYW2
P103	HT		P35S::RBC1ATp:HA:DYW2
P104	HT		P35S::FDHTp:HA:DYW2
P105	HT		P35S::RBC1ATp:HA:DYW2
P106	Wt		P35S::RBC1ATp:HA:DYW2
P107	HT		P35S::RBC1ATp:HA:DYW2
P108	Wt		P35S::RBC1ATp:HA:DYW2
P109	Wt		P35S::RBC1ATp:HA:DYW2
P110	Wt		P35S::RBC1ATp:HA:DYW2
P111	Wt		P35S::FDHTp:HA:DYW2
P112	Wt		P35S::FDHTp:HA:DYW2*
P113	Wt		P35S::FDHTp:HA:DYW2*
P114	HT		P35S::FDHTp:HA:DYW2
P115		HT	P35S::FDHTp:HA:DYW2*
P116	HT		P35S::FDHTp:HA:DYW2
P117	Wt		P35S::RBC1ATp:HA:DYW2*
P118	HT		P35S::RBC1ATp:HA:DYW2
P119	Wt		P35S::RBC1ATp:HA:DYW2
P120	Wt		P35S::FDHTp:HA:DYW2
P121	Wt		P35S::FDHTp:HA:DYW2
P122		HT	P35S::RBC1ATp:HA:DYW2
P123	Wt		P35S::RBC1ATp:HA:DYW2*
P124	HT		P35S::RBC1ATp:HA:DYW2
P125	HT		P35S::RBC1ATp:HA:DYW2
P126	Wt		P35S::RBC1ATp:HA:DYW2
P127	Wt		P35S::RBC1ATp:HA:DYW2*
P128		?	P35S::RBC1ATp:HA:DYW2
P129	Wt		P35S::RBC1ATp:HA:DYW2*
P130	Wt		P35S::RBC1ATp:HA:DYW2*
P131	Wt		P35S::RBC1ATp:HA:DYW2*
P132	Wt		P35S::RBC1ATp:HA:DYW2*
P133	HT		P35S::RBC1ATp:HA:DYW2
P134	HT		P35S::RBC1ATp:HA:DYW2
P135	Wt		P35S::FDHTp:HA:DYW2*
P137	Wt		P35S::RBC1ATp:HA:DYW2
P138	HT		P35S::FDHTp:HA:DYW2
P139	HT		P35S::RBC1ATp:HA:DYW2
P140	Wt		P35S::RBC1ATp:HA:DYW2
P141	Wt		P35S::FDHTp:HA:DYW2
P161	HT		P35S::FDHTp:HA:DYW2
P162	HT		P35S::FDHTp:HA:DYW2
P163	HT		P35S::FDHTp:HA:DYW2
P164	HT		P35S::FDHTp:HA:DYW2

* Insertion à confirmer

Tableau 18. Lignées T1 d'hémi-complémentation des mutants *otp100-1* et *otp100-2*

Code	Génotype		Insertion
	<i>otp100-1</i>	<i>otp100-2</i>	
P201		?	P35S::FDHTp:HA:OTP100
P202	Wt		P35S::FDHTp:HA:OTP100
P203	Ht		P35S::FDHTp:HA:OTP100
P204	Wt		P35S::FDHTp:HA:OTP100
P205	Ht		P35S::FDHTp:HA:OTP100
P206	Wt		P35S::FDHTp:HA:OTP100
P207		Ht	P35S::FDHTp:HA:OTP100
P208		Ho?	P35S::FDHTp:HA:OTP100
P209		Ht	P35S::FDHTp:HA:OTP100
P210		Ht	P35S::FDHTp:HA:OTP100
P211		Ht	P35S::FDHTp:HA:OTP100
P212		Ho?	P35S::FDHTp:HA:OTP100
P213		Wt	P35S::FDHTp:HA:OTP100
P214		Ho?	P35S::FDHTp:HA:OTP100
P215		Ho?	P35S::FDHTp:HA:OTP100
P216		Ht	P35S::RBC1ATp:HA:OTP100
P217		Ho?	P35S::RBC1ATp:HA:OTP100
P218		Ho?	P35S::RBC1ATp:HA:OTP100*
P219		Ht	P35S::RBC1ATp:HA:OTP100*
P220		Ht	P35S::RBC1ATp:HA:OTP100*
P221		Ht	P35S::RBC1ATp:HA:OTP100*
P222		Ht	P35S::RBC1ATp:HA:OTP100*
P223		Ht	P35S::RBC1ATp:HA:OTP100*
P224		Ht	P35S::RBC1ATp:HA:OTP100*
P225		Ht	P35S::RBC1ATp:HA:OTP100*
P226		Ht	P35S::RBC1ATp:HA:OTP100
P227		Ht	P35S::RBC1ATp:HA:OTP100
P228		Ht	P35S::RBC1ATp:HA:OTP100
P229		Ht	P35S::RBC1ATp:HA:OTP100
P230		Ht	P35S::RBC1ATp:HA:OTP100
P231		Ht	P35S::RBC1ATp:HA:OTP100
P232		Ht	P35S::RBC1ATp:HA:OTP100
P233		Ht	P35S::RBC1ATp:HA:OTP100
P234		Ht	P35S::RBC1ATp:HA:OTP100*

* Insertion à confirmer

Tableau 19. Liste de sites d'édition prédits pour être les cibles de la protéine OTP91 utilisant le code de reconnaissance des PPR (Données fournis par M. Takenaka)

AGI	Site prédit	Score	p-value
AT1G06145	clpP_559	9,46581	0,000226
AT1G06145	ccb256_521	8,99105	0,000312
AT1G06145	ccb206_467	6,99707	0,00105
AT1G06145	nad4L_86	5,7627	0,00198
AT1G06145	ccb203_208	5,09803	0,00272
AT1G06145	ccb203_176	5,00308	0,00285
AT1G06145	ccb206_28	4,81318	0,0031
AT1G06145	nad2_1279	4,81318	0,0031
AT1G06145	nad4_74	4,33842	0,00382
AT1G06145	rps12_221	4,33842	0,00382
AT1G06145	nad2_1433	4,14852	0,00415
AT1G06145	ccb382_289	2,81919	0,0071
AT1G06145	cob_325	2,72424	0,00736
AT1G06145	nad2_1309	2,43939	0,00819
AT1G06145	nad1_928	2,15453	0,00909
AT1G06145	ndhB_1251	2,15453	0,00909
AT1G06145	ccb203_391	2,05958	0,0094

Tableau 20. Liste de vecteurs de clonage utilisées dans cette thèse.

Nom	Résistance	Carte
pGEMT-easy	Cb	
pDONR207	Genta	
pGWB1	Kan, Hy, CmR	(--R1-CmR-ccdB-R2--).
pGWB2	Kan, Hy, CmR	(--P35S-R1-CmR-ccdB-R2--).
pGWB13	Kan, Hy, CmR	(--P35S-R1-CmR-ccdB-R2-3xHA--)
pGWB15	Kan, Hy, CmR	(--P35S-3xHA-R1-CmR-ccdB-R2--)

Tableau 21. Vecteurs de destination pGWB15 et pGWB13 remodifiées pour permettre l'hémicomplémentation et la complémentation embryon spécifique à l'aide des promoteurs LEC

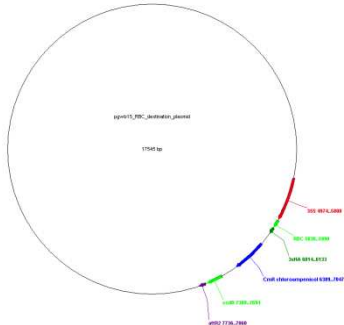
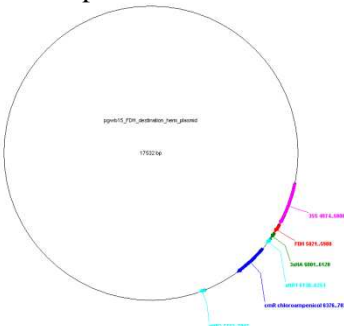
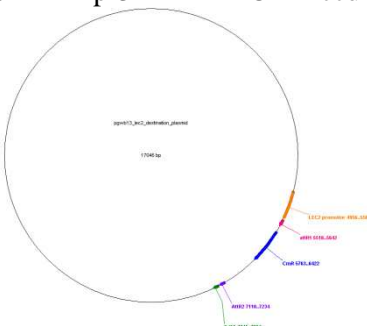
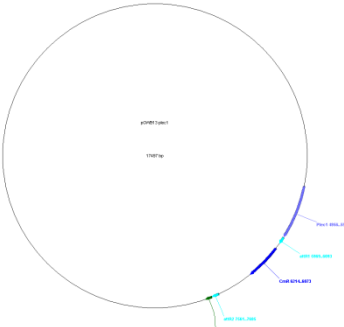
Nom	Résistance	Carte
HemiC_RBC (pGWB15_RBCTp)	Kan, Hy, CmR	 <p>(--P35S-RBCTp-3xHA-R1-CmR-ccdB-R2--)</p>
HemiC_FDHD (pGWB15_FDHTp)	Kan, Hy, CmR	 <p>(--P35S-FDHTp-3xHA-R1-CmR-ccdB-R2--)</p>
Embryo_PLEC2 (pGWB13_PLEC2)	Kan, Hy, CmR	 <p>(--PLEC2-R1-CmR-ccdB-R2-3xHA --)</p>
Embryo_PLEC1 (pGWB13_PLEC1)	Kan, Hy, CmR	 <p>(--PLEC1-R1-CmR-ccdB-R2-3xHA --)</p>

Tableau 22. Vecteurs utilisées pour la stratégie de microRNA artificielles inductibles.

Nom	Résistance	Carte
pRS300	Amp	
pBJ36_AlcA	Amp	
pMLBART_AlcR	Basta	

REFERENCES BIBLIOGRAPHIQUES

- Adl, S.M., Simpson, A.G., Lane, C.E., Lukes, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D.H., McManus, H., Mitchell, E.A., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, R.S., Schoch, C.L., Smirnov, A., and Spiegel, F.W. (2012). The revised classification of eukaryotes. *J Eukaryot Microbiol* 59, 429-493.
- Allen, J.O., Fauron, C.M., Minx, P., Roark, L., Oddiraju, S., Lin, G.N., Meyer, L., Sun, H., Kim, K., Wang, C., Du, F., Xu, D., Gibson, M., Cifrese, J., Clifton, S.W., and Newton, K.J. (2007). Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 177, 1173-1192.
- Andres, C., Agne, B., and Kessler, F. (2010). The TOC complex: preprotein gateway to the chloroplast. *Biochim Biophys Acta* 1803, 715-723.
- Archibald, J.M. (2009). The puzzle of plastid evolution. *Curr Biol* 19, R81-88.
- Archibald, J.M. (2012). Lynn Margulis (1938-2011). *Current Biology* 22, R4-R6.
- Arenas, M.A., Takenaka, M., Moreno, S., Gomez, I., and Jordana, X. (2013). Contiguous RNA editing sites in the mitochondrial nad1 transcript of *Arabidopsis thaliana* are recognized by different proteins. *FEBS Lett* 587, 887-891.
- Armengaud, P. (2009). EZ-Rhizo software: the gateway to root architecture analysis. *Plant Signal Behav* 4, 139-141.
- Aronsson, H., Boij, P., Patel, R., Wardle, A., Topel, M., and Jarvis, P. (2007). Toc64/OEP64 is not essential for the efficient import of proteins into chloroplasts in *Arabidopsis thaliana*. *Plant J* 52, 53-68.
- Asano, T., Miyao, A., Hirochika, H., Kikuchi, S., and Kadowaki, K. (2013). A pentatricopeptide repeat gene of rice is required for splicing of chloroplast transcripts and RNA editing of *ndhA*. *Plant Biotechnology*. 30, 57 - 64.
- Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A. (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol* 42, 603-613.
- Balsera, M., Soll, J., and Bolter, B. (2009). Protein import machineries in endosymbiotic organelles. *Cell Mol Life Sci* 66, 1903-1923.
- Ban, T., Ke, J., Chen, R., Gu, X., Tan, M.H., Zhou, X.E., Kang, Y., Melcher, K., Zhu, J.K., and Xu, H.E. (2013). Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J Biol Chem* 288, 31540-31548.
- Barbrook, A.C., Howe, C.J., Kurniawan, D.P., and Tarr, S.J. (2010). Organization and expression of organellar genomes. *Philos Trans R Soc Lond B Biol Sci* 365, 785-797.
- Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol* 155, 1520-1532.
- Barkan, A., and Small, I. (2014). Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol*.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8, e1002910.

- Baudisch, B., Langner, U., Garz, I., and Klosgen, R.B. (2014). The exception proves the rule? Dual targeting of nuclear-encoded proteins into endosymbiotic organelles. *New Phytol* 201, 80-90.
- 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.
- Belmonte, M.F., Kirkbride, R.C., Stone, S.L., Pelletier, J.M., Bui, A.Q., Yeung, E.C., Hashimoto, M., Fei, J., Harada, C.M., Munoz, M.D., Le, B.H., Drews, G.N., Brady, S.M., Goldberg, R.B., and Harada, J.J. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A* 110, E435-444.
- Ben-Menachem, R., Tal, M., Shadur, T., and Pines, O. (2011). A third of the yeast mitochondrial proteome is dual localized: a question of evolution. *Proteomics* 11, 4468-4476.
- Bendich, A.J. (1993). Reaching for the ring: the study of mitochondrial genome structure. *Curr Genet* 24, 279-290.
- Bendich, A.J. (2004). Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16, 1661-1666.
- Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819-826.
- Bentolila, S., Elliott, L.E., and Hanson, M.R. (2008). Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* 178, 1693-1708.
- Bentolila, S., Oh, J., Hanson, M.R., and Bukowski, R. (2013). Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet* 9, e1003584.
- Bentolila, S., Heller, W.P., Sun, T., Babina, A.M., Friso, G., van Wijk, K.J., and Hanson, M.R. (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A* 109, E1453-1461.
- Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L. (2011). Transcriptional regulation of Arabidopsis *LEAFY COTYLEDON2* involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* 23, 4065-4078.
- Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L. (2012). Transcriptional regulation of Arabidopsis *LEAFY COTYLEDON2* involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* 23, 4065-4078.
- Berglund, A.K., Pujol, C., Duchene, A.M., and Glaser, E. (2009a). Defining the determinants for dual targeting of amino acyl-tRNA synthetases to mitochondria and chloroplasts. *J Mol Biol* 393, 803-814.
- Berglund, A.K., Spanning, E., Biverstahl, H., Maddalo, G., Tellgren-Roth, C., Maler, L., and Glaser, E. (2009b). Dual targeting to mitochondria and chloroplasts: characterization of Thr-tRNA synthetase targeting peptide. *Mol Plant* 2, 1298-1309.
- Binder, S., and Brennicke, A. (2003). Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci* 358, 181-188; discussion 188-189.
- Bionda, T., Tillmann, B., Simm, S., Beilstein, K., Ruprecht, M., and Schleiff, E. (2010). Chloroplast import signals: the length requirement for translocation in vitro and in vivo. *J Mol Biol* 402, 510-523.

- Bischof, S., Baerenfaller, K., Wildhaber, T., Troesch, R., Vidi, P.A., Roschitzki, B., Hirsch-Hoffmann, M., Hennig, L., Kessler, F., Gruissem, W., and Baginsky, S. (2011). Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. *Plant Cell* 23, 3911-3928.
- Blanc, V., Litvak, S., and Araya, A. (1995). RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Lett* 373, 56-60.
- Bock, R. (2007). Structure, function, and inheritance of plastid genomes. In *Cell and molecular biology of plastids* (Springer), pp. 29-63.
- Bonen, L. (2008). Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* 8, 26-34.
- Borner, G.V., Morl, M., Wissinger, B., Brennicke, A., and Schmelzer, C. (1995). RNA editing of a group II intron in *Oenothera* as a prerequisite for splicing. *Mol Gen Genet* 246, 739-744.
- Boussard, C. (2013). Caractérisation moléculaire et fonctionnelle de la protéine DYW1 dans le complexe d'édition chloroplastique d'*Arabidopsis thaliana* (Evry: Université d'Evry Val d'Essonne, University of Western Australia), pp. 283.
- Boussard, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurin, C. (2012). Two Interacting Proteins Are Necessary for the Editing of the NdhD-1 Site in *Arabidopsis* Plastids. *Plant Cell* 24, 3684-3694.
- Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., and Gorlach, J. (2001). Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499-1510.
- Brennicke, A. (2009). The long and winding road to RNA editing in plant mitochondria: the Tübingen-Berlin chapter. *IUBMB Life* 61, 1105-1109.
- Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., and Meinke, D. (2010). Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in *Arabidopsis*. *Plant Physiol* 155, 1678-1689.
- Burger, G., Gray, M.W., and Lang, B.F. (2003). Mitochondrial genomes: anything goes. *Trends Genet* 19, 709-716.
- Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C., and Zhang, L. (2009). LPA66 is required for editing psbF chloroplast transcripts in *Arabidopsis*. *Plant Physiol* 150, 1260-1271.
- Carrie, C., and Small, I. (2013). A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim Biophys Acta* 1833, 253-259.
- Carrie, C., Giraud, E., and Whelan, J. (2009a). Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *Febs J* 276, 1187-1195.
- Carrie, C., Kuhn, K., Murcha, M.W., Duncan, O., Small, I.D., O'Toole, N., and Whelan, J. (2009b). Approaches to defining dual-targeted proteins in *Arabidopsis*. *Plant J* 57, 1128-1139.
- Castandet, B., and Araya, A. (2011). RNA editing in plant organelles. Why make it easy? *Biochemistry (Mosc)* 76, 924-931.
- Castandet, B., Choury, D., Begu, D., Jordana, X., and Araya, A. (2010). Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res* 38, 7112-7121.
- Chang, S., Wang, Y., Lu, J., Gai, J., Li, J., Chu, P., Guan, R., and Zhao, T. (2013). The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *PLoS One* 8, e56502.

- Chateigner-Boutin, A.L., and Hanson, M.R. (2003). Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res* 31, 2586-2594.
- Chateigner-Boutin, A.L., and Small, I. (2010). Plant RNA editing. *RNA Biol* 7, 213-219.
- Chateigner-Boutin, A.L., Colas des Francs-Small, C., Fujii, S., Okuda, K., Tanz, S.K., and Small, I. (2013). The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J* 74, 935-945.
- Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S., and Small, I. (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J* 65, 532-542.
- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-Garcia, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., Jimenez, L.F., Lurin, C., Small, I., and Leon, P. (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56, 590-602.
- Chaudhuri, S., and Maliga, P. (1996). Sequences directing C to U editing of the plastid *psbL* mRNA are located within a 22 nucleotide segment spanning the editing site. *Embo J* 15, 5958-5964.
- Chi, W., Sun, X., and Zhang, L. (2013). Intracellular signaling from plastid to nucleus. *Annu Rev Plant Biol* 64, 559-582.
- Chi, W., Ma, J., Zhang, D., Guo, J., Chen, F., Lu, C., and Zhang, L. (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, 573-584.
- Chi, W., Mao, J., Li, Q., Ji, D., Zou, M., Lu, C., and Zhang, L. (2010). Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in *Arabidopsis* cotyledons. *Plant J* 64, 14-25.
- Chotewutmontri, P., Reddick, L.E., McWilliams, D.R., Campbell, I.M., and Bruce, B.D. (2012). Differential transit peptide recognition during preprotein binding and translocation into flowering plant plastids. *Plant Cell* 24, 3040-3059.
- Chua, N.H., and Schmidt, G.W. (1979). Transport of proteins into mitochondria and chloroplasts. *J Cell Biol* 81, 461-483.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-743.
- Cohen, S., Zmudjak, M., Colas des Francs-Small, C., Malik, S., Shaya, F., Keren, I., Belausov, E., Many, Y., Brown, G.G., Small, I., and Ostersetzter-Biran, O. (2014). nMAT4, a maturase factor required for *nad1* pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in *Arabidopsis* mitochondria. *Plant J*.
- Colas des Francs-Small, C., and Small, I. (2013). Surrogate mutants for studying mitochondrially encoded functions. *Biochimie*.
- Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthome, R., and Lurin, C. (2013). Systematic study of subcellular localization of *Arabidopsis* PPR proteins confirms a massive targeting to organelles. *RNA Biol* 10.
- Coll, N.S., Danon, A., Meurer, J., Cho, W.K., and Apel, K. (2009). Characterization of *soldat8*, a suppressor of singlet oxygen-induced cell death in *Arabidopsis* seedlings. *Plant Cell Physiol* 50, 707-718.

- Consortium., A.I.M. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science* 333, 601-607.
- Courtois, F., Merendino, L., Demarsy, E., Mache, R., and Lerbs-Mache, S. (2007). Phage-type RNA polymerase RPOTmp transcribes the *rrn* operon from the PC promoter at early developmental stages in Arabidopsis. *Plant Physiol* 145, 712-721.
- Covello, P.S., and Gray, M.W. (1989). RNA editing in plant mitochondria. *Nature* 341, 662-666.
- Criscuolo, A., and Gribaldo, S. (2011). Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol Biol Evol* 28, 3019-3032.
- Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* 133, 462-469.
- Cushing, D.A., Forsthoefel, N.R., Gestaut, D.R., and Vernon, D.M. (2005). Arabidopsis *emb175* and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* 221, 424-436.
- Darracq, A., Varre, J.S., Marechal-Drouard, L., Courseaux, A., Castric, V., Saumitou-Laprade, P., Oztas, S., Lenoble, P., Vacherie, B., Barbe, V., and Touzet, P. (2011). Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. *Genome Biol Evol* 3, 723-736.
- de Longevialle, A.F. (2010). Identification des protéines PPR impliquées dans l'épissage des ARN messagers dans les chloroplastes et les mitochondries chez Arabidopsis Thaliana. (Evry: Université d'Evry Val d'Essonne, University of Western Australia), pp. 184.
- de Longevialle, A.F., Small, I.D., and Lurin, C. (2010). Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol Plant* 3, 691-705.
- de Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D. (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial *nad1* Intron 1 in Arabidopsis thaliana. *Plant Cell* 19, 3256-3265.
- de Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H., and Small, I. (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 Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci U S A* 110, 2898-2903.
- 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.
- Delannoy, E., Fujii, S., Colas des Francs-Small, C., Brundrett, M., and Small, I. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol* 28, 2077-2086.
- Delannoy, E., Le Ret, M., Faivre-Nitschke, E., Estavillo, G.M., Bergdoll, M., Taylor, N.L., Pogson, B.J., Small, I., Imbault, P., and Gualberto, J.M. (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNA^{Arg}(ACG) and is essential for efficient chloroplast translation. *Plant Cell* 21, 2058-2071.
- Deshpande, A.P., and Patel, S.S. (2012). Mechanism of transcription initiation by the yeast mitochondrial RNA polymerase. *Biochim Biophys Acta* 1819, 930-938.

- Ding, Y.H., Liu, N.Y., Tang, Z.S., Liu, J., and Yang, W.C. (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, 815-830.
- Dinur-Mills, M., Tal, M., and Pines, O. (2008). Dual targeted mitochondrial proteins are characterized by lower MTS parameters and total net charge. *PLoS One* 3, e2161.
- Doniwa, Y., Ueda, M., Ueta, M., Wada, A., Kadowaki, K., and Tsutsumi, N. (2010). The involvement of a PPR protein of the P subfamily in partial RNA editing of an Arabidopsis mitochondrial transcript. *Gene* 454, 39-46.
- Doolittle, W.F. (2012). Evolutionary biology: A ratchet for protein complexity. *Nature* 481, 270-271.
- Doyle, S.R., Kasinadhuni, N.R., Chan, C.K., and Grant, W.N. (2013). Evidence of evolutionary constraints that influences the sequence composition and diversity of mitochondrial matrix targeting signals. *PLoS One* 8, e67938.
- Dudek, J., Rehling, P., and van der Laan, M. (2013). Mitochondrial protein import: common principles and physiological networks. *Biochim Biophys Acta* 1833, 274-285.
- Duncan, O., Murcha, M.W., and Whelan, J. (2013). Unique components of the plant mitochondrial protein import apparatus. *Biochim Biophys Acta* 1833, 304-313.
- Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. *Science* 304, 253-257.
- Ebhardt, H.A., Tsang, H.H., Dai, D.C., Liu, Y., Bostan, B., and Fahlman, R.P. (2009). Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Res* 37, 2461-2470.
- Estelle, M.A., and Somerville, C. (1987). Auxin-resistant mutants of Arabidopsis thaliana with an altered morphology. *Molecular and General Genetics MGG* 206, 200-206.
- Fabian, M.R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol* 19, 586-593.
- Farre, J.C., Aknin, C., Araya, A., and Castandet, B. (2012). RNA editing in mitochondrial trans-introns is required for splicing. *PLoS One* 7, e52644.
- Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., and Thornton, J.W. (2012). Evolution of increased complexity in a molecular machine. *Nature* 481, 360-364.
- Fisk, D.G., Walker, M.B., and Barkan, A. (1999). Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *Embo J* 18, 2621-2630.
- Freyer, R., Kiefer-Meyer, M.C., and Kossel, H. (1997). Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A* 94, 6285-6290.
- Fuentes, I., Karcher, D., and Bock, R. (2012). Experimental reconstruction of the functional transfer of intron-containing plastid genes to the nucleus. *Curr Biol* 22, 763-771.
- Fujii, S., and Small, I. (2011). The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol* 191, 37-47.
- Fujii, S., Bond, C.S., and Small, I.D. (2011). Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc Natl Acad Sci U S A* 108, 1723-1728.
- Fujii, S., Sato, N., and Shikanai, T. (2013). Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON gradient regulation3. *Plant Cell* 25, 3079-3088.

- Fuss, J., Liegmann, O., Krause, K., and Rensing, S.A. (2013). Green Targeting Predictor and Ambiguous Targeting Predictor 2: the pitfalls of plant protein targeting prediction and of transient protein expression in heterologous systems. *New Phytol.*
- Geerts, P., Toussaint, A., Mergeai, G., and Baudoin, J.P. (2011). Phaseolus immature embryo rescue technology. *Methods Mol Biol* 710, 117-129.
- Giege, P., and Brennicke, A. (1999). RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci USA* 96, 15324-15329.
- Giege, P., Hoffmann, M., Binder, S., and Brennicke, A. (2000). RNA degradation buffers asymmetries of transcription in Arabidopsis mitochondria. *EMBO Rep* 1, 164-170.
- Gobert, A., Gutmann, B., Taschner, A., Gossringer, M., Holzmann, J., Hartmann, R.K., Rossmannith, W., and Giege, P. (2010). A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol* 17, 740-744.
- Gobert, A., Pinker, F., Fuchsbaauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., and Giege, P. (2013). Structural insights into protein-only RNase P complexed with tRNA. *Nat Commun* 4, 1353.
- Goringer, H.U. (2012). 'Gestalt,' composition and function of the Trypanosoma brucei editosome. *Annu Rev Microbiol* 66, 65-82.
- Gothandam, K.M., Kim, E.S., Cho, H., and Chung, Y.Y. (2005). OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. *Plant Mol Biol* 58, 421-433.
- Gray, M.W. (1999). Evolution of organellar genomes. *Curr Opin Genet Dev* 9, 678-687.
- Gray, M.W. (2009). The path to RNA editing in plant mitochondria: the Halifax chapter. *IUBMB Life* 61, 1114-1117.
- Gray, M.W. (2012a). Mitochondrial evolution. *Cold Spring Harb Perspect Biol* 4, a011403.
- Gray, M.W. (2012b). Evolutionary origin of RNA editing. *Biochemistry* 51, 5235-5242.
- Green, B.R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant J* 66, 34-44.
- Grewe, F., Herres, S., Viehover, P., Polsakiewicz, M., Weisshaar, B., and Knoop, V. (2011). A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res* 39, 2890-2902.
- Grienenberger, J.M. (2009). Plant mitochondrial RNA editing: the Strasbourg chapter. *IUBMB Life* 61, 1110-1113.
- Grosche, C., Funk, H.T., Maier, U.G., and Zauner, S. (2012). The chloroplast genome of *Pellia endiviifolia*: gene content, RNA-editing pattern, and the origin of chloroplast editing. *Genome Biol Evol* 4, 1349-1357.
- Grosjean, H. (2005). Fine-tuning of RNA functions by modification and editing (Berlin ; New York: Springer), pp. xxiv, 442 p.
- Gualberto, J.M., Weil, J.H., and Grienenberger, J.M. (1990). Editing of the wheat coxIII transcript: evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. *Nucleic Acids Res* 18, 3771-3776.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., and Grienenberger, J.M. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341, 660-662.
- Gualberto, J.M., Milesheina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F., and Dietrich, A. (2013). The plant mitochondrial genome: Dynamics and maintenance. *Biochimie*.

- Hammani, K., and Barkan, A. (2014). An mTERF domain protein functions in group II intron splicing in maize chloroplasts. *Nucleic Acids Res* 2014.
- Hammani, K., and Giege, P. (2014). RNA metabolism in plant mitochondria. *Trends Plant Sci*.
- Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T., and Small, I. (2009). A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21, 3686-3699.
- Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T., and Giege, P. (2013). Helical repeats modular proteins are major players for organelle gene expression. *Biochimie*.
- Hammani, K., Colas des Francs-Small, C., Takenaka, M., Tanz, S.K., Okuda, K., Shikanai, T., Brennicke, A., and Small, I. (2011). The Pentatricopeptide Repeat Protein OTP87 Is Essential for RNA Editing of nad7 and atp1 Transcripts in Arabidopsis Mitochondria. *J Biol Chem* 286, 21361-21371.
- Hanaoka, M., Kato, M., Anma, M., and Tanaka, K. (2012). SIG1, a sigma factor for the chloroplast RNA polymerase, differently associates with multiple DNA regions in the chloroplast chromosomes in vivo. *International journal of molecular sciences* 13, 12182-12194.
- Handa, H. (2008). Linear plasmids in plant mitochondria: peaceful coexistences or malicious invasions? *Mitochondrion* 8, 15-25.
- Harmel, J., Ruzzenente, B., Terzioglu, M., Spahr, H., Falkenberg, M., and Larsson, N.G. (2013). The leucine-rich pentatricopeptide repeat-containing protein (LRPPRC) does not activate transcription in mammalian mitochondria. *J Biol Chem* 288, 15510-15519.
- Hartel, B., Zehrmann, A., Verbitskiy, D., and Takenaka, M. (2013a). The longest mitochondrial RNA editing PPR protein MEF12 in Arabidopsis thaliana requires the full-length E domain. *RNA Biol* 10.
- Hartel, B., Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2013b). MEF10 is required for RNA editing at nad2-842 in mitochondria of Arabidopsis thaliana and interacts with MORF8. *Plant Mol Biol* 81, 337-346.
- Hashimoto, M., Endo, T., Peltier, G., Tasaka, M., and Shikanai, T. (2003). A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. *Plant J* 36, 541-549.
- Hayes, M.L., Giang, K., and Mulligan, R.M. (2012). Molecular evolution of pentatricopeptide repeat genes reveals truncation in species lacking an editing target and structural domains under distinct selective pressures. *BMC Evol Biol* 12, 66.
- Hayes, M.L., Giang, K., Berhane, B., and Mulligan, R.M. (2013). Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem* 288, 36519-36529.
- Hecht, J., Grewe, F., and Knoop, V. (2011). Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol* 3, 344-358.
- Hedtke, B., Borner, T., and Weihe, A. (2000). One RNA polymerase serving two genomes. *EMBO Rep* 1, 435-440.
- Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A. (1989). RNA editing in plant mitochondria. *Science* 246, 1632-1634.

- Hirose, T., and Sugiura, M. (2001). Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *Embo J* 20, 1144-1152.
- Hoch, B., Maier, R.M., Appel, K., Igloi, G.L., and Kossel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353, 178-180.
- Holec, S., Lange, H., Kuhn, K., Alioua, M., Borner, T., and Gagliardi, D. (2006). Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol Cell Biol* 26, 2869-2876.
- Howard, M.J., Lim, W.H., Fierke, C.A., and Koutmos, M. (2012). Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A* 109, 16149-16154.
- Hricova, A., Quesada, V., and Micol, J.L. (2006). The SCABRA3 nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in Arabidopsis. *Plant Physiol* 141, 942-956.
- Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P.D., Majeran, W., Sun, Q., and van Wijk, K.J. (2013a). Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J Proteome Res* 12, 491-504.
- Huang, S., Taylor, N.L., Whelan, J., and Millar, A.H. (2009). Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiol* 150, 1272-1285.
- Huang, Y.Y., Matzke, A.J., and Matzke, M. (2013b). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* 8, e74736.
- Ichinose, M., Tasaki, E., Sugita, C., and Sugita, M. (2012). A PPR-DYW protein is required for splicing of a group II intron of *cox1* pre-mRNA in *Physcomitrella patens*. *Plant J* 70, 271-278.
- Ichinose, M., Sugita, C., Yagi, Y., Nakamura, T., and Sugita, M. (2013). Two DYW subclass PPR proteins are involved in RNA editing of *ccmF*c and *atp9* transcripts in the moss *Physcomitrella patens*: first complete set of PPR editing factors in plant mitochondria. *Plant Cell Physiol* 54, 1907-1916.
- Ikeda, T.M., and Gray, M.W. (1999). Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol Cell Biol* 19, 8113-8122.
- Inada, M., Sasaki, T., Yukawa, M., Tsudzuki, T., and Sugiura, M. (2004). A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant Cell Physiol* 45, 1615-1622.
- Iorizzo, M., Grzebelus, D., Senalik, D., Szklarczyk, M., Spooner, D., and Simon, P. (2012). Against the traffic: The first evidence for mitochondrial DNA transfer into the plastid genome. *Mob Genet Elements* 2, 261-266.
- Ito, J., Heazlewood, J.L., and Millar, A.H. (2006). Analysis of the soluble ATP-binding proteome of plant mitochondria identifies new proteins and nucleotide triphosphate interactions within the matrix. *J Proteome Res* 5, 3459-3469.
- Jacobs, J., Marx, C., Kock, V., Reifschneider, O., Franzel, B., Krisp, C., Wolters, D., and Kuck, U. (2013). Identification of a chloroplast ribonucleoprotein complex containing trans-

- splicing factors, intron RNA, and novel components. *Mol Cell Proteomics* 12, 1912-1925.
- Jarvis, P. (2008). Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytol* 179, 257-285.
- Jiang, Y., Fan, S.L., Song, M.Z., Yu, J.N., and Yu, S.X. (2012). Identification of RNA editing sites in cotton (*Gossypium hirsutum*) chloroplasts and editing events that affect secondary and three-dimensional protein structures. *Genet Mol Res* 11, 987-1001.
- Kagami, H., Nagano, H., Takahashi, Y., Mikami, T., and Kubo, T. (2012). Is RNA editing implicated in group II intron survival in the angiosperm mitochondrial genome? *Genome* 55, 75-79.
- Kalanon, M., and McFadden, G.I. (2008). The chloroplast protein translocation complexes of *Chlamydomonas reinhardtii*: a bioinformatic comparison of Toc and Tic components in plants, green algae and red algae. *Genetics* 179, 95-112.
- Kanesaki, Y., Imamura, S., Minoda, A., and Tanaka, K. (2012). External light conditions and internal cell cycle phases coordinate accumulation of chloroplast and mitochondrial transcripts in the red alga *Cyanidioschyzon merolae*. *DNA Res* 19, 289-303.
- Kasmati, A.R., Topel, M., Khan, N.Z., Patel, R., Ling, Q., Karim, S., Aronsson, H., and Jarvis, P. (2013). Evolutionary, molecular and genetic analyses of Tic22 homologues in *Arabidopsis thaliana* chloroplasts. *PLoS One* 8, e63863.
- Ke, J., Chen, R.Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.K., Melcher, K., and Xu, H.E. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struct Mol Biol* 20, 1377-1382.
- Keegstra, K., and Cline, K. (1999). Protein import and routing systems of chloroplasts. *Plant Cell* 11, 557-570.
- Keeling, P.J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* 365, 729-748.
- Keeling, P.J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol* 64, 583-607.
- Keren, I., Tal, L., des Francs-Small, C.C., Araujo, W.L., Shevtsov, S., Shaya, F., Fernie, A.R., Small, I., and Ostersetzer-Biran, O. (2012). nMAT1, a nuclear-encoded maturase involved in the trans-splicing of nad1 intron 1, is essential for mitochondrial complex I assembly and function. *Plant J* 71, 413-426.
- Keren, I., Bezawork-Geleta, A., Kolton, M., Maayan, I., Belausov, E., Levy, M., Mett, A., Gidoni, D., Shaya, F., and Ostersetzer-Biran, O. (2009). AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in *Arabidopsis* mitochondria. *RNA* 15, 2299-2311.
- Khrouchtchova, A., Monde, R.A., and Barkan, A. (2012). A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *Rna* 18, 1197-1209.
- Kikuchi, S., Bedard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., and Nakai, M. (2013). Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339, 571-574.
- Kleine, T., Maier, U.G., and Leister, D. (2009). DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol* 60, 115-138.
- Kleinknecht, L., Wang, F., Stube, R., Philippar, K., Nickelsen, J., and Bohn, A.V. (2014). RAP, the Sole Octatricopeptide Repeat Protein in *Arabidopsis*, Is Required for Chloroplast 16S rRNA Maturation. *Plant Cell*.

- Knoop, V. (2011). When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* 68, 567-586.
- Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., and Nakamura, T. (2011). Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res* 40, 2712-2723.
- Kohler, D., Schmidt-Gattung, S., and Binder, S. (2009). The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of *Arabidopsis thaliana*. *Plant Mol Biol*.
- Koprivova, A., des Francs-Small, C.C., Calder, G., Mugford, S.T., Tanz, S., Lee, B.R., Zechmann, B., Small, I., and Kopriva, S. (2010). Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial nad7 transcripts. *J Biol Chem* 285, 32192-32199.
- Kotera, E., Tasaka, M., and Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433, 326-330.
- Kovacs-Bogdan, E., Soll, J., and Bolter, B. (2010). Protein import into chloroplasts: the Tic complex and its regulation. *Biochim Biophys Acta* 1803, 740-747.
- Kubo, T., and Newton, K.J. (2008). Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8, 5-14.
- Kugita, M., Yamamoto, Y., Fujikawa, T., Matsumoto, T., and Yoshinaga, K. (2003). RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Res* 31, 2417-2423.
- Kuhn, K., Richter, U., Meyer, E.H., Delannoy, E., de Longevialle, A.F., O'Toole, N., Borner, T., Millar, A.H., Small, I.D., and Whelan, J. (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21, 2762-2779.
- Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K., Harada, J.J., and Goldberg, R.B. (2010). Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A* 107, 8063-8070.
- Lee, C.P., Eubel, H., O'Toole, N., and Millar, A.H. (2008). Heterogeneity of the mitochondrial proteome for photosynthetic and non-photosynthetic *Arabidopsis* metabolism. *Mol Cell Proteomics* 7, 1297-1316.
- Lee, S., Lee, D.W., Yoo, Y.J., Duncan, O., Oh, Y.J., Lee, Y.J., Lee, G., Whelan, J., and Hwang, I. (2012). Mitochondrial targeting of the *Arabidopsis* F1-ATPase gamma-subunit via multiple compensatory and synergistic presequence motifs. *Plant Cell* 24, 5037-5057.
- Leibovitch, M., Bublak, D., Hanic-Joyce, P.J., Tillmann, B., Flinner, N., Amsel, D., Scharf, K.D., Mirus, O., Joyce, P.B., and Schleiff, E. (2013). The folding capacity of the mature domain of the dual-targeted plant tRNA nucleotidyltransferase influences organelle selection. *Biochem J* 453, 401-412.
- Leister, D. (2012). Retrograde signaling in plants: from simple to complex scenarios. *Front Plant Sci* 3, 135.
- Leister, D., Wang, X., Haberer, G., Mayer, K.F., and Kleine, T. (2011). Intracompartmental and intercompartmental transcriptional networks coordinate the expression of genes for organellar functions. *Plant Physiol* 157, 386-404.

- Lerbs-Mache, S. (2011). Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? *Plant Mol Biol* 76, 235-249.
- Li, H.M., and Chiu, C.C. (2010). Protein transport into chloroplasts. *Annu Rev Plant Biol* 61, 157-180.
- Li, H.M., and Teng, Y.S. (2013). Transit peptide design and plastid import regulation. *Trends Plant Sci* 18, 360-366.
- Li, J.F., Chung, H.S., Niu, Y., Bush, J., McCormack, M., and Sheen, J. (2013). Comprehensive protein-based artificial microRNA screens for effective gene silencing in plants. *Plant Cell* 25, 1507-1522.
- Liere, K., Weihe, A., and Borner, T. (2011). The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J Plant Physiol* 168, 1345-1360.
- Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2013). Human pentatricopeptide proteins: Only a few and what do they do? *RNA Biol* 10.
- Lister, R., Carrie, C., Duncan, O., Ho, L.H., Howell, K.A., Murcha, M.W., and Whelan, J. (2007). Functional definition of outer membrane proteins involved in preprotein import into mitochondria. *Plant Cell* 19, 3739-3759.
- Liu, H., Cui, P., Zhan, K., Lin, Q., Zhuo, G., Guo, X., Ding, F., Yang, W., Liu, D., Hu, S., Yu, J., and Zhang, A. (2011). Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. *BMC Genomics* 12, 163.
- Liu, S., Melonek, J., Boykin, L.M., Small, I., and Howell, K.A. (2013). PPR-SMRs: Ancient proteins with enigmatic functions. *RNA Biol* 10.
- Liu, Y., He, J., Chen, Z., Ren, X., Hong, X., and Gong, Z. (2010). ABA overly-sensitive 5 (ABO5), encoding a pentatricopeptide repeat protein required for cis-splicing of mitochondrial nad2 intron 3, is involved in the abscisic acid response in Arabidopsis. *Plant J* 63, 749-765.
- Lloyd, A.H., and Timmis, J.N. (2011). The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. *Mol Biol Evol* 28, 2019-2028.
- Lung, S.C., and Chuong, S.D. (2012). A transit peptide-like sorting signal at the C terminus directs the *Bienertia sinuspersici* preprotein receptor Toc159 to the chloroplast outer membrane. *Plant Cell* 24, 1560-1578.
- Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089-2103.
- Mackiewicz, P., Bodyl, A., and Gagat, P. (2012). Protein import into the photosynthetic organelles of and its implications for primary plastid endosymbiosis. *Symbiosis* 58, 99-107.
- Maier, U.G., Bozarth, A., Funk, H.T., Zauner, S., Rensing, S.A., Schmitz-Linneweber, C., Borner, T., and Tillich, M. (2008). Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol* 6, 36.
- Malek, O., Lattig, K., Hiesel, R., Brennicke, A., and Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *Embo J* 15, 1403-1411.
- Manthey, G.M., and McEwen, J.E. (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, 4031-4043.
- Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M., and Jacq, C. (2002). Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep* 3, 159-164.
- Marcia, M., Somarowthu, S., and Pyle, A.M. (2013). Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA* 4, 14.
- Margulis, L. (2010). Symbiogenesis. A new principle of evolution rediscovery of Boris Mikhaylovich Kozo-Polyansky (1890–1957). *Paleontological Journal* 44, 1525-1539.
- Marion, J., Bach, L., Bellec, Y., Meyer, C., Gissot, L., and Faure, J.D. (2008). Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of *Arabidopsis* seedlings. *Plant J* 56, 169-179.
- 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.
- Meng, Y., Chen, D., Jin, Y., Mao, C., Wu, P., and Chen, M. (2010). RNA editing of nuclear transcripts in *Arabidopsis thaliana*. *BMC Genomics* 11 Suppl 4, S12.
- Michalovova, M., Vyskot, B., and Kejnovsky, E. (2013). Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. *Heredity (Edinb)* 111, 314-320.
- Millar, A.H., Carrie, C., Pogson, B., and Whelan, J. (2009). Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. *Plant Cell* 21, 1625-1631.
- Mitschke, J., Fuss, J., Blum, T., Høglund, A., Reski, R., Kohlbacher, O., and Rensing, S.A. (2009). Prediction of dual protein targeting to plant organelles. *New Phytol* 183, 224-235.
- Mower, J.P., Case, A.L., Floro, E.R., and Willis, J.H. (2012). Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biol Evol* 4, 670-686.
- Murayama, M., Hayashi, S., Nishimura, N., Ishide, M., Kobayashi, K., Yagi, Y., Asami, T., Nakamura, T., Shinozaki, K., and Hirayama, T. (2012). Isolation of *Arabidopsis* ahg11, a weak ABA hypersensitive mutant defective in nad4 RNA editing. *J Exp Bot* 63, 5301-5310.
- Murcha, M.W., Elhafez, D., Lister, R., Tonti-Filippini, J., Baumgartner, M., Philippar, K., Carrie, C., Mokranjac, D., Soll, J., and Whelan, J. (2007). Characterization of the preprotein and amino acid transporter gene family in *Arabidopsis*. *Plant Physiol* 143, 199-212.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* 104, 34-41.
- Nakamura, T., and Sugita, M. (2008). A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* 582, 4163-4168.
- Nakamura, T., Yagi, Y., and Kobayashi, K. (2012). Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. *Plant Cell Physiol* 53, 1171-1179.

- Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G. (2003). RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. *Eur J Biochem* 270, 4070-4081.
- Nakayama, T., and Archibald, J.M. (2012). Evolving a photosynthetic organelle. *BMC Biol* 10, 35.
- Neupert, W. (1997). Protein import into mitochondria. *Annu Rev Biochem* 66, 863-917.
- Neupert, W., and Herrmann, J.M. (2007). Translocation of proteins into mitochondria. *Annu Rev Biochem* 76, 723-749.
- Nowack, E.C., and Grossman, A.R. (2012). Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci U S A* 109, 5340-5345.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* 25, 1120-1128.
- Okuda, K., and Shikanai, T. (2012). A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic Acids Res* 40, 5052-5064.
- Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006). A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281, 37661-37667.
- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* 104, 8178-8183.
- Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* 21, 146-156.
- Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J* 53, 674-690.
- Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A* 108, 13624-13629.
- Paul, P., Simm, S., Blaumeiser, A., Scharf, K.D., Fragkostefanakis, S., Mirus, O., and Schleiff, E. (2013). The protein translocation systems in plants - composition and variability on the example of *Solanum lycopersicum*. *BMC Genomics* 14, 189.
- Pfalz, J., and Pfannschmidt, T. (2013). Essential nucleoid proteins in early chloroplast development. *Trends Plant Sci* 18, 186-194.
- Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.J., and Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18, 176-197.
- Pfanner, N., and Neupert, W. (1990). The mitochondrial protein import apparatus. *Annu Rev Biochem* 59, 331-353.
- Pfanner, N., Hartl, F.U., and Neupert, W. (1988). Import of proteins into mitochondria: a multi-step process. *Eur J Biochem* 175, 205-212.

- Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci U S A* 108, 415-420.
- Pyle, A.M. (2010). The tertiary structure of group II introns: implications for biological function and evolution. *Crit Rev Biochem Mol Biol* 45, 215-232.
- Pyo, Y.J., Kwon, K.C., Kim, A., and Cho, M.H. (2013). Seedling Lethal1, a pentatricopeptide repeat protein lacking an E/E+ or DYW domain in Arabidopsis, is involved in plastid gene expression and early chloroplast development. *Plant Physiol* 163, 1844-1858.
- Qbadou, S., Becker, T., Mirus, O., Tews, I., Soll, J., and Schleiff, E. (2006). The molecular chaperone Hsp90 delivers precursor proteins to the chloroplast import receptor Toc64. *Embo J* 25, 1836-1847.
- Rackham, O., and Filipovska, A. (2011). The role of mammalian PPR domain proteins in the regulation of mitochondrial gene expression. *Biochim Biophys Acta* 1819, 1008-1016.
- Rajasekhar, V.K., and Mulligan, R.M. (1993). RNA Editing in Plant Mitochondria: [alpha]-Phosphate Is Retained during C-to-U Conversion in mRNAs. *Plant Cell* 5, 1843-1852.
- Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J., and Palmer, J.D. (2013). The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol* 11, 29.
- Rigas, S., Daras, G., Tsitsekian, D., and Hatzopoulos, P. (2011). The multifaceted role of Lon proteolysis in seedling establishment and maintenance of plant organelle function: living from protein destruction. *Physiol Plant* 145, 215-223.
- Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. *Nature* 478, 269-273.
- Rivals, E., Bruyere, C., Toffano-Nioche, C., and Lecharny, A. (2006). Formation of the Arabidopsis pentatricopeptide repeat family. *Plant Physiol* 141, 825-839.
- Rogozin, I.B., Carmel, L., Csuros, M., and Koonin, E.V. (2012). Origin and evolution of spliceosomal introns. *Biol Direct* 7, 11.
- Roslan, H.A., Salter, M.G., Wood, C.D., White, M.R., Croft, K.P., Robson, F., Coupland, G., Doonan, J., Laufs, P., Tomsett, A.B., and Caddick, M.X. (2001). Characterization of the ethanol-inducible alc gene-expression system in Arabidopsis thaliana. *Plant J* 28, 225-235.
- Rudinger, M., Polsakiewicz, M., and Knoop, V. (2008). Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Mol Biol Evol* 25, 1405-1414.
- Rudinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V. (2012). Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J Mol Evol* 74, 37-51.
- Ruwe, H., Castandet, B., Schmitz-Linneweber, C., and Stern, D.B. (2013). Arabidopsis chloroplast quantitative editotype. *FEBS Lett* 587, 1429-1433.
- Saint-Georges, Y., Garcia, M., Delaveau, T., Jourdain, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F., and Jacq, C. (2008). Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One* 3, e2293.
- Salmans, M.L., Chaw, S.M., Lin, C.P., Shih, A.C., Wu, Y.W., and Mulligan, R.M. (2010). Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. *Curr Genet* 56, 439-446.

- Salone, V. (2008). IDENTIFICATION DU FACTEUR CATALYTIQUE DU PROCESSUS D'EDITION DES ARN DES ORGANITES CHEZ LES PLANTES (Evry: Université d'Evry Val d'Essonne, University of Western Australia), pp. 325.
- Salone, V., Rudinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581, 4132-4138.
- Savage, L.J., Imre, K.M., Hall, D.A., and Last, R.L. (2013). Analysis of essential Arabidopsis nuclear genes encoding plastid-targeted proteins. *PLoS One* 8, e73291.
- Schagger, H., Cramer, W.A., and von Jagow, G. (1994). Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* 217, 220-230.
- Schallenberg-Rudinger, M., Kindgren, P., Zehrmann, A., Small, I., and Knoop, V. (2013a). A DYW-protein knockout in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J* 76, 420-432.
- Schallenberg-Rudinger, M., Lenz, H., Polsakiewicz, M., Gott, J.M., and Knoop, V. (2013b). A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol* 10.
- Schleiff, E., and Becker, T. (2011). Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat Rev Mol Cell Biol* 12, 48-59.
- 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.
- Schuster, G., and Stern, D. (2009). RNA polyadenylation and decay in mitochondria and chloroplasts. *Prog Mol Biol Transl Sci* 85, 393-422.
- Schuster, W., Hiesel, R., Wissinger, B., and Brennicke, A. (1990). RNA editing in the cytochrome b locus of the higher plant *Oenothera berteriana* includes a U-to-C transition. *Mol Cell Biol* 10, 2428-2431.
- Schwarzlander, M., Konig, A.C., Sweetlove, L.J., and Finkemeier, I. (2012). The impact of impaired mitochondrial function on retrograde signalling: a meta-analysis of transcriptomic responses. *J Exp Bot* 63, 1735-1750.
- Schwinghammer, K., Cheung, A.C., Morozov, Y.I., Agaronyan, K., Temiakov, D., and Cramer, P. (2013). Structure of human mitochondrial RNA polymerase elongation complex. *Nat Struct Mol Biol* 20, 1298-1303.
- Selosse, M.-A. (2009). endosymbiotiques: de l'avantage d'être petit? *Biofutur* 299, 28.
- Sheppard, A.E., and Timmis, J.N. (2009). Instability of plastid DNA in the nuclear genome. *PLoS Genet* 5, e1000323.
- Sheppard, A.E., Madesis, P., Lloyd, A.H., Day, A., Ayliffe, M.A., and Timmis, J.N. (2011). Introducing an RNA editing requirement into a plastid-localised transgene reduces but does not eliminate functional gene transfer to the nucleus. *Plant Mol Biol* 76, 299-309.
- Shi, L.X., and Theg, S.M. (2013). The chloroplast protein import system: from algae to trees. *Biochim Biophys Acta* 1833, 314-331.

- Shih, P.M., and Matzke, N.J. (2013). Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci U S A* 110, 12355-12360.
- Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* 63, 698-708.
- Shikanai, T., and Fujii, S. (2013). Function of PPR proteins in plastid gene expression. *RNA Biol* 10.
- Sikor, M., Mapa, K., von Voithenberg, L.V., Mokranjac, D., and Lamb, D.C. (2013). Real-time observation of the conformational dynamics of mitochondrial Hsp70 by spFRET. *Embo J* 32, 1639-1649.
- Sloan, D.B. (2013). One ring to rule them all? Genome sequencing provides new insights into the “master circle” model of plant mitochondrial DNA structure. *New Phytologist* 200, 978-985.
- Sloan, D.B., MacQueen, A.H., Alverson, A.J., Palmer, J.D., and Taylor, D.R. (2010). Extensive loss of RNA editing sites in rapidly evolving *Silene* mitochondrial genomes: selection vs. retroprocessing as the driving force. *Genetics* 185, 1369-1380.
- Sloan, D.B., Alverson, A.J., Chuckalovcak, J.P., Wu, M., McCauley, D.E., Palmer, J.D., and Taylor, D.R. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol* 10, e1001241.
- Small, I.D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25, 46-47.
- Small, I.D., Rackham, O., and Filipovska, A. (2013). Organelle transcriptomes: products of a deconstructed genome. *Curr Opin Microbiol* 16, 652-658.
- Solheim, C., Li, L., Hatzopoulos, P., and Millar, A.H. (2012). Loss of Lon1 in *Arabidopsis* changes the mitochondrial proteome leading to altered metabolite profiles and growth retardation without an accumulation of oxidative damage. *Plant Physiol* 160, 1187-1203.
- Sosso, D., Mbello, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., Dauzat, M., Heurtevin, L., Guyon, V., Takenaka, M., and Rogowsky, P.M. (2012). PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* 24, 676-691.
- Steiner, S., Schroter, Y., Pfalz, J., and Pfannschmidt, T. (2011). Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. *Plant Physiol* 157, 1043-1055.
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing. *J Mol Evol* 48, 303-312.
- Stern, D.B., Goldschmidt-Clermont, M., and Hanson, M.R. (2010). Chloroplast RNA metabolism. *Annu Rev Plant Biol* 61, 125-155.
- Stojanovski, D., Bohnert, M., Pfanner, N., and van der Laan, M. (2012). Mechanisms of protein sorting in mitochondria. *Cold Spring Harb Perspect Biol* 4.
- Stoppel, R., and Meurer, J. (2013). Complex RNA metabolism in the chloroplast: an update on the psbB operon. *Planta* 237, 441-449.
- Straub, S.C., Cronn, R.C., Edwards, C., Fishbein, M., and Liston, A. (2013). Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). *Genome Biol Evol* 5, 1872-1885.

- Sugita, M., Ichinose, M., Ide, M., and Sugita, C. (2013). Architecture of the PPR gene family in the moss *Physcomitrella patens*. *RNA Biol* 10.
- Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., and Sugita, M. (2003). Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic Acids Res* 31, 5324-5331.
- Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M.R., and Bentolila, S. (2013). An RNA recognition motif-containing protein is required for plastid RNA editing in *Arabidopsis* and maize. *Proc Natl Acad Sci U S A* 110, E1169-1178.
- Swiatecka-Hagenbruch, M., Emanuel, C., Hedtke, B., Liere, K., and Borner, T. (2008). Impaired function of the phage-type RNA polymerase *RpoTp* in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucleic Acids Res* 36, 785-792.
- Takenaka, M. (2010). MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the *nad7* transcript in mitochondria of *Arabidopsis*. *Plant Physiol* 152, 939-947.
- Takenaka, M., and Brennicke, A. (2003). In vitro RNA editing in pea mitochondria requires NTP or dNTP, suggesting involvement of an RNA helicase. *J Biol Chem* 278, 47526-47533.
- Takenaka, M., and Brennicke, A. (2009). Multiplex single-base extension typing to identify nuclear genes required for RNA editing in plant organelles. *Nucleic Acids Res* 37, e13.
- Takenaka, M., and Brennicke, A. (2012). Using multiplex single-base extension typing to screen for mutants defective in RNA editing. *Nat Protoc* 7, 1931-1945.
- Takenaka, M., Verbitskiy, D., Zehrmann, A., and Brennicke, A. (2010). Reverse genetic screening identifies five E-class PPR-proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J Biol Chem*.
- Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013a). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* 8, e65343.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Hartel, B., and Brennicke, A. (2013b). RNA editing in plants and its evolution. *Annu Rev Genet* 47, 335-352.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelman, M., Hartel, B., and Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc Natl Acad Sci U S A* 109, 5104-5109.
- Tan, X.Y., Liu, X.L., Wang, W., Jia, D.J., Chen, L.Q., Zhang, X.Q., and Ye, D. (2010). Mutations in the *Arabidopsis* nuclear-encoded mitochondrial phage-type RNA polymerase gene *RPOTm* led to defects in pollen tube growth, female gametogenesis and embryogenesis. *Plant Cell Physiol* 51, 635-649.
- Tanz, S.K., Castleden, I., Hooper, C.M., Vacher, M., Small, I., and Millar, H.A. (2012). SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Res* 41, D1185-1191.
- Tasaki, E., Hattori, M., and Sugita, M. (2010). The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J* 62, 560-570.
- Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M., and Sato, N. (2007). The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Mol Biol Evol* 24, 699-709.

- Thiergart, T., Landan, G., Schenk, M., Dagan, T., and Martin, W.F. (2012). An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome biology and evolution* 4, 466-485.
- Tillich, M., Poltnigg, P., Kushnir, S., and Schmitz-Linneweber, C. (2006). Maintenance of plastid RNA editing activities independently of their target sites. *EMBO Rep* 7, 308-313.
- Tillich, M., Hardel, S.L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J.M., Lehwark, P., Leister, D., Small, I.D., and Schmitz-Linneweber, C. (2009). Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A* 106, 6002-6007.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5, 123-135.
- Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G., and Jones, J.D. (1999). Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* 11, 1841-1852.
- Toda, T., Fujii, S., Noguchi, K., Kazama, T., and Toriyama, K. (2012). Rice MPR25 encodes a pentatricopeptide repeat protein and is essential for RNA editing of nad5 transcripts in mitochondria. *Plant J* 72, 450-460.
- Topel, M., Ling, Q., and Jarvis, P. (2012). Neofunctionalization within the Omp85 protein superfamily during chloroplast evolution. *Plant Signal Behav* 7, 161-164.
- Tourasse, N.J., Choquet, Y., and Vallon, O. (2013). PPR proteins of green algae. *RNA Biol* 10.
- Tsuji, J., Frith, M.C., Tomii, K., and Horton, P. (2012). Mammalian NUMT insertion is non-random. *Nucleic Acids Res* 40, 9073-9088.
- Turmel, M., Otis, C., and Lemieux, C. (2002). The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol Biol Evol* 19, 24-38.
- Uniacke, J., and Zerges, W. (2009). Chloroplast protein targeting involves localized translation in *Chlamydomonas*. *Proc Natl Acad Sci U S A* 106, 1439-1444.
- Van Leene, J., Stals, H., Eeckhout, D., Persiau, G., Van De Slijke, E., Van Isterdael, G., De Clercq, A., Bonnet, E., Laukens, K., Remmerie, N., Henderickx, K., De Vijlder, T., Abdelkrim, A., Pharazyn, A., Van Onckelen, H., Inze, D., Witters, E., and De Jaeger, G. (2007). A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. *Mol Cell Proteomics* 6, 1226-1238.
- Van Leene, J., Hollunder, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Stals, H., Van Isterdael, G., Verkest, A., Neiryneck, S., Buffel, Y., De Bodt, S., Maere, S., Laukens, K., Pharazyn, A., Ferreira, P.C., Eloy, N., Renne, C., Meyer, C., Faure, J.D., Steinbrenner, J., Beynon, J., Larkin, J.C., Van de Peer, Y., Hilson, P., Kuiper, M., De Veylder, L., Van Onckelen, H., Inze, D., Witters, E., and De Jaeger, G. (2010). Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Mol Syst Biol* 6, 397.
- Verbitskiy, D., Hartel, B., Zehrmann, A., Brennicke, A., and Takenaka, M. (2011). The DYW-E-PPR protein MEF14 is required for RNA editing at site matR-1895 in mitochondria of *Arabidopsis thaliana*. *FEBS Lett* 585, 700-704.
- Verbitskiy, D., Zehrmann, A., Hartel, B., Brennicke, A., and Takenaka, M. (2012). Two related RNA-editing proteins target the same sites in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 287, 38064-38072.

- Wang, D., and Timmis, J.N. (2013). Cytoplasmic organelle DNA preferentially inserts into open chromatin. *Genome Biol Evol* 5, 1060-1064.
- Wang, D., Lloyd, A.H., and Timmis, J.N. (2012a). Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc Natl Acad Sci U S A* 109, 2444-2448.
- Wang, D., Rousseau-Gueutin, M., and Timmis, J.N. (2012b). Plastid sequences contribute to some plant mitochondrial genes. *Molecular biology and evolution* 29, 1707-1711.
- Wang, J., Jiang, J., Li, X., Li, A., Zhang, Y., Guan, R., and Wang, Y. (2012c). Complete sequence of heterogenous-composition mitochondrial genome (*Brassica napus*) and its exogenous source. *BMC Genomics* 13, 675.
- Weis, B.L., Schleiff, E., and Zerges, W. (2013). Protein targeting to subcellular organelles via mRNA localization. *Biochim Biophys Acta* 1833, 260-273.
- Welchen, E., Garcia, L., Mansilla, N., and Gonzalez, D.H. (2014). Coordination of plant mitochondrial biogenesis: keeping pace with cellular requirements. *Front Plant Sci* 4, 551.
- Werhahn, W., Niemeyer, A., Jansch, L., Kruft, V., Schmitz, U.K., and Braun, H. (2001). Purification and characterization of the preprotein translocase of the outer mitochondrial membrane from *Arabidopsis*. Identification of multiple forms of TOM20. *Plant Physiol* 125, 943-954.
- Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Muller, K.F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76, 273-297.
- Wicke, S., Muller, K.F., de Pamphilis, C.W., Quandt, D., Wickett, N.J., Zhang, Y., Renner, S.S., and Schneeweiss, G.M. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25, 3711-3725.
- 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.
- Williams-Carrier, R., Zoschke, R., Belcher, S., Pfalz, J., and Barkan, A. (2013). A major role for the plastid-encoded RNA polymerase complex in the expression of plastid transfer RNAs. *Plant Physiol* 164, 239-248.
- Woodson, J.D., and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* 9, 383-395.
- Xu, L., Carrie, C., Law, S.R., Murcha, M.W., and Whelan, J. (2013). Acquisition, conservation, and loss of dual-targeted proteins in land plants. *Plant Physiol* 161, 644-662.
- Yagi, Y., and Shiina, T. (2012). Evolutionary aspects of plastid proteins involved in transcription: the transcription of a tiny genome is mediated by a complicated machinery. *Transcription* 3, 290-294.
- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8, e57286.
- Ye, W., Spanning, E., Unnerstale, S., Gotthold, D., Glaser, E., and Maler, L. (2012). NMR investigations of the dual targeting peptide of Thr-tRNA synthetase and its interaction with the mitochondrial Tom20 receptor in *Arabidopsis thaliana*. *Febs J* 279, 3738-3748.

- Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.W., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504, 168-171.
- Yogev, O., and Pines, O. (2011). Dual targeting of mitochondrial proteins: mechanism, regulation and function. *Biochim Biophys Acta* 1808, 1012-1020.
- Yogev, O., Karniely, S., and Pines, O. (2007). Translation-coupled translocation of yeast fumarate into mitochondria in vivo. *J Biol Chem* 282, 29222-29229.
- Yoshida, T., Furihata, H.Y., and Kawabe, A. (2013). Patterns of Genomic Integration of Nuclear Chloroplast DNA Fragments in Plant Species. *DNA Res.*
- Yoshinaga, K., Kakehi, T., Shima, Y., Inuma, H., Masuzawa, T., and Ueno, M. (1997). Extensive RNA editing and possible double-stranded structures determining editing sites in the *atpB* transcripts of hornwort chloroplasts. *Nucleic Acids Res* 25, 4830-4834.
- Yu, W., and Schuster, W. (1995). Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J Biol Chem* 270, 18227-18233.
- Zehrmann, A., van der Merwe, J.A., Verbitskiy, D., Brennicke, A., and Takenaka, M. (2008). Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*. *Mitochondrion* 8, 319-327.
- Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21, 558-567.
- Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J., and Borner, T. (2012). The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24, 123-136.
- Zhu, Q., Dugardeyn, J., Zhang, C., Takenaka, M., Kuhn, K., Craddock, C., Smalle, J., Karampelias, M., Denecke, J., Peters, J., Gerats, T., Brennicke, A., Eastmond, P., Meyer, E.H., and Van Der Straeten, D. (2012). SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J* 71, 836-849.
- Zmudjak, M., Colas des Francs-Small, C., Keren, I., Shaya, F., Belausov, E., Small, I., and Ostersetter-Biran, O. (2013). mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in *Arabidopsis*. *New Phytol* 199, 379-394.
- Zoschke, R., Watkins, K.P., and Barkan, A. (2013). A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* 25, 2265-2275.
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and van Wijk, K.J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* 3, e1994.