

Thèse de doctorat de l'Université d'Evry Val d'Essonne

Présentée par :

Charles Andrés

Pour obtenir le grade de Docteur en Biologie Moléculaire

sur le sujet :

Analyse fonctionnelle de la Famille des Protéines  
Pentatricopeptide Repeat chez *Arabidopsis thaliana*.

Soutenue le 10 mai 2007 devant le jury composé de :

Francis Quetier	Directeur de l'Ecole Doctorale
Claire Lurin	Directeur de thèse
Ian Small	Directeur de thèse
Yves Choquet	Rapporteur
Dominique Gagliardi	Rapporteur
Christian Schmitz-Linneweber	Examinateur

**Institut National de la Recherche Agronomique**  
**Centre National de la Recherche Scientifique**

Ecole doctorale "des génomes aux organismes"

Remerciements :

Parce que ce qui est rare est cher, voici les quelques personnes à qui je veux dire merci pour leur soutien pendant cette thèse:

Céline, Claire et Ian, sans qui.....rien!

Dominique Gagliardi, Yves Choquet et Richard Berthomé, les membres de mon comité de thèse.

Alex., Stéph., Laure et Agnès PARCEQUEEEEEEEEEEEEEEEEE!

Christian, my old abruti!

Andéol, Maud, Anne Laure, Fred. et Etienne parce que loin des yeux mais près du cœur.

Béné et Isa, parce que la Fac!

Michel Caboche pour son accueil

Et puis biensûr Nico S, Nico G., Virgie, Fifi, Antoine !

Philippe et Geneviève

Chantal, Francois et poupette!

Mman, petit frère et minou!

Sinon je ne remercie pas Darth Vador, qui fait que tout le monde se marre quand j'ai de l'asthme ! ☺

Et spéciale dédicace aux monstroplantes...

## Table des matières

<b>CHAPITRE I</b>	<b>INTRODUCTION</b>	<b>I-3</b>
I.1	CHLOROPLASTES ET MITOCHONDRIES :	I-3
I.1.1	<i>La mitochondrie d'Arabidopsis</i>	I-4
I.1.1.1	Structure et dynamique des mitochondries	I-4
I.1.1.2	Métabolismes se déroulant dans les mitochondries	I-5
I.1.2	<i>Le chloroplaste d'Arabidopsis</i>	I-6
I.1.2.1	Structures et dynamique des chloroplastes	I-9
I.1.2.2	Les plastes photosynthétiques	I-9
I.1.2.3	Les plastes non photosynthétiques	I-11
I.1.2.4	Métabolisme	I-12
I.2	LES GENOMES CYTOPLASMIQUES	I-13
I.2.1	<i>Le génome mitochondrial</i>	I-14
I.2.1.1	Composition du génome mitochondrial d'Arabidopsis	I-14
I.2.1.2	Structure du génome mitochondrial	I-17
I.2.2	<i>Le génome chloroplastique</i>	I-20
I.2.2.1	Composition du génome chloroplastique	I-20
I.2.2.2	Structure du génome chloroplastique	I-23
I.3	EXPRESSION DES GENES DANS LES ORGANITES	I-24
I.3.1	<i>La transcription dans la mitochondrie</i>	I-26
I.3.2	<i>La transcription dans le chloroplaste</i>	I-27
I.3.3	<i>Les modifications post-transcriptionnelles</i>	I-30
I.3.3.1	Maturation en 5' et 3'	I-30
I.3.3.2	Cis- et Trans-épissage	I-32
I.3.3.3	Modification de la séquence des ARN	I-33
I.3.3.3.1	Ajout de nucléotides	I-33
I.3.3.3.2	Edition des transcrits	I-33
I.3.4	<i>La traduction</i>	I-34
I.4	COMMUNICATION ENTRE LE NOYAU, LES CHLOROPLASTES ET LES MITOCHONDRIES	I-36
I.4.1	<i>Communication inter-organites, ou transversales</i>	I-36
I.4.2	<i>Les voies rétrogrades</i>	I-37
I.4.3	<i>Les voies antérogrades</i>	I-38
I.5	LES PROTEINES PPR	I-40
I.5.1	<i>Définition d'une protéine PPR</i>	I-40
I.5.2	<i>Ce n'est pas le motif PPR, mais les motifs PPR</i>	I-42
I.5.2.1	Formation de la famille PLS	I-42
I.5.3	<i>Les protéines PPR comportent également des domaines protéiques autre que les motifs PPR</i>	I-42
I.5.4	<i>Au sujet des fonctions des protéines PPR</i>	I-44
I.6	PRESENTATION DU TRAVAIL DE THESE	I-45

<b>CHAPITRE II</b>	<b>RESULTATS.....</b>	<b>II-46</b>
II.1	ANALYSE DE MUTANTS.....	II-46
II.1.1	Article 1 : 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. <i>Plant Cell</i> , 16, 2089-2103.....	II-46
II.1.2	Constitution de la collection de mutants.....	II-62
II.1.3	Analyse des transcriptomes chloroplastiques et mitochondriaux.....	II-64
II.1.4	Article 2 : Andrés C., Falcon de Longevialle A., Lurin C. and Small ID. Rapid low-resolution transcript mapping for identifying mutants affecting organelle gene expression in <i>Arabidopsis thaliana</i> .....	II-65
II.1.5	CLB19, une protéine PPR multifacettes .....	II-80
II.1.5.1	Article 3 : Ramos Vega M., Guevara-García A., Gutiérrez-Nava MdL., Andrés C., Cantero A., Jiménez LF., Lurin C., Small ID. and León P. CLB19, a Novel Pentatricopeptide Repeat Protein that is required for efficient 5' and 3' processing of diverse chloroplast transcripts in <i>Arabidopsis</i> . .....	II-80
II.1.5.2	Analyse du transcriptome chloroplastique de <i>clb19</i> .....	II-109
II.1.5.3	Interaction physique de CLB19 avec des transcrits chloroplastiques .....	II-114
II.1.5.4	Conservation du site ClpP1-1 .....	II-114
II.1.6	Discussion .....	II-115
II.1.6.1	CLB19 agit-elle sur le transcrit de <i>rbcL</i> ? uniquement? .....	II-116
II.1.6.2	Edition de ClpP1-1 .....	II-117
II.1.6.2.1	Les Clp-protéases .....	II-117
II.1.6.2.2	Conservation de la tyrosine 187 .....	II-118
II.2	IDENTIFICATION DE DEFAUTS MOLECULAIRES CHEZ DES MUTANTS PPR A L'AIDE DU CRIBLE.....	II-121
II.2.1	Etude du mutant PPR43 (Andéol Falcon de Longevialle et al, en préparation).....	II-121
II.2.1.1	Analyse du mutant <i>ppr43</i> .....	II-121
II.2.2	PPR_LAGLIDADG (Andéol Falcon de Longevialle et al, en préparation) .....	II-122
II.2.3	Discussion .....	II-123
II.3	ANALYSE DES GENOMES CYTOPLASMIQUES.....	II-124
II.3.1	Comparaison des génomes mitochondriaux de différents écotypes d' <i>Arabidopsis thaliana</i> .....	II-124
II.3.2	Analyse des génomes mitochondriaux de mutants <i>osb1</i> .....	II-126
II.3.3	Discussion .....	II-127
<b>CHAPITRE III</b>	<b>CONCLUSION / DISCUSSION.....</b>	<b>III-128</b>
III.1.1	Article 3 : Andrés C, Lurin C, Small ID (2007)The multifarious roles of PPR proteins in plant mitochondrial gene expression. <i>Physiologia Plantarum</i> 129 (1), 14-22.....	III-129
III.2	PROTEINES PPR PUBLIEES DEPUIS JUIN 2006.....	III-138
III.2.1	Le cas particulier des Kinétoplastides.....	III-139
III.3	A LA RECHERCHE DES DEFAUTS MOLECULAIRES PRESENTS CHEZ LES MUTANTS DE PROTEINES PPR D'ARABIDOPSIS.....	III-142
III.3.1	Première étape: développer un outil de détection.....	III-142
III.3.1.1	La méthode vu par le manipulateur .....	III-142
III.3.1.2	Le crible vu par l'analyste.....	III-143

III.3.1.3	Sensibilité du crible .....	III-144
III.3.1.4	Bilan de l'outil .....	III-144
III.3.2	<i>Seconde étape: collecter des mutants</i> .....	III-145
III.4	ET LES PPR DANS TOUT CELA? .....	III-147
III.4.1	<i>Quel rôle pour les protéines PPR au niveau de l'organisme?</i> .....	III-152
<b>CHAPITRE IV</b>	<b>MATERIELS ET METHODES</b> .....	<b>IV-154</b>
IV.1	MATERIEL BIOLOGIQUE .....	IV-154
IV.1.1	<i>Lignées d'insertions utilisées</i> .....	IV-154
IV.1.2	<i>Méthode de culture d'Arabidopsis</i> .....	IV-157
IV.1.2.1	Culture en terre .....	IV-157
IV.1.2.2	Culture <i>in vitro</i> .....	IV-157
IV.2	METHODES .....	IV-157
IV.2.1	<i>Protocoles relatifs à l'ADN</i> .....	IV-157
IV.2.1.1	Extraction rapide d'ADN génomique : .....	IV-157
IV.2.1.2	Extraction fine d'ADN génomique, méthode dite "au CTAB" .....	IV-158
IV.2.1.3	Electrophorèse .....	IV-159
IV.2.1.4	Amplification d'ADN par PCR .....	IV-159
IV.2.1.4.1	PCR dite "standard" .....	IV-159
IV.2.1.4.2	PCR dite "384 puits" .....	IV-160
IV.2.1.5	Test d'édition par profil de restriction .....	IV-160
IV.2.1.6	Purification d'acides nucléiques .....	IV-161
IV.2.1.7	Séquençage d'ADN .....	IV-161
IV.2.2	<i>Protocoles relatifs aux ARN</i> .....	IV-161
IV.2.2.1	Extraction d'ARN total, méthode dite "au TRIzol <sup>®</sup> " .....	IV-161
IV.2.2.2	Elimination de l'ADN contaminant des extraits d'ARN totaux .....	IV-162
IV.2.2.3	Synthèse de premier brin d'ADN complémentaire d'ARN d'organites .....	IV-162
IV.2.2.4	5' et 3' Race .....	IV-163
IV.2.2.4.1	Synthèse du premier brin d'ADNc .....	IV-163
IV.2.2.4.2	"Rapid Amplification of cDNA Ends" (RACE) .....	IV-164
IV.2.2.4.3	Etapes supplémentaires .....	IV-164
IV.2.2.4.4	Amorces gène spécifique .....	IV-165
IV.2.2.4.5	Amorces généraux et Tampons .....	IV-165
IV.2.3	<i>Protocoles relatifs aux protéines</i> .....	IV-166
IV.2.3.1	Extraction de protéines .....	IV-166
IV.2.3.1.1	Protéines totales .....	IV-166
IV.2.3.1.2	Enrichissement en protéines chloroplastiques .....	IV-166
IV.2.3.2	Dosage des protéines .....	IV-167
IV.2.3.3	Observation sur gel dénaturant .....	IV-167
IV.2.3.4	Western blot .....	IV-168
IV.2.4	<i>Oligonucléotides utilisés lors de cette thèse</i> .....	IV-168
IV.2.4.1	Génotypage de lignées d'insertion .....	IV-168
IV.2.4.2	Analyse de l'éditotype chloroplastique .....	IV-171

<b>CHAPITRE V</b>	<b>REFERENCES .....</b>	<b>V-172</b>
<b>CHAPITRE VI</b>	<b>ANNEXES .....</b>	<b>VI-185</b>
VI.1	DONNEES SUPPLEMENTAIRES DE L'ARTICLE 2, ANDRES C., FALCON DE LONGEVIALLE A., LURIN C. AND SMALL ID. RAPID LOW-RESOLUTION TRANSCRIPT MAPPING FOR IDENTIFYING MUTANTS AFFECTING ORGANELLE GENE EXPRESSION IN ARABIDOPSIS THALIANA.....	VI-185
VI.2	ARTICLE 4 : ZAEGEL, V., GUERMANN, B., LE RET, M., ANDRES, C., MEYER, D., ERHARDT, M., CANADAY, J., GUALBERTO, J. M. AND IMBAULT, P. (2006). THE PLANT-SPECIFIC SSDNA BINDING PROTEIN OSB1 IS INVOLVED IN THE STOICHIOMETRIC TRANSMISSION OF MITOCHONDRIAL DNA IN ARABIDOPSIS. <i>PLANT CELL</i> 18, 3548-63. ....	VI-199
VI.3	ANNEXE A L'ANALYSE DES GENOMES MITOCHONDRIAUX .....	VI-218

**Table des illustrations**

Tableau 1 : Composition du génome mitochondrial d' <i>Arabidopsis</i> .....	I-16
Tableau 2 : Composition du génome chloroplastique d' <i>Arabidopsis</i> (Partie 1).....	I-21
Tableau 3 : Composition du génome chloroplastique d' <i>Arabidopsis</i> Partie 2.....	I-22
Tableau 4 : Top 10 InterPro des familles de protéines chez <i>Arabidopsis</i> .....	I-41
Tableau 5 : PPR publiés avant 2003 .....	I-45
Tableau 6 : Mutants homozygotes .....	II-63
Tableau 7 : Editotype des mutants <i>clb19</i> .....	II-111
Tableau 8 : liste des accessions utilisés dans cette étude .....	II-125
Tableau 9 : Protéines PPR décrites dans la littérature au 1er Avril 2007 .....	III-141
Tableau 10 : Lien entre les sous-famille PPR et les processus moléculaires .....	III-148
Tableau 11 : Liste des lignées d'insertion collectées.....	IV-157
Tableau 12 : liste des amorces utilisées pour le génotypage des lignées d'insertion.....	IV-168
Tableau 13 : Tableau S1 amorces chloroplastiques .....	VI-185
Tableau 14 : Tableau S2 Amorces mitochondriales .....	VI-191
Tableau 15 : Analyse des génomes mitochondriaux complets de 12 écotypes d' <i>Arabidopsis thaliana</i> .....	VI-218
Encart 1 : La stérilité mâle cytoplasmique.....	I-19
Encart 2 : La maturation des ARN ribosomiaux chloroplastiques.....	I-31
Encart 3 : Exemple de contrôle par épistasie de synthèse.....	I-35
Encart 4 : La formation de la sous-famille PLS .....	I-43
Figure 1 : Endosymbiose et évolution des génomes .....	I-2
Figure 2 : La chaîne mitochondriale de transfert d'électrons .....	I-6
Figure 3 : Origine des chloroplastes et évolution des eucaryotes photosynthétiques .....	I-8
Figure 4 : Division des chloroplastes.....	I-9
Figure 5 : Différenciation des plastes .....	I-12
Figure 6 : Photosynthèse et cycle de Calvin .....	I-13
Figure 7 : Réarrangements du génome mitochondrial par recombinaison entre séquences répétées .....	I-17
Figure 8 : Carte du génome mitochondrial d' <i>Arabidopsis</i> .....	I-19
Figure 9 : Carte du génome chloroplastique .....	I-23
Figure 10 : Structure du génome chloroplastique et répllication recombinaison-dépendante .....	I-24
Figure 11 : Composition des complexes photosynthétiques.....	I-25
Figure 12 : Expression des gènes mitochondriaux.....	I-26
Figure 13 : La transcription dans les chloroplastes.....	I-28
Figure 14 : Maturation des ARN de transfert, d'après (Levinger, Morl et al. 2004).....	I-30
Figure 15 : communication entre les organites : transfert de métabolites.....	I-36



Figure 16 : Voies biochimiques de la photo-respiration .....	I-37
Figure 17 : Prédiction de la structure tertiaire d'un motif PPR .....	I-40
Figure 18 : Transcriptomes chloroplastiques A. du mutant <i>clb19</i> et B. de l'écotype Col-0 en faible lumière...	II-108
Figure 19 : Amplification des transcrits dans la région de <i>atpB-rbcL-accD</i> chez <i>clb19</i> et chez le sauvage .....	II-109
Figure 20 : RACE PCR sur le transcrit de <i>rbcL</i> .....	II-110
Figure 21 : Analyse de l'édition du transcrit de <i>clpP1</i> par séquençage .....	II-112
Figure 22 : Analyse de <i>ClpP1-1</i> par digestion enzymatique.....	II-112
Figure 23 : Analyse par Poisoned Primer Extension (PPE) de l'édition de <i>clpP1-1</i> .....	II-113
Figure 24 : Taux d'édition des sites <i>ClpP1-1</i> , <i>AccD-1</i> , <i>NdhG-1</i> et <i>PetL-2</i> .....	II-113
Figure 25 : Conservation de la séquence correspondant au résidu tyrosine 187 de <i>clpP1</i> d' <i>Arabidopsis</i> .....	II-115
Figure 26 : Alignement des séquences protéiques des protéines <i>ClpP/R</i> avec les modèles 1TYF, 1NZY et 1HNO.....	II-120
Figure 27 : Analyse par PCR du mutant <i>osb1-1</i> et de l'écotype Col-0 .....	II-126
Figure 28 : Modèle de fixation de <i>CRP1</i> à ses deux cibles ARN .....	III-150
Figure 29 : <i>Pumilio1</i> humain reconnaissant une séquence d'ARN .....	III-152
Figure 30 : Amorces pour l'analyse par RACE PCR du transcrit <i>rbcL</i> .....	IV-165

Principales abréviations utilisées dans cette thèse :

ATP : Adénosine TriPhosphate

ARN : Acide RiboNucléique

ADN : Acide DésoxyriboNucléique

BET : Bromure d'ETHidium

CLP : CaseinoLytic Protease

CTAB : Cetyl TrimethylAmmonium Bromide

CMS : "Cytoplasmic Male Sterility", stérilité mâle cytoplasmique

DTT : 1,4 DiThioThreitol

DEPC : DiEthylpPyroCarbonate

EDTA : EthyleneDiamineTetraAcetic acid

EGTA : Ethylene Glycol-bis(2-amino ether)-N,N,N',N'-TetraAcetic acid

MOPS : 3-(N-MOrpholino)PropaneSulfonic acid

NAD(P) : forme oxydée de la Nicotinamide Adénine Dinucléotide (Phosphate), NAD(P)H pour la forme réduite

NEP : "Nuclear Encoded Polymerase", polymérase chloroplastique codée par le génome nucléaire

PCR : "Polymerase chain reaction", réaction en chaîne par polymérase

PEP : "Plastidial Encoded Polymerase", polymérase chloroplastique codée par le génome chloroplastique

RIP-CHIP : RNA ImmunoPrecipitation and DNA CHIP hybridization

Rubisco : Ribulose-1,5-bisphosphate

S: Svedberg, unité de mesure du taux de sédimentation

Nomenclature :

Pour un gène nucléaire, la nomenclature suivante est utilisée (Meinke and Koornneef 1997):

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

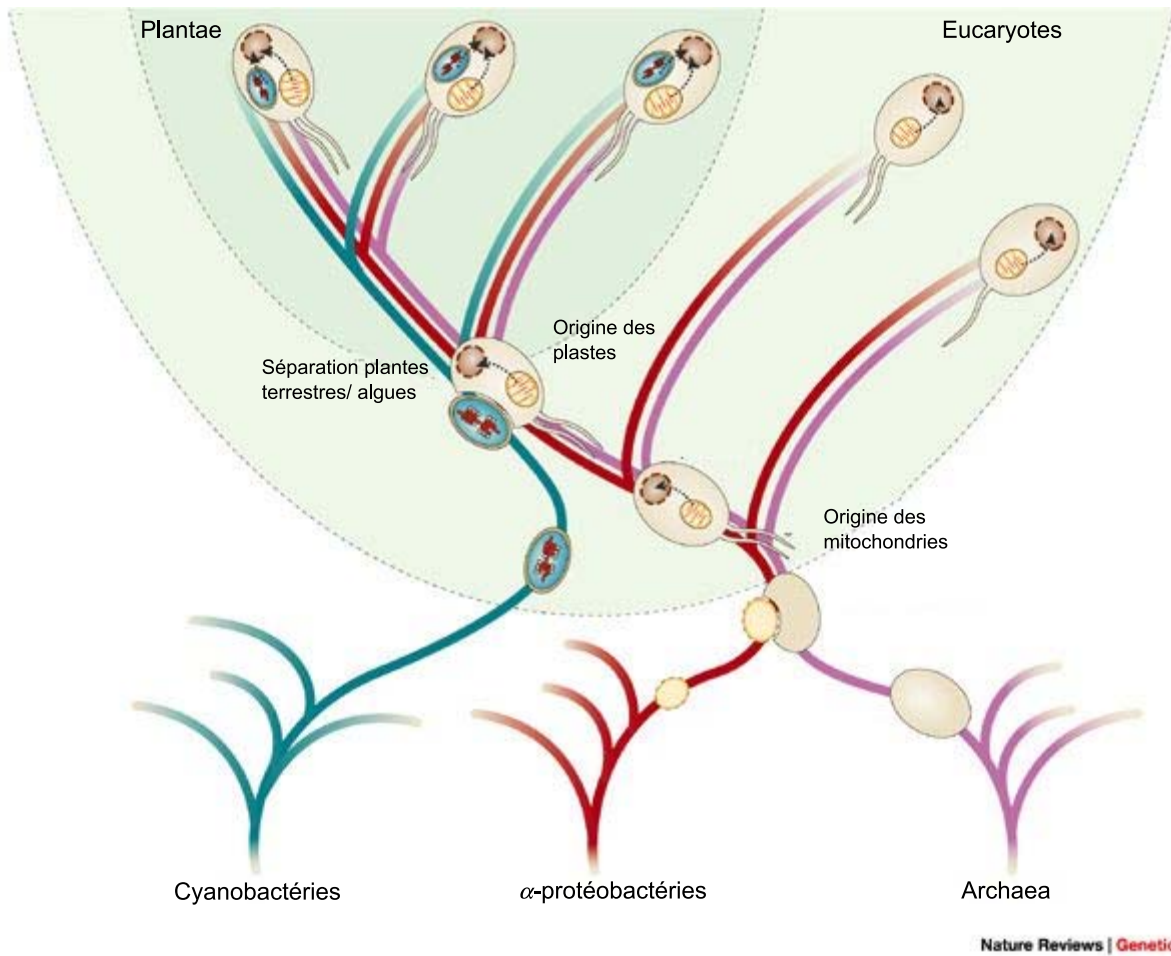
Pour un gène cytoplasmique, la nomenclature bactérienne est adoptée :

Gène : *cyT*, mutant : *cyT*, protéine : CyT

Les sites d'édition sont désignés par le nom de la protéine et un numéro : CyT-2

## **Préambule**

Cette thèse a pour sujet la recherche des fonctions d'une famille de gènes spécifiques aux eucaryotes, les protéines PPR, pour PentatrigoPeptideRepeat. Cette famille produit des protéines quasiment intégralement adressées aux organites, les mitochondries et les chloroplastes, et leur nature suggère une implication dans la gestion des acides nucléiques, en particulier l'ARN. L'introduction de cette thèse aura pour but de présenter la nature des mitochondries et des chloroplastes, leurs positions dans le monde du vivant, leurs origines, leurs histoires, leurs rôles actuels; et de mettre en exergue quelques points importants à leur sujet. Le fonctionnement du système génétique de ces organites sera abordé de manière plus précise, avant de décrire pour le mieux ce qui est connu des protéines PPR.



**Figure 1 : Endosymbiose et évolution des génomes**

Les mitochondries et les chloroplastes proviennent de l'endosymbiose de deux bactéries il y a environ 2 et 1.5 milliards d'années. Cette figure tirée d'une revue de Timmis *et al.* illustre l'incorporation des mitochondries et des chloroplastes dans le royaume des eucaryotes par les événements d'endosymbiose et la co-évolution des trois génomes qui suivit. La nature de l'hôte ayant acquis la première mitochondrie reste à déterminer, pour l'instant la théorie dominante est le rattachement aux Archaea (Timmis *et al.* 2004).

## Chapitre I Introduction

### I.1 Chloroplastes et Mitochondries :

L'une des caractéristiques remarquables des cellules eucaryotes est la présence de sous compartiments : les organites. Parmi ces organites, deux sont d'origine bactérienne : les mitochondries et les chloroplastes (chez les eucaryotes photosynthétiques) (Figure 1). Ces organites sont essentiels à chaque organisme en étant à la base de toute la machinerie énergétique. Ils assurent la majorité de la production d'énergie par divers processus, comme la photosynthèse dans les chloroplastes, ou la production d'ATP dans les mitochondries. Certains eucaryotes ne semblent posséder ni mitochondrie ni plaste, mais la présence d'hydrogénosomes ou de mitosomes dans ces organismes semble être la trace d'ancestrales mitochondries, appuyant ainsi l'idée que les mitochondries font partie de la définition même d'une cellule eucaryote (Rotte *et al.* 2000; van der Giezen *et al.* 2005).

Il est maintenant admis que les mitochondries proviennent de l'endosymbiose d'une  $\alpha$ -protéobactérie il y a environ 2 milliards d'années. Pour les chloroplastes, un consensus existe pour situer leur origine unique par l'endosymbiose d'une cyanobactérie, il y a environ 1,5 ou 1,6 milliard d'années, par contre, différents événements d'endosymbioses secondaires, par exemple un eucaryote photosynthétique dans un eucaryote non-photosynthétique (McFadden *et al.* 1994), ont sans doute eu lieu depuis, créant ainsi une phylogénie complexe des organismes photosynthétiques (Figure 3) (Orphan *et al.* 2002).

Au cours de l'évolution, les génomes des bactéries ancestrales ont été fortement altérés, chaque génome ne conservant qu'une très faible partie du génome originel, mais pour autant de nombreuses fonctions primordiales se déroulent toujours dans les organites (métabolisme de l'ADN, de l'ARN, synthèse protéique, respiration, photosynthèse, etc.). Différentes hypothèses peuvent être avancées sur le processus de réduction de ces génomes, mais on peut considérer que les gènes restants sont en quelque sorte "quasi-impossible" à transférer au noyau (Aubrey 2005). Les deux arguments cités le plus fréquemment sont d'une part les différences de code génétique entre l'hôte et le symbionte et, d'autre part, la forte hydrophobicité de certaines protéines, qui rendrait leur passage à travers les membranes des organites impossible (Aubrey 2005). Depuis longtemps des raisons fonctionnelles à la conservation de ces génomes ont été recherchées. La majorité des gènes conservés dans les génomes des organites font partie des larges complexes photosynthétiques ou de la chaîne respiratoire. Chez *Chlamydomonas*, il a été montré que la synthèse des protéines codées par le

génomique chloroplastique était régulée par le niveau d'assemblage de ces protéines dans les complexes dont elles font partie (Choquet *et al.* 1998; Choquet and Wollman 2002). Ce mode de régulation permet une coordination fine de la synthèse des protéines nucléaires et chloroplastiques, et implique une pression de sélection très forte sur ces gènes.

### **I.1.1 La mitochondrie d'*Arabidopsis***

La mitochondrie est donc un organite fruit de l'endosymbiose d'une  $\alpha$ -protéobactérie et des 2 milliards d'années d'évolution qui ont suivi cet événement. Une conséquence directe de cette origine bactérienne est l'incapacité pour une cellule eucaryote de "créer" une mitochondrie, il doit toujours y avoir au moins une première mitochondrie transmise par le gamète femelle (à part quelques exceptions, les mitochondries sont transmises par la voie maternelle (Korpelainen 2004). Au cours du développement, les mitochondries se multiplient par fission, comme les bactéries. Une autre conséquence est le grand nombre de caractéristiques mi-eucaryotes, mi-bactériennes de cet organite, comme l'illustre le protéome mitochondrial de la levure *Saccharomyces cerevisiae*, dont la moitié des protéines n'a pas d'homologue procaryote évident (Olof Karlberg 2000).

#### **I.1.1.1 Structure et dynamique des mitochondries**

À part quelques exceptions, nous pouvons dire que tous les eucaryotes possèdent des mitochondries, et que chacune d'elles provient du même endosymbionte originel. Par contre toutes les mitochondries ne sont pas semblables, leur état actuel est intimement lié à la nature de l'hôte. Chaque cellule eucaryote possède plusieurs mitochondries, mais leur nombre et leur organisation varie beaucoup d'un organisme à l'autre, d'un type de cellule à un autre, ou encore suivant les stades de développement. Par exemple, l'algue *Chlamydomonas* et la levure *Saccharomyces cerevisiae* ne comportent que quelques mitochondries organisées en réseaux complexes, une cellule humaine compte entre 500 et 1000 mitochondries, également organisées en réseaux, et les cellules végétales comportent de 200 à 2000 mitochondries suivant le tissu.

Les mitochondries sont des organites très dynamiques, la structure générale définie par la présence de deux membranes restant généralement fixe: membrane externe, espace inter-membranaire, membrane interne et matrice; mais la taille, la forme (sphérique, ovoïde, filiforme ou falciforme) ou la localisation dans la cellule changent continuellement. Plusieurs types de formes de mitochondries coexistent dans la même cellule, et la même mitochondrie pourra prendre tour à tour chacune de ces apparences. Les deux membranes n'ont pas les

mêmes caractéristiques. La membrane externe est perméable aux particules de moins de 5kDa, et comporte des transporteurs permettant l'import des protéines dans la mitochondrie (TOM, Translocase of the Outer Membrane). La membrane interne forme de nombreux replis appelés crêtes (ou cristae). Tubulaires, sacculaires, laminaires ou triangulaires, les crêtes peuvent prendre de nombreuses formes différentes au sein d'une même mitochondrie. La membrane interne a une composition un peu différente de l'externe, elle comporte également des transporteurs de protéines (TIM, Translocase of the Inner Membrane).

La régulation de la transmission des mitochondries d'une génération à l'autre peut également beaucoup changer. Chez la levure *Saccharomyces cerevisiae* par exemple, le mode de division par bourgeonnement nécessite un contrôle très fin de la distribution des mitochondries entre la cellule mère et la cellule fille émergente pour éviter que le cytoplasme de la cellule fille ne comporte aucune mitochondrie. Les eucaryotes multicellulaires comme l'homme ou les plantes n'ont pas le même type de contraintes, les mitochondries étant distribuées aléatoirement à chaque division cellulaire, et le gamète femelle procurant l'intégralité de son cytoplasme au nouvel organisme diploïde après fécondation. Néanmoins, la présence de nombreuses formes différentes du génome mitochondrial nécessite un certain contrôle pour éviter la perte de gènes lors du développement (Cf. I.2.12).

### **I.1.1.2 Métabolismes se déroulant dans les mitochondries**

L'évolution des eucaryotes est marquée par de nombreux événements que l'on pourrait qualifier de "sauts", l'apparition des mitochondries en est un principal, l'apparition des chloroplastes en est un second tout aussi important. Ces deux événements ont donné la possibilité aux eucaryotes de coloniser de nouvelles niches écologiques en enrichissant leurs capacités métaboliques, entre autre par l'acquisition de la respiration aérobie. D'un point de vue biochimique, la respiration est une succession de réactions d'oxydoréduction générant une grande quantité d'énergie sous forme d'ATP, elle est réalisée au sein de la chaîne de transfert d'électrons, localisée dans la membrane interne des mitochondries (Figure 2). Cette chaîne est constituée de 5 complexes membranaires comprenant une centaine de sous-unités protéiques différentes, dont seulement 18 sont codées par le génome mitochondrial chez *Arabidopsis*. Chez les plantes, il existe deux composantes additionnelles spécifiques, une oxydase alternative et une NAD(P)H déshydrogénase alternative.

Les métabolismes se déroulant dans les mitochondries ne se limitent pas à la respiration. Les mitochondries sont aussi le siège du cycle de Krebs (ou cycle des acides tricarboxyliques), ce cycle de réactions chimiques est l'étape finale du catabolisme des

carbohydrates, des acides gras et des acides aminés. Le cycle de Krebs fournit également la plupart des formes réduites oxydées pendant la respiration.

La synthèse de nombreuses vitamines et coenzymes est dépendante des mitochondries. Les voies de synthèses des vitamines H, C et B<sub>5</sub> comprennent au moins une étape localisée dans les mitochondries. La voie de synthèse de l'acide folique (vitamine B<sub>9</sub>) comprend des étapes à la fois dans le cytosol, les plastes et les mitochondries. D'autre part, la synthèse de la biotine se déroule dans les mitochondries, alors que la majorité de cette biotine produite sera utilisée dans les chloroplastes et le cytoplasme (pour revue, (Lunn 2007)). Enfin, les mitochondries ont aussi un rôle dans l'homéostasie ionique de la cellule en permettant le stockage des ions Ca<sup>2+</sup> (Lunn 2007).

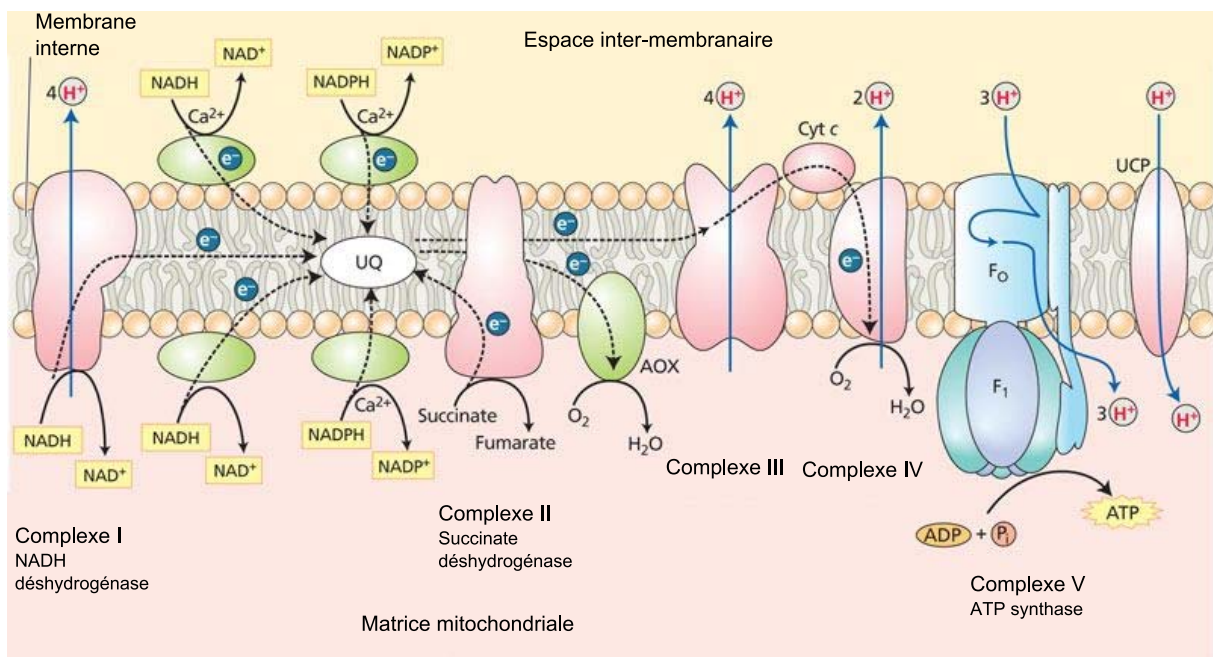


Figure 2 : La chaîne mitochondriale de transfert d'électrons

### I.1.2 Le chloroplaste d'Arabidopsis

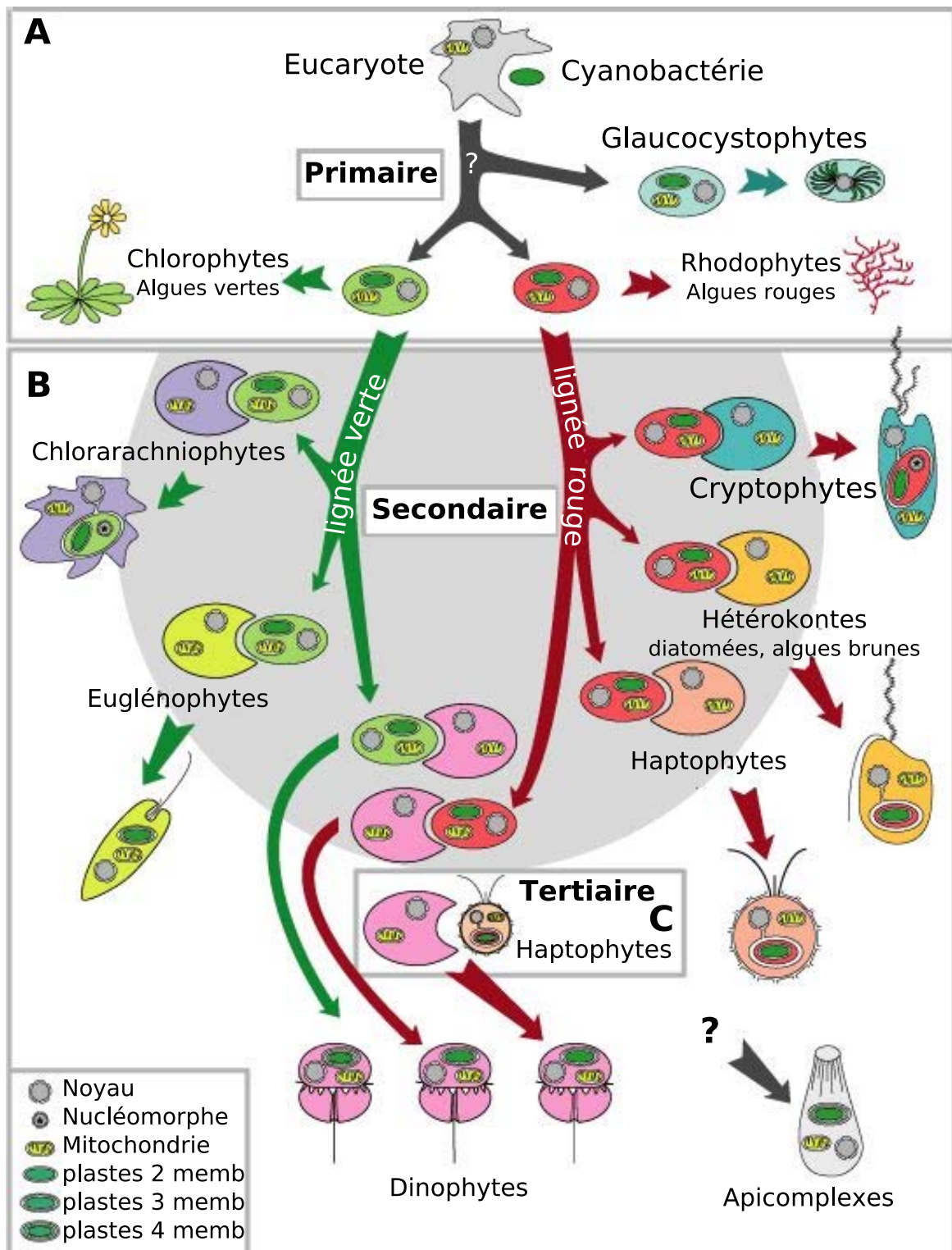
Environ 500 à 600 millions d'années après l'endosymbiose d'une  $\alpha$ -protéobactérie, une cyanobactérie s'est à son tour "fait piéger", donnant naissance au "règne" des PLANTAE ou ARCHEPLASTIDIAE ou encore PRIMOPLANTAE. La classification des espèces photosynthétiques est assez difficile car marquée par plusieurs siècles de biologie systématique basée sur les caractères apparents, ou sur la phylogénie. De ce fait, certains termes comme "plantae" peuvent avoir été utilisés dans différentes classifications pour décrire des divisions ou phyla différents.

Cette division comprend : les rhodophytes (algues rouges), les viridiplantae ou chlorophytes (algues vertes et bryophytes) et les glaucophytes (Figure 3). A l'image des



mitochondries, les chloroplastes ne peuvent être créés *de novo* par une cellule eucaryote, il doit y avoir au moins un chloroplaste transmis de génération en génération. La transmission des chloroplastes est souvent maternelle, mais elle peut également être paternelle, comme chez les gymnospermes, ou même parfois mixte (Korpelainen 2004; Muschner *et al.* 2006).

Comme la mitochondrie, le chloroplaste a grandement divergé de l'endosymbionte originel. Cette divergence est souvent qualifiée de simplification, mais en réalité les plastes ont subi une diversification fonctionnelle inconnue des mitochondries. Il existe une douzaine de "plastés" différents, les plus connus étant les chloroplastes.



**Figure 3 : Origine des chloroplastes et évolution des eucaryotes photosynthétiques**

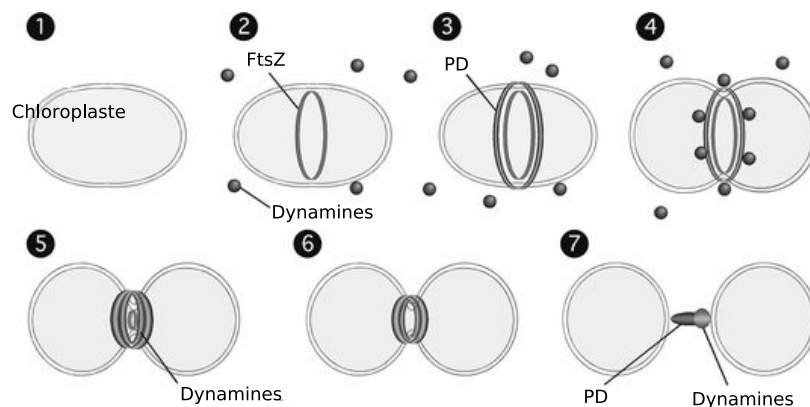
A. L'endosymbiose primaire a donné naissance à la lignée rouge et la lignée verte. B. Les algues vertes proviennent de l'endosymbiose secondaire d'un chlorophyte, alors que les chromophytes, cryptophytes et haptophytes sont les produits de différentes endosymbioses secondaires d'un rhodophyte. C. Diverses endosymbioses tertiaires peuvent être à l'origine des dinophytes. L'origine des Apicomplexes reste énigmatique. (Falkowski *et al.* 2003).

### I.1.2.1 Structures et dynamique des chloroplastes

Les plastes sont caractérisés par la présence de deux membranes héritées de l'endosymbionte originel. Ces deux membranes sont séparées par un espace intermembranaire. Le stroma (intérieur) des chloroplastes comporte un réseau membranaire complexe : les thylakoïdes, c'est là que se déroule la photosynthèse (cf. partie métabolisme). De petites structures lipidiques appelées plastoglobules sont associées aux membranes des thylakoïdes, leur rôle n'est pas encore bien déterminé, mais ces structures semblent très riches en enzymes et pourraient intervenir dans plusieurs métabolismes se déroulant dans les chloroplastes (Kessler *et al.* 1999; Austin *et al.* 2006).

Les chloroplastes sont très riches en différents pigments tels que la chlorophylle ou les caroténoïdes. Ils sont essentiels à la photosynthèse car à la base de la "capture" de l'énergie lumineuse. La protéine la plus abondante sur la planète réside dans les plastes, il s'agit de la rubisco, enzyme essentielle du cycle de Calvin-Benson et donc de la fixation du carbone.

La division des chloroplastes est, semble-t-il, contrôlée par un système combinant des facteurs d'origines procaryotes et eucaryotes. Trois types d'anneaux différents sont impliqués séquentiellement, le "Plastid Dividing Ring", FtsZ et des dynamines (Miyagishima 2005; Lopez-Juez 2007).



**Figure 4 : Division des chloroplastes.**

Séquences d'événements menant à la division des chloroplastes : 1) élongation des chloroplastes, 2) formation d'un anneau FtsZ au site de division, 3) formation séquentielle des anneaux PD, intérieur, intermédiaire et extérieur, la constriction commence, 4) recrutement à la surface des chloroplastes de structures cytosoliques apparentées à des dynamines, 5) formation de l'anneau de dynamines, 6) désassemblage de l'anneau FtsZ, puis des anneaux PD interne et intermédiaire, 7) la division est finie, le reste de l'anneau PD externe se désassemble alors que le reste de l'anneau de dynamines reste temporairement associé à un des chloroplastes (Miyagishima 2005).

### I.1.2.2 Les plastes photosynthétiques

La fonction des plastes la plus connue est la photosynthèse (Figure 11) , mécanisme par

lequel certains organismes convertissent l'énergie solaire en une énergie chimique essentielle au développement. Il existe différents types de plastes dans lesquels peuvent se dérouler la photosynthèse, ils sont chacun propres à un phylum donné.

Les plus connus et les plus étudiés sont les chloroplastes. Caractéristiques des plantes supérieures et des algues vertes, ils sont en fait de plusieurs types en fonction de la lignée considérée. Bien qu'ils soient tous caractérisés par la présence de chlorophylles (a et b ou c) et de thylakoïdes organisés en lamelles, certaines différences existent. Les chloroplastes des chromistes possèdent 4 membranes et un nucléomorphe, une relique du noyau de l'algue rouge capturée par un eucaryote non photosynthétique, celui-ci est situé entre les deux membranes d'origine et les deux nouvelles membranes. Les euglenophytes, qui proviennent de l'absorption d'une algue verte par un eucaryote non photosynthétique, possèdent 3 membranes.

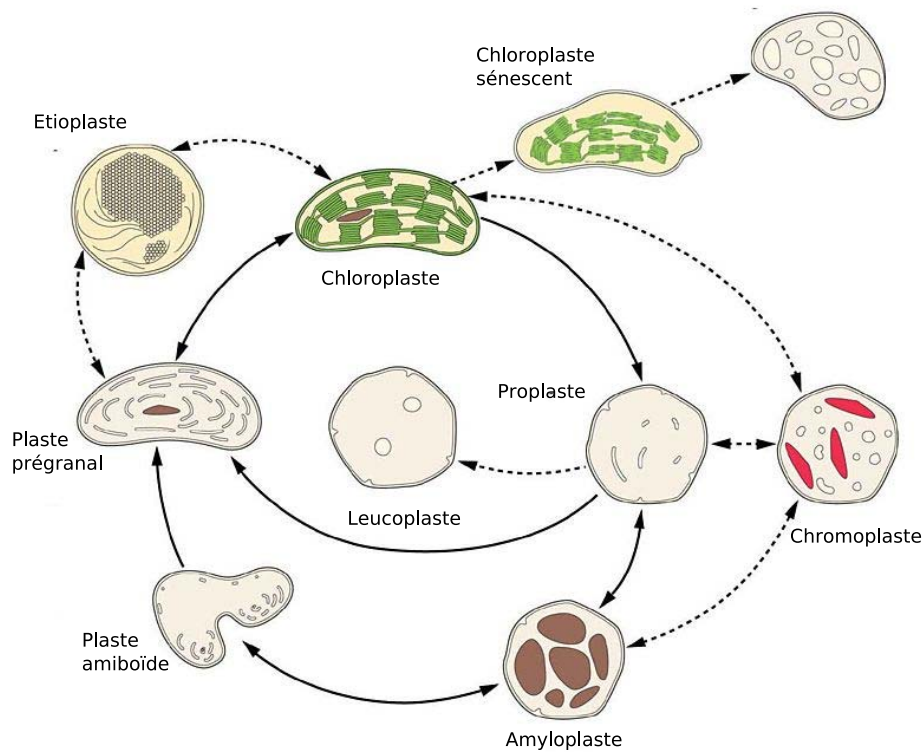
On trouve les muroplastes chez les glaucophytes, ils sont caractérisés par des thylakoïdes libres et concentriques comportant des carboxysomes (regroupement de nombreuses molécules de rubisco). Les muroplastes sont communément considérés comme les plastes les plus proches de la cellule bactérienne originelle.

Chez les rhodophytes, les plastes photosynthétiques sont nommés rhodoplastes, ils sont caractéristiques de la "lignée rouge" des plastes. Les rhodoplastes contiennent de la chlorophylle a (mais ni chlorophylle b ni chlorophylle c), mais leur couleur est généralement donnée par la présence en grande quantité d'autres pigments (phycocyanine et phycoérythrine). Les deux sous-unités de la rubisco des rhodophytes sont codées par le génome chloroplastique, à la différence de la situation chez les chlorophytes. Les rhodoplastes se distinguent aussi par la production d'un amidon modifié, le floridean, stocké dans le cytoplasme.

Il existe un cas de figure très particulier dans lequel un eucaryote "utilise" pendant une longue période un chloroplaste qu'il a "volé" à une algue, ce phénomène est appelé kleptoplastie. Les kleptoplastes ainsi dérobés peuvent être conservés plusieurs mois dans leur "nouvel" hôte, c'est à dire être totalement fonctionnels aussi bien pour la fixation de carbone que la production d'oxygène, et pour la plupart des mécanismes génétiques de l'organite (transcription/traduction). Les mécanismes mis en jeu dans cette "symbiose" sont encore largement incompris mais pourraient donner des informations très importantes pour la compréhension de l'origine des plastes (Rumpho *et al.* 2001).

### I.1.2.3 Les plastes non photosynthétiques

Les plastes ont la particularité d'évoluer et de se différencier au cours du développement de l'organisme considéré. Chaque étape de différenciation est bien définie (Figure 5), ainsi les plastes non différenciés sont les proplast, retrouvés dans les gamètes femelles, les méristèmes et les cellules embryonnaires. Les plastes sont ensuite différenciés pour des fonctions particulières, la photosynthèse dans les chloroplastes, l'assimilation de l'azote dans les nodules, le stockage de métabolites dans les leucoplastes. La différenciation des proplast en chloroplastes peut être interrompue au stade étioplaste dans les tissus à l'obscurité, ces plastes ne constituent pas une classe à proprement parler mais se révèlent être un très bon modèle pour l'étude du développement des chloroplastes. Dans le cas des leucoplastes trois types différents existent suivant le produit concerné : les amyloplast sont spécialisés dans la synthèse et l'accumulation d'amidon, les élaioplastes, spécialisés dans la production de lipides et les protéinoplast, très riches en protéines. Alors que les amyloplast et les élaioplastes sont des plastes différenciés au sens propre puisque leurs métabolismes sont totalement orientés sur leurs fonctions, les protéinoplast n'ont pas encore été montrés comme étant spécifiquement dévoués à la synthèse et le stockage de protéines. Les chromoplastes sont des plastes très riches en différents types de caroténoïdes, ce qui les rend très colorés, on les trouve principalement dans les fruits, les fleurs et les feuilles. Les chromoplastes peuvent provenir de la différenciation de chloroplastes, d'amyloplast ou directement à partir de proplastides. La dernière catégorie de plastes est la plus énigmatique car elle est trouvée dans une famille de vers parasites, les apicomplexas, dont le représentant le plus connu est l'agent de la malaria *Plasmodium falciparum*, il s'agit des apicoplastes. Le rôle et l'origine des apicoplastes sont très obscures, mais une meilleure compréhension du métabolisme de ce plaste un peu particulier pourrait permettre l'élaboration de nouveaux traitements anti-malaria (Waller and McFadden 2005; Vaishnava and Striepen 2006).



**Figure 5 : Différenciation des plastes**

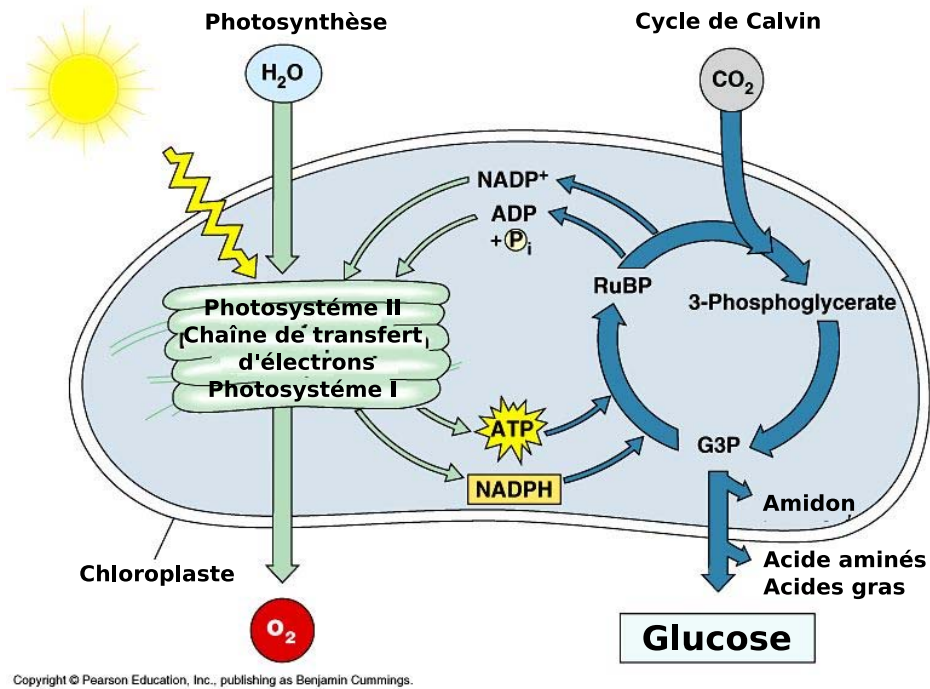
Les différentes possibilités de différenciation des plastes,  
 Buchanan *et al.* 2000 *Biochemistry and Molecular Biology of Plants*

#### I.1.2.4 Métabolisme

Les métabolismes se déroulant dans les chloroplastes sont à la base de l'autotrophie végétale. Au premier plan, la photosynthèse, elle confère aux plantes la capacité de transformer l'énergie lumineuse en énergie chimique directement utilisable, elle permet également la production de dioxygène.

Le cycle de Calvin (ou étape de fixation du carbone) est le second métabolisme essentiel se déroulant dans les chloroplastes. L'énergie, sous forme d'ATP, et un coenzyme d'oxydoréduction NADP, ou NADPH, sont utilisés par le cycle de Calvin pour fixer le carbone prélevé dans le CO<sub>2</sub>. Un intermédiaire du cycle, le glycéraldéhyde-3-phosphate (G3P) est ensuite utilisé pour la synthèse des acides aminés ou acides gras, ou stocké sous forme d'amidon. L'enzyme centrale de ce métabolisme est la rubisco.

L'autotrophie des végétaux est aussi assurée par leur capacité à synthétiser tous les acides aminés, pour cela les plantes sont capables de fixer l'azote et le soufre présents dans le sol, et ainsi synthétiser les acides aminés primaires que sont la glutamine et l'aspartate, puis la cystéine.



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Figure 6 : Photosynthèse et cycle de Calvin

Le lien entre la photosynthèse et le cycle de Calvin est illustré ici par l'utilisation d'ATP et de NADPH (produit par la chaîne de transfert d'électrons) par le cycle de Calvin, celui-ci produisant du NADP<sup>+</sup> et de l'ADP.

## I.2 Les génomes cytoplasmiques

La diminution du nombre de gènes dans les génomes des organites s'explique par un transfert massif de gènes vers le génome de l'hôte, mais il n'est pas évident que tous les gènes ayant "disparus" des génomes cytoplasmiques aient été transférés, une grande partie a simplement été "perdue". En théorie il devrait être possible de reconstituer le génome originel en cumulant les gènes conservés dans les mitochondries et en retrouvant ceux qui ont été transférés vers le noyau, mais cette étape se révèle difficile à effectuer. En effet, les génomes eucaryotes actuels comportent de nombreux gènes de type eubactérien avéré. Certains sont typiques des  $\alpha$ -protéobactéries, ou des cyanobactéries chez les plantes, dans ce cas il est peu risqué d'estimer que ces gènes proviennent des endosymbiontes originaux. Mais des centaines de gènes ne sont spécifiques d'aucune de ces familles, et dans ce cas peuvent tout aussi bien provenir de l'hôte ancestral, s'il était une actinobactérie, comme le suggère Thomas Cavalier-Smith (Cavalier-Smith 2006), ou bien par transfert horizontal. Dans ce cas, de nombreuses combinaisons sont possibles : avant la survenue de l'endosymbiose originelle, par transferts entre eubactéries ou entre eubactérie et archéobactérie ; ou après l'endosymbiose, par transfert entre eubactérie et eucaryote, ou entre archéobactérie et eucaryote ou bien entre eucaryotes (Archibald *et al.* 2003; Embley and Martin 2006).

## I.2.1 Le génome mitochondrial

Le génome mitochondrial des eucaryotes actuels est le vestige du génome de l'endosymbionte initial, une  $\alpha$ -protéobactérie assurément, de la famille des *Rickettsiaceae* très probablement. Cependant il a subi de très nombreuses modifications, de manière flagrante au niveau de la taille puisque le génome d'une bactérie du genre *Rickettsia* fait environ 1 mégabase et comporte près de 1000 gènes, alors que le génome de la mitochondrie par exemple chez l'homme ne fait que 16kb (Ingman *et al.* 2000), et celui d'*Arabidopsis* 367kb (Unsel *et al.* 1997). Les génomes mitochondriaux sont particulièrement hétérogènes, certains comme ceux des chaetognathes (vers sagittés, métazoaires constituant une grande part du plancton) ne conservant qu'une dizaine de gènes sans aucun tRNA sur à peine 11kb (Faure and Casanova 2006), d'autres comme celui du maïs dépassant les 700kb et portant plus de 100 gènes (Fauron *et al.* 1995; Clifton *et al.* 2004).

D'importantes modifications ont également eu lieu au niveau de sa composition fine ou de son fonctionnement. Ainsi la majorité des génomes mitochondriaux connus sont transcrits par une ARN polymérase de type virale et non bactérienne. Seule exception connue à ce jour le génome mitochondrial de *Reclinomonas americana* qui contient les séquences de sous-unités d'une ARN polymérase de type bactérien (Lang *et al.* 1997).

### I.2.1.1 Composition du génome mitochondrial d'*Arabidopsis*

Chez *Arabidopsis thaliana*, le génome mitochondrial fait 367kb, il comporte en tout 60 gènes, 33 gènes codant pour des protéines, 3 gènes codant pour les ARN ribosomiaux et 21 ARN de transfert (Unsel *et al.* 1997). Ces gènes ne couvrent que 10% du génome, 10% supplémentaires sont représentés par 74 ORFs de plus de 300 paires de bases sans homologie particulière avec des gènes connus, et pour lesquels, pour l'instant, aucune protéine produite n'a jamais été observée. Les gènes sont souvent organisés en petites unités de transcription, ce qui confère un aspect très particulier au génome mitochondrial. Celui-ci présente une faible densité de gènes, environ un gène pour 8kb, et le regroupement de ces gènes, sur de petites séquences, crée des zones du génome pouvant aller jusqu'à 27kb ne comportant que des ORFs putatives ou quelques ARNt.

Trois zones répétées couvrent 7% du génome, 8% se retrouvent dans les introns, 5% représentent des traces de transposons ou sont d'origine chloroplastique, laissant donc 60% du génome sans origine ou fonction évidente. La collection d'ARN de transfert n'est pas complète, elle ne couvre que 14 acides aminés et 22 codons, les ARNt correspondant aux six acides aminés restant doivent donc être importés du noyau (alanine, arginine, histidine,



leucine, phénylalanine, thréonine et valine). Il est à noter que sur les 22 ARNt compris dans le génome mitochondrial, 4 sont d'origine chloroplastique (Unsold *et al.* 1997).

Sur les 33 gènes codant pour des protéines, 6 codent pour des protéines ribosomales, 1 pour une maturase, 1 protéine n'a pas encore de fonction connue (*mttB*), tous les autres gènes codent pour des sous-unités de la chaîne de transfert d'électrons ou pour la synthèse du cytochrome C. Il n'y a donc aucun gène relatif à la transcription, ni ARN polymérase ni facteur de transcription, toutes les protéines impliquées dans ces processus doivent donc être importées du noyau.

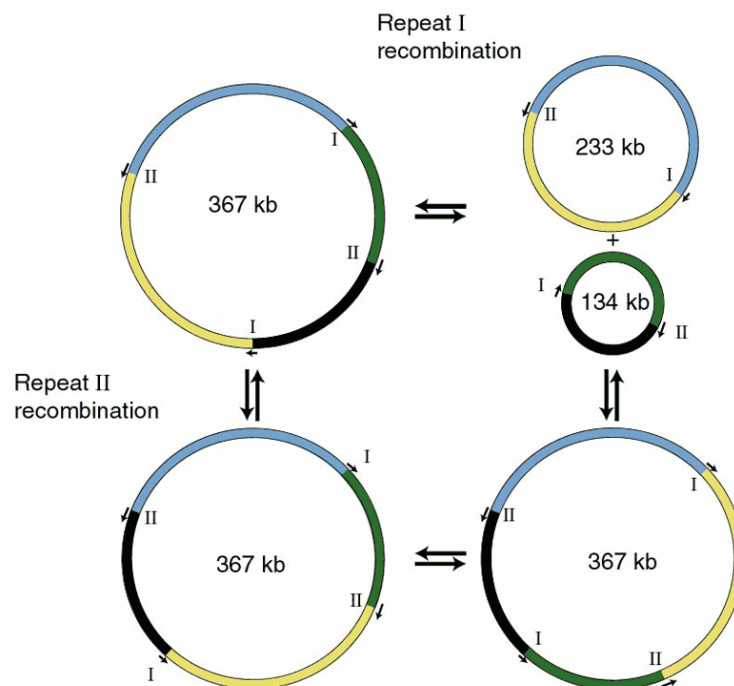
	Gènes	Promoteur	Edition	Epissage		Ref.
				cis	trans	
<b>Complex I</b>	nad1		24	1	3	
	nad2		31	3	1	
	nad3		12			
	nad4		32	3		
	nad4L		9			
	nad5		27	2	2	
	nad6		10			
	nad7		28	4		
	nad9		7			
<b>Complex III</b>	cob		7			
<b>Complex IV</b>	cox1	-355	-			(Kühn <i>et al.</i> 2005).
	cox2	-683 -481 -210	15	1		(Kühn <i>et al.</i> 2005).
	cox3		8			
<b>Complex V</b>	atp1	-1947 -1898	5			(Kühn <i>et al.</i> 2005).
	atp4 (orf25)a		8			
	atp6-1	-916/913 -200 -156	1			(Kühn <i>et al.</i> 2005).
	atp6-2	-436 -507	1			(Kühn <i>et al.</i> 2005).
	atp8 (orfB)b	-999 -710 -228/226 -157	0			(Kühn <i>et al.</i> 2005).
	atp9	-652 -487 -295 -239 -148	4			(Kühn <i>et al.</i> 2005).
<b>Cytochrome-c-biogenesis</b>	ccmB		39			
	ccmC		28			
	ccmFC		16	1		
	ccmFN1		22			
	ccmFN2		12			
<b>Large ribosomal subunit</b>	rpl2		1	1		
	rpl5		10			
	pseudo-rpl16		8			
<b>Short ribosomal subunit</b>	rps3	-1133 -1053	10	1		(Kühn <i>et al.</i> 2005).
	rps4		15			
	rps7		0			
	rps12		8			
	pseudo rps14		1			
<b>Ribosomal RNAs</b>	rrn18S	-424 -353 -156 -69				(Kühn <i>et al.</i> 2005).
	rrn26S	-893				(Kühn <i>et al.</i> 2005).
	rrn5S					
<b>Other ORFs</b>	orfX(mttB)		24			
	matR		9			
<b>ARN de transfert</b>	ARNt-Asn					
	ARNt-Asp					
	ARNt-Cys					
	ARNt-fMet	-574/573 -98				(Kühn <i>et al.</i> 2005).
	ARNt-Gln					
	ARNt-Glu					
	ARNt-Gly					
	ARNt-Ile					
	ARNt-Lys		2			
	ARNt-met					
	ARNt-Pro					
	ARNt-Ser		5			
ARNt-Trp						
ARNt-Tyr		3				

Tableau 1 : Composition du génome mitochondrial d'Arabidopsis

La position des promoteurs est indiquées relativement au codon d'initiation de la traduction, la colonne Edition présente le nombre de site d'édition par gène.

### I.2.1.2 Structure du génome mitochondrial

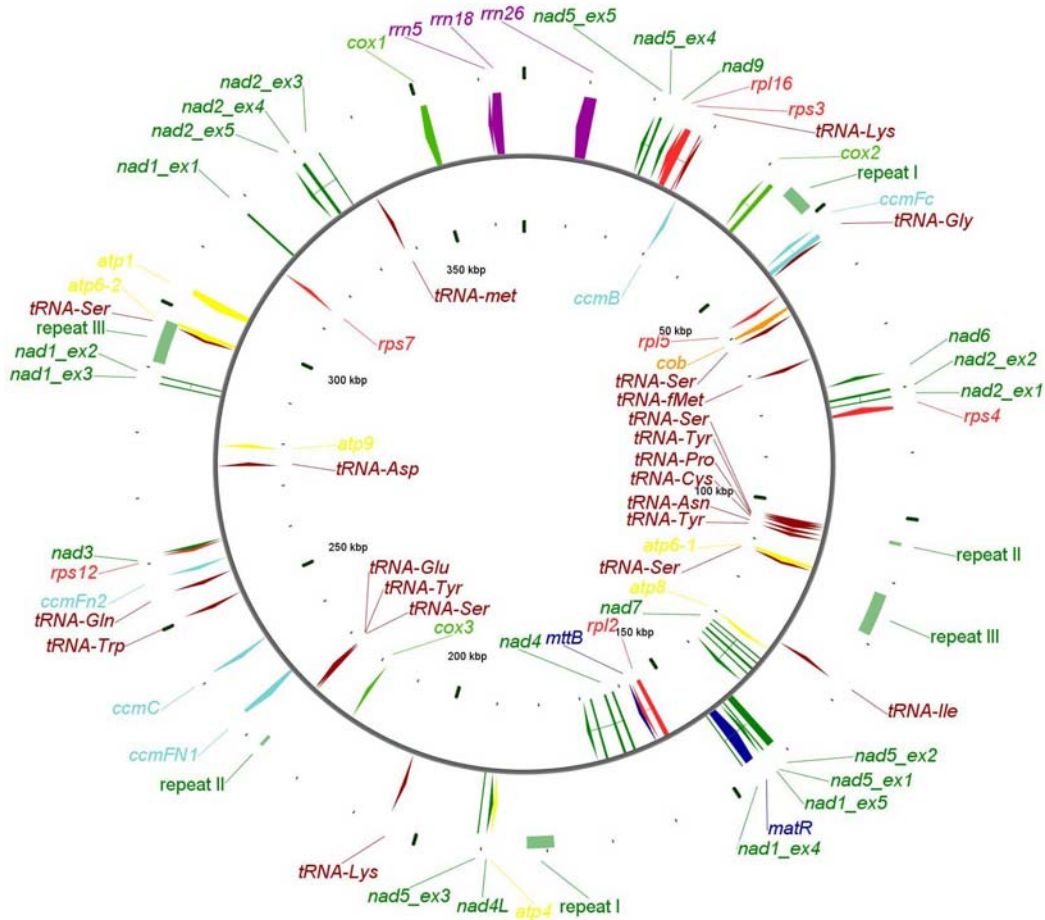
Le génome mitochondrial est extrêmement dynamique, il est majoritairement hétéroplasmique, c'est-à-dire qu'il coexiste différentes formes au sein de la même mitochondrie (Kmiec *et al.* 2006). Il peut être trouvé sous forme circulaire ou linéaire, double ou simple brin. Ces différentes formes sont, entre autre, les produits de la réplication du génome mitochondrial par un mécanisme de cercle roulant, mais aussi d'un mécanisme de réplication recombinaison-dépendant, similaire à la réplication du phage T4 (Backert *et al.* 1997; Backert 2000; Backert and Borner 2000). Les génomes mitochondriaux sont habituellement représentés sous forme circulaire, le "cercle maître" qui correspond à la molécule décrivant le mieux le génome. Chez *Arabidopsis* cette molécule fait 367kb, elle est riche en petites séquences répétées, 22 paires de répétitions identiques de plus de 100pb, dont deux majeures de 6,5 et 4,2kb (écotype C24). Seules ces deux régions sont impliquées dans des événements fréquents de recombinaison, créant 3 formes circulaires de 134, 233 et 367kb, Figure 7 (Unsel *et al.* 1997). L'unité répétée II contient la partie N-terminale du gène *atp6*, il existe donc 2 versions de ce gène avec 2 parties C-terminales différentes. Le ratio entre les différentes formes de génome peut varier, mais généralement une forme est majoritaire, les petites molécules d'ADN qui peuvent être produites lors de ces événements sont appelés sublimons, (Small *et al.* 1989).



**Figure 7 : Réarrangements du génome mitochondrial par recombinaison entre séquences répétées**

Les deux zones répétées impliquées sont indiquées par les chiffres romains I et II

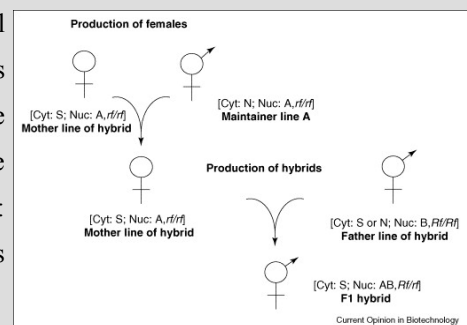
Les événements de recombinaison peuvent également être initiés en dehors des grandes séquences répétées, ces changements peuvent être impliqués dans des phénomènes de stérilité mâle cytoplasmique (Yesodi *et al.* 1995), ou chez les mammifères dans des maladies génétiques (Kajander *et al.* 2000). Chez les plantes, de rapides variations dans les proportions des différentes formes génomiques peuvent se produire, ce phénomène est appelé réarrangement sub-stœchiométrique ("Substoichiometric shifting"). Ces variations peuvent entraîner la perte d'une partie du génome ou au contraire l'amplification de certaines régions. Ceci peut avoir de lourdes conséquences comme l'illustrent les phénomènes de stérilité mâle cytoplasmique (CMS), voir Encart 1. Le contrôle de la transmission des différentes formes génomiques est essentiel à la survie de la plante. Relativement peu de choses sont connues sur cette régulation, il a été montré que, dans des protoplastes, la fission des mitochondries était précédée d'une phase globale de fusion (Sheahan *et al.* 2005). Ce phénomène semble spécifique des cellules en phase de dédifférenciation et permettrait de contrôler la transmission du matériel génétique mitochondrial. La régulation de la transmission des formes génomiques est sous le contrôle de gènes nucléaires, mais pour le moment peu ont été identifiés. MSH1 est important pour la régulation des réarrangements sub-stœchiométriques, mais son mode d'action n'est pas encore compris (Abdelnoor *et al.* 2003). Dans la lignée P2 de maïs, un facteur nucléaire influence la régulation des formes génomiques, et altère ainsi le nombre de copies des gènes mitochondriaux (Kuzmin *et al.* 2005). L'identification d'un nouveau facteur nécessaire à la stabilité du génome mitochondrial, OSB1, identifié par Vincent Zaegel *et al.* est décrite dans ce manuscrit à la fin des résultats (Zaegel *et al.* 2006).



**Figure 8 : Carte du génome mitochondrial d'Arabidopsis**

Les gènes sont indiqués en utilisant un code couleur relatif à la fonction de leur produit. bordeaux : ARNt, violet : ARN ribosomaux, rouge : Protéines ribosomales, vert intense : Complexe I, orange : Complexe III, vert : Complexe IV, jaune : Complexe V, bleu cyan : Synthèse du cytochrome C, vert pâle : Grandes séquences répétées, bleu marine : autres fonctions

La stérilité mâle cytoplasmique est un phénomène par lequel une plante dioïque se retrouve dans l'incapacité de produire des gamètes mâles fertiles. Le terme cytoplasmique évoque le fait que le phénotype est provoqué par un facteur porté par le génome mitochondrial. Ce processus a deux intérêts majeurs en agronomie : faciliter la production de descendance hybride chez des plantes naturellement autogame, et bénéficier ainsi du phénomène de vigueur hybride ou hétérosis, mais aussi pour les semenciers de distribuer des plantes ne pouvant donner une descendance utilisable pour un nouveau semis. Dans la nature le phénomène de stérilité mâle cytoplasmique est contre balancé par l'apparition de facteurs de restauration de la fertilité, codés par le noyau. A l'heure actuelle tous les facteurs connus, sauf un, font partie de la famille des protéines PPR.



**Encart 1 : La stérilité mâle cytoplasmique**

## I.2.2 Le génome chloroplastique

Le génome des chloroplastes est, tout comme celui des mitochondries, extrêmement réduit par rapport à celui de la cyanobactérie à son origine. Mais ses caractéristiques sont par de nombreux points différents de celui des mitochondries. Le génome chloroplastique est généralement plus homogène, réduit aux alentours de 150kb, même si chez certaines algues vertes il peut avoir des tailles extrêmes de 37kb à 220kb (de Koning and Keeling 2006; Pombert *et al.* 2006).

### I.2.2.1 Composition du génome chloroplastique

Les chloroplastes ont généralement conservé une centaine de gènes dans un génome d'une taille comprise entre 100 et 200kb, donnant une densité de gènes nettement plus élevée que dans le génome mitochondrial (Figure 8 et Figure 9). Les principales fonctions couvertes par les gènes codés par le génome chloroplastique sont : l'expression de ce génome (ARN polymérase, ARN ribosomiaux, protéines ribosomiales, ARNt), la photosynthèse (avec de nombreuses sous-unités des photosystème I et II, du cytochrome *b6f*, de l'ATPsynthase et la NADH oxydoréductase) et la fixation du carbone avec la rubisco. Arabidopsis possède une dizaine de gènes complémentaires à ces fonctions : *ycf1* et *ycf2* dont la fonction n'est pas encore connue mais pour lesquels nous savons déjà qu'ils sont essentiels (Drescher *et al.* 2000), *ycf3* et *ycf4* impliqués dans l'assemblage du photosystème I (Boudreau *et al.* 1997; Ruf *et al.* 1997; Naver *et al.* 2001, Boudreau, 1997 #342), *clpP1* ("caseinolytic protein") codant pour une sous-unité catalytique de la protéase ATP-dépendante Clp (complexe ClpP/R), *accD* codant pour une sous-unité de l'Acétyl-CoA carboxylase essentielle au contrôle de la synthèse des acides gras. Le génome chloroplastique a largement conservé la trace de son origine bactérienne avec une organisation des gènes en unités de transcription polycistronique, une quinzaine réparties sur le génome, pouvant comporter jusqu'à sept gènes dans le cas de l'opéron *ndhH-ndhA-ndhI-ndhG-ndhE-psaC-ndhD* (Del Campo *et al.* 2002). Les opérons bactériens représentent généralement une association de gènes co-régulés associés à la même fonction. Dans les chloroplastes cette notion est relativement conservée, mais plusieurs exceptions existent. Dans l'exemple précédent, le gène *psaC*, codant pour une protéine du photosystème I, est co-transcrit avec six sous-unités de la NADH oxydoréductase, mais des mécanismes de contrôle post-transcriptionnel complexes amènent à ce que la quantité de transcrit mature de *psaC* soit nettement supérieure à celle des autres gènes de l'unité de transcription (Del Campo *et al.* 2002). Les génomes chloroplastiques sont également caractérisés au niveau de leur organisation par la présence de deux séquences répétées

inversées de plusieurs kilobases.

	Gènes	Promoteur		Edition	Epissage		Opéron	référence
		PEP	NEP		cis	trans		
<b>PSI</b>	psaA	-188					psaA-psaB-rps14	(Lezhneva and Meurer 2004).
	psaB						psaA-psaB-rps14	(Lezhneva and Meurer 2004).
	psaC						ndhHALGE-psaC-ndhD	(Del Campo <i>et al.</i> 2002).
	psal						<i>psal-ycf4-cemA-petA</i>	
	psaJ						petL-petG-psaJ	(Yamazaki <i>et al.</i> 2004).
<b>PSII</b>	psbA	-77					-	
	psbB						psbB-psbT-psbH-petB-petD	(Meurer <i>et al.</i> 1996).
	psbC						psbD-psbC-psbZ	(Meurer <i>et al.</i> 1996).
	psbD	-950 -550 -190					psbD-psbC-psbZ	(Meurer <i>et al.</i> 1996; Hoffer and Christopher 1997).
	psbE			1			psbE-psbF-psbL-psbJ	
	psbF			1			psbE-psbF-psbL-psbJ	
	psbH						psbB-psbT-psbH-petB-petD	(Meurer <i>et al.</i> 1996).
	psbI						<i>psbK-psbI</i>	
	psbJ						psbE-psbF-psbL-psbJ	
	psbK						<i>psbK-psbI</i>	
	psbL						psbE-psbF-psbL-psbJ	
	psbM						-	
	psbN						-	
	psbT						psbB-psbT-psbH-petB-petD	(Meurer <i>et al.</i> 1996).
psbZ						psbD-psbC-psbZ	(Meurer <i>et al.</i> 1996).	
<b>Cytochrome b6f</b>	petA						<i>psal-ycf4-cemA-petA</i>	
	petB				1		psbB-psbT-psbH-petB-petD	(Meurer <i>et al.</i> 1996).
	petD	-30			1		psbB-psbT-psbH-petB-petD	(Meurer <i>et al.</i> 1996).
	petG						petL-petG-psaJ	(Yamazaki <i>et al.</i> 2004).
	petL			1			petL-petG-psaJ	(Yamazaki <i>et al.</i> 2004).
	petN						-	
<b>ATPase I</b>	atpA						atpI-atpH-atpF-atpA	
	atpB	-520	-318				atpE-atpB	(Swiatecka-Hagenbruch <i>et al.</i> 2007).
	atpE						atpE-atpB	
	atpF			1	1		atpI-atpH-atpF-atpA	
	atpH						atpI-atpH-atpF-atpA	
	atpI	-229					atpI-atpH-atpF-atpA	(Swiatecka-Hagenbruch <i>et al.</i> 2007).
<b>PEP</b>	rpoA			1			<i>rpoA-rps11-rpl36-rps8-rpl14</i>	
	rpoB		-300	3			rpoB-rpoC1-rpoC2	(Swiatecka-Hagenbruch <i>et al.</i> 2007).
	rpoC1				1		rpoB-rpoC1-rpoC2	
	rpoC2						rpoB-rpoC1-rpoC2	
<b>Acetyl coA carboxylase</b>	accD		-252 -172	1			-	(Swiatecka-Hagenbruch <i>et al.</i> 2007).
<b>Inorganic carbon uptake</b>	cemA						<i>psal-ycf4-cemA-petA</i>	
<b>Protease</b>	clpP	-115	-58	1	2		-	
<b>Cyt c biogenesis</b>	cssA						-	
<b>Intron splicing</b>	matK			1			-	
	ycf1	-34	-104 -39				-	
	ycf2						-	
<b>Assembly of PSI</b>	ycf3				2		**	
<b>Assembly of PSI</b>	ycf4						<i>psal-ycf4-cemA-petA</i>	

Tableau 2 : Composition du génome chloroplastique d'Arabidopsis (Partie 1)

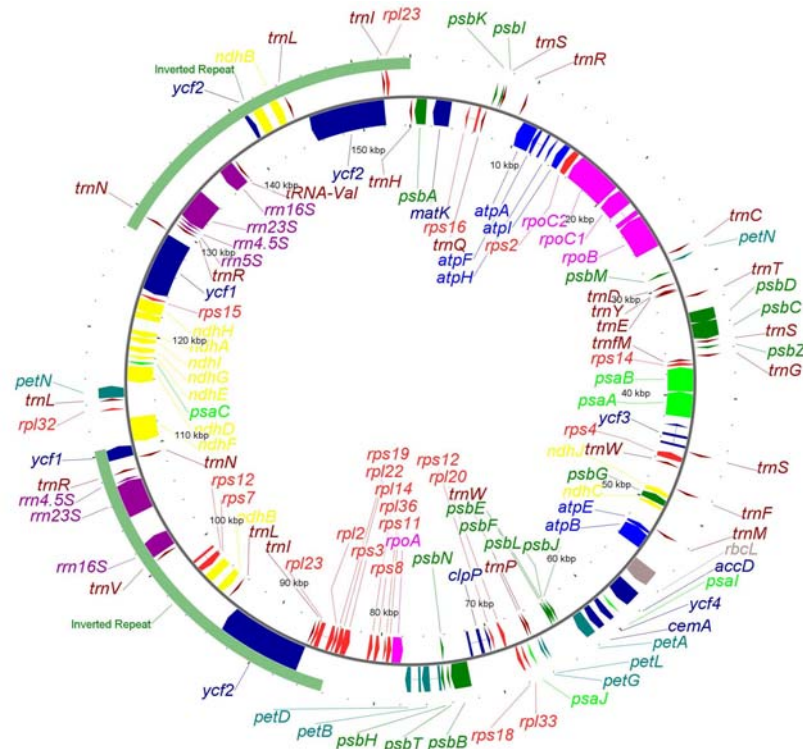
La position des promoteurs est indiquée relativement au codon d'initiation de la traduction, la colonne édition présente le nombre de site d'édition par gène.

	Gènes	Promoteur		Edition	Epissage		Opérons	ref
		PEP	NEP		cis	trans		
<b>Short ribosomal subunit</b>	rps2						-	
	rps3						<i>rpl16-rps3-rpl22-rps19-rpl2-rpl23</i>	
	rps4	-123	-151				-	(Swiatecka-Hagenbruch et al. 2007).
	rps7\$						<i>rps12-rps7-ndhB</i>	(Hashimoto et al. 2003).
	rps8						<i>rpoa-rps11-rpl36-rps8-rpl14</i>	
	rps11						<i>rpoa-rps11-rpl36-rps8-rpl14</i>	
	rps12*\$				1	1	<i>rps12-rps7-ndhB</i>	(Hashimoto et al. 2003).
	rps14			2			<i>psaA-psaB-rps14</i>	
	rps15		-139				-	
	rps16				1		-	
	rps18						<i>rpl33-rps18</i>	(Magee and Kavanagh 2002).
	rps19						<i>rpl16-rps3-rpl22-rps19-rpl2-rpl23</i>	
<b>Large ribosomal subunit</b>	rpl2\$				1		<i>rpl16-rps3-rpl22-rps19-rpl2-rpl23</i>	
	rpl14						<i>rpoa-rps11-rpl36-rps8-rpl14</i>	
	rpl16				1		<i>rpl16-rps3-rpl22-rps19-rpl2-rpl23</i>	
	rpl22						<i>rpl16-rps3-rpl22-rps19-rpl2-rpl23</i>	
	rpl23\$						<i>rpl16-rps3-rpl22-rps19-rpl2-rpl23</i>	
	rpl32						-	
	rpl33						<i>rpl33-rps18</i>	(Magee and Kavanagh 2002).
	rpl36						<i>rpoa-rps11-rpl36-rps8-rpl14</i>	
<b>NADH oxydoreductase</b>	ndhA				1		<i>ndhHAIGE-psaC-ndhD</i>	(Del Campo et al. 2002).
	ndhB		-12	9	1		<i>rps12-rps7-ndhB</i>	(Hashimoto et al. 2003).
	ndhC						<i>ndhC-ndhK-ndhJ</i>	(Steinmuller et al. 1989).
	ndhD		-56 -45 -31	5			<i>ndhHAIGE-psaC-ndhD</i>	(Del Campo et al. 2002).
	ndhE						<i>ndhHAIGE-psaC-ndhD</i>	(Del Campo et al. 2002).
	ndhF			1			-	
	ndhG			1			<i>ndhHAIGE-psaC-ndhD</i>	(Del Campo et al. 2002).
	ndhH						<i>ndhHAIGE-psaC-ndhD</i>	(Del Campo et al. 2002).
	ndhI						<i>ndhHAIGE-psaC-ndhD</i>	(Del Campo et al. 2002).
	ndhJ						<i>ndhC-ndhK-ndhJ</i>	(Steinmuller et al. 1989).
	ndhK						<i>ndhC-ndhK-ndhJ</i>	(Steinmuller et al. 1989).
<b>rubisco</b>	rbcL		-212				-	
<b>ribosomal RNAs</b>	rrn4.5S\$						<i>rrn16S-ARNt-Ile(GAU)-ARNt-Ala(UGC)-rrn23S-rrn4.5s-rrn5S</i>	
	rrn5S\$						<i>rrn16S-ARNt-Ile(GAU)-ARNt-Ala(UGC)-rrn23S-rrn4.5s-rrn5S</i>	
	rrn16S\$	-112	-139				<i>rrn16S-ARNt-Ile(GAU)-ARNt-Ala(UGC)-rrn23S-rrn4.5s-rrn5S</i>	
	rrn23S\$						<i>rrn16S-ARNt-Ile(GAU)-ARNt-Ala(UGC)-rrn23S-rrn4.5s-rrn5S</i>	
<b>ARN de transferts</b>	ARNt-Ala (UGC)\$				1		<i>rrn16S-ARNt-Ile(GAU)-ARNt-Ala(UGC)-rrn23S-rrn4.5s-rrn5S</i>	
	ARNt-Gly (UCC)				1			
	ARNt-Ile (GAU)\$				1		<i>rrn16S-ARNt-Ile(GAU)-ARNt-Ala(UGC)-rrn23S-rrn4.5s-rrn5S</i>	
	ARNt-Leu (CAA)\$				1			
	ARNt-Lys (UUU)\$				1			
	ARNt-Val (UAC)				1			

**Tableau 3 : Composition du génome chloroplastique d'Arabidopsis Partie 2**

Liste des autres ARNt : ARNt-Arg (ACG)\$ ARNt-Arg (UCU) ARNt-Asn (GUU)\$ ARNt-Asp (GUC) ARNt-Cys (GCA) ARNt-fMet (CAU) ARNt-Gln (UUG) ARNt-Glu (UUC) ARNt-Gly (GCC) ARNt-His (GUG) ARNt-Ile (CAU)\$ ARNt-Leu (UAA) ARNt-Leu (UAG) ARNt-Met (CAU) ARNt-Phe (GAA) ARNt-Pro (UGG) ARNt-Ser (GCU) ARNt-Ser (GGA) ARNt-Ser (UGA) ARNt-Thr (GGU) ARNt-Thr (UGU) ARNt-Trp (CCA) ARNt-Tyr (GUA) ARNt-Val (GAC)\$





**Figure 9 : Carte du génome chloroplastique**

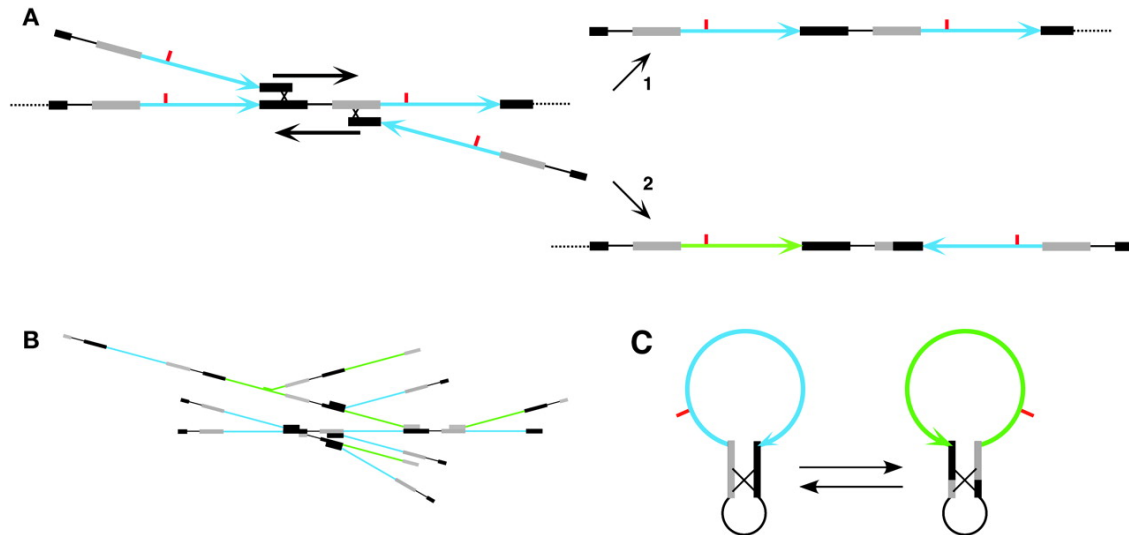
Les gènes sont indiqués en utilisant un code couleur relatif à la fonction de leur produit. bordeaux : ARNt, violet : ARN ribosomiaux, rouge : Protéines ribosomiales, vert vif : Photosystème I, vert foncé : Photosystème II, fuchsia : Polymérase, bleu-vert : Cytochrome *b6f*, jaune : NADH oxydoréductase, gris : rubisco, vert pâle : séquences répétées, bleu marine : autres fonctions

### I.2.2.2 Structure du génome chloroplastique

Une cellule de feuille peut contenir de 50 à 100 chloroplastes, chacun pouvant comporter jusqu'à 100 équivalents génome regroupés en structures appelées nucléoides par analogie aux systèmes procaryotes (Bendich 1987). Le nombre de copies du génome est régulé en fonction des tissus ou de la spécialisation du plaste, jusqu'à 5 fois moins de copies dans un amyloplaste que dans un chloroplaste (Aguettaz *et al.* 1987).

Contrairement à ce que pourrait laisser croire la taille et la simplicité de la séquence du génome, celui-ci est en fait une entité extrêmement complexe aussi bien au niveau de sa structure que de sa répartition. Le génome chloroplastique est généralement représenté sous forme circulaire, mais cela ne représente qu'une infime proportion de l'ADN chloroplastique, et *a priori* plutôt une structure intermédiaire lors de la réplication, la majorité du matériel génétique se retrouvant sous forme linéaire ou ramifiée (Bendich 2004; Oldenburg and Bendich 2004). Le mode de réplication du génome chloroplastique n'est pas encore totalement élucidé. Kolodner et Tewari ont proposé en 1975 que le processus en jeu soit un mécanisme proche du cercle roulant (Kolodner and Tewari 1975). Mais le fait que le génome

chloroplastique comporte plusieurs origines de réplication possibles (Wang *et al.* 2003) et les progrès des méthodes de visualisation tendent à montrer que d'autres modes de réplication doivent exister, comme un mécanisme initié par la recombinaison entre deux molécules linéaires au niveau des zones répétées (Figure 10.)

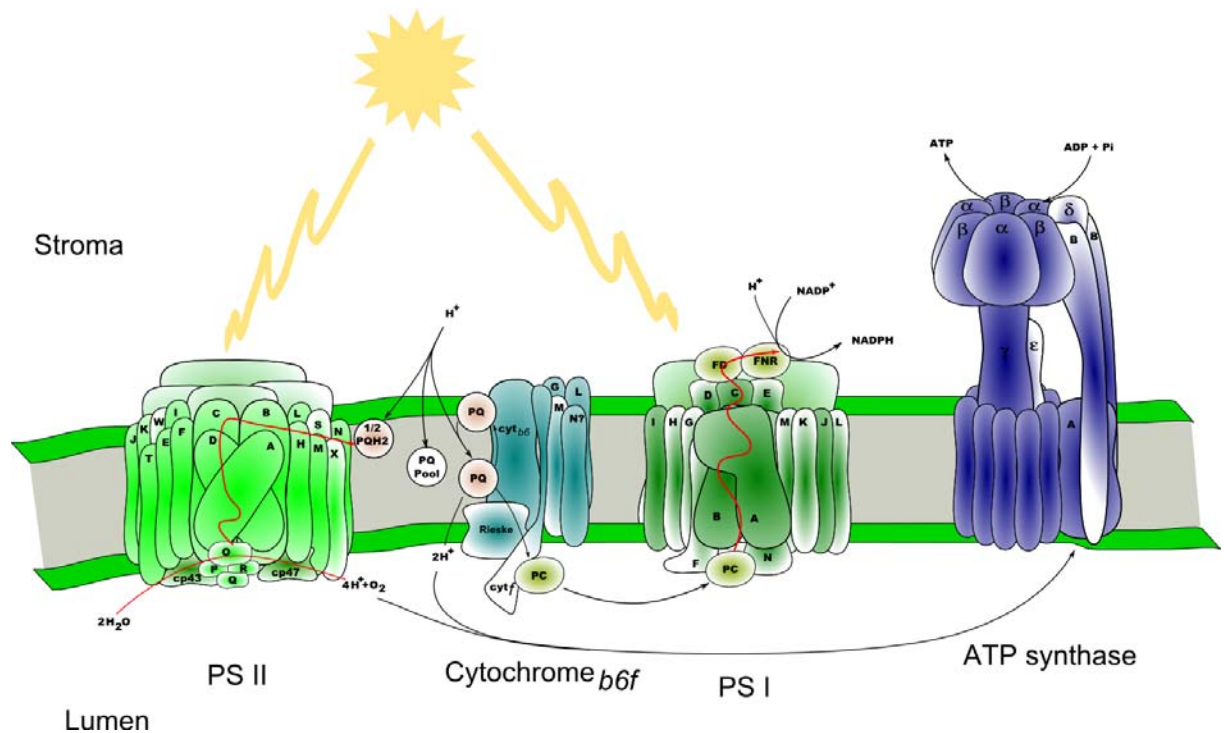


**Figure 10 : Structure du génome chloroplastique et réplication recombinaison-dépendante**

(A) L'initiation de la réplication se fait par la recombinaison entre deux molécules de génome. La réplication se déroule dans le sens du génome, 1. Un second événement de recombinaison initie une réplication en sens inverse créant une molécule contenant une séquence inversée, 2. Les régions répétées inverses sont indiquées par les rectangles gris et noirs. (B) La succession de réplication initiée par recombinaison mène à la formation de structure multigénomique. (C) Des recombinaisons intramolécule créent des formes circulaires du génome (Bendich 2004).

### I.3 Expression des gènes dans les organites

Les organites sont le siège de nombreux métabolismes essentiels pour les plantes, ces métabolismes font intervenir un très grand nombre de facteurs protéiques provenant pour la plupart du noyau, mais certains processus comme la photosynthèse ou la respiration mitochondriale (Figure 12) sont grandement dépendants de l'expression de quelques gènes encore localisés au niveau des organites. L'ensemble des complexes protéiques nécessaires à la photosynthèse comprennent 53 sous-unités protéiques différentes dont seulement 26 sont produites en dehors du chloroplaste (Figure 11). L'assemblage des différents complexes protéiques présents dans les organites est une étape nécessitant la coordination de plusieurs systèmes génétiques différents.

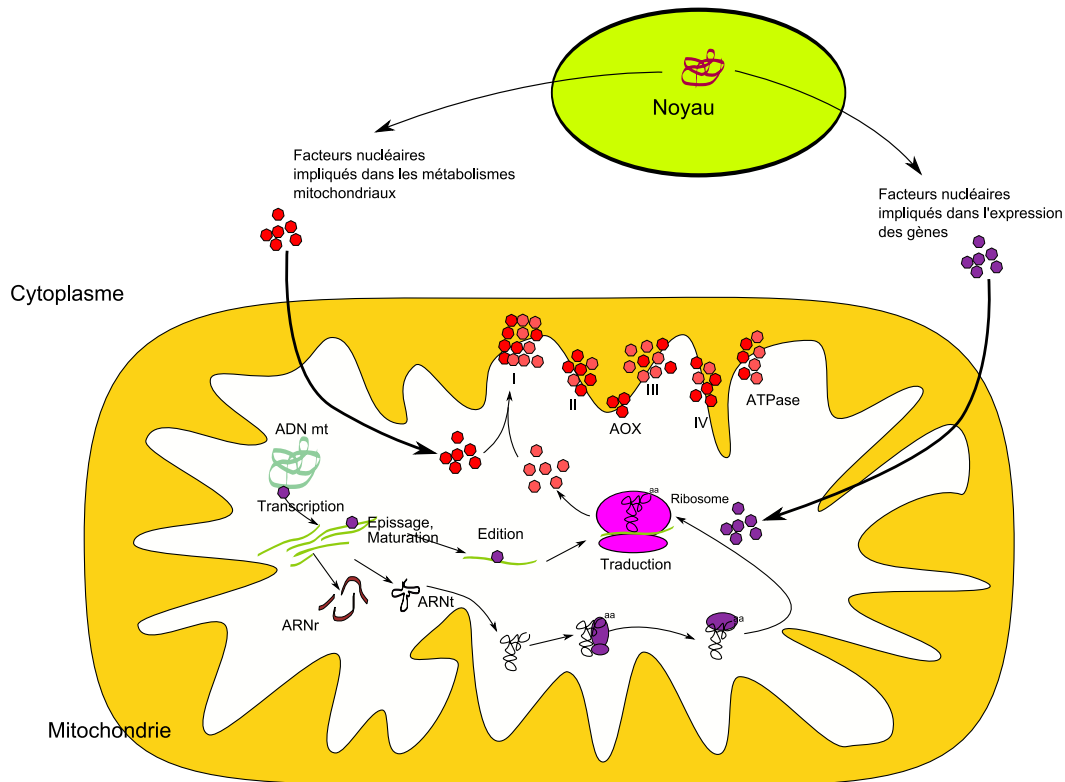


**Figure 11 : Composition des complexes photosynthétiques**

Les protéines codés par le génome chloroplastique sont indiquées en couleurs foncées (PsaA = A vert foncé, PsbA= A vert clair), celles produites par le noyau sont indiqués en couleur claires (PsaH = H vert foncé, PsbO= O vert clair).

La régulation de la production des protéines au niveau des organites se fait essentiellement par l'intermédiaire d'un contrôle précis de l'expression des gènes; celui-ci nécessite un grand nombre de facteurs, pour la plupart codés au niveau du génome nucléaire (Barkan and Goldschmidt-Clermont 2000; Giege and Brennicke 2001). Ces facteurs interviennent à chaque niveau de l'expression des gènes, transcription, édition, épissage, maturation, dégradation, ou encore traduction des ARNm.

Un autre aspect du contrôle de l'expression des gènes peut être illustré par les phénomènes de stérilité mâle cytoplasmique (voir Encart 1), dans ces situations le dynamisme du génome mitochondrial a provoqué l'apparition d'une nouvelle phase codante dans le génome, et le produit de ce nouveau gène est toxique pour la cellule. Ce phénomène illustre la nécessité d'un contrôle très fin au niveau post-transcriptionnel de ce qui peut et doit être traduit en protéines.



**Figure 12 : Expression des gènes mitochondriaux**

Le fonctionnement des mitochondries est totalement dépendant du génome nucléaire. La majorité des facteurs impliqués dans la maintenance et l'expression du génome mitochondrial est d'origine nucléaire (en violet sur le dessin). Au niveau métabolique, toutes les enzymes sont codées par le génome nucléaire, seule une partie des sous-unités des différents complexes de la chaîne respiratoire est encore codée au niveau mitochondrial (en rose), l'assemblage des complexes respiratoires est le fruit d'une coordination fine entre le noyau et les mitochondries.

### I.3.1 La transcription dans la mitochondrie

Un paradoxe important du fonctionnement des mitochondries de plantes est la complexité de la régulation de l'expression des gènes par rapport à la taille de ces génomes. Chez l'homme la transcription des gènes mitochondriaux s'effectue à partir de trois promoteurs de 15 paires de base (1 pour le brin dit léger, et 2 pour le brin dit lourd) à partir desquels de longs ARN poly-cistroniques sont produits, ce processus est effectué par une ARN polymérase de type T7/T3. Les ARN précurseurs sont ensuite coupés de part et d'autre de chaque ARN de transfert pour isoler chaque gène avant la traduction (Fernandez-Silva *et al.* 2003). Chez la levure, le système est un peu différent, la transcription est également effectuée par une polymérase de type T7/T3 (RPO41), mais elle est initiée en plusieurs points (une vingtaine de promoteurs de 9 nucléotides) permettant la production d'ARNm comptant généralement quelques gènes.

Arabidopsis possède trois différentes polymérases de type T7/T3 codées par le noyau

(RpoTm, RpoTp et RpoTmp). RpoTm est exportée dans les mitochondries et RpoTp dans les chloroplastes (Hedtke *et al.* 1997; Hedtke *et al.* 2000; Emanuel *et al.* 2006). Une incertitude demeure sur la réelle localisation de RpoTmp; bien que Kabeya et Sato aient montré en 2005 chez la mousse *Physcomitrella patens* que RpoTmp était uniquement adressée à la mitochondrie, Azevedo *et al.* ont montré l'année suivante une localisation exclusive dans le chloroplaste d'épinard (Azevedo *et al.* 2006); (Kabeya and Sato 2005). Pourquoi les plantes supérieures possèdent 2 polymérase dans leur mitochondrie demeure une énigme, pour l'instant l'analyse de mutant pour RpoTmp n'a pas permis d'identifier de fonction spécifique au niveau mitochondrial (Baba *et al.* 2004).

Comme dans la mitochondrie de levure, la transcription chez les plantes supérieures est initiée à différents points du génome, donnant naissance à des ARN précurseurs comportant un ou plusieurs gènes. L'organisation en unités de transcription n'est pas sans rappeler l'origine bactérienne de ce génome, et les opérons que l'on peut trouver fréquemment chez les bactéries. La différence essentielle est que chez les mitochondries de plantes ces unités transcriptionnelles divergent beaucoup d'un organisme à l'autre et ne comportent pas forcément de gènes ayant des fonctions associées. Une autre particularité est la présence de promoteurs multiples pour beaucoup de gènes : la transcription pouvant être initiée à différents points pour un même gène (Kühn *et al.* 2005) par exemple jusqu'à 5 pour le gène Cox2 (Lupold *et al.* 1999). L'expression des génomes mitochondriaux de plantes est également marquée par la synthèse de transcrits primaires très longs, de longueurs généralement bien supérieures à celle des séquences codantes, mais peu stables sous cette forme, indiquant donc une maturation importante en 5' ou en 3'.

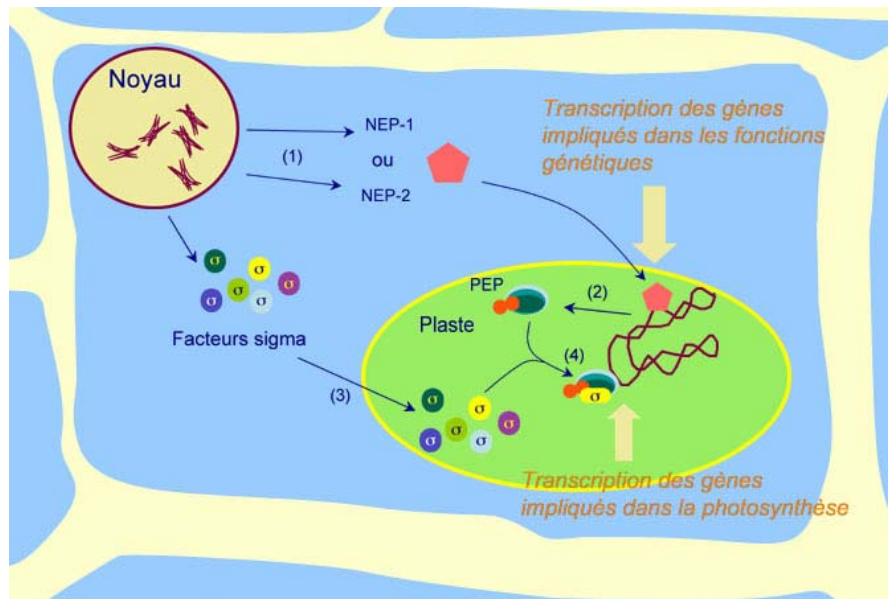
### **I.3.2 La transcription dans le chloroplaste**

Au niveau chloroplastique, la transcription est effectuée par deux types de polymérase différentes, l'une provenant du génome nucléaire (les NEP, pour Nuclear-Encoded plastid ARN Polymerase) et l'autre codée au niveau du génome chloroplastique (Plastid-Encoded plastid ARN Polymerase).

Les NEP représentent en fait les deux polymérase (RpoTp et RpoTmp) codées par le noyau. Les fonctions de RpoTp et RpoTmp ne semblent pas être redondantes puisque une mutation dans l'un des deux gènes entraîne une altération de la transcription chloroplastique et que les deux gènes montrent un profil d'expression différent. Mais les rôles précis de chaque NEP restent encore à déterminer (Baba *et al.* 2004; Emanuel *et al.* 2006).

La PEP est une ARN polymérase de type bactérien constituée de 4 sous-unités, toutes

codées par le génome chloroplastique (*rpoA*, *rpoB*, *rpoC1* et *rpoC2*), la spécificité de cette polymérase requiert l'association de facteurs de type  $\sigma$  codés par le noyau (au nombre de six chez *Arabidopsis*, SIGA, B, C, D et E (Allison 2000; Shiina *et al.* 2005).



**Figure 13 : La transcription dans les chloroplastes**

La transcription dans les chloroplastes est contrôlée par le noyau d'une double manière. La transcription des gènes dits "de ménage" du chloroplaste est effectuée par une (ou deux) polymérase(s) codée(s) par le noyau. Parmi ces gènes, la polymérase en charge de la transcription des gènes de photosynthèse. Cette dernière nécessite l'intervention de facteurs sigma codés au niveau nucléaire.

Les caractéristiques relatives des deux types de polymérases commencent à être comprises: la PEP est essentielle aux fonctions relatives à la photosynthèse comme le montre le phénotype de plantes dénuées d'une des sous-unités, ou de l'un des facteurs sigma associés (Allison *et al.* 1996; Allison 2000; Borukhov and Severinov 2002; Borukhov and Nudler 2003; Kanamaru and Tanaka 2004). Les gènes transcrits par la PEP sont majoritairement précédés par un promoteur semblable au promoteur  $\sigma^{70}$  bactérien, caractérisé par deux séquences consensus -10 (TATAAT) et -35(TTGACA), le niveau de transcription pouvant varier d'un gène à l'autre en fonction de la conservation de la séquence du promoteur. D'un autre côté, la transcription de certains gènes par la PEP se fait à un niveau très élevé, et coïncide avec la présence de promoteurs particuliers. Dans le cas de ces gènes, les deux éléments -10 et -35 sont bien présents mais ne sont pas les seuls éléments requis. Par exemple, le promoteur du gène *psbA* contient une boîte "TATA" entre le -10 et le -35 et un motif "TGn" juste avant le -10, ces deux éléments ne semblent pas être mis en jeu pour la transcription de base de *psbA*, mais être impliqués dans la transcription spécifique induite par la lumière. Ces deux motifs pourraient être reconnus par les facteurs SIGA et/ou SIGB. Les autres gènes

possédant un promoteur PEP particulier sont *psbD*, *rbcL* et les opérons *psaA-psaB-rps14* et ARN ribosomiaux.

Les NEP participent à la transcription de la plupart des gènes, avec quelques spécificités comme *accD* ou l'opéron *rpoBC* exclusivement transcrits par elles. Deux types de promoteurs NEP peuvent être décrits, le type I représente un groupe de promoteurs semblables à ceux trouvés dans la mitochondrie : une séquence consensus d'une dizaine de nucléotides au niveau du site d'initiation de la transcription, avec un motif YRTA essentiel juste avant ce site. Le sous groupe de type Ib possède un motif GAA supplémentaire. Enfin certains gènes comme *clpP1* peuvent être transcrits par une NEP à partir d'un promoteur (*clpP-53*) ne possédant pas de motif particulier, mais dont la conservation chez les plantes suggère un rôle fonctionnel privilégié. La spécificité d'une NEP pour un type de promoteur n'a pas été montrée formellement mais la sur-expression de RpoTp dans le tabac a montré une implication forte pour les promoteurs de Type I et aucune pour le promoteur de type II de *clpP1*.

La fonction d'un système transcriptionnel à trois ADN polymérases n'est pas encore bien comprise. Un modèle de régulation en cascade a été proposé impliquant une transcription basale par la NEP, particulièrement dans les proplastides, et une prise en charge par la PEP dans les chloroplastes matures. Ce modèle s'appuie particulièrement sur les données d'expression des polymérases elles-mêmes. Mais différentes observations viennent compliquer le modèle, à savoir le fait que les ARN ribosomiaux sont tout de même transcrits par la PEP dans les tissus non photosynthétiques, ou encore que tous les gènes chloroplastiques semblent être transcrits dans un mutant déficient en PEP (après analyse par puce à ADN). Dans ce cas, il semblerait que la NEP produise de très long transcrits primaires, couvrant les gènes spécifiques à la PEP, mais qu'il soit impossible d'obtenir les transcrits matures correspondants, certainement dû à une absence de maturation post-transcriptionnelle spécifiquement couplée à chaque ARN polymérase (Krause *et al.* 2000; Legen *et al.* 2002).

En conclusion, à l'exception des cas d'initiation à partir de certains promoteurs particuliers, la transcription dans les organites est un processus présentant peu de modulation, l'essentiel de la régulation se fait au niveau post-transcriptionnel. De nombreux phénomènes interviennent tout au long de la vie des ARNm, comme le clivage des transcrits poly-cistroniques, les épissages en cis ou en trans, l'édition ou encore la polyadénylation des ARN messagers. Tous ces mécanismes sont hautement contrôlés par des facteurs d'origine nucléaire, et n'interviennent pas forcément dans le même ordre pour tous les gènes.

### I.3.3 Les modifications post-transcriptionnelles

#### I.3.3.1 Maturation en 5' et 3'

Dans les organites de plantes, les transcrits primaires sont généralement beaucoup plus longs que la phase codante, pouvant aller jusqu'à plusieurs kilobases de région non codante. On estime que tous les transcrits doivent être maturés en 3', et qu'une grande partie doivent également l'être en 5'. Ces phénomènes de maturation sont souvent nécessaires pour le bon déroulement des étapes postérieures, comme l'initiation ou la terminaison de la traduction, ou encore le contrôle de la stabilité du transcrit.

La longueur en 3' des transcrits primaires semble due à une absence de contrôle efficace de la terminaison de la transcription. La famille de gène mTERF (Mitochondrial transcription TErmination Factor) est retrouvée chez les animaux et les plantes (Fernandez-Silva *et al.* 1997; Linder *et al.* 2005). Chez *Chlamydomonas*, un mutant rattaché à cette famille, MOC1, montre un phénotype fort au niveau du contrôle de la transcription du génome mitochondrial (Schonfeld *et al.* 2004). Mais la transcription mitochondriale chez *Chlamydomonas* se rapproche plus du système humain, avec la transcription de deux transcrits primaires, que du système présent chez *Arabidopsis*. Pour l'instant aucun rôle précis n'a été observé pour les gènes homologue à mTERF chez *Arabidopsis* (35 loci). Au niveau de la maturation des fins de transcrits, plusieurs cas de figure sont observés selon la nature des ARN (ARN de transferts, ARN ribosomiaux ou ARN messagers).

La maturation des ARNt est relativement bien connue chez les plantes. Ils sont transcrits sous forme de longs précurseurs, puis réduits à la bonne taille par deux coupures endonucléolytiques successive en 5' par la RNase P (Hanic-Joyce and Gray 1990; Marchfelder *et al.* 1990) et en 3' par la RNase Z (Kunzmann *et al.* 1998; Schiffer *et al.* 2002).

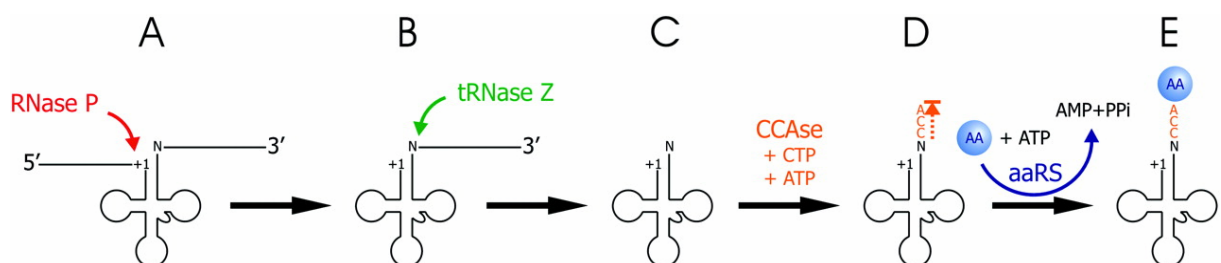


Figure 14 : Maturation des ARN de transfert, d'après (Levinger *et al.* 2004).

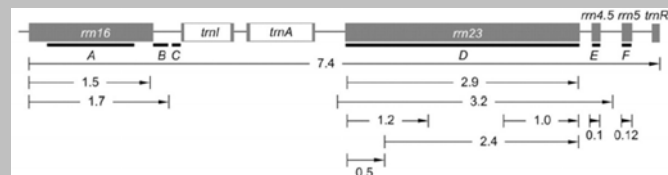
Après transcription A., les ARNt sont réduits en 5' et 3' par deux coupures endonucléolytiques B. et C., le triplet CCA essentiel au chargement de l'acide aminé est ajouté par une ARNt nucléotidyltransférase (CCAse) D. l'acide aminé est chargé grâce à une aminoacyl-tRNA synthétase (aaRS).



La maturation des ARN messagers n'est pas encore bien comprise, mais il existe certainement plusieurs systèmes parallèles permettant la maturation de ces transcrits.

Ainsi, certains transcrits comportent en 3' des séquences répétées inversées qui pourraient correspondre à des signaux nécessaires à la bonne maturation des transcrits (Dombrowski *et al.* 1997). Il est vraisemblable que des exonucléases soient en charge de la dégradation de la partie en aval de ces structures mais cela n'a pas encore été mis en évidence. D'autre part, tous les gènes mitochondriaux ne possèdent pas de telles séquences dans leur région 3' non traduite, ceci implique qu'au moins un second mécanisme existe, celui-ci pourrait impliquer l'action d'un facteur de spécificité guidant une machinerie non spécifique vers chaque ARN messager nécessitant une maturation de ses extrémités.

Dans le chloroplaste les ARN ribosomiaux sont produits à partir d'un transcrit polycistronique primaire comprenant les 4 ARN ribosomiaux et



2 ARNt. L'analyse fine des produits de maturation montre qu'il existe au moins 10 molécules produites à partir du transcrit primaire. Ce processus nécessite plusieurs événements de coupure endonucléolotique ainsi que la maturation en 3' de plusieurs intermédiaires dans un processus dépendant de RNR1. (Bollenbach *et al.* 2005).

#### Encart 2 : La maturation des ARN ribosomiaux chloroplastiques

La maturation des ARN dans les organites passe également par la polyadénylation en 3' des transcrits. Ce mécanisme est largement connu dans le monde eucaryote et est habituellement associé à la stabilisation des ARN messagers. Chez les bactéries et les organites, la polyadénylation est au contraire associée à la dégradation des ARN messagers (Dreyfus and Regnier 2002; Gagliardi *et al.* 2004).

La transcription dans les organites de plantes a la particularité d'être un processus intensif. Le taux très élevé de transcription dans les organites peut même aboutir à la transcription de nombreuses régions non codantes. Le contrôle de la quantité des ARN dans les organites passe en grande partie par la régulation de leur dégradation. La polyadénylation des transcrits est un processus essentiel de ce contrôle. Dans les organites de plantes, la polyadénylation est un signal reconnu par une 3'-5' exonucléase, la PNPase (PolyNucléotide Phosphorylase). La PNPase est essentielle pour la maturation en 3' et la dégradation de certains ARN des mitochondries (Walter *et al.* 2002; Perrin *et al.* 2004; Perrin *et al.* 2004; Holec *et al.* 2006).

### I.3.3.2 Cis- et Trans-épissage

Chez *Arabidopsis*, 9 gènes mitochondriaux et 12 gènes chloroplastiques doivent être épissés pour que la traduction puisse avoir lieu, représentant respectivement 18 et 14 cis-épissages et, 5 et 1 trans-épissages (Tableaux 1, 2 et 3). Tous ces introns font partie de la famille des introns de groupe II.

La biochimie de ces épissages reste semblable à celle des introns nucléaires pris en charge par le spliceosome, suggérant ainsi une origine commune à ces introns. L'excision des introns comporte deux étapes de trans-estérification, une première attaque nucléophile en 5' de l'intron par le groupement 2'-OH d'une adénosine du domaine VI de l'intron, puis les deux exons sont réunis par formation de la liaison 3'-5' phosphodiester. L'intron est libéré sous la forme caractéristique de "lariat" ou lasso. Les introns épissés en "trans" semblent provenir de la cassure d'un intron épissé en "cis", cet événement est intervenu à plusieurs reprises durant l'évolution des génomes des organites (Qiu and Palmer 2004).

Généralement décrits comme auto-catalytiques, ces introns requièrent en fait l'intervention de facteurs protéiques, ces facteurs sont *a priori* d'origine nucléaire. Seules exceptions, les gènes *matK*, codé par le génome chloroplastique, et *matR*, situé dans l'intron 4 de *nadI* de la mitochondrie (sous-unité 1 de la NADH déshydrogénase). Ces gènes comportent une séquence codante homologue aux ARN maturases (Wahleithner *et al.* 1990; Liere and Link 1995). Dans le génome nucléaire, plusieurs gènes présentent une identité de séquence importante avec ces maturases (Mohr and Lambowitz 2003) et l'un d'eux a été montré comme impliqué dans le contrôle de l'épissage du gène *nad4* dans la mitochondrie d'*Arabidopsis* (Nakagawa and Sakurai 2006).

Les introns de groupes II nécessitent l'intervention de nombreux facteurs protéiques pour leur épissage qui sont pour la plupart codés par le noyau. Dans les chloroplastes de maïs, trois de ces facteurs nucléaires ont été identifiés et appartiennent à la famille des protéines CRM (Jenkins *et al.* 1997; Jenkins and Barkan 2001; Ostheimer *et al.* 2003; Ostersetzer *et al.* 2005; Barkan *et al.* 2007). La fonction de ce domaine protéique, trouvé chez les Archaea, les bactéries et les plantes, n'est pas encore élucidée, mais la plupart des protéines le contenant présentent la capacité de se lier à l'ARN.

Le fait que chaque transcrit doive subir un ou plusieurs épissages nécessite l'intervention de nombreux facteurs codés par le noyau. Par exemple, l'assemblage de l'ARN de *psaA* (2 trans-épissages) chez *Chlamydomonas* nécessite l'intervention des produits d'au moins 14 loci différents (Goldschmidt-Clermont *et al.* 1990).

### I.3.3.3 Modification de la séquence des ARN

#### I.3.3.3.1 Ajout de nucléotides

Chez *Arabidopsis*, les gènes d'ARNt ne possèdent pas le triplet CCA terminal nécessaire à la réaction d'estérification par la synthétase pour accrocher l'acide aminé correspondant. Ces 3 nucléotides sont ajoutés par une ARNt nucléotidyltransférase, Figure 14 (Dietrich *et al.* 1992; Schürer *et al.* 2001). Ce triplet se retrouve également rajouté en 3' du transcrit de *rps12* chez le maïs (Williams *et al.* 2000) éventuellement par la même activité que pour les ARNt.

#### I.3.3.3.2 Edition des transcrits

L'édition est une modification des ARN très importante car elle modifie la séquence codante existant au niveau ADN. Elle peut se dérouler pendant la transcription ou de manière post-transcriptionnelle. Ce terme recouvre trois événements différents : l'addition de nucléotide(s), le remplacement d'une base ou la modification d'une base au niveau de l'ARN. Jusqu'à présent, chez les plantes, l'édition n'a été observée que dans les organites.

Chez la mitochondrie d'*Arabidopsis*, 441 sites d'édition ont été décrits. La plupart des gènes sont concernés, à l'exception des ARN ribosomiaux et de trois gènes (*atp8*, *rps7* et *cox1*). Certains transcrits mitochondriaux comptent plus de 30 sites (*ccmB*, *nad4*) (Giege and Brennicke 1999). Le génome chloroplastique comporte généralement beaucoup moins de sites avec par exemple 29 sites décrits, chez *Arabidopsis*. Neuf d'entre eux étant dans le gène *ndhB*, (Tableaux 2 et 3) (Tillich *et al.* 2005). Différentes transitions sont possibles, chez les organites d'*Arabidopsis* seules les transitions Cytidine vers Uridine ont été observées. Cependant la transition inverse ayant été trouvée dans les mitochondries de blé ou les chloroplastes de mousses, on ne peut exclure qu'elle existe chez *Arabidopsis* (Gualberto *et al.* 1990; Yoshinaga *et al.* 1996). Des phénomènes d'édition d'ARN ont également été décrits chez les mammifères ou les trypanosomes mais, d'un point de vue évolutif, les différents systèmes d'édition ne semblent pas liés entre eux. L'apparition de l'édition dans les organites de plantes est habituellement placée au moment de la colonisation des milieux aériens par les plantes, car à l'exception des *Marchantia* (Hépatophytes) (Siegfried *et al.* 1999), toutes les plantes terrestres possèdent ce mécanisme (Malek *et al.* 1996; Freyer *et al.* 1997). La conservation des sites d'édition entre espèces est assez hétérogène d'un site à l'autre. Certains sites sont fortement conservés (cf. résultats), alors que des divergences entre espèces proches peuvent être observées au niveau d'autres sites (Schmitz-Linneweber *et al.* 2002). Ceci a d'ailleurs suggéré l'idée que l'édition soit un facteur de spéciation (Schmitz-Linneweber *et al.*

2002; Herrmann *et al.* 2003). La forte pression de sélection sur les séquences protéiques des gènes codés dans les génomes des organites de plantes joue certainement un rôle dans l'apparition des sites d'édition. Dans de nombreux cas, les sites d'édition permettent de restaurer une séquence protéique très conservée. Ainsi, des codons sont systématiquement restaurés par édition, il s'agit en général de codons relatifs à des acides aminés situés au niveau de sites catalytiques ou de codons d'initiation ou de terminaison de la traduction (Maier *et al.* 1995; Hirose and Sugiura 1997; Lutz and Maliga 2001; Tsudzuki *et al.* 2001). Dans le cas des ARN de transfert, l'édition permet une expression correcte (Marchfelder *et al.* 1996; Kunzmann *et al.* 1998), ou bien permet de rétablir une structure tridimensionnelle adéquate (Fey *et al.* 2002).

La machinerie cellulaire responsable de l'édition est toujours inconnue chez les plantes, le seul gène montré, pour l'heure comme étant essentiel à l'édition est un gène PPR, *CRR4* (Okuda *et al.* 2006).

### **I.3.4 La traduction**

Chez les plantes supérieures, les deux organites utilisent le code génétique standard, à la différence des mitochondries humaines ou de levure. Relique de leur origine bactérienne, les chloroplastes et les mitochondries ont conservé leur propre machinerie de synthèse de protéines. Néanmoins celle-ci est dépendante du génome nucléaire. Par exemple 65% des protéines ribosomales chloroplastiques chez *Arabidopsis* sont codées par le noyau.

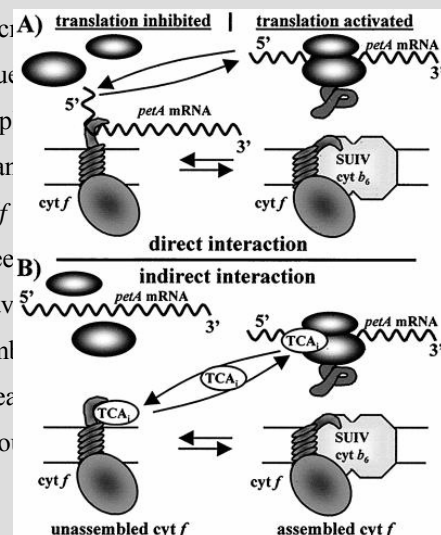
La structure globale des ribosomes est homologue à celle des bactéries : deux sous-unités de 30 et 50S composées à la fois de protéines et d'ARN (ARNr). Le ribosome chloroplastique comporte 60 éléments orthologues à *Escherichia coli* : 21 protéines (*rps1* à *rps21*) et 1 ARNr (16S) au niveau de la petite sous-unité et 36 protéines (*rpl1* à *rpl36*) et 2 ARNr (5S et 23S) au niveau de la grande sous-unité (Yamaguchi and Subramanian 2000; Yamaguchi *et al.* 2000); ainsi que 6 protéines spécifiques aux chloroplastes (PSRP1 à 6) (Yamaguchi and Subramanian 2003). La mitochondrie connaît quelques variations au niveau des ARNr, elle semble avoir perdu l'ARNr de 4,5S impliqué dans le SRP (Signal Recognition Particle), et les séquences des ARNr ont été altérées donnant un ARNr18S et un ARNr26S à la place des 18S et 23S habituels. Les modifications de séquence dans l'ARNr18S sont importantes car elles incluent en particulier l'absence d'une séquence anti-Shine-Dalgarno à son extrémité 5'.

Cette observation est en accord avec le fait que les gènes codés par le génome mitochondrial ne possèdent pas ces séquences Shine-Dalgarno dans leur séquence 5' non traduite, impliquant que l'initiation de la traduction se fait d'une manière différente de celle

observée chez les bactéries. Au niveau chloroplastique, le tableau est intermédiaire : le système d'initiation de la traduction semble en partie semblable au système bactérien avec des séquences Shine-Dalgarno en amont de la plupart des gènes. Cependant, ces séquences montrent quelques divergences dans leur séquence ou leur position et d'autre part, en amont de certains gènes, il n'a pas été observé de telles séquences (Fargo *et al.* 1998; Hirose *et al.* 1998; Hirose and Sugiura 2004; Hirose and Sugiura 2004).

La régulation de la traduction est un concept complexe qui recouvre de nombreuses fonctions très différentes, donc des facteurs très différents. Nombre de ces facteurs agissent au niveau de l'ARN où ils peuvent promouvoir ou inhiber l'initiation de la traduction, ou encore permettre la terminaison de celle-ci. Le contrôle par épistasie de synthèse observé chez *Chlamydomonas* (CES, illustrée en encart) représente un système évolué de coordination de la traduction de protéines dans les organites en fonction de la présence de leurs partenaires produits par le noyau (Choquet *et al.* 1998). Chez *Chlamydomonas*, ce système représente même la règle en matière de contrôle de l'assemblage des différents complexes photosynthétiques (Choquet *et al.* 2001; Choquet and Wollman 2002). La synthèse des sous-unités des complexes photosynthétiques est autorégulée par l'assemblage de celles-ci dans les complexes. Ce système peut s'effectuer de manière directe par fixation de la protéine sur son propre ARN, ou par l'intermédiaire d'un facteur ternaire (Choquet and Vallon 2000; Wostrickoff *et al.* 2001; Choquet *et al.* 2003; Wostrickoff *et al.* 2004; Minai *et al.* 2006).

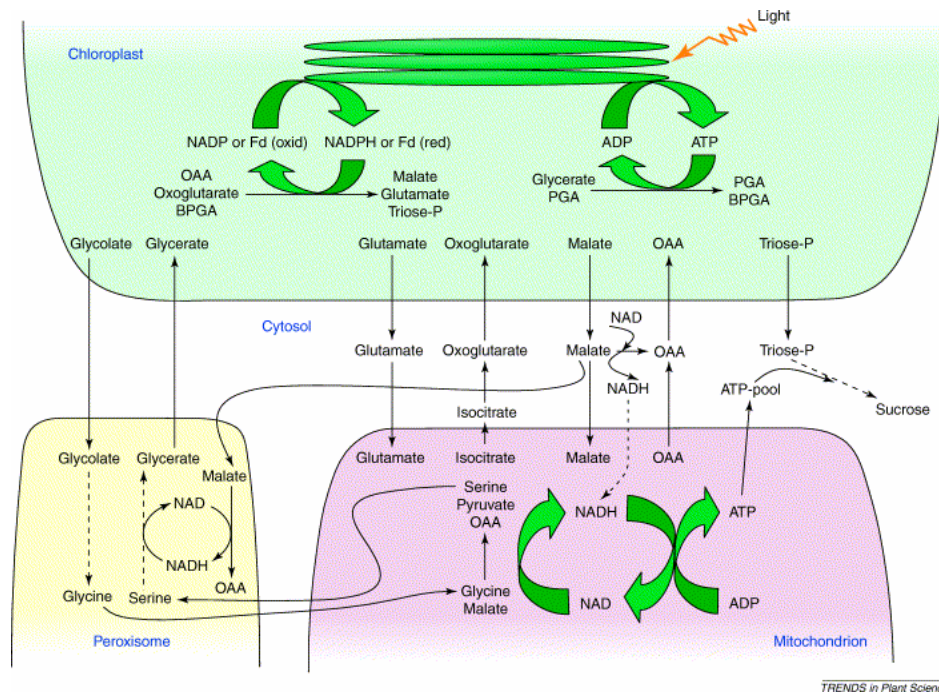
Le processus CES (Contrôle par Epistasie de Synthèse) décrit la régulation de la synthèse de certaines protéines chloroplastiques en fonction de la présence de leur partenaire au sein de complexes protéiques. Ici le cas de la traduction du gène *petA* dans le chloroplaste de *Chlamydomonas*. A) tant que le cytochrome *b6f* n'est pas totalement assemblé la traduction de *petA* est inhibée par l'interaction directe de la partie C-terminale du cytochrome *f* avec la partie 5' non traduite de l'ARN m de *petA*. B) Tant que l'assemblage du cytochrome *b6f* n'est pas complet, TCA1 est retenu au niveau du cytochrome *f*, après assemblage il est relâché et peut ainsi promouvoir la traduction de *petA*. (Choquet *et al.* 1998).



Encart 3 : Exemple de contrôle par épistasie de synthèse

Tout récemment la première description d'un système CES chez une plante supérieure a été publiée (Wostrickoff and Stern 2007). La rubisco est l'enzyme centrale du cycle de Calvin, elle est composée de deux sous-unités présentes en 8 exemplaires chacune. La large sous-unité

(LSU) est codée par le gène chloroplastique *rbcL*, alors que la petite sous-unité (SSU) est codée par le gène nucléaire *RBCS*. La traduction de RbcL est régulée par la présence de larges sous-unités non assemblées, celles-ci se fixeraient au transcrit *rbcL* et empêcheraient la traduction (Wostrikoff and Stern 2007).



**Figure 15 : communication entre les organites : transfert de métabolites**

Vue d'ensemble de la communication métabolique entre les différents organites.

## I.4 Communication entre le noyau, les chloroplastes et les mitochondries

Le noyau, les chloroplastes et les mitochondries sont trois compartiments inter-dépendants dans la cellule végétale, ceci implique une "communication" développée et précise entre eux. Il existe trois sens de communication entre ces compartiments : les voies antérogrades du noyau aux organites, les voies rétrogrades des organites au noyau, et les voies transversales entre les organites.

### I.4.1 Communication inter-organites, ou transversales

Au niveau de l'ADN, aucun transfert de la mitochondrie vers le chloroplaste n'a été démontré, dans l'autre sens 1% à 6% du génome mitochondrial des plantes supérieures est d'origine chloroplastique, mais seuls quelques ARNt représentant une infime partie seulement de ces séquences sont codés par ces régions d'origine chloroplastique (Unsel *et al.* 1997; Kubo *et al.* 2000; Notsu *et al.* 2002).

D'un point de vue métabolique, les deux organites sont intimement liés, le chloroplaste

utilise différentes voies pour dissiper les excès de formes réduites produits par la photosynthèse (NADPH et ATP), toutes sont dépendantes de la mitochondrie, Figure 15, pour revue voir (Raghavendra and Padmasree 2003). Le rôle de la mitochondrie est également essentiel dans la prévention des phénomènes de photo-inhibition, entre autre par le processus de photo-respiration (Figure 16).

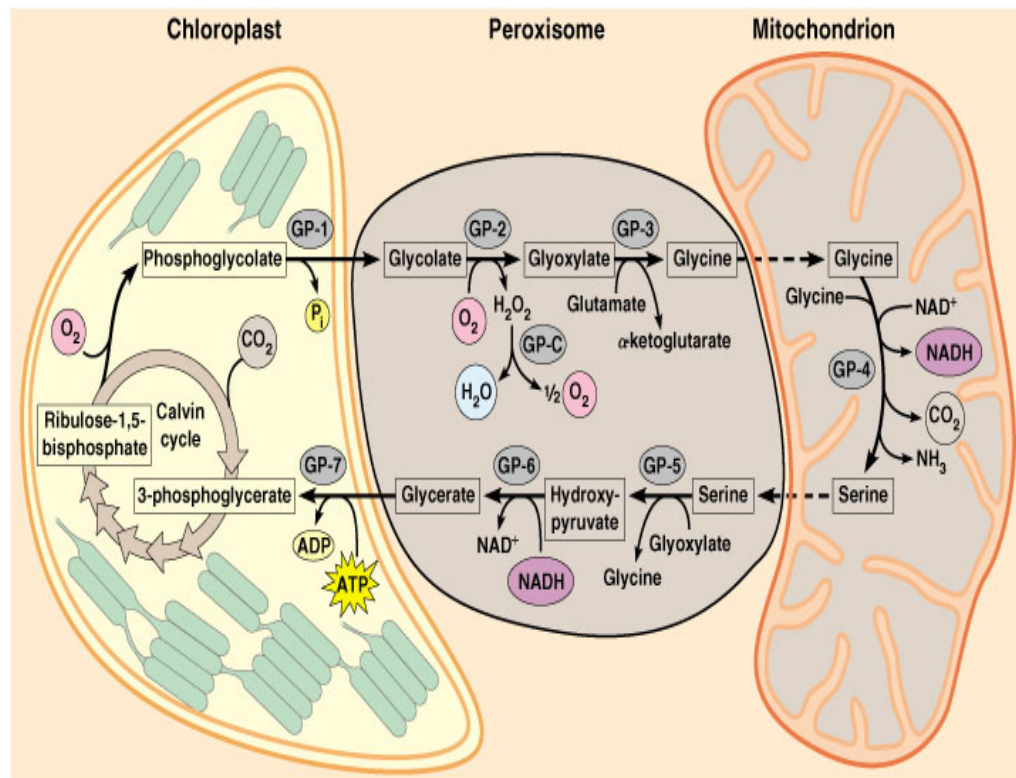


Figure 16 : Voies biochimiques de la photo-respiration

Voie de synthèse alternative de glyceraldéhyde 3-phosphate par la rubisco. L'oxydation de la rubisco mène à la formation de phosphoglycolate, celui-ci est exporté vers le peroxysoxe où il est transformé en glycine, elle-même transformée en sérine dans la mitochondrie. La sérine est transformée en glycerate dans le peroxysoxe avant d'être importé dans le chloroplaste et réintègre le cycle de Calvin sous forme de 3-phosphoglycérate. Tiré de la 4<sup>ème</sup> édition de Plant Physiology.

## I.4.2 Les voies rétrogrades

Les deux principaux centres réactionnels des organites (les photosystèmes et la chaîne mitochondriale de transfert d'électrons) sont constitués de plusieurs dizaines de protéines produites à la fois au niveau des organites et du noyau, cela requiert donc une coordination poussée de la production de ces protéines. La majorité de ce contrôle est effectuée par le noyau sur les organites (

Encart 3 et Figure 12). Néanmoins, l'utilisation d'inhibiteurs spécifiques des mitochondries ou des chloroplastes a permis de mettre en évidence l'existence d'un contrôle inverse, des organites vers le noyau (Oelmüller and Mohr 1986). De plus, plusieurs études ont montré que l'état d'oxydoréduction du pool de plastoquinones dans le chloroplaste influait sur la régulation de l'expression de gènes nucléaires (Escoubas *et al.* 1995; Streatfield *et al.* 1999;

Tullberg *et al.* 2000; Yang *et al.* 2001).

Les voies rétrogrades chloroplaste-noyau sont particulièrement étudiées. Elles sont *a priori* de trois natures : (i) médiée par la voie de synthèse des tétrapyrroles (pour revue voir (Nott *et al.* 2006), (ii) par le niveau d'oxydoréduction des chloroplastes ou (iii) dépendante de l'expression des gènes chloroplastiques. Cependant la compréhension de ces mécanismes et des facteurs impliqués est encore loin d'être complète. Très récemment une protéine PPR a été montrée comme impliquée dans ces voies rétrogrades, GUN1 (Genome uncoupled) une protéine PPR chloroplastique. Les signaux en provenance des trois voies rétrogrades sont intégrés dans le chloroplaste, une voie unique transmet ensuite le message jusqu'au noyau. GUN1 est essentiel à la génération ou la transmission de ce signal (Koussevitzky *et al.* 2007).

Les signaux rétrogrades à partir des mitochondries sont nettement moins connus. Néanmoins, de nombreuses évidences montrent que certaines formes d'oxygène actif libérées dans le cytoplasme ont un effet sur la transcription des gènes nucléaires (Rhoads and Subbaiah 2007)., d'autre part cet effet a aussi été observé dans le cas de plusieurs cas de systèmes CMS (stérilité mâle cytoplasmique, Encart 3) (Teixeira *et al.* 2005; Carlsson *et al.* 2007; Fujii *et al.* 2007).

### **I.4.3 Les voies antérogrades**

Les voies antérogrades sont principalement représentées par les protéines codées par des gènes situés au niveau du génome nucléaire et importées dans les organites, elles sont essentielles à leur fonctionnement. Au cours de l'évolution, de nombreuses séquences nucléotidiques ont été transférées entre les 3 génomes, la grande majorité des protéines nécessaires au fonctionnement des organites étant maintenant produite au niveau du noyau, et ensuite redirigée vers leur lieu d'action. Dans le chloroplaste d'*Arabidopsis*, on estime la taille du protéome entre 2500 et 3000 protéines, quasiment toutes produites dans le noyau, et dont au moins la moitié sont d'origine cyanobactérienne (Richly and Leister 2004). Dans certains cas, les gènes transférés ont été détournés par le génome les recevant, comme c'est le cas pour plusieurs ARNt, ou encore redirigés vers un autre organite, comme certaines protéines ribosomales codées au niveau du noyau mais redirigées vers un autre organite que l'originel (Gallois *et al.* 2001; Adams *et al.* 2002).

Les voies antérogrades peuvent être régulées au niveau de l'import des protéines dans les deux organites, par les complexes Tic/Toc et Tim/Tom, mais les mécanismes de régulation à ce niveau sont encore largement incompris (pour revue (Lister *et al.* 2005; Kessler and Schnell 2006).



Les voies antérogrades contrôlent un grand nombre de processus comme le contrôle de la division des chloroplastes et mitochondries, les processus métaboliques comme la respiration ou la photosynthèse, ou encore le niveau de transcription des gènes dans les organites et les événements post-transcriptionnels. Ils apparaît donc que les voies antérogrades impliquent un grand nombre de protéines codées par le noyau et acheminées vers les organites, certaines de ces protéines proviennent de l'ancêtre endosymbiotique. Néanmoins, il existe des familles de protéines typiquement eucaryotes qui se sont spécialisées au fil de l'évolution dans le contrôle des organites, parmi celle-ci nous trouvons la famille PPR.

## I.5 Les protéines PPR

Les protéines PPR ont été identifiées indépendamment par deux groupes suite au séquençage du génome complet d'*Arabidopsis* (Aubourg *et al.* 2000; Small and Peeters 2000). Cette famille de gènes est spécifique aux eucaryotes, et avec 450 membres, elle représente l'une des plus grandes familles de gènes chez *Arabidopsis*.

### I.5.1 Définition d'une protéine PPR

Les Protéines PPR sont définies par la présence d'un motif protéique de 35 acides aminés répété en tandem, entre 2 et 26 fois chez *Arabidopsis* (article Lurin *et al.*, II-46). La structure tridimensionnelle de ces motifs semble être constituée de deux hélices  $\alpha$  anti-parallèles (Figure 17).

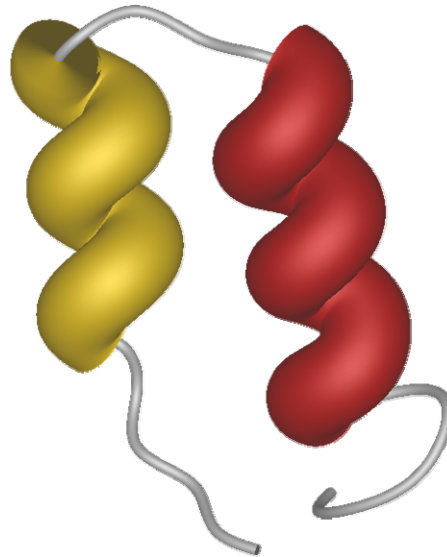


Figure 17 : Prédiction de la structure tertiaire d'un motif PPR

Les deux hélices  $\alpha$  ont été colorées en jaune pour la première, en rouge pour la seconde, le même code couleur est utilisé sur la séquence consensus du motif: **dvvtyntllsglckaGrlecAlefecMkekGlaP**

Des gènes de la famille PPR ont été trouvés dans tous les génomes eucaryotes séquencés (hormis *Giardia*). En revanche, à trois exceptions près, ils n'ont pas été observés dans les génomes procaryotes. Ceci laisse supposer qu'ils sont ubiquitaires chez les eucaryotes et absents chez les procaryotes. Les trois exceptions à cette règle sont : deux protéines PPR trouvées chez deux bactéries, *Ralstonia solanacearum* et *Chlorobium limicola* (Interpro Q8XTC0\_RALSO et Q44LX2\_CHLLI), mais l'absence totale de PPR dans les 436 génomes bactériens publiés plaide pour un transfert horizontal dans ces deux cas.

Une protéine PPR a été trouvée lors du séquençage d'un cosmide d'Archaea (Interpro

Q64EN5\_9ARCH). Pour l'instant, il est très difficile de faire des hypothèses à partir de cette donnée. En effet le cosmide séquencé a une composition typique des archaea, mais le faible nombre de génomes Archaea séquencés (37, <http://www.genomesonline.org/>) ne couvre qu'une infime partie de la diversité de ce royaume.

Les protéines PPR ont connu une expansion unique en nombre et en structure dans le monde végétal par rapport aux autres eucaryotes. Alors que les métazoaires ne comportent au maximum qu'une dizaine de PPR, les génomes de plantes supérieures en comportent plusieurs centaines. La famille PPR représente entre 1 et 2 % du protéome total d'*Arabidopsis* et forme la 9<sup>ème</sup> famille la plus grande en classant par la présence de domaines protéiques connus, InterPro, Tableau 4. D'autre part, les protéines PPR trouvées chez les plantes supérieures sont divisées en plusieurs sous-familles (cf. ci-dessous) alors que les autres eucaryotes n'en comportent qu'une seule .

Rang	Nombre de protéines	Nom	InterPro
1	1287 (3.8%)	Protéines kinase-like	IPR011009
2	1245 (3.7%)	Protéines kinases	IPR000719
3	858 (2.5%)	Sérine/thréonine kinases	IPR008271
4	730 (2.2%)	Cyclin-like F-box	IPR001810
5	667 (2.0%)	TPR	IPR011990
6	640 (1.9%)	LRR	IPR001611
7	598 (1.8%)	Protéines prényltransférase	IPR008940
8	538 (1.6%)	Zinc finger, RING-type	IPR001841
9	522 (1.5%)	PPR	IPR002885
10	496 (1.5%)	Homeodomain-like	IPR009057

**Tableau 4 : Top 10 InterPro des familles de protéines chez *Arabidopsis***

Les famille de protéines sont constituées par la présence d'un domaine protéique défini dans la base InterPro, <http://www.ebi.ac.uk/integr8/>. Les résultats correspondent aux nombres de protéines possédant le domaine indiqué dans la dernière colonne. Entre parenthèse est indiqué le pourcentage par rapport au nombre de protéines totales dans la base utilisée.

Les protéines PPR sont apparentées aux protéines TPR (TetraTricopeptideRepeat), connues pour être impliquées dans des interactions protéine-protéine (Blatch and Lassle 1999). Par analogie, les protéines PPR ont été proposées par Small et Peeters pour interagir avec l'ARN (Small and Peeters 2000). En effet, à la différence des protéines TPR, les chaînes latérales des acides aminés situés dans le sillon central sont quasiment uniquement hydrophiles, la base de ce sillon étant lui même chargé positivement. Ces caractéristiques sont trouvées chez des protéines se liant à des protéines acides et/ou phosphorylées (Huber *et al.* 1997). Cependant le nombre de motifs PPR par protéine et les connaissances sur les quelques protéines PPR connues à ce moment là supportaient l'idée d'une reconnaissance d'acides

nucléiques, idée souvent démontrée par la suite.

## **I.5.2 Ce n'est pas le motif PPR, mais les motifs PPR**

Les protéines PPR sont divisées en deux grandes sous-familles, les PPR-P (pour Pure) et les PPR-PLS (pour Pure, Long et Short). Cette deuxième sous-famille, est elle même divisée en sous-groupes suivant la nature du motif C-terminal présent. Cette sous-famille correspond aux protéines AtPCMP décrite par Aubourg *et al.* (Aubourg *et al.* 2000). Elle est caractérisée d'une part par la présence de motifs PPR modifiés, les motifs P<sup>2</sup>, L, L<sup>2</sup>, S et S<sup>2</sup>, organisés quasi systématiquement en bloc PLS ou P<sup>2</sup>L<sup>2</sup>S<sup>2</sup> et d'autre part par la présence de motifs non PPR en partie C-terminale.

### **I.5.2.1 Formation de la famille PLS**

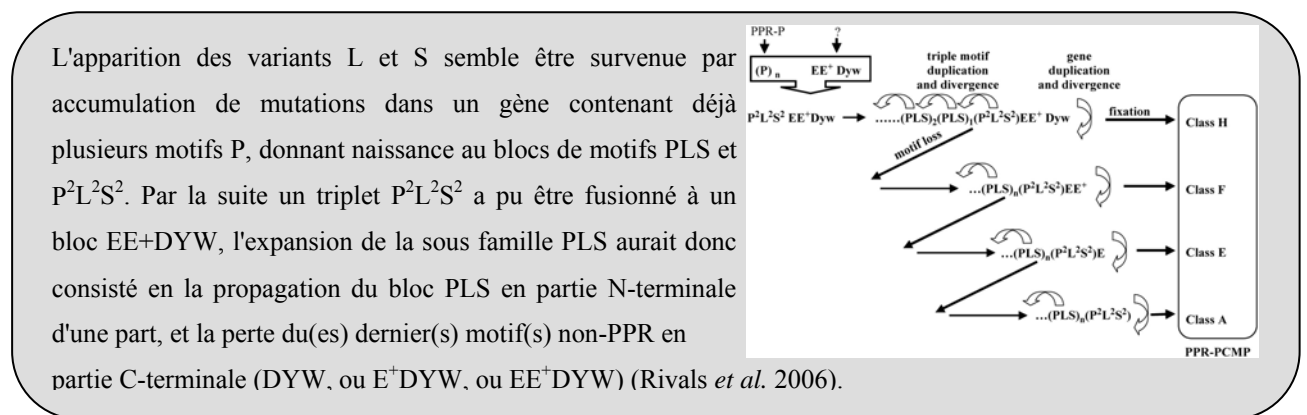
Un schéma de base de l'évolution de la famille PPR peut être donné en fonction de trois données principales (i) les protéines PPR pures sont retrouvées chez tous les eucaryotes, (ii) la sous-famille PLS est spécifique aux plantes terrestres, même chez les mousses, (iii) Arabidopsis et le riz comportent plusieurs centaines de gènes de chaque famille (Lurin *et al.* 2004; Rivals *et al.* 2006); O'toole *et al.* Soumis). Ces trois données suggèrent que i) l'apparition du motif PPR est arrivée très tôt dans l'évolution des eucaryotes, ii) l'apparition des variants L et S coïncide avec l'émergence des plantes terrestres et iii) la vaste expansion de la famille PPR est intervenue avant la séparation entre monocotylédones et dicotylédones.

La taille des effectifs de protéines PPR chez les plantes supérieures suggère une expansion massive et rapide dans ce phylum, le faible nombre d'introns dans les gènes de PPR est en faveur d'une expansion de la famille par reverse transcription, probablement en association avec un rétrotransposon. Cette théorie est confortée par le biais observé dans la localisation des introns dans les rares gènes en comportant, ceux-ci sont préférentiellement retrouvés dans la partie 5' des gènes, partie souvent perdue lors de la reverse transcription des ARNm (O'toole *et al.* Soumis).

## **I.5.3 Les protéines PPR comportent également des domaines protéiques autre que les motifs PPR**

En addition aux motifs P, L et S, les protéines peuvent comporter différents autres motifs, les principaux sont les motifs E, E+ et DYW spécifiques à la sous-famille PLS (à l'exception du produit du gène At1g47580 constitué quasi exclusivement du motif DYW). Ces motifs représentent potentiellement la partie catalytique des protéines PPR, les motifs PPR étant

limités à la reconnaissance des cibles ARN. Cependant, la quasi inexistence d'homologie des motifs E et E<sup>+</sup> avec d'autres motifs catalytiques connus, et la diversité des rôles attribués aux quelques protéines PPR ne permet pas l'élaboration d'un modèle d'action simple des protéines PPR. Le motif DYW présente une ressemblance avec certaines protéines impliquées dans l'édition. Par ailleurs, une corrélation existe entre la présence de ce domaine et l'existence du phénomène d'édition de transcrits cytoplasmiques chez les Plantae. Ainsi, il est probable que le motif DYW soit nécessaire à l'édition dans les organites de plantes (V. Salone, communication personnelle). Les Protéines PPR de la famille PLS ressemble à des poupées russes, les combinaisons possibles sont uniquement celles-ci: PLS, PLS-E, PLS-EE<sup>+</sup>, PLS-EE<sup>+</sup>DYW.



#### Encart 4 : La formation de la sous-famille PLS

Certaines protéines PPR contiennent d'autres motifs. C'est le cas du produit du gène At2g15820, qui comporte à son extrémité C-terminale un motif LAGLIDADG. Cette protéine sera présentée dans la partie résultats (II.2.2).

Le produit du gène At5g04810 comporte un motif RRM en partie N-terminale, cette structure est semblable à celle de PPR4 chez le maïs discuté en conclusion (Schmitz-Linneweber *et al.* 2006).

Le gène AtC401 (At5G21222) code pour une protéine contenant à la fois 12 motifs PPR et un motif serine/thréonine kinase. Cependant, le fait que l'identité de séquence avec les gènes homologues chez le riz ou *Pharbitis nil* soit limitée aux motifs PPR, et que le locus AtC401 présente des identités de séquence significatives avec deux autres locus dans le génome, l'un correspondant à une protéine kinase, l'autre à une protéine PPR, plaident pour une fusion de gène récente. Néanmoins, il peut aussi être envisagé que cette fusion soit en fait le résultat de problèmes d'annotation.

21 autres motifs protéiques ont été décrits comme présents dans des protéines PPR (Liu *et al.* 2005), mais l'absence de validation par identification d'ADNc pleine longueur

d'homologue chez le riz semble indiquer qu'il s'agit plutôt de problèmes d'annotation (O'toole et al. soumis).

Une dizaine de protéines PPR porte en partie C-terminale un domaine apparenté au domaine terminal des protéines bactérienne MutS2 ou SMR, la fonction de ce domaine n'est pas connue, mais pourrait intervenir pour des interactions protéine-protéine ou porter une fonction nucléase (Moreira and Philippe 1999; Malik and Henikoff 2000; Fukui *et al.* 2007).

#### **I.5.4 Au sujet des fonctions des protéines PPR**

Quand j'ai débuté ma thèse très peu de données étaient disponibles concernant la fonction des protéines PPR. La synthèse des données actuelles sera présentée en conclusion.

En 2003, quelques indices existaient sur la structure, le nombre et la localisation des protéines PPR. Ceux-ci nous laissaient penser qu'elles étaient impliquées dans des processus localisés dans les chloroplastes ou mitochondries, et que ces processus avaient trait aux acides nucléiques, et plus précisément aux ARN. Les premières expériences de biochimie avaient confirmé l'affinité des protéines PPR pour des acides nucléiques, mais leur rôle précis ne pouvait être compris sans approche fonctionnelle, pour cela l'étude des mutants affectant des protéines PPR était essentielle. C'est dans ce contexte que je débutai mon DEA puis ma thèse.... (Lurin *et al.* 2004).

## I.6 Présentation du travail de thèse

Ainsi, avant 2003, très peu de protéines PPR avaient été caractérisées, et aucune ne l'avait été chez *Arabidopsis*. Il s'agissait uniquement de protéines de la sous-famille P, pour la plupart identifiées par des cribles génétiques, les processus moléculaires mis en jeu n'étaient pas précisément compris (Tableau 5).

Gene	Organisme	processus physiologique	processus moléculaire	cible	PPR	Loc.	ref
<b>Rf</b>	Pétunia	CMS	ARN dégradation	-	P	Mt	(Bentolila <i>et al.</i> 2002)
<b>MCA1</b>	Chlamydomonas	Photosynthèse	ARN stabilisation	petA	P	Cp	(Lown <i>et al.</i> 2001)
<b>CRP1</b>	Mais	Photosynthèse	Traduction	petA, psaC	P	Cp	(Fisk <i>et al.</i> 1999)
<b>LRPPRC, hLRP130</b>	Humain	Respiration	ARN stabilisation	CoxI CoxIII	P	Mt	(Liu and McKeehan 2002; Tsuchiya <i>et al.</i> 2002; Mili and Pinol-Roma 2003; Tsuchiya <i>et al.</i> 2004)
<b>Cya5</b>	Neurospora	Respiration	Traduction	CoxI	P	Mt	(Coffin <i>et al.</i> 1997)
<b>Pet309</b>	Levure	Respiration	ARN stabilisation, translation	CoxI	P	Mt	(Manthey and McEwen; Manthey <i>et al.</i> )
<b>Bsf</b>	Drosophile	-	ARN stabilisation	bicoid 3'	P	C	(Mancebo <i>et al.</i> )
<b>P67</b>	Radis	-	ARN binding	-	P	Cp	(Lahmy <i>et al.</i> 2000)
<b>P63</b>	Blé	-	Transcription	CoxII	P	Mt	(Ikeda and Gray 1999)

**Tableau 5 : PPR publiés avant 2003**

Il était essentiel pour comprendre le rôle de cette famille d'accumuler des informations sur tous les types de protéines PPR, pour cela j'ai entrepris un projet de recherche de mutants PPR chez *Arabidopsis thaliana*. Il fallait d'une part accumuler des lignées de disruption de gènes PPR (Article 1 : Lurin *et al.* 2004), et d'autre part développer un moyen d'identifier la cible de la PPR en question (Andrés *et al.*, en préparation, Article 2 de cette thèse).

Cette thèse présente le développement d'un crible de défauts de maturation des ARN codés par les organites, la sélection de lignées d'insertion pour des gènes codant pour des protéines PPR et la caractérisation de certaines de ces lignées.

## **Chapitre II Résultats**

### **II.1 Analyse de mutants**

Le premier objectif de ma thèse a été de collecter un nombre suffisant de mutants PPR pour avoir la vue la plus large possible sur la famille. L'article suivant présente les premiers résultats obtenus par l'équipe et ses collaborateurs. Il décrit la famille PPR dans son ensemble, la nature des motifs, les gènes dans différents organismes, et il s'attarde plus particulièrement sur la famille chez *Arabidopsis*. Le résultat d'analyse de données d'expression est aussi présenté. D'un point de vue fonctionnel, cet article présente les résultats d'un test d'affinité pour les acides nucléiques ainsi que les expériences de localisation sub-cellulaire effectuées sur un échantillon de 48 protéines PPR.

Ma participation à cet article a essentiellement consisté en l'analyse de collections de mutants.

**II.1.1 Article 1 : 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.**



# Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis <sup>W</sup>

Claire Lurin,<sup>a</sup> Charles Andrés,<sup>a</sup> Sébastien Aubourg,<sup>a</sup> Mohammed Bellaoui,<sup>a</sup> Frédérique Bitton,<sup>a</sup> Clémence Bruyère,<sup>b</sup> Michel Caboche,<sup>a</sup> Cédric Debast,<sup>c</sup> José Gualberto,<sup>d</sup> Beate Hoffmann,<sup>a</sup> Alain Lechamy,<sup>b</sup> Monique Le Ret,<sup>d</sup> Marie-Laure Martin-Magniette,<sup>a,e</sup> Hakim Mireau,<sup>c</sup> Nemo Peeters,<sup>c,1</sup> Jean-Pierre Renou,<sup>a</sup> Boris Szurek,<sup>a</sup> Ludivine Taconnat,<sup>a</sup> and Ian Small<sup>a,2</sup>

<sup>a</sup>Unité de Recherche en Génomique Végétale (Institut National de la Recherche Agronomique/Centre National de la Recherche Scientifique/Université d'Evry Val d'Essonne), CP 5708, 91057 Evry Cedex, France

<sup>b</sup>Institut de Biotechnologie des Plantes (Centre National de la Recherche Scientifique/Université Paris-Sud), Université Paris-Sud, 91405 Orsay Cedex, France

<sup>c</sup>Station de Génétique et Amélioration des Plantes, Institut National de la Recherche Agronomique, Versailles, 78026 Versailles Cedex, France

<sup>d</sup>Institut de Biologie Moléculaire des Plantes (Centre National de la Recherche Scientifique), 67084 Strasbourg Cedex, France

<sup>e</sup>Unité Mixte de Recherche Institut National Agronomique Paris-Grignon/Ecole Nationale du Génie Rural et des Eaux et Forêts/Institut National de la Recherche Agronomique, Biométrie et Intelligence Artificielle 518, 75231 Paris, Cedex 05, France

**The complete sequence of the *Arabidopsis thaliana* genome revealed thousands of previously unsuspected genes, many of which cannot be ascribed even putative functions. One of the largest and most enigmatic gene families discovered in this way is characterized by tandem arrays of pentatricopeptide repeats (PPRs). We describe a detailed bioinformatic analysis of 441 members of the Arabidopsis PPR family plus genomic and genetic data on the expression (microarray data), localization (green fluorescent protein and red fluorescent protein fusions), and general function (insertion mutants and RNA binding assays) of many family members. The basic picture that arises from these studies is that PPR proteins play constitutive, often essential roles in mitochondria and chloroplasts, probably via binding to organellar transcripts. These results confirm, but massively extend, the very sparse observations previously obtained from detailed characterization of individual mutants in other organisms.**

## INTRODUCTION

The *Arabidopsis thaliana* genome sequence uncovered many previously undescribed and often unsuspected genes. The initial analysis of the genome sequence estimated that 31% of Arabidopsis genes were too dissimilar to genes of known function to be attributed putative functions (Arabidopsis Genome Initiative, 2000). Many of the genes in this class fall into families that have greatly expanded in plants or are entirely plant specific. A major challenge for the plant science community is to

discover the functions of these genes, given that little or no help can be expected from studies on model microorganisms or metazoa, so useful for more widely conserved gene families. Amongst these families, the largest (~450 members) and perhaps the most mysterious is defined by the so-called pentatricopeptide repeat (PPR) (Small and Peeters, 2000). PPR proteins make up a significant proportion (~6%) of the unknown function proteins in Arabidopsis. Only a few very recent articles describe the functional analysis of individual Arabidopsis PPR genes (Hashimoto et al., 2003; Meierhoff et al., 2003; Yamazaki et al., 2004).

Although individual PPR genes were described from mutant studies in yeast and *Neurospora* many years ago (Manthey and McEwen, 1995; Coffin et al., 1997), the existence of a large family of similar proteins only became apparent with the sequencing of the Arabidopsis genome (Arabidopsis Genome Initiative, 2000). Two independent partial and sometimes contradictory descriptions of the family and the characteristic motif(s) defining it were offered before the completion of the genome sequence. Small and Peeters coined the name PPR for the major motif based on its similarity to the better-known tetratricopeptide repeat (TPR) motif and briefly described a large protein family containing it

<sup>1</sup>Current address: Laboratoire des Interactions Plantes Micro-organismes, Centre National de la Recherche Scientifique-Institut National de la Recherche Agronomique, Chemin de Borde Rouge, BP27, 31326 Castanet-Tolosan Cedex, France.

<sup>2</sup>To whom correspondence should be addressed. E-mail small@evry.inra.fr; fax 33-1-60874510.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Ian Small (small@evry.inra.fr).

<sup>W</sup>Online version contains Web-only data.

Article, publication date, and citation information can be found at www.plantcell.org/cgi/doi/10.1105/tpc.104.022236.

(Small and Peeters, 2000). Aubourg et al. described a large Arabidopsis protein family containing multiple modules of different repeated motifs (Aubourg et al., 2000). Subsequently it became obvious that these two families overlapped considerably and that several of the motifs described by Aubourg et al. correspond to PPR motifs or variants thereof. One of the goals of this article is to provide a unified and detailed description of the family from the complete Arabidopsis genome sequence to provide a coherent, stable platform on which future studies can be based.

The few data available point to an involvement of PPR proteins in posttranscriptional processes in organelles. The mutants *pet309* (Manthey and McEwen, 1995; Manthey et al., 1998) and *cya-5* (Coffin et al., 1997) in yeast and *Neurospora crassa* are affected in stability and translation of mitochondrial *cox1* transcripts. A similar PPR gene in humans has been linked to a genetic disease characterized by COX1 deficiency (Mootha et al., 2003). In a landmark study, Fisk et al. recognized the similarity of the fungal genes to a maize (*Zea mays*) gene, *CRP1*, implicated by genetic analysis in processing and translation of plastid *pet* transcripts (Fisk et al., 1999). Similar effects on plastid transcripts have been subsequently observed in other mutants from *Chlamydomonas* (Lown et al., 2001), Arabidopsis (Hashimoto et al., 2003; Meierhoff et al., 2003; Yamazaki et al., 2004), and maize (Williams and Barkan, 2003). Further evidence for a role of PPR proteins in organelle gene expression has come from positional cloning of cytoplasmic male sterility (CMS) restorer genes. *Rf1* from petunia (*Petunia hybrida*) (Bentolila et al., 2002), *Rfk* and *Rfo* from radish (*Raphanus sativus*) (Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003), and *Rf-1* from rice (*Oryza sativa*) (Kazama and Toriyama, 2003; Komori et al., 2004) have all been shown to encode PPR proteins. Although the mechanism of action of these proteins is not known, they all share the ability to prevent expression of the protein encoded by the corresponding mitochondrial CMS inducer gene.

The phenotypes of these mutants strongly suggest that PPR proteins act directly or indirectly on RNA, but evidence for RNA binding by these proteins remains very sketchy. Radish p67 and *Drosophila* bicoid stability factor were both purified as sequence-specific RNA binding proteins (Lahmy et al., 2000; Mancebo et al., 2001), although doubts remain as to whether the bait RNAs used are physiological targets for these proteins. Wheat (*Triticum aestivum*) p63 (Ikeda and Gray, 1999) and mouse LRP130 (Tsuchiya et al., 2002) have been shown to bind DNA, although LRP130 has also been convincingly shown to bind RNA (Mili and Pinol-Roma, 2003). *Hcf152* has been shown to bind the *pet* transcript affected in the *hcf152* mutant and is the best documented case to our knowledge where a PPR protein has been demonstrated to bind its probable natural target RNA (Meierhoff et al., 2003; Nakamura et al., 2003).

Hence, whereas the functional data available on PPR proteins are rather coherent, they remain very sparse and, as we shall demonstrate, almost completely confined to one subgroup of the whole family. The second goal of the work described here was to obtain family-wide data, using high-throughput functional genomics technology where possible, to expand our fragmentary knowledge of the roles of these proteins.

## RESULTS

### Identification of the PPR-Encoding Genes in the Arabidopsis Genome

With the completion of the sequencing of the Arabidopsis genome, we were able to search for all the PPR genes in Arabidopsis. We first used the alignment of 2357 previously identified PPR motifs (Small and Peeters, 2000) to obtain a consensus model using the HMMER package. This PPR signature matrix was subsequently used to search for PPR motifs in the six reading frames of the whole Arabidopsis genome. Because PPR motifs are highly degenerate, a fairly large number of false positive hits were expected. However, because PPR motifs are almost always observed in proteins as tandem repeats, we could remove most random uninteresting matches by omitting all orphan PPR motifs that were found farther than 200 nucleotides from any other PPR motif. The 6371 PPR motifs retained formed 526 clusters, each of which comprises a putative PPR gene. Each PPR motif cluster was then investigated in detail by manually analyzing the positions and reading frames of the PPR motifs compared with (1) Arabidopsis Genome Initiative (AGI) open reading frame (ORF) models in the same region, (2) sequenced cDNAs or ESTs matching to the same region, and (3) predicted N-terminal mitochondrial and plastid targeting sequences within potential coding sequence in the region. From this analysis, 441 putative PPR ORF models were constructed (i.e., ~80 motif clusters were discarded or fused with other clusters). As of Release 5.0 of The Institute for Genomic Research (TIGR) annotations of Arabidopsis proteins, 447 of the AGI ORFs are annotated as PPR-containing proteins based on matches with the PFAM profile PF01535 or Interpro profile IPR002885. Three other AGI models are annotated as homologs of known PPR proteins. Of the two sets of ORF models, 326 are identical (i.e., our manual analysis agreed with the AGI model) (see Supplemental Table S1 online). Twenty-four AGI models do not have equivalents in our set; nine because we do not consider them to be PPR-containing proteins by our criteria (they lack tandem motifs matching our HMMER matrices) and 15 because we consider them to be probable pseudogenes (they would require multiple unlikely introns and/or frameshifts to produce a model resembling other members of the family). Six of our models have no AGI equivalent and correspond to genes apparently overlooked during the AGI annotation or considered to be pseudogenes. In three more cases (At1g62910, At5g08310, and At5g40410), we split AGI models into two to form six new models because we felt that the TIGR annotations had fused tandemly duplicated PPR genes. We also found six other cases of probable fusions of ORF models, this time between PPR genes and members of completely unrelated families. In all, 115 of our models differ in at least some respects from the corresponding AGI model. It should be noted that in very few of these cases are molecular data available that can be used to decide between the two discordant models. Our choice has been generally made by comparison with other genes in the family and a general familiarity with these proteins. This manually annotated version of the PPR family from the Arabidopsis genome is available in the supplemental data online.

## The PPR-Related Motifs

The PPR motifs are usually present in proteins as tandem arrays of about a dozen repeats. In about one half of the family, these PPR motifs are adjacent to each other without gaps between the motifs. In the other half of the family, we observed regular gaps between PPR motifs of ~65 to 70 amino acids. We extracted and aligned the fragments found separating PPR motifs, and these alignments were used to define models using the HMMER package. This analysis allowed us to define two new motifs of 31 and 35 to 36 amino acids. These consensus sequences were compared with PPR consensus (Figure 1) showing that the two motifs are clearly related to the PPR motif. In particular, the secondary structure predictions for the two motifs strongly suggested a pair of  $\alpha$ -helices as already proposed for the PPR motif (Small and Peeters, 2000). These two motifs were named the PPR-like S (for short) and PPR-like L (for long) motifs. The L motifs in particular are variable in size and sequence, and several variants could be defined. The two major variants, described in Figure 1, have been used in this study. The HMMER matrices for these motifs were used to search for PPR-like motifs in the six frames of the Arabidopsis genome (as described earlier for the PPR motif). Because all these PPR-like motifs are related, many overlapping or coincident matches were obtained. In these cases, the arrangement of nonoverlapping motifs giving the highest cumulative hmmsearch scores was retained (see Supplemental Table S2 online). After removing redundant motifs in this way, 806 L-type and 1304 S-type motifs were found in >300 proteins to add to the 3340 P-type classical PPR motifs. It is noticeable that these motifs were only found in previously identified PPR proteins, reinforcing the idea that these motifs are highly related to PPR motifs. By contrast, these motifs were not found in all the PPR proteins identified in Arabidopsis. Slightly less than half of the PPR proteins present an organized pattern of triple motifs (P-L-S) repeated along the protein sequence. This subclass of the PPR family corresponds to the Arabidopsis plant combinatorial and modular protein (AtPCMP) family, previously discovered by bioinformatic analysis of the Arabidopsis genome (Aubourg et al., 2000). They described 127 proteins character-

```

PF01535   vtYntLIsgycknGkLeeAlelfeeMkek.GikPdv
S         dvvvynaLIdmYaKcGdleeArkvFdeMper
PPR (P)  dvvtYntLIsglckaGrLeeAlelfeeMkek.GiaP
L1       defTlasvLkACaslgaLslGkqiHgyviKs.Gfds
L2       devTFIqVLSACSHsGLVeeGleyFesMkekYGIeP

```

**Figure 1.** Alignment of the Four Most Abundant PPR-Related Motifs in Arabidopsis in Comparison with the PFAM PF01535 PPR Motif.

Consensus sequences were obtained using the HMMER package based on alignments of thousands of Arabidopsis motifs. The PPR consensus used by PFAM (and obtained using motifs from a variety of organisms) is almost identical to our consensus except shifted by two amino acids, such that it overlaps the first helix of the following motif. Residues in capital letters are more highly conserved within each motif. Residues in bold are conserved between PPR-related motifs. The underlined sequences indicate the correspondences to the motifs A (underlined three times), B (underlined once), and C (underlined twice) described by Aubourg et al. (2000). Motif C overlaps adjacent S and P motifs. The shaded boxes indicate the maximum extent of the predicted  $\alpha$ -helical regions.

ized by a combinatorial organization of A, B, and C motifs. These motifs, identified on the basis of sequence similarity rather than structural considerations, match to parts of the PPR and PPR-related motifs that we describe here (Figure 1). The A, B, and C motifs correspond to the N-terminal part of the L motif, an internal peptide of the S motif, and a peptide overlapping with S and PPR motifs, respectively. Aubourg et al. considered that the sequence similarity between PCMP proteins and PPR proteins was not significant, but the structural and organizational similarities highlighted by this genome-wide study strongly suggest that they are related and belong to the same superfamily.

PPR proteins contain from 2 to at least 26 copies of the PPR motif or its variants, with an average of 12 motifs per protein, usually in a single unbroken tandem array that forms at least two-thirds of the protein sequence. There are clear indications that polymorphism within the family can be because of deletion or duplication of tandem repeats, and this, coupled with the difficulty in obtaining unambiguous alignments given the number of repetitions of similar sequence, makes sequence homology-based phylograms of the family difficult or impossible to interpret.

## Identification of C-Terminal Domains

Many PPR proteins contain sequences unrelated to PPR motifs either before or after the tandem arrays of PPR motifs. The sequences before are mostly comprised of organelle targeting sequences and show little sequence similarity. However, the sequences following the PPR motifs show considerable sequence conservation in many cases. To characterize C-terminal motifs in the PPR family, we aligned the polypeptide sequences C-terminal of the last PPR-related motif from all of the Arabidopsis PPR proteins. This alignment allowed us to define three motifs (E, E+, and DYW) that were defined in more detail using the HMMER package (Figure 2). The E, E+, and DYW motifs were found in 191, 145, and 87 Arabidopsis proteins, respectively. They roughly correspond to the E, F-G, and H motifs defined by Aubourg et al. (2000) from 120 AtPCMP proteins. Despite the fact that these new motifs are not related in sequence to the PPR motifs, the three motifs were only found in members of the PPR family and not in any other Arabidopsis protein with the exception of At1g47580, which contains a lone DYW motif. Interestingly, they were only found in the subfamily defined by P-L-S repeats. The relative organization of the three C-terminal motifs follows several quasi-absolute rules: (1) the motifs are never observed in multiple copies in the same protein; (2) when the motifs are observed in the same protein they are ordered colinearly E – E+ – DYW, DYW being the C-terminal tripeptide; (3) proteins carrying the DYW motif almost always have preceding E and E+ motifs; similarly, proteins carrying E+ motif always have a preceding E motif. The identification of these motifs allowed us to define four subclasses in the P-L-S subfamily: (1) proteins that do not have any of the three C-terminal motifs, (2) proteins with the E motif alone, (3) proteins with E and E+ motifs, and (4) proteins with the E, E+, and DYW motifs (Figure 3), named the P-L-S, E, E+, and DYW subclasses, respectively (see Supplemental Table S2 online). These subclasses of the Arabidopsis PPR family contain 6, 47, 60, and 87 proteins, respectively (Figure 3).

**E**  
 1 VWgaLLgACrCriHgnvnellaeaAakkLl 30  
 31 eLePendgsYeeGensGnYVLLSNiYAsaG 60  
 61 rWeevakvRklMkdrekGvkKePGcSWIEv 90  
 91 k

**E+**  
 1 GkvHeFvaGDkkeSHPeseeIyekLeeLie 30  
 31 emk 33

**DYW**  
 1 EaGYvPdtssEvvlhdveEEeKEellRyH 30  
 31 **SEKLA**iAFGLiStpegrkspgtptnpIRim 60  
 61 **KNL**RvCgD**CH**naiKLI*S*kitgREIivRDrn 90  
 91 **RFH**HfkdGs**CS**CgDYW 106

**Figure 2.** HMMER-Derived Consensus Sequences of C-Terminal Motifs Present in PPR Proteins.

The alignments employed for the E, E+, and DYW motifs contained 184, 148, and 85 sequences, respectively. The best conserved residues are in capital letters; bold, underlined amino acids are completely invariant. For the DYW motif, the DYW triplet (or a closely related sequence) forms the C terminus of the protein.

In contrast with PPR motifs that appear to be TPR-related motifs (Small and Peeters, 2000), the three C-terminal motifs do not resemble any previously described polypeptide motif. The E and E+ motifs are highly degenerate motifs, but the high conservation in the amino acid sequences and the presence of invariant Cys and His in the DYW motif may imply a catalytic function for this domain (Aubourg et al., 2000).

### Genomic Organization

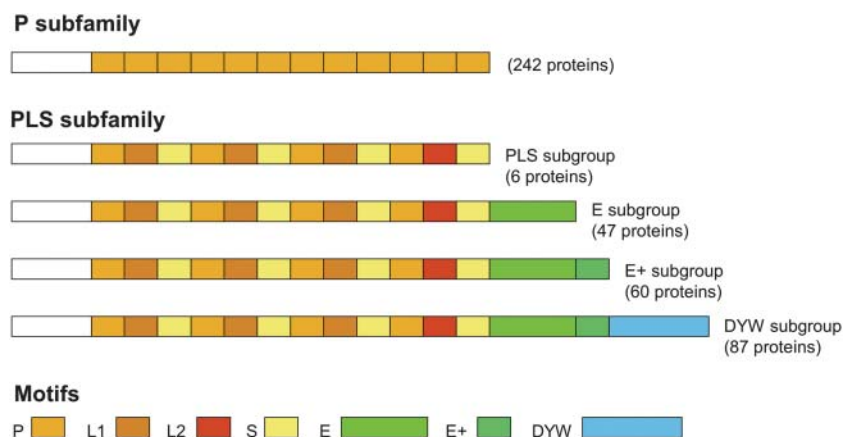
PPR genes are fairly evenly distributed throughout the 10 chromosome arms, with little in the way of obvious clusters. The densest grouping of PPR genes lies on the long arm of chromosome 1 (around 23 megabases) where 19 genes and

several probable pseudogenes lie within little more than one megabase (data not shown). Many of the genes in this cluster are closely related, and they include the closest Arabidopsis homologs of the recently cloned CMS restorer genes from petunia, radish, and rice. The radish restorer *Rfo* (*Rfk1*) lies in a region that is very clearly syntenic to this Arabidopsis cluster (Desloire et al., 2003).

A noticeable characteristic of PPR genes is that they rarely contain introns within the coding sequence. More than 80% (363/441) of our PPR ORF models contain a single exon, and only 7% (31/441) contain more than one intron. These figures are astonishingly low compared with the average of five introns or more for other Arabidopsis genes (Arabidopsis Genome Initiative, 2000). This characteristic explains why PPR genes are relatively short (on average <2 kb from ATG to stop codon) despite the fact that PPR proteins are comparatively large proteins (642 amino acids on average).

### PPR and PPR-Related Motifs in Other Genomes

To characterize the PPR family in a variety of other organisms, we collected data on several fully sequenced and annotated genomes (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Cyanidioschyzon merolae*, *Escherichia coli*, *Ralstonia solanacearum*, *Rickettsia prowazekii*, and *Synechocystis* sp 101) or almost completely annotated genomes (human, rice, and *Trypanosoma brucei*). Using the HMMER matrices defined using Arabidopsis PPR, E, and DYW motifs, we searched for these motifs in the collected genomes (Table 1). The criteria for the motif search were (1) an hmmsearch E-value < 10 and (2) at least one pair of PPR or PPR-like repeats in any protein hit. Using these criteria, no PPR proteins were identified in any of the prokaryotic genomes that we analyzed except for a single protein from *R. solanacearum*. This PPR-like protein appears to be a fairly recent capture by horizontal transfer (Salanoubat et al., 2002). By contrast, PPR motifs were found in all eukaryotes analyzed but with an extraordinary discrepancy in numbers between plant and



**Figure 3.** Motif Structure of Arabidopsis PPR Proteins.

Typical structures of proteins from each of the principal subfamilies and subgroups are shown. The structures are purely indicative, and the number and even order of repeats can vary in individual proteins. The number of proteins falling into each subgroup is shown.

**Table 1.** PPR Proteins Are Specific to Eukaryotes

Organism	Sequences	PPR Hits	E Hits	DYW Hits
<i>Homo sapiens</i>	37,490	6	0	0
<i>Drosophila melanogaster</i>	17,087	2	0	0
<i>Caenorhabditis elegans</i>	20,673	2	0	0
<i>Schizosaccharomyces pombe</i>	5,010	6	0	0
<i>Saccharomyces cerevisiae</i>	6,304	5	0	0
<i>Trypanosoma brucei</i>	16,757	19	0	0
<i>Cyanidioschyzon merolae</i>	4,772	10	0	0
<i>Arabidopsis thaliana</i>	28,581	470	193	87
<i>Oryza sativa</i>	74,385	655	292	104
<i>Ralstonia solanacearum</i>	5,118	1	0	0
<i>Rickettsia prowazekii</i>	834	0	0	0
<i>Synechocystis</i> sp	3,169	0	0	0

Searches for PPR and PPR-associated motifs were conducted on complete or quasicomplete predicted proteomes from a range of species using HMMER. The figures indicate the number of proteins found to contain two or more PPR motifs with an E-value of <10 or one or more E or DYW motifs with a score of >0.

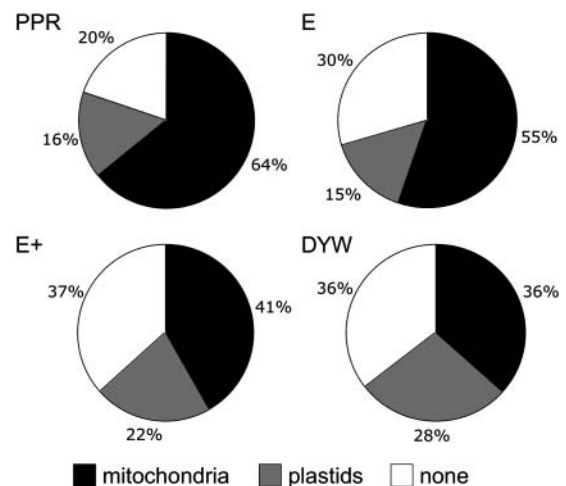
nonplant organisms. Thus, whereas *Arabidopsis* and rice have several hundreds of PPR genes in their genome, the human genome encodes only six putative PPR proteins. Trypanosomes, recently linked to the plant kingdom (Hannaert et al., 2003; Martin and Borst, 2003) and the red alga *C. merolae* (Matsuzaki et al., 2004), have more PPR proteins than humans but many fewer than the two plant genomes. The list of putative nonplant PPR proteins is given in Supplemental Table S3 online. We also looked for E and DYW motifs (Table 1) and for L and S variants of PPR motifs (data not shown), but we were unable to find conclusive examples of any of them outside the higher plant genomes, indicating that the P-L-S half of the family is strictly restricted to plants.

### Subcellular Localization of the PPR Proteins

Most of the few known PPR proteins have roles in mitochondria or plastids, and the PPR family was discovered while systematically screening *Arabidopsis* proteins for those predicted to be targeted to mitochondria or chloroplasts (Small and Peeters, 2000). We therefore checked for predicted targeting signal peptides at the N terminus of PPR proteins using the TargetP (Emanuelsson et al., 2000) and Predotar (Small et al., 2004) prediction programs. Both programs assigned most of the proteins of the family to either mitochondria or plastids (see Supplemental Table S4 online). Using the Predotar program and a threshold probability of 0.25, 54 and 19% of the PPR proteins were predicted to be targeted to mitochondria or plastids, respectively (Figure 4). TargetP gave similar results (see Supplemental Table S4 online). Differences were observed in the proportion of the different targeting predictions between the various subclasses (Figure 4). For example, whereas pure PPR proteins are mostly predicted to be targeted to mitochondria, DYW proteins are almost as likely to be predicted to be targeted to plastids as to mitochondria. Furthermore, the percentage of predicted untargeted proteins in the E, E+, and DYW subclasses is significantly higher than in the pure PPR subfamily (35%

versus 20%,  $P < 0.001$ ,  $\chi^2$  test). These differences might reflect fundamental differences in the function of these proteins, but they may simply be because of major differences in ORF model quality, members of the pure PPR subfamily being better known and more expressed than the other subclasses (see below). Overall, the percentage of observed false negative results using Predotar or TargetP is ~20 to 30% (Emanuelsson et al., 2000; Small et al., 2004) (i.e., not very different than the number of apparently untargeted PPR proteins). Furthermore, because less than half of the PPR ORF models could be validated by cDNA sequences (see below) and because small exons were found at the 5'-end of a few cDNA-validated PPR ORF models, some of the models are undoubtedly lacking the correct initiation codon and thus are missing a potential targeting peptide. Although it is not possible to definitively conclude, our targeting prediction results are consistent with the idea that very few, if any, PPR proteins are localized outside the organelles of plant cells.

To further characterize the subcellular localization of PPR proteins, we experimentally analyzed the localization of several PPR-green fluorescent protein (GFP) or PPR-DsRed2 fluorescent fusion proteins. A set of 48 PPR proteins representing each subclass of the PPR family was chosen for the targeting experiments. The corresponding full-length ORFs (from start codon to the last codon before the stop codon) and the N-terminal targeting signal coding sequences (arbitrarily fixed as being the first 300 bp of the coding region of the genes) were systematically cloned into a pDONR207 Gateway entry vector. This recombinational cloning system was subsequently used to further subclone PPR ORFs and PPR presequences into several GFP or DsRed2 destination vectors (see Methods for details). pOL-type vectors (Peeters et al., 2000) were used for electroporation of tobacco protoplasts. *Agrobacterium* carrying binary vectors were used for agroinfiltration of *Nicotiana benthamiana* leaves. The

**Figure 4.** Predicted Subcellular Localization of PPR Proteins.

The proportions of each PPR subclass predicted by Predotar to be targeted to mitochondria (black segments) or plastids (gray segments) or to lack targeting signals (white segments) are indicated. PPR and E proteins are mostly predicted to be mitochondrial; E+ and DYW proteins are predicted to be more evenly distributed between the two organelles.

subcellular localization of 45 PPR proteins was analyzed in these experiments. We were able to observe fluorescence for 26 out of these 45 PPR protein fusions (Table 2). For the remaining 19 PPR proteins, we were unable to detect any interpretable fluorescence signal for any of the constructs tested (data not shown), the fusion proteins either aggregating or not accumulating to detectable levels.

When several different constructs or transformation methods (Table 2) were used for the experimental localization of the same PPR protein, the results were always consistent. In summary, as shown in Table 2, 18 PPR proteins were shown to be targeted to mitochondria, and eight were shown to be targeted to plastids. The results were in general consistent with the bioinformatics predictions for the proteins tested,

**Table 2.** Subcellular Localization of Selected PPR Proteins

AGI Code	PPR Subclass	Electroporation Experiments	Agroinfiltration Experiments	Organellar Proteomics	TargetP Prediction	Predotar Prediction
At1g11630	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At1g26460	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At1g55890	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Possibly mitochondria
At1g59720	DYW	ND	Mitochondria <sup>b,c</sup>		Plastid	None
At1g60770	PPR	ND	ND	Mitochondria <sup>a</sup>	None	Possibly mitochondria
At1g61870	PPR	Mitochondria <sup>c,d,e</sup>	Mitochondria <sup>c</sup>	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At2g01390	PPR	Mitochondria <sup>c</sup>	ND		Plastid	Possibly mitochondria
At2g19280	PPR	Mitochondria <sup>b</sup>	ND		Mitochondria	Possibly plastid
At2g20540	E+	No signal <sup>d</sup>	Mitochondria <sup>b</sup>		None	None
At2g34370	DYW	Mitochondria <sup>c,e</sup>	ND		Mitochondria	Mitochondria
At3g02010	DYW	Mitochondria <sup>b,d,e</sup>	ND		Mitochondria	None
At3g11460	DYW	Mitochondria <sup>d,e</sup>	Mitochondria <sup>b</sup>		Plastid	Plastid
At3g13160	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At3g13770	DYW	Mitochondria <sup>b</sup>	No signal <sup>c</sup>		Mitochondria	Mitochondria
At3g13880	E+	Mitochondria <sup>b</sup>	No signal <sup>c</sup>		Mitochondria	Possibly mitochondria
At3g24000	DYW	Mitochondria <sup>d</sup>	Mitochondria <sup>b</sup>		Mitochondria	Mitochondria
At3g26780	DYW	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At3g60960	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Possibly mitochondria
At4g02750	DYW	Mitochondria <sup>b</sup>	ND		Mitochondria	None
At4g20090	PPR	ND	Mitochondria <sup>b</sup>		Plastid	Possibly ER
At4g35850	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At4g36680	PPR	ND	ND	Mitochondria <sup>a</sup>	Mitochondria	Mitochondria
At5g09950	DYW	ND	Mitochondria <sup>c</sup>		Mitochondria	Plastid
At5g13230	DYW	Mitochondria <sup>b</sup>	ND		Mitochondria	Mitochondria
At5g14770	PPR	Mitochondria <sup>b</sup>	No signal <sup>c</sup>		Mitochondria	Mitochondria
At5g15340	DYW	Mitochondria <sup>b</sup>	ND		Mitochondria	Mitochondria
At5g55740	E+	Mitochondria <sup>b</sup>	Mitochondria <sup>b</sup>		Plastid	Plastid
At1g74850	PPR	ND	ND	Plastid <sup>f</sup>	Plastid	Possibly plastid
At2g01860	PPR	Plastids <sup>b</sup>	ND		Plastid	Possibly plastid
At2g03880	DYW	Plastids <sup>b</sup>	ND		Mitochondria	Mitochondria
At2g29760	DYW	Plastids <sup>b</sup>	ND		Plastid	Plastid
At3g04760	PPR	ND	ND	Plastid <sup>f</sup>	Plastid	Possibly plastid
At3g23020	PPR	ND	ND	Plastid <sup>f</sup>	None	Possibly plastid
At3g42630	PPR	ND	ND	Plastid <sup>f</sup>	None	Possibly mitochondria
At3g46870	PPR	ND	ND	Plastid <sup>f</sup>	Plastid	Possibly mitochondria
At4g16390	PPR	Plastids <sup>d,e</sup>	ND		Plastid	Possibly plastid
At4g31850	PPR	Plastids <sup>c,e</sup>	ND		Plastid	Plastid
At5g13270	DYW	Plastids <sup>b</sup>	ND		Plastid	Plastid
At5g52630	DYW	Plastids <sup>c,e</sup>	Plastids <sup>b</sup>		None	None

<sup>a</sup>Organellar proteomics data from Heazlewood et al. (2004).

Localization was studied by fusion of presequences to the following:

<sup>b</sup>DsRed2,

<sup>c</sup>GFP, or

<sup>d</sup>full-length ORFs to DsRed2.

<sup>e</sup>The sequence of the entry clone used for the LR reaction was verified.

<sup>f</sup>Organellar proteomics data from Kleffmann et al. (2004).

Experimentally proven localizations were compared to bioinformatics predictions from TargetP and Predotar. The first and second parts of the table correspond to mitochondrial- and plastid-localized PPR proteins, respectively.

which implies that the predictions for the family as a whole are likely to be not far wrong. Whereas some PPR proteins were predicted to be nonorganellar, mitochondria and plastids were the only subcellular localization that we observed in our experiments. In addition, no cases of dual targeting to mitochondria and plastids were observed for any of the fluorescent fusions.

### Analysis of PPR Gene Expression

One of the reasons why such a huge family of proteins was discovered only very recently in Arabidopsis is that few ESTs corresponding to PPR genes were previously identified. Among the 210,388 cDNA and EST sequences (including many sequences from normalized and subtracted libraries) available from FLAGdb<sup>++</sup> (Samson et al., 2004), only 843 transcripts are from PPR genes, corresponding to only 215 out of the 441 genes (48.7%) described here (Table 3; see Supplemental Table S5 online). Using the same database, 69.1% of all predicted Arabidopsis genes have proof of expression from EST or cDNA data. This implies that PPR genes are expressed at low levels and could even mean that many are not expressed at all. To investigate this further, we analyzed PPR transcripts in RNA from two organs expected to show high expression of genes for mitochondrial and plastid proteins, namely flowers and leaves. This was performed by systematic RT-PCR amplification using 372 primer pairs from the Complete Arabidopsis Transcriptome MicroArray (CATMA) project (Hilson et al., 2004), each corresponding to a different PPR transcript, and also by microarray experiments using the full set of CATMA gene-specific probes. Each technique detected expression of ~40 to 50% of PPR genes, but taken together, evidence of expression of the vast majority of PPR genes could be obtained (Table 3). Interestingly, all three methods of transcript analysis detected a much higher proportion of transcripts from PPR subfamily genes than from P-L-S subfamily genes. This difference is highly significant ( $\chi^2$  test,  $P < 10^{-11}$ ). The same tendency toward higher expression of PPR subfamily genes was found in a quantitative analysis of the microarray data (Figure 5A). We found that members of the pure PPR subclass were in general expressed at higher levels than the other members of the family (Wilcoxon rank sum test,  $P < 10^{-15}$ ). Although hybridization signals between different genes on microarrays are not strictly comparable because of variations in probe

efficiency, the large number of genes in the family allows broad conclusions to be drawn. Taken together, these results confirm the low overall level of expression of the PPR gene family and the extremely low expression levels of the P-L-S subfamily. We looked at various experiments comparing different plant organs, and we never observed any obvious deviation of the PPR cloud from the central diagonal, indicating that these genes are expressed generally constitutively. This is illustrated in Figures 5B and 5C, where the genes encoding predicted plastid-targeted PPR proteins fail to show the bias toward higher expression in leaves exhibited by the bulk of genes encoding predicted chloroplast proteins and in fact behave indistinguishably from genes encoding predicted mitochondrial PPR proteins.

### Analysis of the Mutant Collections

Reverse genetics and especially sequenced insertion mutant collections are valuable resources for obtaining information about the function of unknown genes. The collections from Versailles (Samson et al., 2002), from GABI (Li et al., 2003), and from the Salk Institute (Alonso et al., 2003) were screened to look for insertions in Arabidopsis PPR genes. Not surprisingly, given the large size of the family, many putative insertion mutants are listed in these databases. We further characterized some of these mutants corresponding to different subclasses of the PPR family. Among the 25 mutants (corresponding to 21 genes) we have genetically characterized so far, we were unable to find homozygous plants for six mutants (five genes) when we looked at progeny from heterozygous plants (Table 4). Dissection of the siliques of heterozygous mutant plants showed that about one-fourth of the embryos aborted at an early stage (data not shown), indicating that the homozygous mutation was lethal in early stages of embryo development. Furthermore, in direct genetic screens for embryo-lethal mutations, the Arabidopsis SeedGenes project (Tzafir et al., 2003) has identified many PPR mutants (included in Table 4). Taken together, these results indicated that many PPR genes have an essential function in plant embryos that leads to the abortion of the embryos if the genes are mutated. For the 19 remaining PPR mutants, homozygous plants were found. Most of them did not have any obvious macroscopic phenotype (when looking at growth and development of plants in soil or in vitro). Five mutant lines were found with a clear visible phenotype on homozygous plants: three of these present short siliques containing many deformed embryos. Two of these lines carry an insertion in the same gene (At3g25970). A fourth line with a clear slow growth phenotype carries an insertion in a gene (At1g20230) that when knocked out by an independent insertion gives rise to an embryo-lethal phenotype. Twelve out of fourteen genes encoding proteins predicted to be targeted to plastids give rise to a phenotype when the mutation is homozygous (usually embryo lethality) but 7 of the 12 mutations in genes encoding predicted mitochondrial proteins generate no observable macroscopic phenotype.

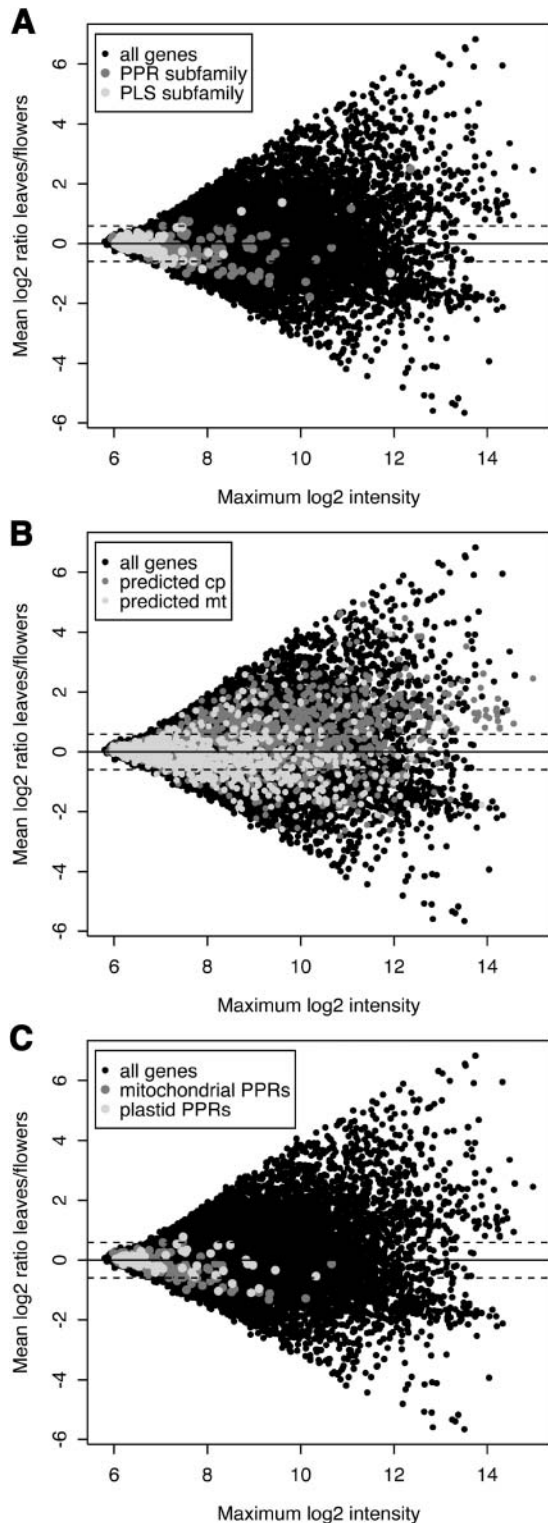
**Table 3.** Expression of PPR Genes Detected by Different Methods

Subfamily	Detection of Gene Expression			
	ESTs	RT-PCR	Microarrays	By at Least One Technique
PPR	148 (61.4%)	93 (50.3%)	128 (69.2%)	228 (94.6%)
P-L-S	67 (33.5%)	63 (35.0%)	58 (32.2%)	139 (69.5%)
Total	48.8%	41.9%	50.0%	83.2%

The number of genes whose expression could be detected by different techniques is given for both subfamilies (the percentage detected in the subfamily is given in parentheses). EST data is from FLAGdb<sup>++</sup>; RT-PCR experiments were conducted on Arabidopsis leaf and flower mRNA with 372 gene-specific primer pairs from the CATMA collection; microarray hybridization results are from the data in Figure 5, covering 384 PPR genes.

### RNA Binding Activity of Some PPR Proteins

As a preliminary study of the RNA binding capacity of the PPR protein family, the proteins At3g25970, At5g13270, At5g12100,



**Figure 5.** Expression of Sets of Genes as Measured by Microarray Hybridization Data.

RNA was extracted from *Arabidopsis* rosette leaves or flowers and hybridized to CATMA arrays containing 24,576 gene-specific probes.

and At1g79540, representing different subclasses of the family, were synthesized *in vitro* and incubated with homoribopolymers, ssDNA, and dsDNA. All four PPR proteins preferentially bound poly(G) homoribopolymers (Figure 6A), whereas the previously characterized RNA recognition motif-containing mRBP2b protein (Vermel et al., 2002), which was used as a positive control in this study, bound poly(U) and poly(G) RNA. Consistent with these data, competition experiments indicated that for each of the four PPR proteins analyzed, the addition of 50  $\mu$ g of either poly(G) RNA or total RNA, competed for their poly(G) binding capacity (Figure 6B; data not shown). In conclusion, these results demonstrate that the four PPR proteins tested show a much higher affinity to poly(G) RNA than to the other three homoribopolymers and DNA. The biological significance of these findings remains to be elucidated in future experiments, hopefully via the identification of physiological RNA targets.

## DISCUSSION

### A Model for the Action of PPR Proteins

At first sight, the genetic analysis of the PPR mutants is quite surprising. How can a mutation in a single gene of a 441-gene family lead to a phenotype as dramatic as the death of the embryo? How can a mutation in another very similar gene of the same family give no obvious phenotype?

The data accumulated so far concerning PPR proteins are mostly consistent with the view that they are RNA binding proteins involved in posttranscriptional processes (RNA processing and translation) in mitochondria and chloroplasts.

Data from four independent two-color hybridizations (comprising two dye swaps) were corrected and averaged as described in Methods. The scales are logarithmic ( $\log_2$ ), representing the mean signal ratio (leaves/flowers) against the maximum mean signal intensity (leaves or flowers). The dotted lines indicate the ratios (0.46) above or below that which the statistical analysis indicates the genes to be differentially expressed in this set of experiments.

**(A)** Expression of PPR genes. Data points corresponding to the PPR and P-L-S subfamilies are depicted in dark gray or light gray, respectively. In general, the PPR subfamily is more highly expressed (Wilcoxon rank sum test,  $P < 10^{-15}$ ).

**(B)** Expression of genes predicted to encode plastid or mitochondrial proteins. Data points corresponding to genes encoding Predotar-predicted plastid or mitochondrial proteins (cutoff 0.5) are depicted in dark gray or light gray, respectively. The predicted plastid set shows a strong bias toward higher expression in leaves (Wilcoxon rank sum test,  $P < 10^{-15}$ ). cp, chloroplast; mt, mitochondria.

**(C)** Expression of genes predicted to encode plastid or mitochondrial PPR proteins. Data points corresponding to genes encoding Predotar-predicted mitochondrial or plastid PPR proteins (cutoff 0.25) are depicted in dark gray or light gray, respectively. The two sets do not show significantly different distributions of leaf/flower expression ratios (Wilcoxon rank sum test,  $P > 0.75$ ) and are much less biased toward expression in leaves than the complete predicted plastid set shown in **(B)** (Wilcoxon rank sum test,  $P < 10^{-15}$ ) while being slightly less biased toward expression in flowers than the complete mitochondrial set (Wilcoxon rank sum test,  $P < 0.04$ ).



**Table 4.** List of Arabidopsis T-DNA Mutants with Insertions in PPR Genes

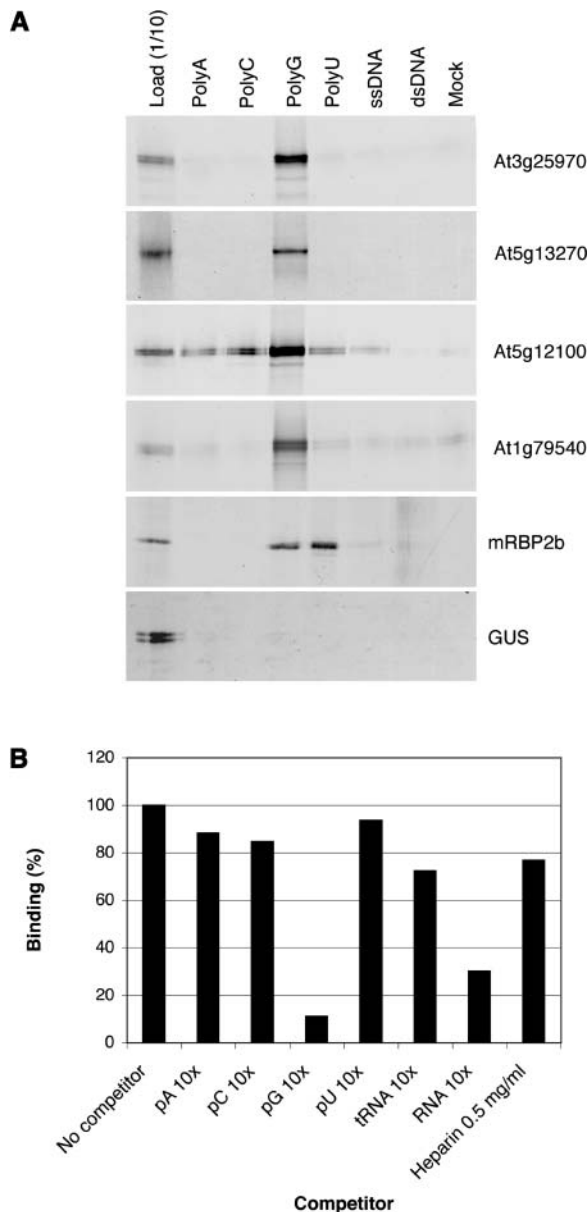
Mutant ID	AGI Code	Subclass	Targeting	Homozygotes	Phenotype of Homozygotes
DNR2	At1g14470	PLS	?	Yes	Dwarf, reduced fertility
EAN42	At1g20230	DYW	?	No	– <sup>a</sup>
Salk_033409	At1g20230	DYW	?	Yes	Dwarf, late flowering
DVM10	At2g40720	E+	?	Yes	–
DDU146	At3g25970	E+	?	Yes	Dwarf, reduced fertility
Salk_127777	At3g25970	E+	?	Yes	Dwarf, reduced fertility
EMB 2744	At5g39680	DYW	?	No	Embryo defective
DJU9	At1g06710	PPR	Mitochondria	Yes	–
T243	At1g22830	E	Mitochondria	Yes	–
DYK12	At1g51965	PPR	Mitochondria	Yes	–
ABP1	At1g80550	PPR	Mitochondria	No	–
Salk_024306	At1g80550	PPR	Mitochondria	No	–
DNP1	At2g20540	E+	<b>Mitochondria</b>	Yes	–
DXZ45	At3g13150	PPR	Mitochondria	Yes	–
Salk_015469	At3g49140	DYW	Mitochondria	Yes	–
EMB 1796	At3g49240	PPR	Mitochondria	No	Embryo defective
EMB 1025	At4g20090	PPR	<b>Mitochondria</b>	No	Embryo defective
EMB 2745	At5g39710	PPR	Mitochondria	No	Embryo defective
Salk_033891	At5g46460	DYW	Mitochondria	Yes	–
DSN8	At5g62370	PPR	Mitochondria	No	–
DZY11	At5g65570	DYW	Mitochondria	Yes	–
EMB 1444	At1g06150	E+	Plastids	No	Embryo defective
Salk_067029	At1g09410	DYW	<b>Plastids</b>	Yes	–
Gabi_104E05	At1g09410	DYW	<b>Plastids</b>	Yes	–
EMB 2279	At1g30610	PPR	Plastids	No	Embryo defective
CYK8	At2g01860	PPR	<b>Plastids</b>	No	Embryo defective
<i>hcf152</i>	At3g09650	PPR	<b>Plastids</b>	Yes	Defective processing of <i>petB</i>
EMB 1270	At3g18110	PPR	Plastids	No	Embryo defective
Salk_040629	At3g22150	E+	Plastids	Yes	–
<i>crr2</i>	At3g46790	DYW	Plastids	Yes	Defective processing of <i>ndhB</i>
Salk_046131	At3g46790	DYW	Plastids	Yes	Defective processing of <i>ndhB</i>
Gabi_425F06	At3g46790	DYW	Plastids	Yes	Defective processing of <i>ndhB</i>
EMB 2261	At3g49170	DYW	Plastids	No	Embryo defective
<i>pgr3</i>	At4g31850	PPR	<b>Plastids</b>	Yes	Cytochrome <i>b<sub>6</sub>f</i> and/or NDH-deficient
CRI6	At4g31850	PPR	<b>Plastids</b>	Yes	Yellow leaves
EMB 2453	At4g39620	PPR	Plastids	No	Embryo defective
EMB 1899	At5g03800	DYW	Plastids	No	Embryo defective
DYH216	At5g27270	PPR	Plastids	No	Embryo defective
EMB 1006	At5g50280	PPR	Plastids	No	Embryo defective

<sup>a</sup> Not determined.

The mutants are named according to the nomenclature of their collections of origin: Salk\_xxx mutants are from the Salk collection (Alonso et al., 2003), Gabi\_xxx mutants are from the GABI-Kat collection (Li et al., 2003), EMB xxx mutants are from the SeedGenes project (Tzafirir et al., 2003), and the other mutants are from the Versailles collection (Samson et al., 2002). The targeting column indicates predictions based on Predotar and TargetP scores (see Supplemental Table S3 online), or if the protein localization has been experimentally determined (Table 2), the organelle is indicated in bold. For each mutant, we indicate whether or not homozygotes could be obtained, and if so, the phenotype of the homozygous plants. When homozygotes could not be obtained, in some cases the siliques of heterozygous plants were examined for evidence of embryo abortion, and the phenotype is indicated as embryo defective. Several alleles of the mutants *hcf152*, *crr2*, and *pgr3* have been described (Hashimoto et al., 2003; Meierhoff et al., 2003; Yamazaki et al., 2004). Unlike the other mutants listed here, these are mostly point mutations and are included for comparison.

Posttranscriptional processes in organelles are incredibly complex, involving many highly specific processing events (trimming of 5' and 3' ends, *cis*- and in some cases *trans*-splicing, stabilization or destabilization of transcripts, etc.) and a relatively poorly understood control of translation initiation at internal AUG

codons (Barkan and Goldschmidt-Clermont, 2000; Giege and Brennicke, 2001). The PPR family could potentially be involved in many of these functions. Plant cells are not viable without mitochondrial respiration, and in addition, mutations in genes encoding essential plastid proteins lead to an embryo-lethal



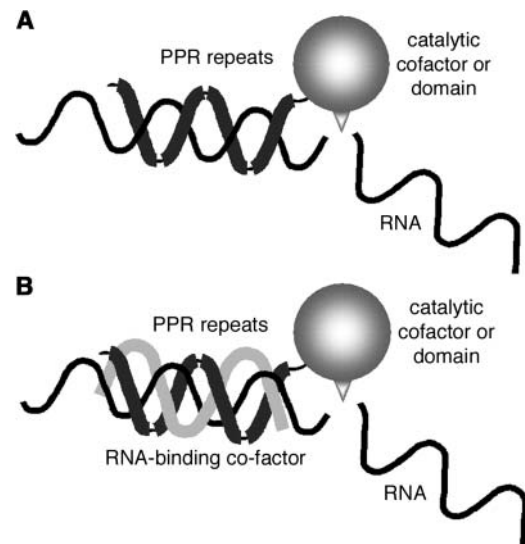
**Figure 6.** RNA Binding Assay for Four Representative Arabidopsis PPR Proteins.

Two of these proteins (At1g79540 and At5g12100) are P subfamily members. At3g25970 is an E+ protein, and At5g13270 is a DYW protein. **(A)** Radioactively labeled protein retained on Sepharose columns carrying various polyribonucleotides, single-stranded DNA (ssDNA), or double-stranded DNA (dsDNA). The left lane was loaded with one-tenth of the protein loaded on the columns. The far right lane shows binding to a Sepharose column lacking added nucleotides. mRBP2b is a previously characterized RNA binding protein of the RNA recognition motif family, used here as a positive control. The bottom panel shows binding of  $\beta$ -glucuronidase (GUS) as a negative control.

**(B)** Competition assays for At5g12100. Labeled proteins were preincubated with competitor RNA or heparin before being loaded on a poly(G) Sepharose column and binding quantified by a phosphorimager, using binding in absence of competitor as 100%.

phenotype in Arabidopsis (Budziszewski et al., 2001). The most frequent phenotype observed for T-DNA insertions in PPR genes (i.e., a lack of homozygous mutants) is therefore consistent with a role for PPR proteins in organelle biogenesis. The genetic results also imply that there is little or no redundancy of function between many PPR proteins, despite the huge size of the family. Furthermore, the very limited primary defects observed in the few mutants that have been studied, the specificity of action of the cloned CMS restorer genes, and the low expression levels of the majority of the genes in the family all imply that these proteins are for the most part targeted to specific transcripts rather than being general unspecific RNA binding factors. The large majority of these proteins lack any obvious domains likely to have catalytic activity; thus, we imagine their role as adaptors for directing the action of other factors (Figure 7). In this case, the phenotype of the mutants will be entirely dependent on the identity of the target transcripts; for example, the macroscopic phenotypes of the mutants *crr2*, *hcf152*, and *crp1* range from wild type to lethal despite the fact that the molecular functions of these three PPR proteins are apparently very similar.

This model is probably basically correct in its broad details, but several points still need to be investigated. The RNA binding activity of PPR proteins now seems to be beyond doubt (Lahmy et al., 2000; Mancebo et al., 2001; Meierhoff et al., 2003; Mili and Pinol-Roma, 2003; Nakamura et al., 2003), but there is very little evidence showing binding specificity, in particular to known target transcripts. This may be attributable to technical difficulties (PPR proteins are difficult to produce in soluble form; C. Lurin and H. Mireau, unpublished data) but may also be explained if,



**Figure 7.** A Model for PPR Protein Action.

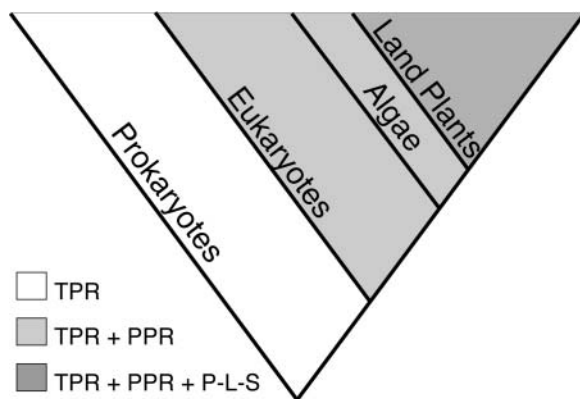
We assume that the putative superhelix formed by tandemly repeated PPR motifs forms a sequence-specific RNA binding surface either alone **(A)** or in the presence of an additional factor **(B)**. The resulting protein-RNA complex recruits one or more other transactors to a specific site on the RNA target (in this case an endonuclease). We assume that in most cases the catalytic site is in the partner protein; for the DYW class of PPR proteins, it may lie in the C-terminal domain itself.

in vivo, PPR proteins bind in concert with one or more other RNA binding proteins (Figure 7B). Such complexes are relatively common, one of the best-studied examples being the binding of Pumilio, Nanos, and Brat to the 3' UTR of *hunchback* mRNA (Sonoda and Wharton, 1999; Edwards et al., 2003). This brings in the more general question of which other proteins are associated with PPR proteins. Several PPR proteins have been shown to be in protein complexes in vivo (Fisk et al., 1999; Liu et al., 2002; Liu and McKeehan, 2002; Tsuchiya et al., 2002; Williams and Barkan, 2003), but in very few cases to date have any of the partners been identified.

### Evolution of the Family and Putative Functions

Probably the most striking feature of the PPR family is its spectacular expansion in plants (Table 1). The Arabidopsis genome codes for 70 times more PPR proteins than the human genome, and all the evidence so far from genome sequencing and large-scale EST projects suggests that other plant genomes encode even more. The P-L-S subfamily is entirely specific to plants but accounts for only about half of the extra plant proteins. From EST data, this subfamily can be traced as far as bryophytes (e.g., *Physcomitrella patens* ESTs AW599911 and AW561409) but is apparently absent from Chlamydomonas. This unusual distribution poses at least three major questions. (1) Where did the plant-specific proteins originate? (2) How did the family grow at least 70-fold in the plant lineage? (3) Why do plants need so many PPR proteins?

It seems more than probable that the pure PPR subfamily in plants has arisen by expansion of the small family of pure PPR genes in all eukaryotes, given the similarities in structure and function (Figure 8). The PPR-like repeats of the plant-specific P-L-S subfamily probably arose by mutation of classical PPR repeats, but the origins of the characteristic C-terminal E, E+, and DYW motifs are mysterious. Many plant-specific genes can be traced to the endosymbiotic ancestor of plastids, but in the case of PPR genes there is no evidence for this whatsoever. As



**Figure 8.** Order of Appearance and Likely Evolutionary Relationships between PPR Families Based on Phylogenetic Distribution.

TPR proteins are ubiquitous, whereas PPR proteins are only found in eukaryotes and the P-L-S subfamily only in land plants.

far as we can tell, PPR proteins are absent from almost all prokaryotes that we have looked at, including the closest living relatives of the organellar endosymbionts. What is more, the lack of these plant-specific proteins in trypanosomes and Chlamydomonas suggests that they were acquired long after the acquisition of plastids.

If the origin of many of these proteins remains obscure, the expansion of the family may be easier to explain. PPR genes stand out from the vast majority of other plant genes in their paucity of introns. When introns are present, they are generally outside the zone of the repeats or in proteins rather divergent from the bulk of the family. This is characteristic of other large gene families with repeated motifs, such as Leu-rich repeat proteins that can evolve rapidly via unequal crossing-over events (Michelmore and Meyers, 1998; Meyers et al., 2003). Such events can fuse, split, duplicate, or delete all or part of a gene (Jelesko et al., 1999). However, a general consequence of this type of evolution is the generation of clusters of related genes, again as typified by disease resistance genes. Apart from the restorer gene-related cluster on chromosome 1, PPR genes do not show such obvious clustering in Arabidopsis. An alternative hypothesis would be that PPR genes have been amplified by one or more bursts of reverse transcription/integration (Lecharny et al., 2003), perhaps by fortuitous association with a retrotransposon. This could also explain the lack of introns while accounting for a wide chromosomal distribution. However, if this is the true explanation, no obvious trace of these events persists.

Our major goals, however, are to discover the functions of PPR proteins in plants to understand why the family has grown so large. Genetic data from yeast and Chlamydomonas strongly suggest that many specific *trans*-acting factors are required for proper expression of each and every transcript in mitochondria and plastids, leading to estimates that several hundred of these factors must exist (Barkan and Goldschmidt-Clermont, 2000; Costanzo et al., 2000). A fair number of these factors have been identified from these single-celled organisms, but only a few are PPR proteins. Clearly hundreds of PPR proteins are not a prerequisite for organelle biogenesis outside the plant kingdom. This leaves two explanations for the huge increase in the PPR family in plants: either PPR proteins play roles in plants that are performed by other proteins in other organisms or PPR proteins perform novel functions in plants absent from animals, fungi, and algae. The first explanation is likely to be correct to some extent because Arabidopsis lacks recognized homologs to most of the transcript-specific mitochondrial factors identified in yeast. The second explanation is very tempting, especially for the plant-specific half of the family. However, the gene content of land plant organelles is not significantly higher than that of other organisms, and their posttranscriptional processes are not notably more complex, with one exception, RNA editing. The posttranscriptional conversion of specific cytosines to uracil in mitochondrial and plastid transcripts is unique to land plants (Steinhauser et al., 1999). Several hundred editing sites in Arabidopsis mitochondria (Giege and Brennicke, 1999) and 30 or so in plastids (Tsudzuki et al., 2001) have been cataloged. These sites appear to require many specific *trans* factors to provide the observed editing specificity (Chateigner-Boutin and Hanson, 2002, 2003). It now seems accepted that these factors are probably proteins (Hirose

et al., 1999; Miyamoto et al., 2002, 2004), but their identity remains unknown. The evolutionary expansion of the PPR family and the appearance of the plant-specific P-L-S subfamily correlates with the appearance of C-to-U RNA editing within the coarse resolution permitted by the EST data. The only mutants in this subfamily so far described are the *Arabidopsis crr2* mutants defective in processing of the plastid *ndhB* transcript (Hashimoto et al., 2003). However, editing of *ndhB* transcripts is not obviously affected in these mutants (data not shown), and it should be stressed that there is as yet no physical evidence for an involvement of PPR proteins in RNA editing.

## Conclusion

We have performed an exhaustive, systematic bioinformatics analysis of *Arabidopsis* PPR genes and proteins. The resulting descriptions of the family and its members should provide a solid and unified platform on which future genetic and functional studies can be based. The preliminary functional characterization of a range of family members, together with the sparse data on these proteins published, allows us to propose a plausible model (Figure 7) in which PPR proteins play the role of sequence-specific adaptors for a variety of other RNA-associated proteins. This model makes two testable predictions: PPR proteins should (directly or indirectly) associate with specific RNA sequences and with defined effector proteins. Future work needs to be directed toward the identification of these factors to elucidate the precise functions of one of the largest and least understood protein families in plants.

## METHODS

### Bioinformatics

*Arabidopsis thaliana* genomic sequences, predicted protein sequences, and functional annotations were obtained from TIGR ([ftp://ftp.tigr.org/pub/data/a\\_thaliana/](ftp://ftp.tigr.org/pub/data/a_thaliana/)). ClustalW version 1.81 (Thompson et al., 1994) was used for aligning multiple peptide sequences to construct alignments as input to the HMMER package versions 2.2g (Eddy, 1998), which was used to construct models of different PPR motifs and then to search peptide sequences for these motifs. The programs hmmbuild and hmmcalibrate were used to construct matrices describing the motifs, and hmmsearch was used for discovering motifs within protein sequences. The default parameters were used in all cases.

A modified version of the FLAGdb<sup>++</sup> database and interface (Samson et al., 2004) was used for much of the visualization needed for manual annotation, including display of PPR motifs, AGI ORF models, known mRNAs, and predicted mitochondrial and plastid targeting sequences. Predotar version 1.03 (Small et al., 2004) and TargetP version 1.01 (Emanuelsson et al., 2000) were used for prediction of organelle targeting from protein sequences. A modified version of Predotar was used to predict regions encoding putative targeting sequences in genomic DNA. Secondary structure predictions of PPR motifs were performed with the PredictProtein server (Rost and Liu, 2003) using the alignments of motifs described previously.

The complete predicted proteomes of various organisms used to create Table 1 were obtained from the National Center for Biotechnology Information RefSeq database, apart from *Trypanosoma brucei* sequences (from [ftp://ftp.sanger.ac.uk/pub/databases/T.brucei\\_sequences/](ftp://ftp.sanger.ac.uk/pub/databases/T.brucei_sequences/)) and

*Cyanidiodschyzon merolae* sequences (from <http://merolae.biol.s.u-tokyo.ac.jp/>).

The alignment files, hmm matrices, and search results are available in the supplemental data online.

### Transcriptome Studies

The microarray analysis has been performed with the CATMA array containing 24,576 gene-specific tags (GSTs) from *Arabidopsis* (Hilson et al., 2004). The GST amplicons were purified on Multiscreen plates (Millipore, Bedford, MA) and resuspended in TE-DMSO at 100 ng/ $\mu$ L. The purified probes were transferred to 1536-well plates with a Genesis workstation (TECAN, Männedorf, Sweden) and spotted on UltraGAPS slides (Corning, Corning, NY) using a Microgrid II (Genomic Solution, Huntingdon, UK). The CATMA array printed at the Unité de Recherche en Génomique Végétale consists of three metablocks, each composed of 64 blocks of 144 spots. A block is a set of spots printed with the same print tip. In these arrays, a print tip is used three times to print a block in each metablock. For the transcriptome studies, eight sets of 25 *Arabidopsis* Columbia-0 plants were grown on horticultural potting soil (Tref substrate with NFU 44-571 fertilizer; BAAN, Vulaines, France) under cool white light at 100  $\mu$ mol m<sup>-2</sup> s<sup>-1</sup> with a 16-h photoperiod at 22°C, 50% humidity. Four pooled samples of young leaves were harvested at stage 3.9, and four pooled samples of flower buds harvested at stage 6.0, the material being collected 5 to 7 h into the light period. RNA was extracted from these samples using TRIzol extraction (Invitrogen, Carlsbad, CA) followed by two ethanol precipitations and then checked for RNA integrity with the Bioanalyzer from Agilent (Waldbronn, Germany). cRNAs were produced from 2  $\mu$ g of total RNA from each sample with the Message Amp aRNA kit (Ambion, Austin, TX). Then 5  $\mu$ g of cRNAs were reverse transcribed in the presence of 300 units of SuperScript II (Invitrogen), cy3-dUTP, and cy5-dUTP (NEN, Boston, MA) according to Puskas et al. (2002) for each slide. Samples were combined, purified, and concentrated with YM30 Microcon columns (Millipore). Slides were prehybridized for 1 h and hybridized overnight at 42°C in 25% formamide. Slides were washed in 2 $\times$  SSC + 0.1% SDS for 4 min, 1 $\times$  SSC for 4 min, 0.2 $\times$  SSC for 4 min, and 0.05 $\times$  SSC for 1 min and dried by centrifugation. Four hybridizations (two dye swaps) were performed. The arrays were scanned on a GenePix 4000A scanner (Axon Instruments, Foster City, CA), and images were analyzed by GenePix Pro 3.0 (Axon Instruments).

For RT-PCR validation, mRNAs were purified from the extracts described above with Dynabeads (Dyna, Oslo, Norway) according to the manufacturer's specifications. Fifty nanograms of mRNA per 20- $\mu$ L reaction were reverse transcribed in the presence of 10 units of SuperScript II (Invitrogen). PCR reactions for each sample were performed in 96-well plates each with 0.5  $\mu$ L of the RT reaction and the corresponding set of specific CATMA primers for each of 384 PPR genes in presence of 2.5 units of Platinum Taq DNA polymerase (Invitrogen). The PCR conditions were 94°C for 1 min, followed by 10 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 30 s (with a 1°C decrease of the annealing temperature at each cycle), followed by 25 cycles of 94°C for 30 s, 68°C for 30 s, 72°C for 30 s, and finally 72°C for 7 min. Additional PCR reactions were performed on RT minus samples as negative controls. The RT-PCR products were validated by size verification after electrophoresis on agarose gels.

### Statistical Analysis of Microarray Data

The statistical analysis was based on two dye swaps (i.e., four arrays each containing 24,576 GSTs and 384 controls). The controls were used for assessing the quality of the hybridizations but were not included in the statistical tests or the graphical representation of the results shown in the Results. For each array, the raw data comprised the logarithm of median feature pixel intensity at wavelengths 635 nm (red) and 532 nm (green).

No background was subtracted. In the following description, log ratio refers to the differential expression between leaves and flowers. It is either  $\log_2(\text{red/green})$  or  $\log_2(\text{green/red})$  according to the experiment design. An array-by-array normalization was performed to remove systematic biases. First, we excluded spots that were considered badly formed features. Then we performed a global intensity-dependent normalization using the loess procedure (see Yang et al., 2002) to correct the dye bias. Finally, for each block, the log-ratio median calculated over the values for the entire block was subtracted from each individual log-ratio value to correct print tip effects on each metablock.

To study the expression of PPR genes, we first tested if the distribution of the maximum signal intensity of the leaf and the flower samples differed between PPR subfamilies. For each array, and for each gene, we calculated according to the method of Yang and Thorne (2003) a normalized logarithm of median feature pixel intensity. We kept the maximum of these normalized intensities for each gene. The statistical test used was a one-tailed Wilcoxon rank sum test.

To determine differentially expressed genes, we performed a paired *t* test on the log ratios. The number of observations per spot varies between two and four and is inadequate for calculating a gene-specific variance. For this reason, we assume that the variance of the log-ratios is the same for all genes, and we excluded 256 spots displaying extremes of variance (too small or too large). The raw *P* values were adjusted by the Bonferroni method, which controls the Family Wise Error Rate (FWER) (Ge et al., 2003). A total of 6566 genes (26.7%) were found to be differentially expressed between flowers and leaves taking a FWER of 5%. Amongst these, only 45 are PPR genes, out of the 384 represented by GSTs in the array (i.e., only 11.7% of the PPR genes are differentially expressed in these experiments). The relative stability of PPR gene expression was confirmed by applying Wilcoxon rank sum tests on the normalized log-ratios (Figure 5C) to determine whether the expression differs or not between leaves and flowers.

After the removal of badly formed features, we retained four replications of the normalized log-ratios for 24,391 GSTs, three observations for 67 GSTs, and two observations for 118 GSTs. The absolute value of log-ratio threshold corresponding to a FWER of 5% is 0.46 when four observations are available. It increases respectively to 0.54 and 0.66 when three and two observations are available. In Figure 5, we indicate the thresholds of the log-ratios corresponding to a FWER of 5% when four observations are available.

### Construction of a Collection of PPR ORFs

The ORFs of 48 intronless PPR genes were PCR amplified from Arabidopsis (ecotype Columbia-0) genomic DNA using *Pfx* DNA polymerase and primers flanked with *attB1* and *attB2* sites and cloned into the entry vector pDONR207 by Gateway recombination (Invitrogen). For targeting experiments, either the whole ORF or just the first 100 codons were cloned in this way. For the nucleic acid binding assays, the ORFs were cloned starting from the predicted start of the mature protein to remove most or all of the putative targeting sequences. TargetP was used for predicting the extent of the targeting sequences. In all cases, the 5' primer included Kozak and Shine-Dalgarno consensus sequences to maximize translation in eukaryotic or prokaryotic systems. Details on the cloning procedures, including the primer sequences, are available in the supplemental data online.

### Subcellular Localization of Proteins

Two types of experiments were performed. The sequences coding for precursor PPR proteins or putative targeting sequences were shuttled into Greengate or Redgate vectors (<http://www.evry.inra.fr/public/projects/ppr/greengate.html>) from the corresponding pDONR207 vector

using Gateway recombination (Invitrogen). These constructs were used to express GFP or DsRed2 fusion proteins in electroporated tobacco protoplasts using previously described protocols (Akashi et al., 1998; Peeters et al., 2000).

Binary destination vectors (p0229-GFP and p0229-RFP) were created by transferring the GFP or DsRed2 cassettes, respectively, from Greengate or Redgate into pGreenII 0229 (Hellens et al., 2000), and the sequences coding for precursor PPR proteins or putative targeting sequences were shuttled into these vectors from the corresponding pDONR207 as before. In this case GFP- and RFP-fusion proteins were transiently expressed in *Nicotiana benthamiana* leaves via agroinfiltration. *Agrobacterium tumefaciens* strain GV3101 harboring the transformation helper pSoup (Hellens et al., 2000) and the GFP or DsRed2 constructs were grown overnight and resuspended in infiltration media (10 mM Mes, pH 5.3, 5 mM MgCl<sub>2</sub>, and 150 mM acetosyringone) at an OD of 0.5. Protoplasts were prepared from leaf material (harvested 48 h after infiltration) cut into small pieces and incubated in enzyme solution (Chupeau et al., 1974) at 30°C for 2 to 3 h. Fluorescent proteins were visualized using an Eclipse TE2000S inverted microscope (Nikon, Kanagawa, Japan), a GFP-B filter block (Nikon), and a custom filter block for detection of DsRed2 (exciter HQ546/12, emitter HQ605/75, beamsplitter Q560lp; Chroma Technology, Rockingham, VT).

### Nucleic Acid Binding Assay

The sequences coding for the mature PPR proteins and  $\beta$ -glucuronidase were shuttled into pDEST17 from the corresponding pDONR207 vector using Gateway recombination (Invitrogen). Proteins were synthesized by coupled transcription-translation in the presence of [<sup>35</sup>S]Met (TNT coupled reticulocyte lysate; Promega, Madison, WI). Homoribopolymers [poly(A), poly(U), poly(G), and poly(C)] cross-linked to agarose beads were obtained from Sigma-Aldrich (St. Louis, MO). Concentrations were equilibrated to 0.5 mg per mL packed gel by dilution with the appropriate volume of Sepharose 4B (Amersham Biosciences, Little Chalfont, UK). Single- and double-stranded DNA cellulose were obtained from Sigma-Aldrich. The nucleic acid binding assay was performed as described by Vermel et al. (2002). A 10-fold excess of competitor RNA was used for competition experiments. The bound fraction of proteins was analyzed by SDS-PAGE, and the relative protein amounts were calculated by a phosphor imager (Fujix Bas 1000; Fuji Photo Film, Tokyo, Japan).

### Mutant Analyses

Arabidopsis plants were grown on soil in a growth chamber under the following conditions: photoperiod 16 h day (100  $\mu\text{mol m}^{-2} \text{s}^{-1}$ )/8 h night; temperature 22°C day/16°C night; humidity 80%.

Plants were genotyped by PCR. Two specific primers from each side of the insertion site were used to detect wild-type loci. One primer specific to the genomic sequence near the insertion site and one primer from the T-DNA border were used to amplify a fragment flanking each insertion. These fragments were sequenced for precise localization of the insert. In all mutants described here, the insertion is within the coding sequence. Plants from at least 10 seeds of each initial stock were genotyped to obtain at least one heterozygous plant. Twenty-four seeds from each heterozygote were grown for genotyping to detect plants homozygous for the T-DNA insertion.

The complete set of microarray data from these experiments has been submitted to ArrayExpress (accession numbers A-MEXP-60 and E-MEXP-94).

Received March 3, 2004; accepted April 22, 2004.

## REFERENCES

- Akashi, K., Grandjean, O., and Small, I.** (1998). Potential dual targeting of an *Arabidopsis* archaeobacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *FEBS Lett.* **431**, 39–44.
- Alonso, J.M., et al.** (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657.
- Arabidopsis Genome Initiative.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Aubourg, S., Boudet, N., Kreis, M., and Lechamy, 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., and Goldschmidt-Clermont, M.** (2000). Participation of nuclear genes in chloroplast gene expression. *Biochimie* **82**, 559–572.
- Bentolila, S., Alfonso, A.A., and Hanson, M.R.** (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl. Acad. Sci. USA* **99**, 10887–10892.
- Brown, G.G., Formanova, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., Zhang, J., Cheung, W.Y., and Landry, B.S.** (2003). The radish *Rfo* restorer gene of *Ogura* cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* **35**, 262–272.
- Budziszewski, G.J., et al.** (2001). *Arabidopsis* genes essential for seedling viability: Isolation of insertional mutants and molecular cloning. *Genetics* **159**, 1765–1778.
- Chateigner-Boutin, A.L., and Hanson, M.R.** (2002). Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol. Cell. Biol.* **22**, 8448–8456.
- 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.
- Chupeau, Y., Bourgin, J.P., Missonier, C., Dorion, N., and Morel, G.** (1974). Préparation et culture de protoplastes de divers *Nicotiana*. *C. R. Acad. Sci. Paris* **278**, 1565–1568.
- Coffin, J.W., Dhillon, R., Ritzel, R.G., and Nargang, F.E.** (1997). The *Neurospora crassa* *cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COX1 protein. *Curr. Genet.* **32**, 273–280.
- Costanzo, M.C., Bonnefoy, N., Williams, E.H., Clark-Walker, G.D., and Fox, T.D.** (2000). Highly diverged homologs of *Saccharomyces cerevisiae* mitochondrial mRNA-specific translational activators have orthologous functions in other budding yeasts. *Genetics* **154**, 999–1012.
- Desloire, S., et al.** (2003). Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep.* **4**, 588–594.
- Eddy, S.R.** (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755–763.
- Edwards, T.A., Wilkinson, B.D., Wharton, R.P., and Aggarwal, A.K.** (2003). Model of the brain tumor-Pumilio translation repressor complex. *Genes Dev.* **17**, 2508–2513.
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G.** (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.
- 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.
- Ge, Y., Dudoit, S., and Speed, T.P.** (2003). Resampling-based multiple testing for microarray data analysis. *TEST* **12**, 1–44.
- 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., and Brennicke, A.** (2001). From gene to protein in higher plant mitochondria. *C. R. Acad. Sci. III* **324**, 209–217.
- Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J.P., Rigden, D.J., Michels, P.A., and Opperdoes, F.R.** (2003). Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl. Acad. Sci. USA* **100**, 1067–1071.
- Hashimoto, M., Tsuyoshi, E., 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.
- Heazlewood, J.L., Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J., and Millar, A.H.** (2004). 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* **16**, 241–256.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M.** (2000). pGreen: A versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* **42**, 819–832.
- Hilson, P., et al.** (2004). Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: Transcript profiling and reverse genetics applications. *Genome Res.*, in press.
- Hirose, T., Kusumegi, T., Tsudzuki, T., and Sugiura, M.** (1999). RNA editing sites in tobacco chloroplast transcripts: Editing as a possible regulator of chloroplast RNA polymerase activity. *Mol. Gen. Genet.* **262**, 462–467.
- 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.
- Jelesko, J.G., Harper, R., Furuya, M., and Grissem, W.** (1999). Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**, 10302–10307.
- Kazama, T., and Toriyama, K.** (2003). A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. *FEBS Lett.* **544**, 99–102.
- Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Grissem, W., and Baginsky, S.** (2004). The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* **14**, 354–362.
- Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S., and Imamura, J.** (2003). Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant J.* **34**, 407–415.
- Komori, T., Ohta, S., Murai, N., Takakura, Y., Kuraya, Y., Suzuki, S., Hiei, Y., Imaseki, H., and Nitta, N.** (2004). Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant J.* **37**, 315–325.
- Lahmy, S., Barneche, F., Derancourt, J., Filipowicz, W., Delseny, M., and Echeverria, M.** (2000). A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett.* **480**, 255–260.
- Lechamy, A., Boudet, N., Gy, I., Aubourg, S., and Kreis, M.** (2003). Introns in, introns out in plant gene families: A genomic approach of the dynamics of gene structure. *J. Struct. Funct. Genomics* **3**, 111–116.
- Li, Y., Rosso, M.G., Strizhov, N., Viehoveer, P., and Weisshaar, B.** (2003). GABI-Kat SimpleSearch: A flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in *Arabidopsis thaliana*. *Bioinformatics* **19**, 1441–1442.

- Liu, L., Amy, V., Liu, G., and McKeehan, W.L.** (2002). Novel complex integrating mitochondria and the microtubular cytoskeleton with chromosome remodeling and tumor suppressor RASSF1 deduced by in silico homology analysis, interaction cloning in yeast, and colocalization in cultured cells. *In Vitro Cell. Dev. Biol. Anim.* **38**, 582–594.
- Liu, L., and McKeehan, W.L.** (2002). Sequence analysis of LRPPRC and its SEC1 domain interaction partners suggests roles in cytoskeletal organization, vesicular trafficking, nucleocytoplasmic shuttling, and chromosome activity. *Genomics* **79**, 124–136.
- Lown, F.J., Watson, A.T., and Purton, S.** (2001). *Chlamydomonas* electron mutants that fail to assemble respiratory or photosynthetic electron transfer complexes. *Biochem. Soc. Trans.* **29**, 452–455.
- Mancebo, R., Zhou, X., Shillinglaw, W., Henzel, W., and Macdonald, P.M.** (2001). BSF binds specifically to the bicoid mRNA 3' untranslated region and contributes to stabilization of bicoid mRNA. *Mol. Cell. Biol.* **21**, 3462–3471.
- 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.
- Manthey, G.M., Przybyla-Zawislak, B.D., and McEwen, J.E.** (1998). The *Saccharomyces cerevisiae* Pet309 protein is embedded in the mitochondrial inner membrane. *Eur. J. Biochem.* **255**, 156–161.
- Martin, W., and Borst, P.** (2003). Secondary loss of chloroplasts in trypanosomes. *Proc. Natl. Acad. Sci. USA* **100**, 765–767.
- Matsuzaki, M., et al.** (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657.
- 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.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W.** (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–834.
- Michelmore, R.W., and Meyers, B.C.** (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**, 1113–1130.
- Mili, S., and Pinol-Roma, S.** (2003). LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. *Mol. Cell. Biol.* **23**, 4972–4982.
- Miyamoto, T., Obokata, J., and Sugiura, M.** (2002). Recognition of RNA editing sites is directed by unique proteins in chloroplasts: Biochemical identification of *cis*-acting elements and *trans*-acting factors involved in RNA editing in tobacco and pea chloroplasts. *Mol. Cell. Biol.* **22**, 6726–6734.
- Miyamoto, T., Obokata, J., and Sugiura, M.** (2004). A site-specific factor interacts directly with its cognate RNA editing site in chloroplast transcripts. *Proc. Natl. Acad. Sci. USA* **101**, 48–52.
- Mootha, V.K., et al.** (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc. Natl. Acad. Sci. USA* **100**, 605–610.
- 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.
- Peeters, N.M., Chapron, A., Giritch, A., Grandjean, O., Lancelin, D., Lhomme, T., Vivrel, A., and Small, I.** (2000). Duplication and quadruplication of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *J. Mol. Evol.* **50**, 413–423.
- Puskas, L.G., Zvara, A., Hackler, L., Jr., and Van Hummelen, P.** (2002). RNA amplification results in reproducible microarray data with slight ratio bias. *Biotechniques* **32**, 1330–1334, 1336, 1338, 1340.
- Rost, B., and Liu, J.** (2003). The PredictProtein server. *Nucleic Acids Res.* **31**, 3300–3304.
- Salanoubat, M., et al.** (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**, 497–502.
- Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M., and Lechamy, A.** (2002). FLAGdb/FST: A database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Res.* **30**, 94–97.
- Samson, F., Brunaud, V., Duchene, S., De Oliveira, Y., Caboche, M., Lechamy, A., and Aubourg, S.** (2004). FLAGdb++: A database for the functional analysis of the *Arabidopsis* genome. *Nucleic Acids Res.* **32** (Database issue), D347–D350.
- Small, I., Peeters, N., Legeai, F., and Lurin, C.** (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581–1590.
- 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.
- Sonoda, J., and Wharton, R.P.** (1999). Recruitment of Nanos to hunchback mRNA by Pumilio. *Genes Dev.* **13**, 2704–2712.
- Steinhauser, S., Beekert, S., Capesius, I., Malek, O., and Knoop, V.** (1999). Plant mitochondrial RNA editing. *J. Mol. Evol.* **48**, 303–312.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J.** (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- Tsuchiya, N., Fukuda, H., Sugimura, T., Nagao, M., and Nakagama, H.** (2002). LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1. *Eur. J. Biochem.* **269**, 2927–2933.
- Tsudzuki, T., Wakasugi, T., and Sugiura, M.** (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* **53**, 327–332.
- Tzafir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., Patton, D., and Meinke, D.** (2003). The *Arabidopsis* SeedGenes project. *Nucleic Acids Res.* **31**, 90–93.
- Vermel, M., Guermann, B., Delage, L., Grienberger, J.M., Marechal-Drouard, L., and Gualberto, J.M.** (2002). A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc. Natl. Acad. Sci. USA* **99**, 5866–5871.
- Williams, P.M., and Barkan, A.** (2003). A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J.* **36**, 675–686.
- Yamazaki, H., Tasaka, M., and Shikanai, T.** (2004). PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*. *Plant J.* **38**, 152–163.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P.** (2002). Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15.
- Yang, Y.H., and Thorne, N.P.** (2003). Normalization for two-colour cDNA microarray data. In *Science and Statistics: A Festschrift for Terry Speed*, D.R. Goldstein, ed (Bethesda, MD: IMS), pp. 403–418.

## II.1.2 Constitution de la collection de mutants

Plusieurs voies étaient possibles pour parvenir à comprendre le rôle et les modes d'action de la famille PPR. Il existait un faisceau de présomptions sur le rôle de certaines sous-familles (DYW cf. discussion), mais cela ne pouvait correspondre à l'ensemble des protéines PPR. Nous avons choisi d'utiliser une approche par génétique inverse sur un grand nombre de gènes de PPR, et ainsi d'essayer d'avoir une vue d'ensemble sur les fonctions des protéines PPR chez *Arabidopsis*. Pour cela, nous avons profité de l'existence de plusieurs collections de mutants d'insertion (SALK, GABI, SAIL et FST) et du développement du projet Agrikola au sein de notre laboratoire (Hilson *et al.* 2004).

Etant donné la taille et la diversité de la famille des gènes de PPR chez *Arabidopsis*, il était difficile de fixer un choix sur telle ou telle protéine pour en effectuer une analyse détaillée. Les critères de choix des mutants étudiés ont été de deux types : i) nous avons choisi des membres de chaque sous-famille afin de couvrir la diversité de la famille PPR et ii) nous avons préféré des gènes pour lesquels il existait au moins deux lignées d'insertion indépendantes. Ainsi nous avons réduit les risques de biais dans la phase de choix des candidats et rendu dispensables certaines étapes de vérification des phénotypes. En effet, si deux lignées d'insertion indépendantes présentent le même phénotype, il est fort probable que celui-ci soit dû à la disruption du gène en question.

Ainsi, 126 lignées d'insertion différentes, couvrant 56 gènes PPR uniques ont été identifiées et collectées à partir des différentes collections de lignées d'insertion (SALK, GABI, FST ou SAIL). 38 lignées ont été criblées pour la présence de T-DNA. A la réception des graines en provenance des différents stocks, 8 graines ont été semées et génotypées à la recherche de plantes homozygotes pour l'insertion, ou hétérozygotes. Lorsque seules des plantes hétérozygotes étaient identifiées, 24 graines issues d'une de ces plantes étaient semées. Si aucune plante homozygote ne pouvait être obtenue de la descendance d'une plante hétérozygote, nous avons conclu que le gène ciblé était très probablement essentiel.

Parmi ces 38 lignées, 14 n'ont jamais pu être isolées à l'état homozygote, 24 l'ont été, représentant 22 gènes. Les autres lignées n'ont pas dépassé le stade de la culture de la première génération et de la recherche de la présence du T-DNA, trouvé dans 36 d'entre elles. En parallèle Anne Marmagne et Andéol Falcon de Longevialle ont dans le cadre du projet Agrikola recherché des lignées RNAi pour les mêmes gènes (Hilson *et al.* 2004).



Mutant	code AGI	classe	Localisation	Phénotypes	RNAi	crible
DNR2	At1g14470	PLS	?	Nain fertilité réduite	Inflorescence retardée	-
Salk_071712	At3g16890	P	?	Inflorescence retardée		
DNP1	At2g20540	E+	?			-
DVM10	At2g40720	E+	?			-
DDU146	At3g25970	E+	?	Nain fertilité réduite	Nain ou normal	-
Salk_127777	At3g25970	E+	?	Nain fertilité réduite	Nain ou normal	-
Salk_033409	At1g20230	DYW	?	Inflorescence retardée		
DYK12	At1g51965	PPR	Mitochondrie			-
DXZ45	At3g13150	PPR	Mitochondrie			-
DJU9	At1g06710	P	Mitochondrie			-
SAIL_867_F07	At1g74900	P	Mitochondrie	Culture In vitro seulement	réduit	
Salk_024306	At1g80550	P	Mitochondrie	Inflorescence sur-abondante	Grosse rosette	
T243	At1g22830	E	Mitochondrie			-
Salk_067026	At1g09410	DYW	Mitochondrie	Inflorescence retardée	multiple	
Gabi_104E05	At1g09410	DYW	Mitochondrie	pale, inflorescence retardée	multiple	
Salk_148973	At1g68930	DYW	Mitochondrie		Grosse rosette, inflorescence retardée	
Salk_119701	At3g02010	DYW	Mitochondrie	nain	Meurt ou reste très réduit	
Salk_015469	At3g49140	DYW	Mitochondrie			
Salk_033891	At5g46460	DYW	Mitochondrie			
DZY11	At5g65570	DYW	Mitochondrie			-
CRI6	At4g31850	P	Chloroplaste	Vert pâle	Meurt ou reste très réduit	+/-
Clb19-3	At1g05750	E+	Chloroplaste	albinos	Multiple	+
Clb19-1	At1g05750	E+	Chloroplaste	albinos	Multiple	+
Clb19-2	At1g05750	E+	Chloroplaste	albinos	Multiple	+
Pde247	At1g05750	E+	Chloroplaste	albinos	Multiple	+
Salk_040629	At3g22150	E+	Chloroplaste		Meurt ou reste très réduit	
Salk_012455	At1g59720	DYW	Chloroplaste		Croissance retardée	
Salk_046131	At3g46790	DYW	Chloroplaste	Défaut de maturation de <i>ndhB</i>	très réduit	+/-
Gabi_425F06	At3g46790	DYW	Chloroplaste	Défaut de maturation de <i>ndhB</i>	très réduit	+/-
EFS132	At3g63370	DYW	Chloroplaste			-
Salk_102445	At3g63370	DYW	Chloroplaste			

Tableau 6 : Mutants homozygotes

Liste des lignées d'insertion pour lesquelles nous avons isolé des plantes homozygotes, les localisations indiquées sont le produit de la prédiction par les logiciels Predotar et TargetP, un "?" indique soit un désaccord entre les deux méthodes de prédiction soit une localisation autre que organite. La dernière colonne indique les mutants dont l'analyse des transcriptomes chloroplastiques et mitochondriaux a été entreprise, "-" pas de défauts détectés, "+" défauts identifiés, "vide" analyse à effectuer, +/- défaut publiés difficile à identifier.

### II.1.3 Analyse des transcriptomes chloroplastiques et mitochondriaux

En parallèle à la recherche de mutants affectés dans la production de protéines PPR, il était nécessaire de développer une méthode d'identification des cibles des protéines PPR. En effet, la compréhension du fonctionnement des protéines PPR passe obligatoirement par l'identification de leurs cibles moléculaires.

Partant de l'idée que les protéines PPR avaient un rôle dans les événements affectant les ARN après la transcription, il était essentiel de choisir une méthode d'analyse pouvant nous renseigner sur l'état de chaque transcrit dans chacun des organites. Les transcrits mitochondriaux et chloroplastiques sont affectés par de nombreux processus de maturation, clivage, épissage, édition, maturation en 5' ou 3'. Au début de cette thèse (et c'est toujours le cas!), il était impossible de prédire la fonction ou la cible d'une protéine PPR directement à partir de sa séquence. Seule la localisation sub-cellulaire pouvait être soupçonnée grâce aux logiciels de prédiction. La méthode de crible devait donc être exhaustive, tout du moins s'efforcer de l'être.

Une méthodologie basée sur l'hybridation de sondes spécifiques, comme les puces à ADN ou le *northern blot*, a été envisagée. Cependant les oppositions étaient de plusieurs ordres, aussi bien d'un point de vue faisabilité au sein du laboratoire que du point de vue biologique.

Premièrement plusieurs processus auraient été difficile, voir impossible, à détecter par une méthodologie reposant sur l'hybridation comme les puces à ADN. Un défaut d'épissage nécessiterait un grand nombre de sondes différentes pour être identifié. D'autre part, rechercher les défauts de maturations en 5' et 3' par hybridation de sondes impliquerait la connaissance de chaque fin de transcrit.

Le *northern blot* demeurait la méthode de prédilection pour l'analyse des ARN, mais cette méthode était difficilement adaptable à un crible haut-débit. En effet, cette méthode nécessite plusieurs microgrammes d'ARN pour chaque mutant à analyser, ce qui aurait été limitant dans le cas de plantes très affectées dans leur croissance.

Les puces ADN sont des ressources coûteuses, difficilement utilisables comme outil de crible, particulièrement sans *a priori*. De plus, les technologies de marquage utilisées pour les puces à ADN sont généralement des marquages chimiques après amplification des ARN. Cette méthode repose sur la poly-adénylation des ARN. Dans le cas des organites, l'hétérogénéité des populations d'ARN poly-adénylés aurait rendu les résultats obtenus difficiles à interpréter. Une méthode alternative de marquage des ADNc par ajout de nucléotides biotynilés existait bien, mais demeurait assez coûteuse. Enfin, les puces

CATMA, que nous aurions pu utiliser au laboratoire, ne comprenaient pas alors de sondes pour les organites. Il aurait donc fallu développer de nouvelles puces.

Nous avons choisi de baser notre crible sur l'utilisation d'amorces spécifiques en RT-PCR. Cela a nécessité l'établissement d'une collection de primers couvrant toutes les régions potentiellement transcrites dans les deux organites. L'idée était de pouvoir obtenir, en une expérience de RT-PCR (deux plaques PCR de 384 puits), une information sur l'état de chaque transcrit présent dans les organites.

La conception de la collection est décrite dans l'article suivant:

**II.1.4 Article 2 : Andrés C., Falcon de Longevialle A., Lurin C. and Small ID.  
Rapid low-resolution transcript mapping for identifying mutants affecting  
organelle gene expression in *Arabidopsis thaliana***

# Rapid low-resolution transcript mapping for identifying mutants affecting organelle gene expression in *Arabidopsis thaliana*

Charles Andrés<sup>1</sup>, Andéol Falcon de Longevialle<sup>1,2</sup>, Claire Lurin<sup>1</sup> and Ian D. Small<sup>1,2§</sup>

<sup>1</sup>UMR Génomique Végétale, 2 rue Gaston Cremieux CP5708, 91057 Evry, France

<sup>2</sup>ARC Centre of Excellence for Plant Energy Biology, 4th Floor MCS Building, M316, University of Western Australia, 35 Stirling Highway, Crawley 6009, Western Australia

§Corresponding author

Email addresses:

CA: [andres@evry.inra.fr](mailto:andres@evry.inra.fr)

AF: [falcon@cyllene.uwa.edu.au](mailto:falcon@cyllene.uwa.edu.au)

CL: [lurin@evry.inra.fr](mailto:lurin@evry.inra.fr)

IDS: [iansmall@cyllene.uwa.edu.au](mailto:iansmall@cyllene.uwa.edu.au)

# Abstract

## Background

Plant organellar genomes in mitochondria and chloroplasts contain more than 150 genes and require hundreds of nuclear genes for correct expression. Various post-transcriptional processes within organelles, including RNA processing and splicing, are thought to need specific factors that remain for the most part uncharacterised. Researchers in this field have lacked easy methods for systematically screening mutant plants for defects in organelle gene expression; RNA gel blots are time-consuming and labour-intensive, whilst array-based approaches are inadequate for analysing transcriptomes dominated by post-transcriptional processing rather than simply transcript accumulation. To solve these problems, we set out to develop a method to quickly screen all organellar transcripts for processing defects.

## Results

The method is based on a collection of 731 primer pairs covering the gene-containing regions of both organellar genomes. Amplification of organellar cDNA using these primer pairs indicates the presence of significant amounts of contiguous RNA between the binding sites for the two primers, whereas a lack of amplification indicates little or no such contiguous RNA exists. This simple approach can detect semi-quantitative alterations in RNA processing, splicing or stability and is adapted to high-throughput screens. After validating the approach on a known chloroplast RNA processing mutant, we screened previously uncharacterised mutants to discover a new factor involved in splicing of a specific mitochondrial intron, as well as a factor involved in correct expression of *rbcL* transcripts. A very similar approach can be used to study DNA rearrangements.

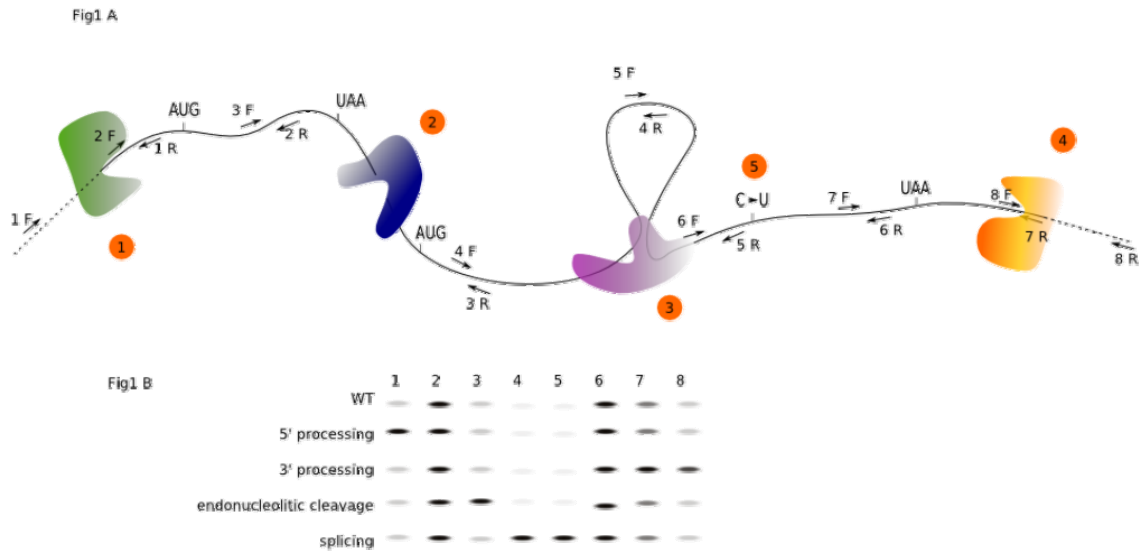
## Conclusions

We have established a technique to rapidly generate a low-resolution map of organellar transcriptomes. The usefulness of the approach has been demonstrated by identifying the molecular defects in previously uncharacterised *Arabidopsis* mutants with insertions in three different genes of the pentatricopeptide repeat family. The extension of this type of screen to other mutants putatively affected in organelle gene expression should allow the rapid identification of the RNA targets of the hundreds of factors of this type that remain to be characterised.

## Background

The screening of *Arabidopsis* mutant collections is a powerful tool for discovering the role of unknown function genes, of which there remain thousands [1]. The abundance of mutants and the availability of high-throughput RNAi technology [2] make it theoretically possible to access the function of all *Arabidopsis* genes. Many tools for high-throughput screening of these collections are available, such as automated image analysis of morphological characteristics or metabolite analysis, as well as many specialized screens using resistance to chemicals or markers for gene activation. However, not all possible molecular or biochemical functions are amenable to the currently available screens. Examples include genes involved in organellar gene expression.

The plant chloroplast and mitochondrial genomes are the remnants of endosymbiotic prokaryotes. In the billions of years since their endosymbiotic capture, the two genomes have been comprehensively altered, with the majority of the genes being transferred to the nucleus,



### Figure 1 - Detection of post-transcriptional processing in organelles

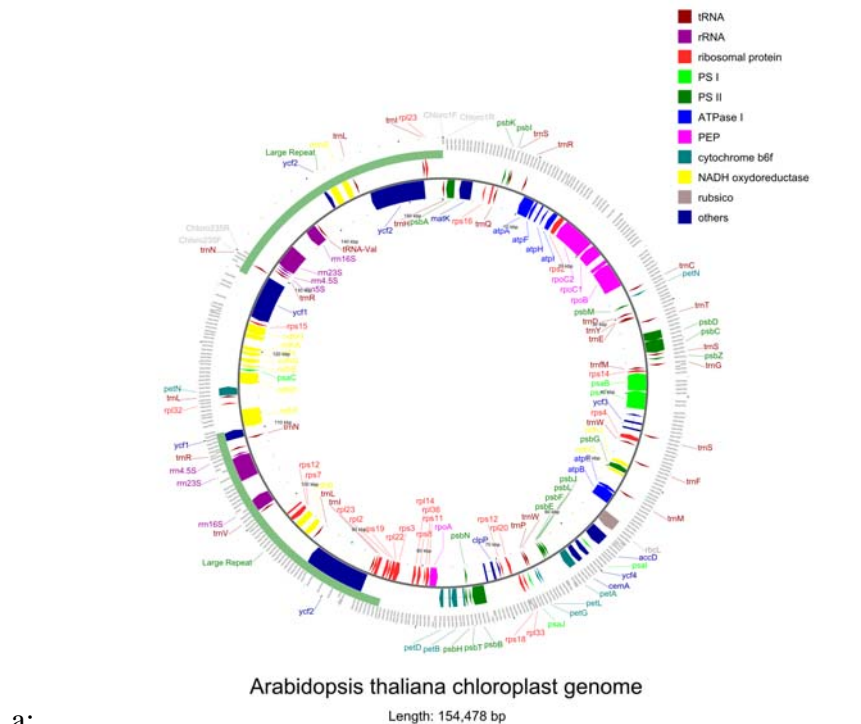
**A** In plant organelles, mRNAs are generally synthesised as long, often polycistronic molecules that are subsequently processed to give mature mRNA. The most common events for mRNA processing are: 5' end processing (1), endonucleolytic cleavage (2), splicing (3), 3' end processing (4) and in some case editing (5) as shown here on a hypothetical but typical transcription unit. The arrows indicate primers placed to reveal eventual processing defects (F forward, R reverse). **B** Hypothetical RT-PCR results obtained from wild-type and mutant plants showing characteristic patterns of amplification for defects in each potential processing step. The numbering of the lanes corresponds to the primer pairs indicated in **A**.

creating the two current genomes, containing 87 [3] and up to 57 [4] protein-coding genes in Arabidopsis plastids and mitochondria respectively. Much of the transcription in organelles is carried out by nucleus-encoded phage-type RNA polymerases [5], and only chloroplasts have conserved a eubacterial-type DNA-dependent RNA polymerase [6]. In both organelles steady-state RNA levels are high and relatively constant, with chloroplasts exhibiting some level of specific transcriptional regulation [7], particularly in response to light and developmental stage, leading to differentiation of plastids in different tissues. But for the most part, the control of organellar gene expression is made at the post-transcriptional level, via regulation of 5' and 3' end processing, editing, *cis*- and *trans*-splicing, mRNA stability and initiation of translation [8, 9]. All mRNAs in both organelles are processed via multiple successive or parallel events, many of which are transcript-specific, and this therefore implies large numbers of independent factors. These factors are almost invariably encoded by the nucleus as the organelle genomes code for hardly any of them. From bioinformatics predictions, the total mitochondrial plus plastid proteome is estimated at around 4000 proteins, and many of these can be predicted to be implicated in organellar gene expression [10]. To give an idea of the scale of the processes involved, just one protein family, the pentatricopeptide repeat (PPR) family contains 450 members in Arabidopsis [11] the vast majority of which are thought to be organellar post-transcriptional control factors.

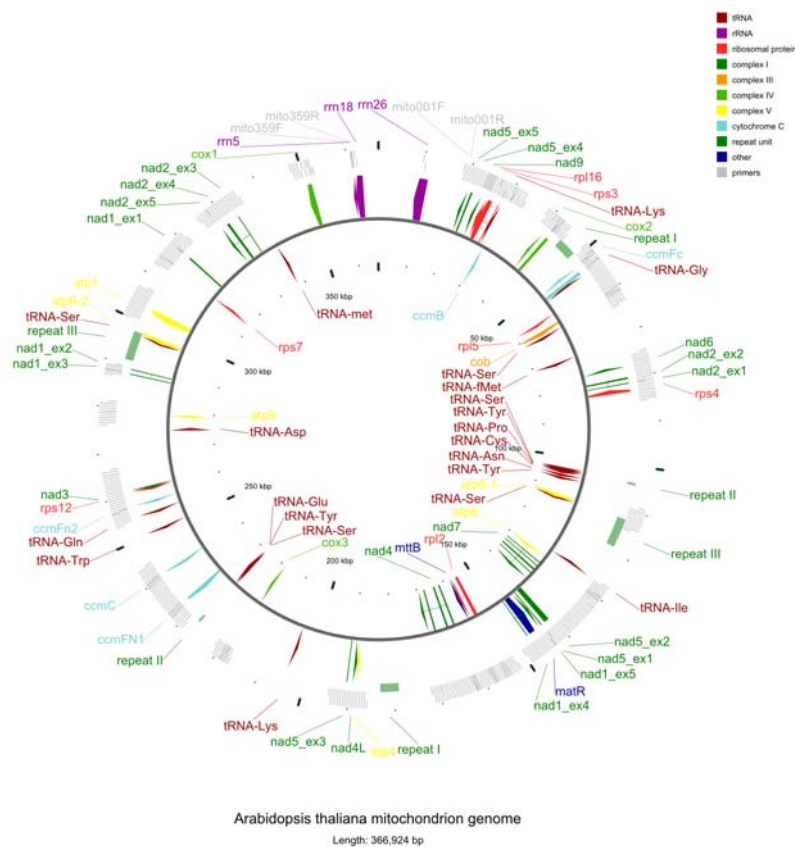
The PPR proteins are specific to eukaryotes and have undergone a huge expansion in land plants, from about 10-20 genes in protozoa and algae, to 450 or more in Arabidopsis and rice [11]. With one exception [12], all known plant PPR proteins are implicated in post-transcriptional processes in mitochondria ([13] and for review [14]) or chloroplasts([15-24]). Several Arabidopsis PPR mutants have been described, such as *hcf152*, with defects in splicing and the processing of *petB* in chloroplasts [24]; *crr2*, defective in processing of chloroplast *ndhB* transcripts [25]; *crr4*, defective in editing of the start codon of *ndhD* in chloroplasts [18, 20]; *pgr3*, defective in stabilization and translation of the *petL* operon in chloroplasts [21]. Without exception, all of these mutants have been discovered by forward genetic screens looking for specific physiological defects in organelle activity, not defects in organelle gene expression and represent only the tip of the iceberg of potential mutants in these processes. To directly screen for mutants affected in organelle gene expression, we have developed a method to rapidly look for post-transcriptional defects in both organelles. RNA gel blots are an excellent method for revealing the pattern of transcripts present for a particular gene, but are not adaptable to high-throughput screening of over one hundred genes in hundreds of mutants. Array-based methods are more easily adapted to high-throughput screens, but do not give information on RNA length or integrity. As the majority of the post-transcriptional processes affect mRNA sequence integrity, without necessarily affecting RNA stability, we chose to use a PCR-based method, with an array of primer pairs arranged across the transcribed regions of both organelles.

## Results

The principle of the screen is quite simple; reverse-transcription of RNA into cDNA and amplification of the cDNA using specific primers will lead to obvious accumulation of amplification products whenever there is contiguous cDNA that both primers can bind to. As shown in Figure 1, judicious placement of such primer pairs can lead to the analysis of all of the possible post-transcriptional processes for each transcription unit. The Arabidopsis cytoplasmic genomes are respectively 154kb and 367kb for the chloroplast and the mitochondria [3, 4]. For the chloroplast we can consider that virtually all of the genome is potentially transcribed into physiologically relevant transcription units, but for the mitochondria, much of the genome lacks any conserved coding sequence. We therefore chose to totally cover the chloroplast genome (except that we did not duplicate primer pairs on both copies of the large inverted repeat, figure 2)



a:



b:

## Figure 2 - Maps of the organelle genomes showing the regions covered by the PCR primers

**A** Circular map of the chloroplast genome according to [3], with the following genes renamed, following more recent literature; *ycf5* to *ccsA*, *ycf6* to *petN*, *ycf7* to *petL*, *ycf9* to *psbZ* [35-38]. The primers used in this work are indicated as short grey radial lines around the outside of the genome (the forward primers are the more external). A full table of primer sequences and genome positions is given as Table S1 in the supplementary data.

**B** Circular map of the mitochondrial genome according to Unsel et al. The primers used in this work are indicated as in **A**. A full table of primer sequences and genome positions is given as Table S2 in the supplementary data.



and the potentially coding regions of the mitochondrial genome (i.e. 60% of the genome, supplemental data). We produced two “tiling arrays” of primer pairs of less than 384 pairs each, to be able to perform a complete analysis in one PCR run (two 384-well PCR plates). We defined potentially transcribed regions as extending 2kb upstream and downstream of each open reading frame (ORF). Within these regions, primer pairs were designed to cover potential cleavage, splicing and editing sites. The size range for the amplicons is 350 to 700bp, with a preferred size of 400bp in plastids and 500bp in mitochondria. The melting temperature of each primer/cDNA hybrid was designed to be between 58 and 62°C, with no more than 2 degrees difference between each primer in a pair. This allows the use of the primers in standard PCR, real time PCR, or for generating amplicons for spotting on microarrays.

### **Analysis of wild type plants**

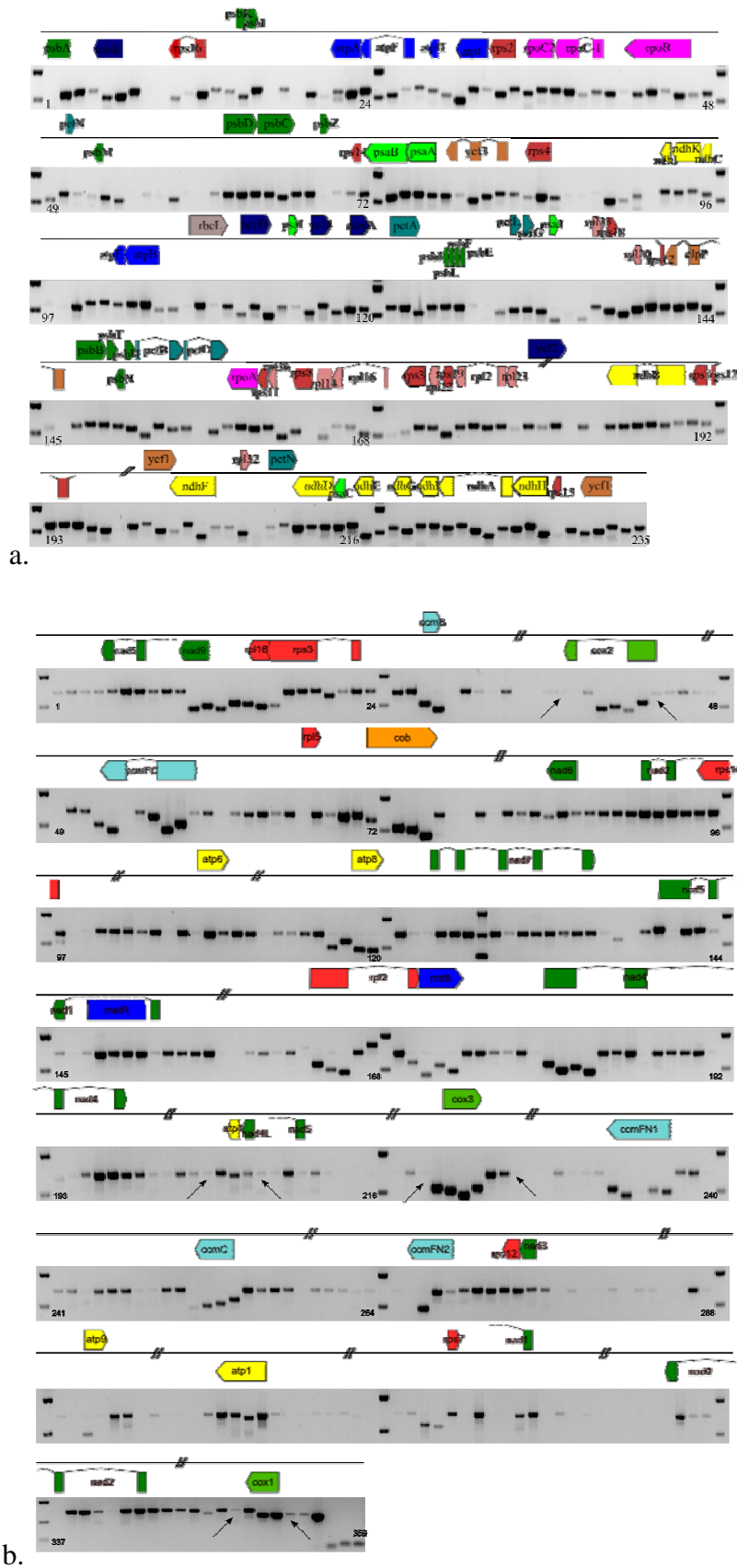
The sensitivity and the reproducibility of the method were verified by the analysis of wild type plants from Col-0, Ws and Ler accessions. A low-resolution map of organelle transcripts was established (Figure 3), and was found to be very stable between accessions and under different growth conditions (not shown). The mitochondrial RT-PCR map reveals a pattern of mostly discrete transcripts separated by regions where no contiguous transcripts are detected. Mapping the known end points of Arabidopsis transcripts [26, 27] to the RT-PCR map shows that they correspond, as would be expected, to the first unamplified amplicon extending away from the coding sequence (Figure 3A). The chloroplast RT-PCR map abundantly demonstrates the polycistronic nature of many chloroplast transcripts. Where transcript ends are known, in the case of long polycistronic mRNA precursors we can often observe a fainter band for the amplicon covering the cleavage site. Some new putative transcription units can be detected such as the predicted *atpI-atpH-atpF-atpA* operon (Figure 3 and Table 1).

### **Screening of transcriptional defects in Arabidopsis mutants**

To verify the use of this approach as a screen for mutants in organellar RNA processing, we tested previously uncharacterised mutant alleles of the *CRR2* gene described by Hashimoto et al [25]. *CRR2* is a PPR protein from the DYW subfamily, shown to be essential for processing of *ndhB* transcripts in Arabidopsis chloroplasts. The mature, translatable *ndhB* mRNA is obtained from a polycistronic precursor containing exons 2 and 3 of *rps12*, the single *rps7* exon and the two exons of *ndhB*. Several processing events take place in this transcript, such as *trans*-splicing between exons 1 and 2 of *rps12* prior to *cis*-splicing between exons 2 and 3. Hashimoto et al observed that two endonucleotidic cleavages occur between the *rps7* coding sequence and the first exon of *ndhB*, respectively 180nt and 12nt before the *ndhB* start codon. They have shown that only the -12 cleavage point is under control of *CRR2*, as a longer mRNA corresponding to the -180 cleavage point accumulates in the mutant. The *crr2* mutant examined by Hashimoto et al was an EMS-induced point mutation. We screened two T-DNA insertion mutants in *CRR2* from the Salk and the GABI-KAT collections (Salk\_046131 and GABI\_425F06) using our plastid primer collection. As shown in Figure 4, the defect in *ndhB* processing is detectable in both insertion mutants, confirming that this defect is due to a lack of *CRR2*, but more importantly, we could show that no other transcript seems to be affected in this mutant.

Having confirmed that the approach was capable of detecting expected processing defects, we attempted to uncover alterations in RNA processing in other PPR mutants for which we had no prior knowledge of the functions affected. In all, 16 mutants were screened and 3 mutants gave altered transcriptome maps that prompted further study (the detailed study of each of these mutants will be published elsewhere - Ramos et al., in preparation; Falcon de Longevialle et al., in preparation).

Two examples are of particular interest. CLB19 (encoded by At1g05750) is a PPR protein from the E+ subfamily consisting of 10 PPR motifs and two additional C-terminal domains termed E and E+[11]. CLB19 is localized in the chloroplast and T-DNA knock-out mutants exhibit severe defects in chloroplast biogenesis (Ramos et al., in preparation). The lack of CLB19 is lethal at the early seedling stage in soil, but plants can be rescued *in vitro* on sucrose and under low light conditions grow similarly to wild type plants under the same conditions. The RT-PCR transcriptome map of *clb19* mutants revealed that several transcripts were affected, in particular the region containing the gene for *rbcL*



**Figure 3 - Typical RT-PCR products from wild-type plants**

RT-PCR products obtained with the entire collection of primers were visualized on ethidium bromide gels. As the products are arranged in the order that they are found in the genome, the sequences amplified correspond approximately to the transcripts represented schematically above the gel pictures. **A**, Amplification with plastid primers; **B**, amplification with mitochondrial primers. In **B**, arrows indicate known transcript ends from mapping experiments (Schuster..).

(Rubisco large subunit), *atpE* and *atpB* (ATP synthase subunits E and B) and *accD* (Acetyl-coA carboxylase beta subunit). In *clb19* mutants we observe the accumulation of mRNA containing the intergenic region between *atpB* and *rbcL* (Figure 5). This accumulation is due to a defect in processing the 5' end of *rbcL* transcripts, confirmed by RNA gel blots and RACE analysis (Ramos et al., in preparation).

The second example is the gene At1g74900, encoding a PPR protein containing 11 PPR motifs predicted to be addressed to mitochondria. Mutants with decreased expression of this gene were first identified in a large-scale RNAi screen [28] as exhibiting delayed development and flowering. We analysed the RT-PCR map of the mitochondrial transcriptome of a T-DNA insertion mutant (Sail867\_F07) from the Sail T-DNA collection [29]. The phenotype of this putative null mutant is extremely severe – most homozygous mutant seeds fail to germinate and the few homozygous plants recovered grow extremely slowly (Falcon de Longevialle et al., in preparation). Only one reproducible defect could be observed (Figure 6); the mutant plants lack all trace of the fully spliced mitochondrial *nad1* transcript. Further experiments have shown that the *trans*-splicing between the exons 1 and 2 is abolished in the mutant, and that the defect can be restored by transient expression of the PPR protein encoded by At1g74900 (Falcon de Longevialle et al., in preparation). *nad1* encodes the primary subunit of the mitochondrial Complex I, the largest of the respiratory electron transport complexes.

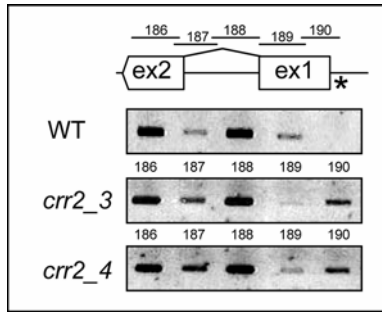
## Discussion

The two energy organelles power most of primary metabolism in plants and the proteins they encode play critical roles in photosynthesis and respiration. Understanding how the synthesis of the organellar proteins is regulated is a topic of major importance, and yet the total understanding of post-transcriptional control of gene expression in plant organelles is a long way off, given the huge number of predicted factors and the very small number of them with identified functions. Here we describe an approach to quickly, easily and cheaply screen *Arabidopsis* mutants for defects in organellar transcriptomes. The screening method is based on a collection of primers covering the two organelle genomes, allowing the analysis of numerous post-transcriptional events in both mitochondria and chloroplasts. Both processing and splicing defects can be detected as illustrated in Figs. 2-4 and the addition of a sequencing step could identify editing mutants as well. The screening of mutants requires no *a priori* knowledge of the process in which the mutant gene is implicated and requires very little plant material.

From a technical point of view, this method only requires cheap, fast and simple RT-PCR reactions. The abundance of organellar transcripts and the sensitivity of RT-PCR means that very little cDNA is needed (8  $\mu$ l of a standard 20 $\mu$ l Superscript II reaction with 3 $\mu$ g of total RNA). This approach can provide results that are inaccessible by other methods. In comparison, RNA gel blots would require a hundred-fold more plant material to screen as many transcripts and orders of magnitude more effort. Array-based techniques would permit a similar throughput, but at much greater cost, and simple arrays cannot easily detect the length or structure of RNAs; transcripts for many organellar genes exist as multiple precursor and mature forms that cannot be distinguished by hybridisation to a standard array. A high-density tiling array of the type recently constructed for *Arabidopsis* [30] would probably be a useful tool, but at a prohibitive cost for large-scale screening.

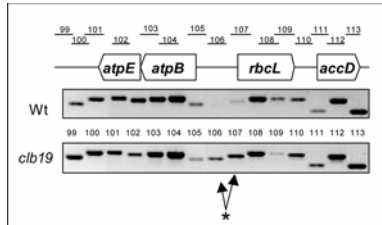
Not all post-transcriptional processes can be detected by this approach. Editing differences cannot be observed in the way described here without further analysis of the amplification products, neither can polyadenylation or RNA methylation. In addition, processing events involving tRNA precursors appear to be completed too fast for the precursors to be detectable under the conditions used here. In addition, the approach is no better than semi-quantitative as many amplification products are nearing saturation under conditions optimal to detect most transcripts. Hence small changes in steady state RNA levels will not be detected.

For these reasons, the approach is not intended to provide a final, definitive result on the structure of the transcripts being analysed, and once putative differences are observed, other more targeted approaches



**Figure 4 – Analysis of *crr2* insertion mutants**

Excerpts of chloroplast transcriptome screens of wild-type *Arabidopsis* Col0 (WT) and two independent T-DNA insertion mutants in the gene At3g46790 (*crr2-3*, *crr2-4*) are shown. The region shown corresponds to the gene *ndhB*, known not to be correctly processed in two EMS mutants of *CRR2* (ref). In the two *crr2* mutants, amplification product 190 accumulates, whereas it does not in wild-type as the template RNA is efficiently cleaved within the region bracketed by the primers (at the point marked with an asterisk).



**Figure 5 - Analysis of a *clb19* insertion mutant**

Excerpts of chloroplast transcriptome screens of wild-type *Arabidopsis* Col0 (Wt) and a T-DNA insertion mutant in the gene Atxxx (*clb19-1*) are shown. RT-PCR amplification was done from equal amounts of RNA from 15 day-old plants grown under low light conditions ( $5 \mu\text{mol m}^2 \text{sec}^{-1}$ ). The amplified products of the transcripts covering the *atpE*, *atpB*, *rbcL* and *accD* genes are shown. Differences in the amplified products between the Wt and mutant plants are marked by arrows and an asterisk.

such as northern blotting, RNase protection or RACE need to be employed. However, as an initial screen it is unsurpassed. Repeating the analysis on an independent sample is usually sufficient to eliminate most ‘false positive’ findings due to loading or dilution errors.

### **Alternative uses for the primer collection**

A collection of primers of this kind can be used for many other purposes than the transcriptome mapping described above. Recently we used it to check the integrity of the mitochondrial genome in the OSB1 mutant [31]. OSB1 participates in controlling the stoichiometry of alternative mtDNA forms generated by recombination. This regulation may take place in gametophytic tissues in order to insure transmission of a functional mitochondrial genome into the female gamete. We verified the integrity of the mitochondrial genome in the progeny of an *osb1* mutant, and were able to find several altered regions in the mitochondrial genome that could be confirmed by southern blot.

The amplicons described here have also been used to spot probes on microarray slides to produce a low-resolution tiling array of the organelle genomes. Such arrays are useful for quantitative analysis of transcript levels and for RIP-chip experiments to identify targets of RNA-binding proteins [17, 19].

## **Conclusions**

We have developed and tested a novel approach for screening *Arabidopsis thaliana* mutants potentially affected in organelle gene expression. The approach is well-adapted to high-throughput screening and effective at uncovering new mutants, as demonstrated by the identification of a mutant affected in processing of plastid *rbcL* and a mutant affected in splicing of mitochondrial *nad1*.

## **Methods**

### **Primer design**

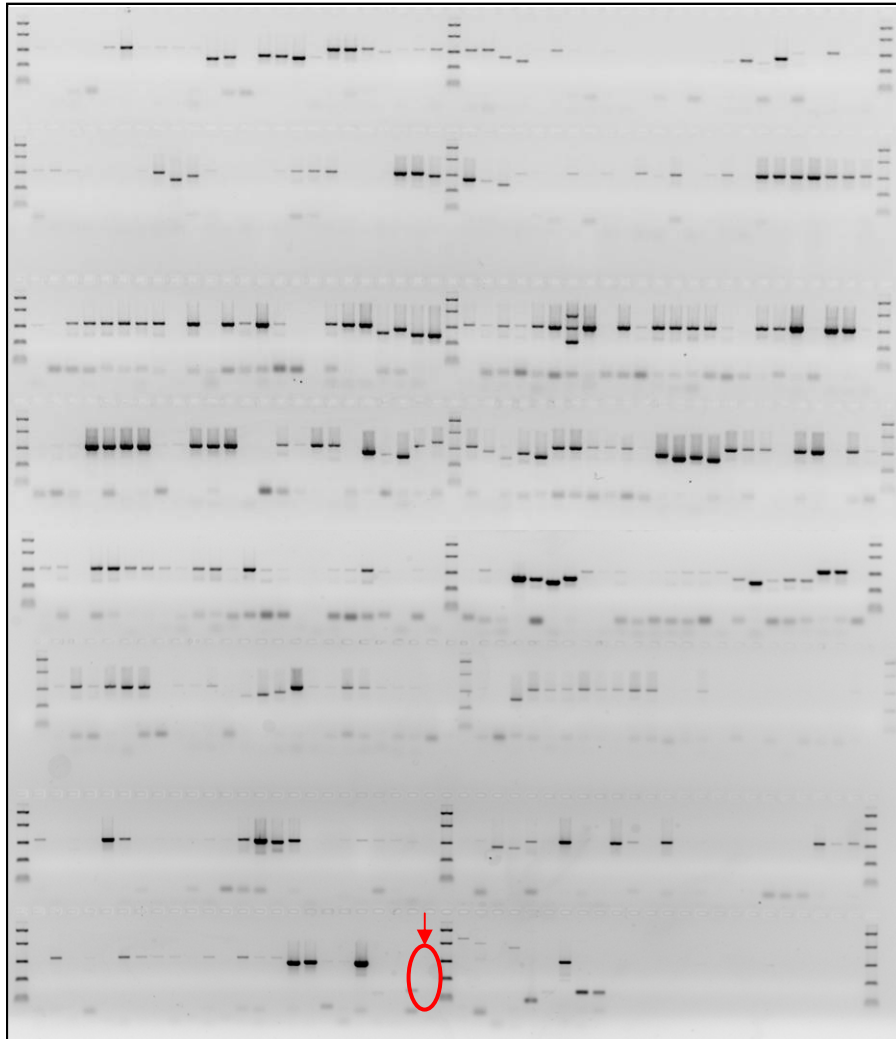
The primers have been designed using the Primer3 software [32] combined with a script to automate the design. Different lists of primer pairs were produced and the final choice for the most effective collection was made by hand using Artemis software [33] to visualize the primers on the organellar genomes.

The script to design the primers functioned by first defining a 700bp nucleotide sequence in the target genome (overlapping the previous region by 50bp), then generating the primers with the software Primer3 [32] using the following parameters: primer size between 15-31nt; amplicon size 350-550bp; primer Tm minimum 58°C, maximum 62°C, delta max 2°C; maximum Poly-X 3; GC\_clamp 0; Target [250,1]; Excluded region [50,360].

A master list of primers for the two organelles was generated, containing 1124 pairs for the chloroplast (729 amplicons of 500bp and 395 amplicons of 350bp) and 2187 for the mitochondria (752 amplicons of 550bp, 887 of 500bp and 548 of 350bp). For the final selection of primer pairs, we used the genome browser Artemis [33], to display the primers directly on the chloroplast and mitochondrial genome annotation. The pairs have been chosen with the following objectives: at least one amplicon in each ORF and amplicons covering at least 2kb of potential 5’ and 3’ UTR. Additional amplicons were designed to explicitly amplify cDNA sequences across spliced exons. The final collection contains 355 and 376 primer pairs for chloroplasts and mitochondria respectively. Primer sequences and positions are given in Table S1 in the supplementary data.

### **RNA preparation**

RNA extraction was performed with RNA NOW<sup>TM</sup> (Biogenex, Seabrook, TX), derived from the single step RNA isolation method by [34]. The suppliers instructions were followed with the following additional step: after the first centrifugation supernatant was transferred into a new tube before addition of chloroform. Total RNA was treated with DNaseI (AM2222, Ambion®) at 37°C for an hour to



**Figure 6- PPR43**

Low resolution map of the mitochondrial transcript for PPR43 mutant, no reproducible difference with wild type can be observe, except specific amplification of the mature *nad1* transcript.

eliminate DNA contamination. First strand synthesis was performed using SuperscriptIII Reverse Transcriptase (Invitrogen™), using the supplier's protocol and 3µg of RNA, with 100ng of random hexamer.

### PCR analysis

An initial PCR experiment was performed with 4 primer couples targeting two genes in each genome, respectively *atpA* and *rpoB* for the chloroplast, *cox2* and *cox1* for the mitochondria. This step was to check for the absence of DNA contamination and determine the appropriate dilution for the full analysis. PCR was performed in a 20µl reaction volume (0.1 µM primers, 0.2 mg/ml cresol red, 60 mg/ml sucrose), with 1µl of 10X serial dilution of cDNA and negative control (without Superscript II), and 0.4 units Taq DNA polymerase (Biolabs), using the following program: 5 min at 94°C, then 35 cycles of the sequence 20s at 94 ° C, 20s at 52°C and 60s at 72°C, and a final step of 5 min at 72°C. PCR products were visualized by ethidium bromide staining after migration in 2% agarose gels.

The full PCR analyses were performed in 384-well plates in a final volume of 10µl (0.1 µM primers, 0.2 mg/ml cresol red, 60 mg/ml sucrose) with 0.2 units Taq DNA polymerase (Biolabs) with 1µl of the appropriate cDNA dilution determined in the setup PCR, using same program. PCR products were visualized by ethidium bromide staining after migration in 2% agarose gels. Gel images were analyzed with Genetools® software from Syngene.

### Authors' contributions

CA carried out the primer design, the development of the screening protocol, the analysis of the *crr2* and *clb19* mutants, planned the manuscript and prepared the figures. AF carried out the analysis of the *nad1* splicing mutant. CL conceived and participated in the design of the study. IS participated in the design and coordination of the study and wrote the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

This work was supported by the Fondation Recherche Medical and the French Ministry of Research for Charles Andrés' PhD fellowship. AF, IS and CL were supported by EU Framework 5 grant AGRIKOLA (QLRT-2001-01741) and AF and IS by an Australian Research Council Centre of Excellence grant (CE0561495).

### References

1. Alonso JM, Ecker JR: **Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis.** 2006, **7**:524-536.
2. Hilson P, Small I, Kuiper MT: **European consortia building integrated resources for Arabidopsis functional genomics.** *Curr Opin Plant Biol* 2003, **6**:426-429.
3. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S: **Complete structure of the chloroplast genome of Arabidopsis thaliana.** *DNA Res* 1999, **6**:283-290.
4. Unseld M, Marienfeld JR, Brandt P, Brennicke A: **The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides.** *Nat Genet* 1997, **15**:57-61.
5. Hedtke B, Borner T, Weihe A: **Mitochondrial and Chloroplast Phage-Type RNA Polymerases in Arabidopsis** **10.1126/science.277.5327.809.** *Science* 1997, **277**:809-811.
6. Maliga P: **Two plastid RNA polymerases of higher plants: an evolving story.** *Trends in Plant Science* 1998, **3**:4-6.
7. Allison LA: **The role of sigma factors in plastid transcription.** *Biochimie* 2000, **82**:537-548.

8. Holec S, Lange H, Kuhn K, Alioua M, Borner T, Gagliardi D: **Relaxed Transcription in Arabidopsis Mitochondria Is Counterbalanced by RNA Stability Control Mediated by Polyadenylation and Polynucleotide Phosphorylase.** *Mol Cell Biol* 2006, **26**:2869-2876.
9. Bollenbach TJ, Schuster G, Stern DB: **Cooperation of Endo- and Exoribonucleases in Chloroplast mRNA Turnover**  
**Progress in Nucleic Acid Research and Molecular Biology.** In. Volume 78 edition. Edited by Moldave K: Academic Press; 2004: 305-337
10. Millar AH, Whelan J, Small I: **Recent surprises in protein targeting to mitochondria and plastids.** *Current Opinion in Plant Biology*  
*Cell Biology / Edited by Laurie G Smith and Ulrike Mayer* 2006, **9**:610-615.
11. Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast 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**:2089-2103.
12. 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**:815-830.
13. Gutierrez-Marcos JF, Dal Pra M, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Tatout C, Paul W, Perez P, et al: **empty pericarp4 Encodes a Mitochondrion-Targeted Pentatricopeptide Repeat Protein Necessary for Seed Development and Plant Growth in Maize.** *Plant Cell* 2007.
14. Andrés C, Lurin C, Small ID: **The multifarious roles of PPR proteins in plant mitochondrial gene expression**  
**doi:10.1111/j.1399-3054.2006.00766.x.** *Physiologia Plantarum* 2007, **129**:14-22.
15. Kobayashi K, Suzuki M, Tang J, Nagata N, Ohyama K, Seki H, Kiuchi R, Kaneko Y, Nakazawa M, Matsui M, et al: **LOVASTATIN INSENSITIVE 1, a Novel Pentatricopeptide Repeat Protein, is a Potential Regulatory Factor of Isoprenoid Biosynthesis in Arabidopsis.** *Plant Cell Physiol* 2007, **48**:322-331.
16. Hattori M, Miyake H, Sugita M: **A pentatricopeptide repeat protein is required for RNA processing of clpP pre-mRNA in moss chloroplasts.** *J Biol Chem* 2007.
17. 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 rps12 pre-mRNA.** *Plant Cell* 2006, **18**:2650-2663.
18. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T: **A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing.** *J Biol Chem* 2006, **281**:37661-37667.
19. Schmitz-Linneweber C, Williams-Carrier 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**:2791-2804.
20. Kotera E, Tasaka M, Shikanai T: **A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts.** *Nature* 2005, **433**:326-330.
21. Yamazaki H, Tasaka M, Shikanai T: **PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in Arabidopsis.** *Plant J* 2004, **38**:152-163.
22. Williams PM, Barkan A: **A chloroplast-localized PPR protein required for plastid ribosome accumulation.** *Plant J* 2003, **36**:675-686.
23. Nakamura T, Meierhoff K, Westhoff P, Schuster G: **RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA.** *Eur J Biochem* 2003, **270**:4070-4081.
24. 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-petB-petD RNAs.** *Plant Cell* 2003, **15**:1480-1495.



25. 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**:541-549.
26. Giese A, Thalheim C, Brennicke A, Binder S: **Correlation of nonanucleotide motifs with transcript initiation of 18S rRNA genes in mitochondria of pea, potato and Arabidopsis.** *Mol Gen Genet* 1996, **252**:429-436.
27. Brennicke A, Zabaleta E, Dombrowski S, Hoffmann M, Binder S: **Transcription signals of mitochondrial and nuclear genes for mitochondrial proteins in dicot plants** **10.1093/jhered/90.3.345.** *J Hered* 1999, **90**:345-350.
28. Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, et al: **Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications.** *Genome Res* 2004, **14**:2176-2189.
29. Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, et al: **A high-throughput Arabidopsis reverse genetics system.** *Plant Cell* 2002, **14**:2985-2994.
30. Stolc V, Li L, Wang X, Li X, Su N, Tongprasit W, Han B, Xue Y, Li J, Snyder M, et al: **A Pilot Study of Transcription Unit Analysis in Rice Using Oligonucleotide Tiling-path Microarray.** *Plant Molecular Biology* 2005, **59**:137-149.
31. Zaegel V, Guermann B, Le Ret M, Andres C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P: **The Plant-Specific ssDNA Binding Protein OSB1 Is Involved in the Stoichiometric Transmission of Mitochondrial DNA in Arabidopsis.** *Plant Cell* 2006, **18**:3548-3563.
32. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365-386.
33. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**:944-945.
34. Chomczynski P, Sacchi N: **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.** *Anal Biochem* 1987, **162**:156-159.
35. Xie Z, Merchant S: **The Plastid-encoded ccsA Gene Is Required for Heme Attachment to Chloroplast c-type Cytochromes** **10.1074/jbc.271.9.4632.** *J Biol Chem* 1996, **271**:4632-4639.
36. Hager M, Biehler K, Illerhaus J, Ruf S, Bock R: **Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome b(6)f complex.** *Embo J* 1999, **18**:5834-5842.
37. Takahashi Y, Rahire M, Breyton C, Popot JL, Joliot P, Rochaix JD: **The chloroplast ycf7 (petL) open reading frame of Chlamydomonas reinhardtii encodes a small functionally important subunit of the cytochrome b6f complex.** *Embo J* 1996, **15**:3498-3506.
38. Swiatek M, Kuras R, Sokolenko A, Higgs D, Olive J, Cinque G, Muller B, Eichacker LA, Stern DB, Bassi R, et al: **The Chloroplast Gene ycf9 Encodes a Photosystem II (PSII) Core Subunit, PsbZ, That Participates in PSII Supramolecular Architecture** **10.1105/tpc.13.6.1347.** *Plant Cell* 2001, **13**:1347-1368.

### **II.1.5 CLB19, une protéine PPR multifacettes**

Durant ma thèse, nous avons entrepris une collaboration avec Patricia León de l'université de Mexico. Elle avait effectué un crible de mutants affectés dans le développement des chloroplastes. Lors de celui-ci elle avait isolé un mutant de protéines PPR, et avait réalisé la caractérisation physiologique de la lignée d'insertion. Au niveau moléculaire, quelques transcrits avaient été analysés par *northern* blot, et des défauts avaient été observés pour les gènes *atpB*, *rbcL* et *accD*. Mon travail a consisté en l'isolement de deux nouveaux allèles mutants, et en la caractérisation des transcriptomes chloroplastiques de tous les mutants. L'approche globale du transcriptome chloroplastique, permit par le crible, de confirmer les défauts observés, et d'indiquer s'ils étaient les seuls présents.

La caractérisation du mutant est décrite dans l'article "**CLB19, a Novel Pentatricopeptide Repeat Protein that Regulates the Expression of the rubisco Large Subunit in Arabidopsis**", l'analyse moléculaire de ce mutant est reprise par la suite.

**II.1.5.1 Article 3 : Ramos Vega M., Guevara-García A., Gutiérrez-Nava MdL., Andrés C., Cantero A., Jiménez LF., Lurin C., Small ID. and León P. CLB19, a Novel Pentatricopeptide Repeat Protein that is required for efficient 5' and 3' processing of diverse chloroplast transcripts in Arabidopsis.**

# **CLB19, a novel pentatricopeptide repeat protein that is required for transcript accumulation and efficient 5' and 3' processing of diverse chloroplast transcripts in *Arabidopsis*.**

Maricela Ramos Vega<sup>1</sup>, Arturo Guevara-García<sup>1</sup>, María de la Luz Gutiérrez-Nava<sup>1§</sup>, Charles Andrés<sup>2,3</sup>, Araceli Cantero<sup>1</sup>, Luis Felipe Jiménez<sup>4</sup>, Claire Lurin<sup>2</sup>, Ian Small<sup>2,3</sup> and Patricia León<sup>1\*</sup>

<sup>1</sup>Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62271, México

<sup>2</sup> Unité de Recherche en Génomique Végétale (Institut National de la Recherche Agronomique/Centre National de la Recherche Scientifique/Université d'Evry Val d'Essonne), CP 5708, 91057 Evry Cedex, France.

<sup>3</sup> ARC Centre of Excellence in Plant Energy Biology, Molecular and Chemical Sciences Building (M316), University of Western Australia, 35 Stirling Highway Crawley, Perth 6009 WA, Australia

<sup>4</sup> Laboratorio de Microscopía Electrónica, Facultad de Ciencias, Universidad Nacional Autónoma de México, México D.F. 04510, México

\* Corresponding author: Patricia León,

Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001 Chamilpa, Cuernavaca, Morelos 62271, México

Telephone number 52 (55) 5622 7856

Fax number 52 (777) 313 9988

Email: [patricia@ibt.unam.mx](mailto:patricia@ibt.unam.mx)

**Key Words:** yellow mutants, chloroplast development, PPR proteins, chloroplast RNA processing, *Arabidopsis*

## Summary

It has been demonstrated that post-transcriptional processes are major regulators of chloroplast gene expression and differentiation. The experimental data has shown the complexity underlying these mechanisms and the diversity of molecules that participate, but the vast majority of proteins involved still remain to be elucidated. Although biochemical analysis has provided advances in this field, recent genetic approaches through mutant isolation and characterization have permitted the identification of new components and revealed unexpected functions. Here, we present the characterization of *CLB19*, a novel gene essential for the correct end processing of various chloroplast transcripts. Loss of function mutants in *CLB19* result in a yellow phenotype with impaired chloroplast development and early seedling lethality under normal conditions. CLB19 is a chloroplast protein and a member of the PPR family. CLB19 has profound effects on the processing pattern of various plastidial transcripts, affecting transcript accumulation and protein levels. One of the transcripts severely affected in the *clb19* mutants is that encoding the large subunit of Rubisco (*rbcL*), which has alterations in abundance and processing at both ends of the transcript. Our results show that CLB19 is fundamental for transcript accumulation and correct processing of various chloroplast transcripts, suggesting a more general function for CLB19 than other PPRs previously characterized.

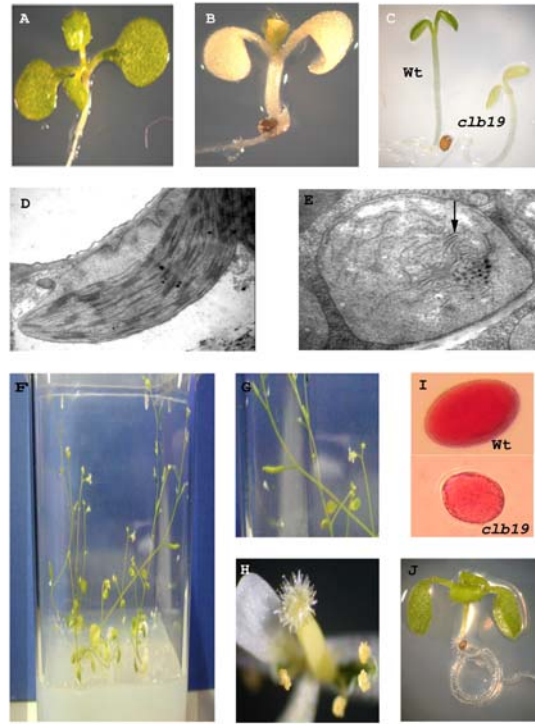
## Introduction

One of the most striking characteristics of plants is their capacity to manufacture their own carbon skeletons through photosynthesis. This essential process, which includes the fixation of CO<sub>2</sub> and the biosynthesis of carbon skeletons, takes place in the chloroplasts. The biogenesis of chloroplasts is a complex process. During the early differentiation of the mesophyll cells, the undifferentiated proplastids coordinately mature into chloroplasts (Lopez-Juez and Pyke, 2005; Vothknecht and Westhoff, 2001). Chloroplast differentiation is modulated by environmental cues such as light, and includes the proliferation of a complex internal membrane system and the expression of a specific set of genes.

Similar to the case for mitochondria, the plastome encodes a minimal number of genes. The majority of genes required for chloroplast functions are in the nucleus and their encoded proteins transported post-translationally into the organelle. Chloroplast proteins encoded in the nucleus are also largely responsible for the expression and regulation of the organelle genes. In recent years, the unique role that post-transcriptional and translational regulation has in plastid gene expression has become apparent (Herrin and Nickelsen, 2004; Nakamura et al., 2004; Stern et al., 2004). Plastids appear to have profoundly exploited post-transcriptional processes as a major way to regulate gene expression. Many chloroplast transcripts are transcribed as polycistronic units. To generate mature transcripts, chloroplast transcripts are subject to extensive internal processing. However, processing is not exclusive to polycistronic units, as various monocistronic genes also undergo processing on their 5' and 3' non-coding regions (Bollenbach et al., 2003; Herrin and Nickelsen, 2004). In addition, other post-transcriptional events including RNA splicing, trans-splicing and RNA editing also directly affect chloroplast gene expression. RNA processing has also been shown to be important for transcript stability and for translation (Bollenbach *et al.*, 2003; Nickelsen, 2003; Stern and Gruissem, 1987). Although progress has been made in the identification of some of the proteins involved in these regulatory events, most of them are still unknown. Moreover, based on the available evidence it appears that several of these regulatory proteins do not have obvious counterparts in bacteria, and represent novel acquisitions in evolution. As the vast majority of the chloroplast proteins that orchestrate these regulatory events are encoded in the nucleus, the processing and protein synthesis of chloroplast genes relies on their correct expression and import into the organelle.

The isolation of mutants that block chloroplast biogenesis has been a good strategy to identify regulatory proteins essential for organelle function (Aluru *et al.*, 2006; Barkan and Goldschmidt-Clermont, 2000; Gutiérrez-Nava *et al.*, 2004). The recent characterization of various chloroplast biogenesis mutants has shown that a significant proportion affect chloroplast RNA processing events (Bollenbach *et al.*, 2003; Gothandam *et al.*, 2005; Krause *et al.*, 2000; Meierhoff *et al.*, 2003; Rochaix, 1996; Williams and Barkan, 2003; Yamazaki *et al.*, 2004), splicing or translation. Several of these regulatory proteins are members of the pentatricopeptide repeat (PPR) family, characterized by tandem repeats of a degenerate 35 amino acid structural motif (Lurin *et al.*, 2004; Small and Peeters, 2000).

The PPR family is one of the largest protein families in plants, with approximately 450 members in *Arabidopsis* (Lurin et al., 2004). PPR proteins are also present in other



**Figure 1. *clb19* mutant phenotype.**

Morphology of 15 day-old Wt (A) and *clb19-1* (B) seedlings grown under high light (A and B) or low light (C) conditions. Transmission electron micrograph of plastids from Wt (D), and *clb19-1* (E) 15 day-old leaves. (F) Morphology of 60 day-old *clb19-1* mutant grown in GM media supplemented with 1% sucrose in which pale green leaves, siliques (G), and flowers (H) are seen. (I) Acetocarmine-stained pollen grains from Wt and *clb19-1*. (J) Molecular complementation of the *clb19* mutant is shown in a representative *clb19-1* hygromycin-resistant 15 day-old transgenic seedling (TL2), containing the Wt *CLB19* cDNA expressed from the 35SCaMV promoter. The arrows in E point to the internal membrane structures observed in the mutant plant and in G to the siliques.

**Table 1. Chlorophyll and carotenoid content in different light conditions\***

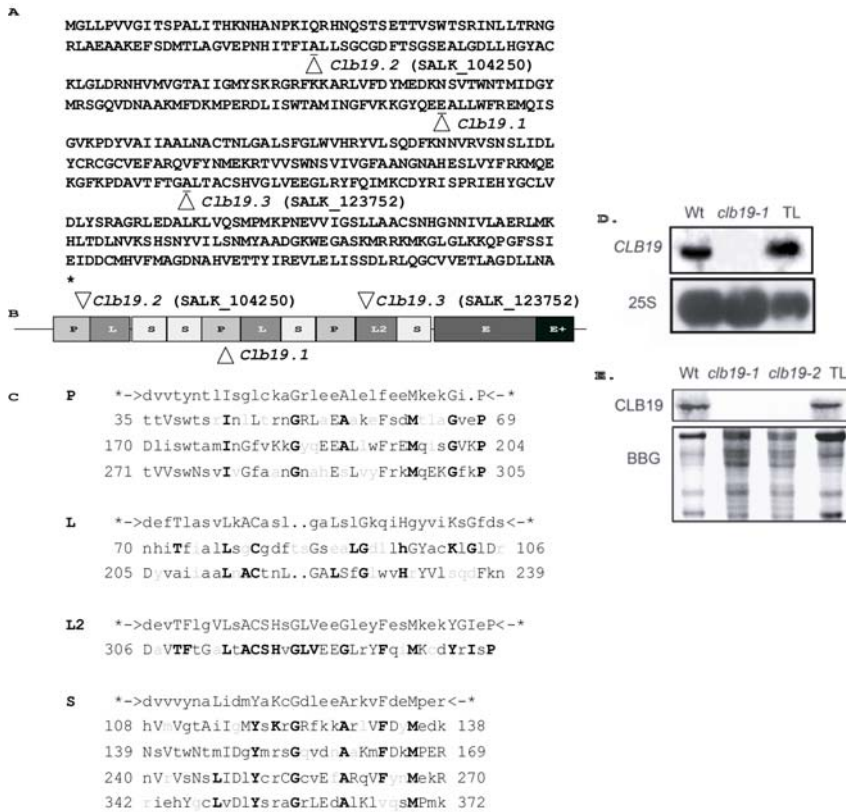
Light Condition	Plant	PIGMENTS			
		Chlorophyll A	Chlorophyll B	Total Chlorophyll	Total Carotenoids
Standard	<i>Col-0</i> Wt	70.38 ± 20.32	31.58 ± 7.29	104.93 ± 29.52	34.44 ± 9.90
	<i>clb19</i>	22.23 ± 7.55	15.01 ± 4.63	37.16 ± 12.14	12.58 ± 4.07
Low	<i>Col-0</i> Wt	31.79 ± 4.91	11.29 ± 1.88	42.95 ± 6.75	12.93 ± 2.02
	<i>clb19</i>	3.97 ± 1.81	5.94 ± 3.27	9.90 ± 5.01	8.11 ± 2.60

\*Pigments were extracted from 18-d-old seedlings. Values are expressed micrograms pigment per grams fresh weight of seedling tissue. Standard (120  $\mu\text{M m}^{-2} \text{sec}^{-1}$ ) or Low (5 $\mu\text{M m}^{-2} \text{sec}^{-1}$ ). The average of three replicates is shown, with  $\pm$  SE

organisms, including animals and fungi, but in considerably smaller number. For example, the human genome contains only 6 members of this family. This family has been divided in two subgroups based on their structural characteristics (Lurin et al., 2004; Rivals et al., 2006). The P subfamily is composed by proteins formed exclusively by PPR repeats, whereas the PLS subgroup, in addition to the classical PPR motifs also contains PPR-related motifs (L and S) and in some cases other characteristic motifs at their carboxy-termini (Lurin et al., 2004). Although the PPR subgroup is found in several eukaryotes, the PLS subfamily appears to be specific to plants.

Computational analysis predicts that more than 70% of the plant PPR proteins are targeted to either mitochondria or plastids (Small *et al.*, 2004; Small and Peeters, 2000) but this subcellular localization has been corroborated for relatively few members (Lurin et al., 2004). Structural analysis suggests that PPRs might function as RNA-binding proteins (Small and Peeters, 2000; Williams and Barkan, 2003). This binding activity has been corroborated *in vitro* for some members (Lurin *et al.*, 2004; Mancebo *et al.*, 2001; Meierhoff *et al.*, 2003; Nakamura *et al.*, 2003; Okuda *et al.*, 2006). More recently, *in vivo* interaction between these proteins and particular RNAs has been shown through direct co-immunoprecipitation of the specific targets with antibodies against a particular PPR (Schmitz-Linneweber et al., 2005; Schmitz-Linneweber et al., 2006). The molecular function of the PPR family has been characterized in detail only in few members. Some PPR proteins have been found to restore male fertility in different plants carrying mitochondria that confer cytoplasmic male sterility (Klein *et al.*, 2005; Koizuka *et al.*, 2003; Wang *et al.*, 2006). This restoration is due to the suppression of the accumulation of aberrant proteins responsible for the sterility phenotype, by affecting post-transcriptional processes required for their expression (Bentolila et al., 2002; Desloire et al., 2003; Koizuka et al., 2003; Wang et al., 2006). In chloroplasts, the CRP1 protein of maize is implicated in processing or translation of the plastid *petA* and *petD* transcripts (Fisk et al., 1999). Similar roles in plastid transcript processing have been documented for other chloroplast PPR proteins (Barkan et al., 1994; Hashimoto et al., 2003; Meierhoff et al., 2003; Williams and Barkan, 2003; Yamazaki et al., 2004). The participation of PPR proteins in other post-transcriptional events like chloroplast RNA editing and trans splicing has also been shown. Recent evidence also suggests that PPR proteins might be involved in transcriptional processes .

Despite the high number of genes comprising the PPR family and the sequence similarity between some of them, mutations in specific members often result in specific and severe phenotypes. For example, these proteins have been found to be essential for embryogenesis and/or plastid development (Ding *et al.*, 2006; Gothandam *et al.*, 2005; Gutierrez-Marcos *et al.*, 2007; Lurin *et al.*, 2004). This demonstrates that a high degree of specificity for particular targets or functions exist for members of this family. Currently, we know little about how this specificity is achieved and the targets for each particular PPR . A better understanding of these aspects requires the detailed characterization of more members of the family. In this study, we report the characterization of *clb19*, an Arabidopsis pale yellow mutant that affects a gene essential for chloroplast development. Molecular analysis shows that *CLB19* is a member of the E<sup>+</sup> PPR gene family subgroup. *CLB19* is localized in chloroplasts and expressed at higher levels in photosynthetic tissues. Our results demonstrate that this protein is required for the RNA accumulation and for the correct processing of various chloroplast transcripts. In particular, *CLB19* is required for efficient processing of the *rbcl* transcript. Our analyses demonstrate that mis-processing occurs at



**Figure 2. CLB19 gene structure.**

Amino acid sequence of the *CLB19* gene (A). The putative plastid transit peptide sequence (54 amino acids) identified by TargetP (<http://www.cbs.dtu.dk/services/TargetP>) is underlined; the T-DNA insertion sites in the *clb19-1*, *clb19-2* and *clb19-3* alleles are indicated with open triangles. (B) Diagrammatic representation of the PPR motifs within the *CLB19* protein. The designation of PPR motif; P, L, L2 and S, and the carboxy-terminal motifs (E and E+) correspond to the classification proposed by (Lurin et al., 2004). Prediction was done using Hmmssearch. cTP refers to the plastid transit peptide shown in A and the position of *clb19* mutant alleles are also shown. (C) Alignment of each of the PPR motifs in *CLB19* against the HMMER-derived consensus sequences (Lurin et al., 2004). The sequence on the top of each motif type represents the predicted consensus; capital letters correspond to conserved residues, capital bold letters in *CLB19* correspond to residues identical to the consensus.

(D) Transcript expression profile of *clb19-1* and the *clb19-1* allele complemented with the Wt *CLB19* gene expressed from the CAMV35S promoter (TL). Northern blot analysis was done using 10 µg of total RNA from 15 d-old seedlings of Col (Wt), *clb19-1* mutant, and a representative kanamycin/hygromycin resistant transgenic line (TL). Hybridization was performed using the complete *CLB19* cDNA as a probe. Hybridization to the 25S rRNA is shown as a loading control. Both ends of affected transcripts, suggesting a more general function for *CLB19* than other PPRs previously characterized.

(E) Accumulation of the *CLB19* protein in Wt, *clb19-1*, *clb19-2* plants and in a representative TL transgenic line. Western blot analysis was done using 10 µg of total protein from 15 d-old seedlings, and immunodetection was done using *CLB19* polyclonal antibodies. The Coomassie brilliant blue-stained gels (BBG) are shown as a loading control. Membranes are representative of at least two biologically independent experiments.



## Results

### *Phenotypic characterization of the *clb19-1* mutant*

A major interest of our group is to identify genes and factors required for proper development of proplastids into chloroplasts. One strategy is genetic screening for mutants accumulating low levels of photosynthetic pigments. In *Arabidopsis*, proplastids mature into chloroplasts in the cotyledons during embryogenesis. Therefore, mutagenized plants segregating for pale embryos, referred to as *chloroplast biogenesis (clb)*, were selected for further characterization. The *clb19-1* mutant develops pale yellow cotyledons and seedling leaves when grown under standard light conditions ( $120 \mu\text{mol m}^{-2} \text{sec}^{-1}$ ) and dies shortly afterwards (Figure 1A and 1B). Under these conditions, *clb19-1* accumulates low levels of chlorophylls (40% of normal) and carotenoids (38% of normal) (Table I). The development of *clb19-1* plants was also assessed. Under standard light conditions and supplemented by sucrose as a carbon source, *clb19-1* continues development albeit its low pigment levels, producing numerous leaves (Fig. 1F), an inflorescence stem (Fig. 1G), flowers (Fig. 1G and 1H), siliques (Fig. 1G), viable pollen (Fig. 1I), and even viable seeds (data not shown). Despite this normal appearance, growth and development of *clb19-1* takes considerably longer than for Wt plants.

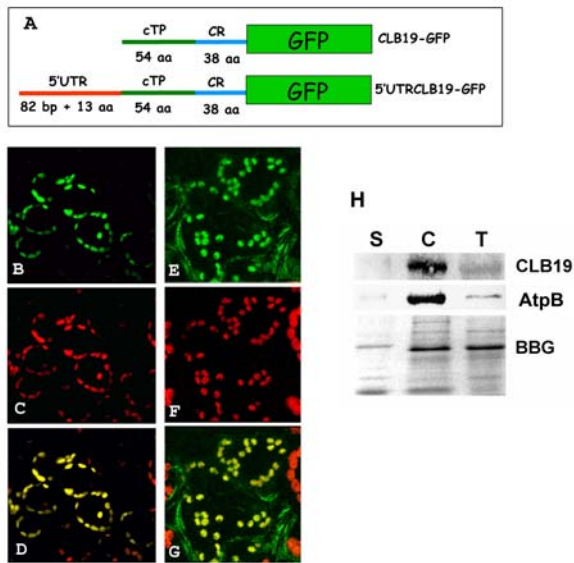
Neither viability nor pigmentation of this mutant are restored under low light ( $2\text{-}5 \mu\text{mol m}^{-2} \text{sec}^{-1}$ ) growing conditions (Figure 1C and Table I), that reduce photo-oxidative damage. Therefore, this photosynthetic pigment phenotype is a direct consequence of the *clb19-1* mutation and is not a pleiotropic effect caused by the lack of protective pigments. Thus, the gene affected in this mutant was a good candidate to be required for chloroplast biogenesis and so was chosen for further analysis.

The examination by electron microscope of *clb19-1* indicated that the leaf plastids are irregularly shaped (Figure 1E) and smaller than wild-type (Wt) chloroplasts (Figure 1D). The presence of short linear appressed internal membranes is evident, but no membrane stacking is observed. These features indicate that organelle maturation is arrested at an early stage of chloroplast biogenesis.

### *The *CLB19* gene encodes a PPR protein*

The yellow phenotype of the *clb19-1* mutant is inherited as a single recessive (3:1 ratio Wt:mutant) allele, through several generations. This mutant was isolated from a T-DNA insertion collection. Analysis of T3 families derived from backcrosses with Col Wt showed a 100% cosegregation between the mutant phenotype and the kanamycin resistance marker. Sequences flanking the T-DNA insertion site were isolated from a partial *clb19-1* genomic library screened with the left border of the T-DNA. The 125 bp genomic sequence adjacent to the T-DNA was identical to the central region of gene At1g05750 in the *Arabidopsis* genome (<http://www.arabidopsis.org/>).

The At1g05750 locus encodes a predicted ORF of 500 amino acids that lacks introns (Figure 2A). In the *clb19-1* mutant the T-DNA was inserted at amino acid 230 (Figure 2A). Two additional mutant alleles (SALK\_104250 and SALK\_123752) were identified from the TAIR insertion mutant collection (Alonso et al., 2003), and designated as *clb19-2* and



**Figure 3. Subcellular localization of CLB19.**

(A) Schematic representation of the two green fluorescent protein (GFP) fusion constructs used in subcellular localization assays. The CLB19-GFP construct contains the putative chloroplast transit peptide (cTP) and a part of the mature CLB19 (CR). In addition to the sequence present in CLB19-GFP the 5'UTRCLB19-GFP construct contains a non-coding segment immediately upstream of the *CLB19* ATG codon. This 121 bp addition comprises part of the 5'UTR.

Localization of GFP in Arabidopsis 15 day-old leaf cells bombarded with CLB19-GFP (B-D) or with 5'UTRCLB19-GFP (E-G) constructs. Cells were examined using a confocal microscope for GFP expression (B and E) and chlorophyll autofluorescence (C and F). The merged images for each construct are shown in D and G.

(H) CLB19 accumulates in the chloroplast-enriched protein fraction of Arabidopsis leaves. Five  $\mu\text{g}$  of proteins from an enriched chloroplast fraction (C), the depleted supernatant (S), or total protein extracts (T) were resolved by SDS polyacrylamide gel. Immunodetection was done with polyclonal antibodies to CLB19 or the beta subunit of ATP synthase (AtpB). The BBG blue-stained gel is shown as a loading control.

*clb19-3*. Phenotypic characterization of these lines demonstrated that seedlings with a pale yellow phenotype, indistinguishable from the phenotype of the *clb19-1* reference allele (data not shown), also segregate in a 3:1 ratio. The insertion sites for *clb19-2* and *clb19-3* were corroborated by PCR using primers for the T-DNA and *CLB19* (data not shown) and are indicated in Figures 2A and 2B. These results strongly imply that the *CLB19* (At1g05750) gene was indeed responsible for the yellow phenotype in these mutants.

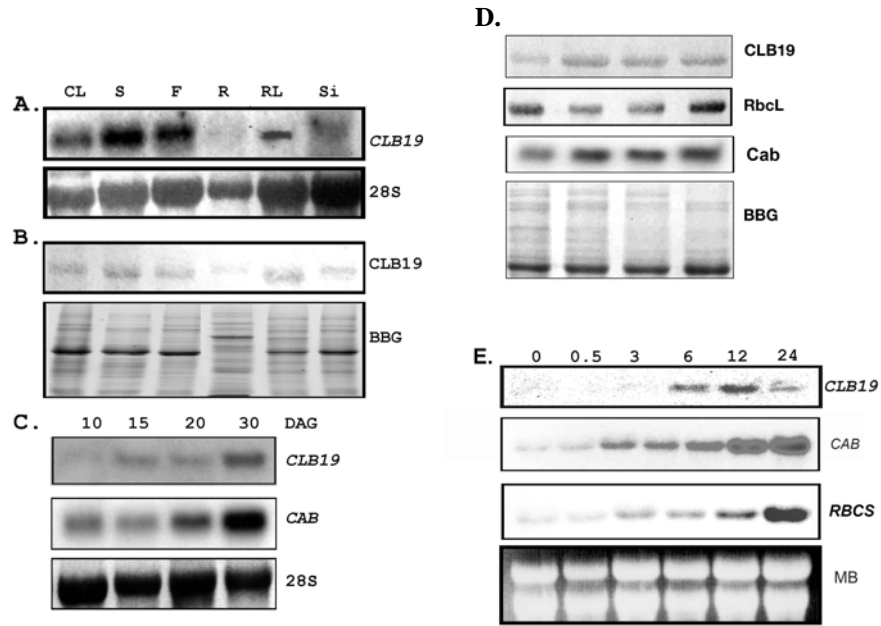
Final confirmation of gene identity was obtained from molecular complementation of the yellow phenotype of *clb19-1* with the Wt gene. A full-length *CLB19* cDNA expressed under the control of the CaMV35S constitutive promoter was introduced into the *clb19-1* mutant background via *Agrobacterium tumefaciens* mediated transformation. A representative transgenic T2 plant derived from the complementation test is shown in Figure 1J, displaying a normal green phenotype. PCR analysis corroborated the genotype of the complemented lines using primers for the *CLB19* gene, the kanamycin marker (inserted in the *clb19-1* locus) as well as for the hygromycin gene present in the *35S::CLB19* transgene (data not shown). No plants with yellow phenotype segregated in the following generations on the progeny of the hygromycin- kanamycin-resistant individuals. Therefore, we conclude that the presence of the Wt *CLB19* gene is sufficient to rescue the yellow phenotype of the *clb19-1* mutant.

The *CLB19* gene is a member of the PPR family. The CLB19 sequence contains 10 PPR or PPR-like motifs (P, L, L2 or S) present in tandem (Lurin et al., 2004). These motifs are present in a triple motif arrangement (P-L-S), similar to other PLS subfamily PPRs (Figure 2C). In addition to the PPR motifs, the carboxy terminus of CLB19 contains two extra motifs, E and E<sup>+</sup>. Thus, based on this structure, CLB19 belongs to the E/E<sup>+</sup> subclass as proposed by Lurin et al. (2004).

Transcript expression of *CLB19* was analyzed in the *clb19* mutant alleles and compared to the Wt. As shown in Figure 2D, the transcript of the *CLB19* gene is undetectable in the *clb19-1* mutant in contrast to Wt and the phenotypically complemented transgenic plants. For the analysis of the corresponding CLB19 protein, specific antibodies were raised and used to examine total protein extracts from Wt and from two independent mutant alleles. As shown in Figure 2E, neither *clb19-1* nor *clb19-2* have detectable levels of the CLB19 protein. These data demonstrate that the *clb19-1* and *clb19-2* mutants are null alleles. The CLB19 antibody does not show significant cross-reactivity with other proteins of the family based on the analysis of the mutant alleles (data not shown).

### *The CLB19 protein is targeted to plastids*

The prediction programs TargetP and Predotar (Emanuelsson et al., 1999; Small et al., 2004) predict that CLB19 is targeted to plastids. According to Target P, the 54 amino acids at the N-terminus of CLB19 could function as a putative plastid transit peptide (Figure 2A). To validate the subcellular localization of CLB19, the putative chloroplast transit peptide (cTP) that includes 54 amino acids (aa) plus 38 aa of CLB19 coding region was fused to a GFP reporter gene. This fusion (CLB19-GFP) expressed under the control of the constitutive CaMV35S promoter was introduced into *Arabidopsis* cells by DNA bombardment (Figure 3A). A second construct that includes an additional 131 bp upstream of the ATG codon of the *CLB19* open reading frame was also generated (5UTRCLB19-GFP), as recent reports indicated that 5' UTR sequences can affect targeting (Christensen et al., 2005; Millar et al., 2006). Analysis by confocal imaging of *Arabidopsis* leaves after cell



**Figure 4. Expression analysis of the *CLB19* transcript and *CLB19* protein.**

A. Northern-blot analysis of the *CLB19* transcript in Wt seedling organs. Each lane contains 10  $\mu\text{g}$  of total RNA extracted from cauline leaves (CL), inflorescence stem (S), flower (F), roots (R), rosette leaves (RL), or siliques (Si). Hybridization was performed using the complete *CLB19* cDNA as a probe. The membrane was then rehybridized with the 25S rRNA (25S) as a loading control.

B. Western-blot analysis of *CLB19* expression in plant organs. Total protein extracts were isolated from the Wt organs listed in A. Immunoblots were performed using *CLB19* polyclonal antibodies. Each lane contains 10  $\mu\text{g}$  of total protein extract. The BBG stained gel is shown as a loading control.

C. *CLB19* transcript expression during development. Each lane contains 10  $\mu\text{g}$  of total RNA extracted from whole plants on the specified day after germination (DAG). Hybridization was performed with *CLB19* and *CAB* specific probes. The membrane was rehybridized with the 25S rRNA (25S) as a loading control.

D. Developmental expression of the *CLB19* protein. Ten  $\mu\text{g}$  of total protein extracts from Wt seedlings at the same developmental stages as in C. Immunodetection was performed using polyclonal *CLB19*, *RbcL* and *CAB* antibodies. The BBG stained gel is shown as a loading control.

E. Induction of the *CLB19* transcript by light. Northern-blot analysis was performed with 10  $\mu\text{g}$  of total RNA from 15 day-old Wt whole seedlings grown in a 16:8 light:dark cycle for 12 day, then transferred to darkness for 3 day. Plants were returned to the light for different time periods (h) as specified in each lane. Hybridization was performed using fragments specific for *CLB19*, *CAB* and *RBCS* as probes. The methylene blue-stained membrane (MB) is shown as a loading control.

bombardment with either the CLB19-GFP (Figure 3B and 3D) or with 5UTRCLB19-GFP (Figure 3E and 3G) constructs, showed fluorescence exclusively in chloroplasts. The GFP colocalizes with chlorophyll autofluorescence (Figures 3C, 3D, 3F and 3G); confirming that CLB19 is exclusively targeted to chloroplasts.

The chloroplast localization of CLB19 was further corroborated by immunoblot analysis after chloroplast enrichment. As shown in Figure 3H, CLB19 is detected in total leaf protein extracts (T) and enriched in the chloroplast protein fractions (C). This parallels other chloroplast proteins such as AtpB (Figure 3H). CLB19 is undetectable in the non-chloroplast fraction (S), where cytoplasmic and mitochondrial proteins are located. Together, these data indicate that CLB19 is imported into plastids to perform its function and that 52 amino acids from the amino terminus are sufficient for this subcellular localization.

### *Expression analysis of the CLB19 gene*

To further characterize the *CLB19* gene, its transcript and protein profiles were analyzed. The *CLB19* transcript was detected in all analyzed tissues (Figures 4A); its abundance is highest in photosynthetic tissues and flowers. In roots, the *CLB19* transcript accumulates to low but detectable levels, suggesting that the function of this protein is required in other plastid types. *CLB19* is also present at various developmental stages after germination (Figure 4C) consistent with a persistent role of *CLB19* in photosynthetic tissues. Our data are confirmed by microarray data analyzed by Genevestigator which show low expression in root and flower tissues compared with leaves and stems (data not shown). The protein profile of CLB19 parallels the transcript levels; CLB19 accumulates in green tissues, including stem, rosette and cauline leaves. The lowest levels of CLB19 protein are found in roots (Figure 4B), which is consistent with the low transcript accumulation. Protein analysis during seedling development confirms that the level of CLB19 is low in young seedlings (10 d) followed by a significant accumulation in 15 day-old seedlings, and it is maintained at similar levels up to 30 days (Figure 4D).

Accumulation of the *CLB19* transcript is positively regulated by light (Figure 4E). In RNA blot analysis of 18 day-old Wt seedlings acclimated in the dark, *CLB19* transcripts are almost undetectable. Upon illumination, *CLB19* transcripts in seedlings became detectable after 3 h, reaching their highest level after 12 h of light exposure (Figure 4E). The up-regulation of this transcript by light parallels what is observed for photosynthetic genes such as *RBCS* and *CAB1*, although the latter transcripts accumulate at higher levels after 24 h of light exposure whereas *CLB19* has a more transient response (Figure 4E).

### *CLB19 protein is required for efficient processing of various chloroplast transcripts*

To address the role that CLB19 protein plays in chloroplast development, the expression patterns of various plastidial nuclear- and chloroplast-encoded genes were investigated. Some of these genes are known to be expressed at different stages of chloroplast development and constitute good markers for the differentiation process. Similar to other chloroplast biogenesis mutants (e.g. *clb6*; Gutierrez-Nava et al. 2004), the expression of several nuclear photosynthetic genes is reduced in the *clb19* mutant (Figure 5A), when these plants are grown under normal light conditions ( $120 \mu\text{mol m}^{-2} \text{sec}^{-1}$ ). However, this difference in expression is basically undetectable when the plants are grown under low light

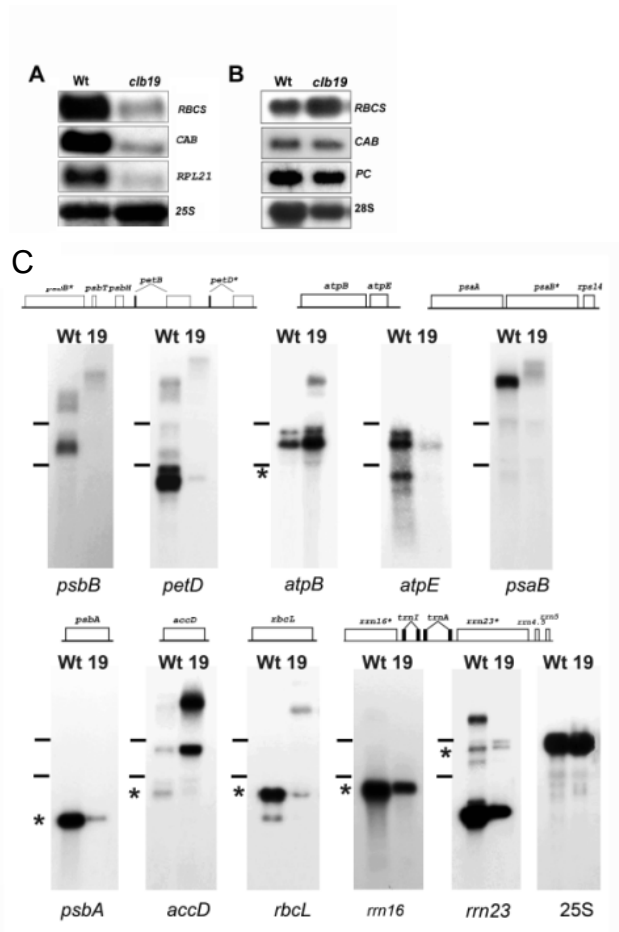


Figure 5. Expression analysis of the nuclear- and chloroplast-encoded genes in the *clb19* mutant.

Total RNA from Wt (Wt) or *clb19-1* seedlings grown in the 16:8 light:dark cycle at  $120 \mu\text{mol m}^{-2} \text{seg}^{-1}$  light intensity (A) or at  $5 \mu\text{mol m}^{-2} \text{seg}^{-1}$  (B) for 15 days was size fractionated on an agarose gel, transferred to nylon membrane, and hybridized with the following nuclear-encoded gene probes: *RBCS*, *CAB1*, *RPL21*, *PC*, and *25S* as a loading control.

(C) Expression pattern of chloroplast-encoded genes. Five  $\mu\text{g}$  of total RNA from 15 day-old Wt (Wt) and *clb19* (19) plants grown under low light intensity were fractionated on agarose gels. Hybridization was done with the following genes: *psbB* and *petD* that belong to the same transcriptional unit as shown in the diagram, *atpB* and *atpE* (dicistronic unit), *psaB* (tracistronic), *rrn16S* and *rrn23S* (polycistrons) and the monocistronic genes *psbA*, *accD* and *rbcL*. In the upper part of the gels is shown a diagram of the transcriptional units where each of the genes used as probes belongs. The asterisks in the diagrams indicate the gene used in each case. The nuclear-encoded *25S* is showed as loading control.

(5  $\mu\text{mol m}^{-2} \text{sec}^{-1}$ ) (Figure 5B). As this condition greatly reduces photooxidative damage, our initial observations likely reflect a consequence of photooxidative damage of the immature plastids rather than a direct effect of the *clb19* mutation.

Previous reports documented that some PPR proteins participate in the metabolism of organellar transcripts, probably through direct interaction with RNAs (Lurin et al., 2004). To investigate if CLB19 might affect the expression of chloroplast genes, we analyzed the consequence of its absence on overall chloroplast gene expression through a transcriptome profile covering around 80% of the Arabidopsis plastid genome. RT-PCR amplification was carried out with RNA from 15 day-old Wt and *clb19* seedlings grown under non-photooxidative conditions and using specific oligonucleotides throughout the plastome. This oligo set generates a low-resolution transcript map from which qualitative or large quantitative differences in transcripts between mutant and Wt plants can be inferred (Andres et al, in preparation). From this analysis we observed that several chloroplast transcripts are affected in the *clb19* mutant. One of the genomic regions that showed reproducible transcript differences corresponds to the *atpE-atpB* operon and the adjacent *rbcL* gene (data not shown). In *clb19*, the accumulation of transcripts covering the intergenic region between the *atpB* and *rbcL* genes was observed (Andres et al, in preparation).

To corroborate and validate the results of this systematic survey, the transcript profile of various chloroplast-encoded genes present in several locations of the plastome were compared between the *clb19* mutant and the Wt plants. This analysis confirmed quantitative and qualitative differences in several chloroplast transcripts. Three different behaviors among the genes analyzed were observed (Figure 5D): genes like *atpE*, *psbA* and *rrn16*, showed a decrease in their corresponding transcript levels but displayed a pattern basically indistinguishable from that of the Wt plant, which suggest that either their transcription or their transcript stability might be affected. The transcripts for other genes belonging to polycistronic units (*psbB - petD*, *psaB* and *rrn23*) or to monocistronic genes (*psbA* and *rbcL*) also showed a substantial decrease in their accumulation of the typical mature transcripts, but in addition, the presence of novel transcripts for these genes were observed. These novel transcripts probably represent unprocessed species that aberrantly accumulate in the *clb19* mutant. Finally, transcripts like *atpB* and *accD* also display alterations in their transcript level and transcript pattern, but in contrast to the previous examples, a significant over-accumulation of transcripts is observed (Figure 5D). We also observed that not all the genes in a polycistronic unit are equally affected by the absence of CLB19. For example, although changes are observed in the *rrn23* transcript profile, we did not detect substantial differences in the *rrn16* processing pattern, even though both genes belong to the same polycistronic unit. Thus, CLB19 facilitates processing of several, but not all transcripts.

To characterize more precisely the type of alterations caused by the absence of CLB19, we focused on a chloroplast region that contains two of the monocistronic units whose transcriptional pattern was affected in *clb19* plants; the *rbcL* and *accD* genes. RNA hybridization experiments were conducted using sequential probes that cover the coding sequences as well as the 5' and the 3' untranslated regions of these genes (Figure 6A). As shown in Figure 6B, we confirmed both qualitative and quantitative differences in the transcripts of both genes in the mutant. In the case of *rbcL*, the mature transcript of around 1.5 kb that normally accumulates in Wt plants is practically nonexistent in young *clb19* seedlings. However, a longer transcript of approximately of 4 kb in length is detected,

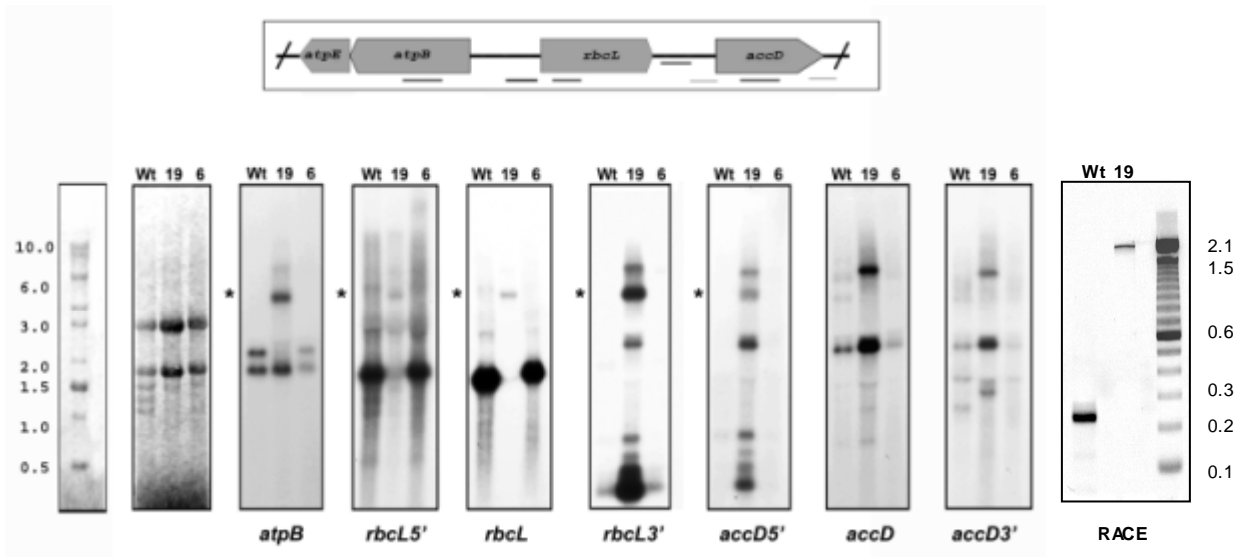


Figure 6. Expression analysis of the *atpE-atpB*, *rbcL*, and *accD* plastid region in the *clb19* mutant.

Schematic representation of the chloroplast region used for the expression analysis. The black lines underneath the diagram indicate the regions used as probes for the northern blot analyses in B, designated with a specific number (probes 1-7). The position and the region of major transcripts present in the *clb19-1* (19) for the *atpB*, *rbcL* and *accD* genes are shown by arrows. The longest species found for *rbcL* (\*) and *accD* (+) are marked. The solid line in the arrows represents the region analyzed with the probes (1-7), and the dashed lines correspond to the transcript regions not analyzed but estimated by the size and 5' RACE analysis. The diagram and the lines are not to scale.

Northern analysis. Each lane contains 10  $\mu\text{g}$  of total RNA isolated from 15 day-old Wt (Wt), *clb19* (19), or *clb6* (6) seedlings grown under low light conditions ( $5\mu\text{M m}^2 \text{seg}^{-1}$ ). The probes used in each case are indicated in A by the numbers and include: *atpB* ORF (1), *rbcL* 5' UTR (2), *rbcL* ORF (3), *rbcL* 3' UTR (4), *accD* 5' UTR, *accD* ORF (6) and *accD* 3' UTR. RNA molecular weight markers are shown (MW) and the methylen blue-stained membrane (MB) is shown as a loading control.

5' RACE analysis from wild-type (Wt) or *clb19* mutant plants (19). The bands are nested PCR products obtained using *rbcL*-specific primers and SMART kit primers designed to amplify the 5' extremities of cDNAs. Molecular weight markers are shown on the right (MW). Sequencing of the products indicates that the shorter product (in Wt) derives from transcripts starting at -179 with respect to the initiation codon.



albeit in lower levels (Figure 6B). The sequential probes that hybridize with the 3' UTR of *rbcL* and with the 5' UTR of *accD* demonstrated that the 4 kb transcript expands into the *rbcL-accD* intergenic region. Hybridization to this large transcript is not detected with an *accD* coding sequence probe, demonstrating that this transcript does not extend this far (Figure 6B). These data suggest that the 3' terminus of this aberrant 4 kb transcript ends near the region where the *accD* coding sequence begins. Our analysis also demonstrates that the large *rbcL*-containing transcript includes sequences that extend the 5'UTR of *rbcL* and also include the entire complementary sequence of the *atpB* gene (Figure 6B). This data has been corroborated using a strand-specific probe (data not shown) and 5' RACE analysis (Figure 6C) that clearly demonstrates a processing defect of *rbcL* transcripts in *clb19*. Sequencing of the 5' RACE products indicates that the major, highly abundant *rbcL* transcript in Col Wt starts at -177 with respect to the AUG start codon. In *clb19*, *rbcL* transcripts extend almost 2kb upstream. This is entirely consistent with the low resolution mapping data from both the northern hybridizations (Fig. 6B) and with the RT-PCR experiments (Andrés et al, in preparation). Thus, this analysis demonstrates that in the *clb19* mutant the *rbcL* transcript has processing defects at both its 5' and 3' ends.

We can also show that the *accD* gene transcript has processing defects. Similar to *rbcL*, a long transcript (more than 5 kb) not present in the Wt accumulates in *clb19* plants. This transcript contains sequences covering the intergenic *rbcL-accD* region, continues into the *accD* coding region and includes the 3' end of the *accD* gene (Figure 6B). Based on this data, *accD* transcription starts within or immediately downstream of the *rbcL* gene and extends several kb downstream of *accD*. Again processing defects exist at the 5' and 3' termini of the transcript when compared to the Wt pattern. In contrast to what we observed for *rbcL*, the *accD* transcripts that normally accumulate in the Wt plant are also present in the mutant and at higher levels (Figure 6B). Finally, this analysis also demonstrates that these processing defects are unique to *clb19* and not a consequence of the overall chloroplast arrest as they are not observed in the unrelated albino *clb6* mutant (Figure 6B) that also has undifferentiated plastids. Collectively, these results demonstrate that one of the functions of CLB19 relates to the proper 5' and 3' processing of various chloroplast transcripts.

### *Chloroplast protein profile of the clb19 plants*

To evaluate the impact that the transcript processing alterations have over the accumulation of the corresponding proteins, immunoblot analysis was carried out for some proteins whose transcripts are altered. As some PPR proteins have global effects on chloroplast translation, we also included representative subunits for additional photosynthetic complexes such as the cytochrome  $b_6f$  complex (*petA*) and photosystem I (*psaD*). Immunoblot analysis showed that most of the representative protein subunits analyzed are basically undetectable in the *clb19* mutant (Figure 7A), when grown under normal light conditions ( $120 \mu\text{mol m}^{-2} \text{sec}^{-1}$ ). Although higher accumulation of these proteins is observed when the mutant is grown under non-photooxidative conditions ( $5 \mu\text{mol m}^{-2} \text{sec}^{-1}$ ), the levels of most of them are still considerably lower than in Wt plants (Figure 7A, LL). The only exception is AtpB, which accumulates to Wt levels particularly under low light conditions. In contrast, as already obvious from stained protein gels (Fig. 3C and 7A), *clb19* contains very low levels of RbcL, even under low light conditions. Based on these results we conclude that the transcript accumulation and processing defects in *clb19* are

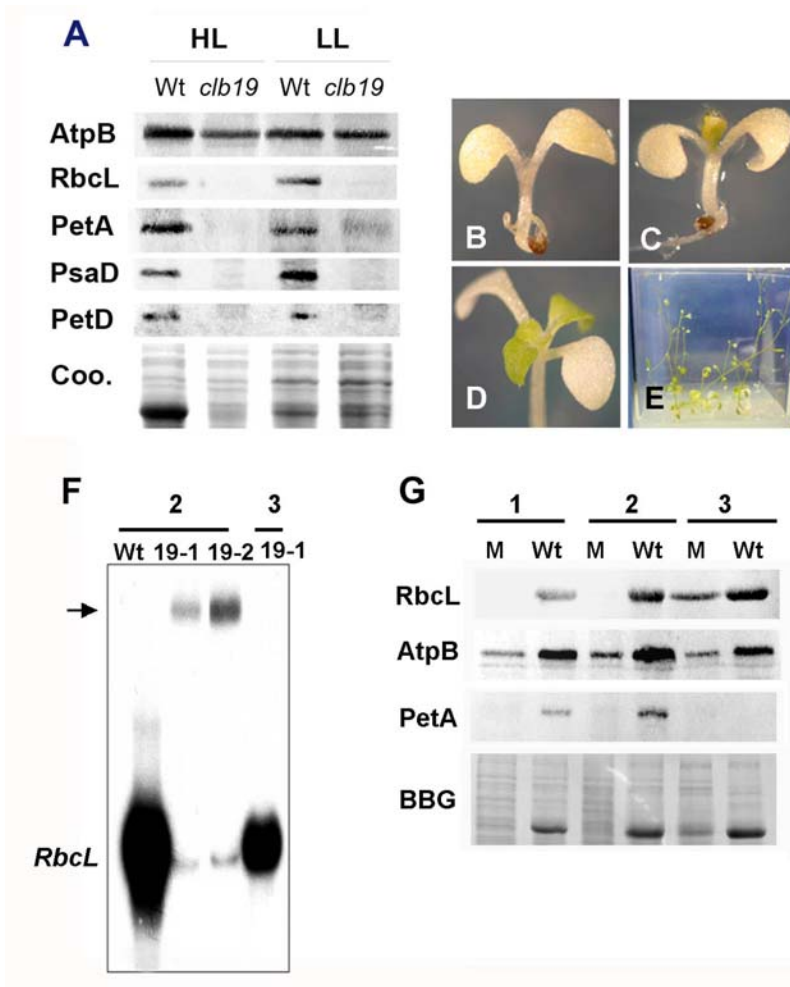


Figure 7. Phenotypes of *clb19* during development.

(A) Western blot of chloroplast proteins in Wt and in *clb19* mutant. Total protein extracts were obtained from 15 day-old seedlings grown in standard light (HL) or low light (LL) conditions. Immunoblots were performed using antibodies against the subunit B of ATP synthase (AtpB), the large subunit of RuBISCO (RbcL), a subunit of the cytochrome  $b_6/f$  complex (PetA), a subunit of photosystem I (PsaD) and a subunit of the cytochrome  $b/f$  (PetD). Each lane contains 10  $\mu\text{g}$  of total protein extract. The BBG stained gel is shown as a loading control.

(B) Phenotype of *clb1-19* plants at 8 day-old (1), 15 day-old (2), 30 day-old seedling (3) and 60 day-old (4) grown in media supplemented with sucrose.

(C) Northern blot quantification of *rbcL* transcript in the *clb19-1* mutant during development. 10  $\mu\text{g}$  of total RNA from 15 day-old seedlings (2) or 60 day-old plants (3), Wt or *clb19-1* (19-1) or *clb19-2* (19-2) mutants. The *rbcL* coding region was used as probe.

(D) Immunoblot analysis of chloroplast proteins during *clb19-1* development. Ten  $\mu\text{g}$  of total protein extracts from Wt (Wt) and *clb19-1* (M) plants at 15 day-old (2), 30 day-old (3) and 60 day-old (4) plants. Protein levels were assessed with antibodies specific to RbcL, AtpB, and PetA. A stained membrane (BBG) is shown as the loading control.

associated with severe effects on the accumulation of several plastid-encoded proteins, notably RbcL. However, general protein translation is not affected as other plastid-encoded proteins such as AtpB are easily detectable.

### *The requirement for CLB19 changes during development*

The yellow phenotype of *clb19* during early seedling development (Figure 7B panels 1 and 2) is followed by a gradual greening during later stages of developmental under *in vitro* growth conditions with standard light conditions. As shown in Figure 7B (panel 3), a pale green color is clearly evident in the second pair of true leaves by day 30. Greening is more evident in older stages of the mutant plants (Figure 7B panel 4). To address whether this apparent phenotypic reversion correlates with changes in chloroplast transcript processing, the transcript pattern of the *rbcL* gene was analyzed at different developmental stages in the mutant plant. As shown in Figure 7C, a dramatic increase in the abundance of the 1.5 kb mature *rbcL* transcript was observed in older developmental stages. The accumulation of the mature *rbcL* transcript correlates with the disappearance of the longer 4 kb transcript (Figure 7C stage 4), predominant in the 15 day-old seedlings (Figure 7C stage 2) and the correct accumulation of correctly 5' processed *rbcL* transcripts beginning at -177 (data not shown). This phenotypic reversion also correlates with the accumulation of RbcL protein (Figure 7D lanes 4). Considering these results together, there is a close correlation between correct *rbcL* transcript accumulation, processing, *rbcL* protein accumulation and progressive reversion of the yellow phenotype in older *clb19* plants.

## Discussion

There are increasing examples of diverse post-transcriptional mechanisms that regulate the expression of the chloroplast-encoded genes. Most chloroplast RNAs in plants are represented by multiple transcripts and undergo various processing events. Several nuclear genes have been demonstrated to participate in aspects of plastid RNA processing and RNA stability in higher plants and in *Chlamydomonas reinhardtii* (Barkan and Goldschmidt-Clermont, 2000; Herrin and Nickelsen, 2004; Nickelsen, 2003). The majority of these proteins play specific functions mediating particular intramolecular cleavage, end terminal processing (either 5' or 3'), splicing, editing and translation of one or few targets. A few others have more pleiotropic functions towards several genes or processes without apparent sequence specificity (Jenkins *et al.*, 1997; Nakamura *et al.*, 2004; Nickelsen, 2003; Rymarquis *et al.*, 2006; Walter *et al.*, 2002). In the case of *C. reinhardtii* several of the nuclear factors that play important roles in post-transcriptional processing act through the 5' or 3' UTR and many of them belong to the TPR protein family. In higher plants, many examples demonstrate the participation of nuclear proteins for the correct expression of particular plastidial genes through post-transcriptional processes. These factors include TPR proteins, novel proteins (Lennartz *et al.*, 2006; Lezhneva and Meurer, 2004) and several PPR proteins.

In this work we present the identification and characterization of CLB19, a novel PPR protein that plays an essential role in proper chloroplast development in *Arabidopsis thaliana*. PPR proteins are present in many organisms, but in plants represent one of the largest gene families (Aubourg *et al.*, 2000; Small and Peeters, 2000). Computational analysis has shown that PPR proteins are formed by the repetition of PPR or PPR-related

motifs (PLS). In the case of CLB19, three PLS clusters are present. In addition, at its C-terminal region CLB19 contains sequences characteristic of the E/E<sup>+</sup> subfamily defined by Lurin et al. (2004). In *Arabidopsis*, the E/E<sup>+</sup> subgroup consists of more than 100 members of which an important proportion are predicted to be targeted to mitochondria or to plastids. The molecular functions of almost none of this subgroup of proteins is known. Available experimental evidence demonstrates that other members of the PPR protein family play active roles in different aspects of post-transcriptional processes in organelles. Several PPRs are required for the proper maturation of organellar transcripts (Fisk et al., 1999; Hashimoto et al., 2003; Meierhoff et al., 2003); as well as splicing, editing and translation (Barkan et al., 1994; Fisk et al., 1999; Kotera et al., 2005; Schmitz-Linneweber et al., 2006; Williams and Barkan, 2003).

In spite of the number of PPR proteins in the *Arabidopsis* genome, the lethal phenotype caused by the absence of CLB19 demonstrates its unique role during early stages of seedling development. Our analyses showed that CLB19 is targeted to chloroplasts where it has a non-redundant function for the development of these organelles. Mutations that eliminated the expression of CLB19 result in yellow seedlings that cannot survive photoautotrophically. RNA and protein analyses demonstrate that CLB19 is expressed in most tissues of the plant, but is most abundant in green tissues. The expression pattern agrees with the requirement of this protein for chloroplast function, as does the positive regulation that light has over the expression of CLB19, similarly to other photosynthetic genes. *CLB19* transcript and protein are detected in root tissues at low levels, suggesting that this protein also has a role in the expression of genes in other plastid types. This result is consistent with the finding that CLB19 is required for the correct maturation of 23S rRNA, whose function is known to be required in all plastid types .

We conducted various analyses to identify the primary defects in *clb19* mutants. Our analyses show that the absence of CLB19 results in decreased amounts of mature transcripts for several chloroplast genes, and in the accumulation of longer RNA species that probably represent unprocessed or partially processed precursor transcripts. In addition, the level of several plastid proteins decreases in *clb19*. As it is known that processing, stability and translation are often coupled in plastids , CLB19 could participate in one or more of these mechanisms. Phenotypes similar to *clb19* have been found in PPR mutants that affect general chloroplast translation in maize . However, our data shows that general translation is not impaired in *clb19*, since transcripts such as *atpB* are translated in this mutant at similar levels to Wt plants. Our data also show that it is unlikely that the processing defects observed in the mutant could be attributed to a role of CLB19 in general editing or splicing, because several of the transcripts affected in *clb19*, such as *rbcl*, do not contain introns and are not edited. Nevertheless, the absence of CLB19 affects the expression level, the processing and the protein accumulation of several chloroplast genes. This could be because CLB19 interacts directly with multiple target transcripts or because a highly specific interaction with one target leads to pleiotropic effects on many others. Such apparently relatively unspecific effects have been observed for other chloroplast PPR proteins such as PGR3 (Yamazaki et al., 2004) and CRP1 (Schmitz-Linneweber et al., 2005). In the case of CRP1, the protein has been shown to bind *in vivo* to at least two different transcripts, but the expression of other chloroplast genes is affected too (Fisk et al., 1999; Schmitz-Linneweber et al., 2005). The exact molecular function of CLB19 is still an open question because in spite of multiple efforts, we have not been able to convincingly demonstrate binding of this protein to specific transcript(s) *in vitro* or *in vivo*. Although

various evidence supports that PPR proteins commonly act as RNA-binding proteins, we cannot eliminate the possibility that this particular PPR does not directly interact with its RNA targets or, more likely, that additional proteins are required for such interactions.

We have shown that CLB19 is required not only in the processing of transcripts from polycistronic units but it is also for the correct processing of monocistronic transcripts, including *rbcL* and *accD*. In both cases we observed that the absence of CLB19 results in the presence of longer transcripts with processing defects at both the 5' and 3' termini, suggesting that CLB19 is required at both ends of these transcripts, or that processing at each end is linked in some way. In *Chlamydomonas* chloroplasts it is known that some factors, such as MCD1, have the capacity to participate in the processing and protection of both ends of a transcript (Drager et al., 1999; Murakami et al., 2005). MCD1 appears to interact with additional proteins (MCD3 and MCD4) in a multiprotein complex predicted to be involved in the endonucleolytic cleavage required during both 5' and 3' end processing. These proteins have pleiotropic effects over the processing and stability of both ends of the transcripts, similar to CLB19. Although CLB19 does not exhibit sequence similarity with any of the known *Chlamydomonas* factors, (and indeed *Chlamydomonas* lacks PLS family PPR proteins completely) the functional similarities provide an interesting avenue for future analysis and also highlights important parallelisms between *Chlamydomonas* and higher plants.

Besides the alterations in RNA processing, our expression analysis in the *clb19* mutant also demonstrated major differences in the transcript levels for several chloroplast genes. This analysis showed substantial decrease in the transcript accumulation of several genes that could result from changes in the RNA stability of the mis-processed transcripts. However, in this analysis we also observed the over-accumulation of some transcripts including *atpB*, *accD* and *clpP* (data not shown) in comparison to Wt plants. This varied pattern of transcript abundance in *clb19* shows similarities to the pattern observed in mutants of the *rpo* genes, impaired in the function of the plastid polymerase (PEP) and also with mutants in components of the transcriptionally active chromosome (TAC) complex (Pfalz et al., 2006; Serino and Maliga, 1998). Similar to those mutants, the transcripts of genes expressed by the PEP polymerase (*rbcL*, *psbA*, *psaB*) are less abundant in *clb19*. In contrast, transcripts from genes such as *atpB*, *accD* and *clpP*, whose expression depends partially or wholly on the NEP polymerase, are more abundant in *clb19* than in wild-type plants. The mapping analysis of *rbcL* transcript also supports defects at the transcriptional level as we demonstrated that in *clb19* the major *rbcL* transcript is transcribed from a distant promoter located 2Kb upstream of the gene and not from the reported PEP promoter as in the Wt. The precise role of CLB19 in this respect is still an open question. It is possible that CLB19 could directly participate as part of the multimeric TAC complex and thus influence both transcription and RNA processing. Such functions have been proposed for TAC-associated components including a novel PPR protein unrelated to CLB19. However, it is equally possible that the effect of CLB19 on transcription could be indirect, for example by affecting the correct processing of one or more components of the PEP polymerase.

Rubisco is one of the most abundant proteins in nature and it has been target of numerous studies. Translation of the chloroplast *rbcL* gene has been shown to be regulated by light and redox potential (Kim and Mayfield, 2002; Kim and Mullet, 2003). In *Chlamydomonas* it was found that specific elements in the *rbcL* 5' UTR are important determinants of transcript stability (Anthonisen et al., 2006; Salvador et al., 2004). Despite its importance,

little is known of the regulatory events that affect the accumulation of *rbcL* transcripts in higher plants, and therefore the *clb19* mutants provide an important resource for understanding the control of Rubisco expression in land plants.

## Experimental procedures

### *Plant material and growth conditions*

*Arabidopsis thaliana* (L) Heyhn. ecotypes Landsberg *erecta* (*Ler*) and Columbia (Col) were used in this study. Growth under sterile conditions was done using surface-sterilized seeds plated on germination medium (GM) containing 1X Murashige and Skoog basal salts (Gibco BRL, Grand Island, NY), 2% (w/v) sucrose, 1X B5 vitamin solution (Gamborg's, Sigma, Inc., St Louis, MO), 0.05% (w/v) MES [2-(N-morpholino) ethanesulfonic acid], solidified with 0.8% (w/v) phytoagar. Adult plants were grown in Metro-Mix 200 (Grace Sierra, Milpitas, CA). Seedlings were grown under 16 h light: 8 h dark cycle at standard conditions ( $120 \mu\text{M m}^{-2} \text{sec}^{-1}$ ) or at  $5\mu\text{M m}^{-2} \text{sec}^{-1}$  for low light (LL), at 22°C in growth chambers. In all experiments seeds were vernalized at 4°C for 5 d to break dormancy prior to germination. The T-DNA pools from which the *clb19-1* mutant was selected were generously provided by Chris Somerville (Carnegie Institute of Washington, Stanford, CA) as previously described. *clb19-2* (SALK\_104250) and *clb19-3* (SALK\_123752) were obtained from the ABRC stock center. The positions of the T-DNA in the *clb19-2* and *clb19-3* alleles were corroborated by amplifying the genomic and T-DNA borders using specific primers for the T-DNA: LBb1 (5'-GCGTGGACCGCTTGCTGCAACT-3') and LBa1 (5'-TGGTTCACGTAGTGGGCCATCG-3'), and a *CLB19* specific primer: clb19-5'UTR (5'-AATCTTAAAACCGG-3').

### *Pigment Determination*

Total carotenoid and chlorophyll levels were obtained from frozen tissue of 18 day-old Wt or *clb19* seedlings grown under HL or LL conditions. Pigments were extracted for 5 h at room temperature with a methanol:dicloromethane (2:1, v/v) mixture in glass vials protected from light. Pigment concentration was estimated in a Beckman DUR650 spectrophotometer at 645 nm (chlorophyll a), at 663nm (chlorophyll b), and at 450 nm (carotenoids).

### *Gene identification*

The genomic sequence adjacent to the T-DNA insertion from *CLB19* was obtained by generating a partial genomic library from the *clb19-1* mutant. 4  $\mu\text{g}$  of total genomic DNA digested with *HindIII* was size fractionated by gel and the region that contained fragments that included left T-DNA were purified. These fragments were cloned into lambda ZAP Express vector (Stratagene La Jolla, CA). Specific clones carrying T-DNA sequences were identified through hybridization using a probe from the left border of the T-DNA. Positive clones were isolated and the region adjacent to T-DNA analyzed by sequencing. The site of insertion was found through *in silico* analysis against the *Arabidopsis* genomic sequences.

### Mutant Complementation

The complete CLB19 open reading frame (from ATG to TGA) was amplified by PCR from total DNA by standard PCR, using the oligonucleotides 5'-CAATCGAATGGGTCTCCTTCCCGTC-3' (forward primer) and 5'-TCAAGCATTGAGGAGATCACCAGCAA-3' (reverse primer). The resulting 1510 bp PCR fragment was cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) and checked by restriction enzyme and sequence. The CLB19 gene was subcloned into the pFGC1008 binary vector (<http://ag.arizona.edu>) between the CaMV35S promoter and the ocs 3' terminator and used to transform heterozygous *clb19-1* mutant plants. Hygromycin-B T0 resistant (50 µg/mL) transgenic lines were selected to obtain T1 and T2 plants for further analysis. Homozygous independent transgenic lines for the transgene were identified by 100% segregation of hygromycin-B and kanamycin resistance derived from the transgene and the T-DNA insertion in subsequent generations. The presence of the transgene in a mutant background was verified by PCR from total mRNA using primers specific for the CaMV35S promoter (5'-CCAACCACGTCTTCAAAGC-3') and for the *CLB19* coding sequence (5'-GGCTTCACTTCCG-3').

### Gene expression

Total RNA was isolated from frozen tissue using TRIZOL reagent (Invitrogen, Carlsbad, CA). For RNA gel blot analysis, the RNA was fractionated by electrophoresis on 1.2% agarose 6% v/v formaldehyde gels and transferred onto Hybond N+ nylon membranes (Amersham Biosciences, Buckinghamshire, UK). Hybridizations and washes were performed at high stringency conditions according to standard procedures using 32P-radiolabeled probes with Megaprime DNA labelling system (Amersham Biosciences, Buckinghamshire, UK). The fragments used as probes are as follows: for *CLB19* (1.5 Kb) generated by PCR with the F: 5'-CAATCCCATGGGTCTCCTTCCCGTC-3' and R: 5'-CAAGCATTGAGGAGATCACCAGCAAC-3' primers from genomic DNA. For *accD* coding region (589 bp) using the primers *accD* F: 5'-ATGGAAAATCGTGG-3' and *accD* R: 5'-CACGAATCTTTCCGC-3'; for *accD* 5'UTR (354 bp) with the primers *accD*5' F: 5'-GCATGTTGTCCTTTTC-3' and *accD*5' R: 5'-CTGGCCTCTATTTAAATG-3' and for the *accD* 3' untranslated region (413 bp) using the primers *accD*3' F: 5'-CGGTTAGTTTATCAGAA-3' and *accD*3' R: 5'-GTATCATAATCCCCTTG-3'. For *rrn23S* probe (400 bp) with the primers *rrn23S* F: 5'-GCCCGGGCGGGACTGTTTAC-3' and *rrn23S* R: 5'-CTCACTGATGGCTCGGGC-3'; for the *rrn16S* the primers *rrn16S* F: 5'-AAATGCGTAGAGATCGGAAAG-3' and *rrn16S* R: 5'TTCATGCAGGCGAGTTG-3'; and for the *atpB* probe (455 bp) with the primers *atpB* F: 5'-ATGAGAACAAATCC-3' and *atpB* R: 5'-CCTGTTTCAAAAATCG-3'. For *rbcL* gene: the *rbcL*3' untranslated sequence probe (307 bp) the primers *rbcL*3'UTR F: 5'-GATTTAGTAATTCACG-3' and *rbcL*3'UTR R: 5'-GGATGGGGTAGAAGAAG-3'; for the *rbcL*5' untranslated leader probe (321 bp) with the primers *rbcL*5'UTR F: 5'-CAACAGATATTACTGTCAAGAGTC-3' and *rbcL*5' UTR R: 5'-GTCCCTCCCTACAAGTCATG-3'. Specific fragments were obtained from the *Arabidopsis* *PC* (At1g76100), *CAB1* (At1g29930), *RBCS* (At1g67090.), *RPL21* (At1g33680); from the spinach *atpE* (GenBank accession no. CAB88735), *psaB* (GenBank accession no. CAB88725), *psbB* (GenBank accession no. CAB88753), *petD* (GenBank accession no. CAB88758) and *psbA* (GenBank accession no. CAB88705) and from the pea

*rbcL* gene (GenBank accession no. X03850). RACE analysis employed the SMART RACE cDNA Amplification kit (Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France) using the manufacturer's instructions. The first round of PCR used the *rbcL*-specific primer *rbcL5'*: 5'-ACCAGCTTTGAACCCAACAC-3' whilst the second nested PCR used the primer *rbcL5UTR*: 5'-TGACATAAGTCCCTCCCTACAA-3'.

### *Antibody Preparation*

Polyclonal antibodies were generated against a His 6-tagged -CLB19 fusion protein. To construct the fusion protein the entire coding sequence of the *CLB19* gene was cloned in the PROEX vector (Gibco BRL, Grand Island, NY). The integrity of the chimeric genes was verified by direct sequencing. The IPTG-induced protein was purified from *E. coli* crude extracts by affinity chromatography using Ni-NTA Agarose (Qiagen, Inc. Valencia, CA) under native conditions according to the manufacturer's recommendation. For antibody generation, 50 µg of purified His 6-CLB19 protein in 1 ml of PBS (140 mM NaCl, 2.8 mM NaH<sub>2</sub>PO<sub>4</sub>, 7.2 mM Na<sub>2</sub>HPO<sub>4</sub> pH 7.4) and complete Freund's adjuvant (Sigma-Aldrich, Inc., St Louis, MO) were subcutaneously injected as 1:1 (v:v) emulsion in 8 week-old female New Zealand rabbits. Four additional injections (50 µg each) with incomplete Freund's adjuvant were administered at 10 day intervals starting 15 days after the initial injection. The serum obtained by total bleeding of the rabbit was collected 3 days after the last injection and the CLB19 antibody titre determined.

### *Protein gel blot analysis*

Intact chloroplasts were purified from 3 week-old Arabidopsis leaves by two-step Percoll gradients as described . Total protein samples or intact chloroplasts were obtained from frozen tissue in SDS sample buffer (0.125 M Tris-Cl pH 6.8, 20% v/v glycerol, 4% w/v SDS, 2% v/v 2-mercaptoethanol). Protein concentration was determined with Bradford reagent (Bio-Rad, Hercules, CA) using BSA as a standard. Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto nitrocellulose Hybond C membrane (Amersham Pharmacia Biotech, Piscataway, NJ) by electroblotting for 1h at 350 mA in 25 mM Tris, 0.2 M glycine, and 20% (v/v) methanol. To verify equal protein loading, a parallel gel was run and stained with Coomassie Brilliant Blue R-250. Immunodetection was performed using the following dilutions of polyclonal antibody: 1:1000 for CLB19; 1:100,000 for RbcL; 1:500 for PsaD; 1:1000 for PetD; 1:1000 for PetA and 1:10,000 for AtpB. Detection of the primary antibodies was done with goat anti-rabbit IgG-alkaline phosphatase using the BCIP/NBT substrate Kit (Zymed Laboratories, San Francisco, CA).

### *Subcellular localization*

For subcellular localization two different CLB19-GFP translational fusions were constructed. One of them (CLB19-GFP), contained the 92 amino acids from the N-terminal of the CLB19 protein and the other 5'UTRCLB19-GFP, contains the same fragment as the CLB19-GFP plus 121 pb from the 5'UTR. The fragments were produced by PCR and fused in-frame to the N-terminal of the GFP in the pHBT vector using the following primers: 5'-CGGATCCGAATGGGTCTCC-3' (CLB19 forward), 5'-CGGATCCAATCTTAAAACCGG-3' (5'UTRCLB19 forward) and 5'-



CCCATGGGCTTCACTTCCG-3' (reverse for both). *A. thaliana* Col-0 transient transformation was performed by bombardment of 2-week-old plantlets with DNA-coated tungsten particles (Tungsten M17, Bio-Rad, Hercules, CA). GFP expression was monitored in leaf tissue by confocal microscopy after 1 week from bombardment.

### *Confocal, Light, and Electron Microscopy*

Confocal images were obtained in a Carl Zeiss LSM510 META laser scanning microscopy (Carl Zeiss AG, Deutschland), equipped for excitation with an Ar2 488 nm laser. GFP and chlorophyll emissions were isolated using BP500-530 nm and BP650-710 nm filters respectively. Transmission electron micrographs were obtained exactly as described in Mandel et al. (1996). All images were processed using Adobe Photoshop (Mountain View, CA).

### *Acknowledgements*

Many thanks to Dr. Alice Barkan for kindly providing the antibodies for AtpB, PetA, PsaD and PetD proteins. We thank Dr. Patricia Dupree for her assistance with pollen staining and Dr. Rafael Vazquez Dualt for his advice on pigment quantification. We also thank Carolina San Roman and Martha Trujillo for technical support, Alma Lidia Martínez and Arturo Ocadíz for computer support and Andrés Saralegi Amaro in the confocal microscopy. We gratefully acknowledge the Salk T-DNA insertion project for the production of sequenced T-DNA insertion lines and the ABRC for providing seeds. This work was supported by Consejo Nacional de Ciencia y Tecnología (31791-N), Dirección General de Asuntos para el Personal Académico-UNAM (IN218007), BASF, Australian Research Council (CE0561495) and Howard Hughes Medical Institute grants.

## **References**

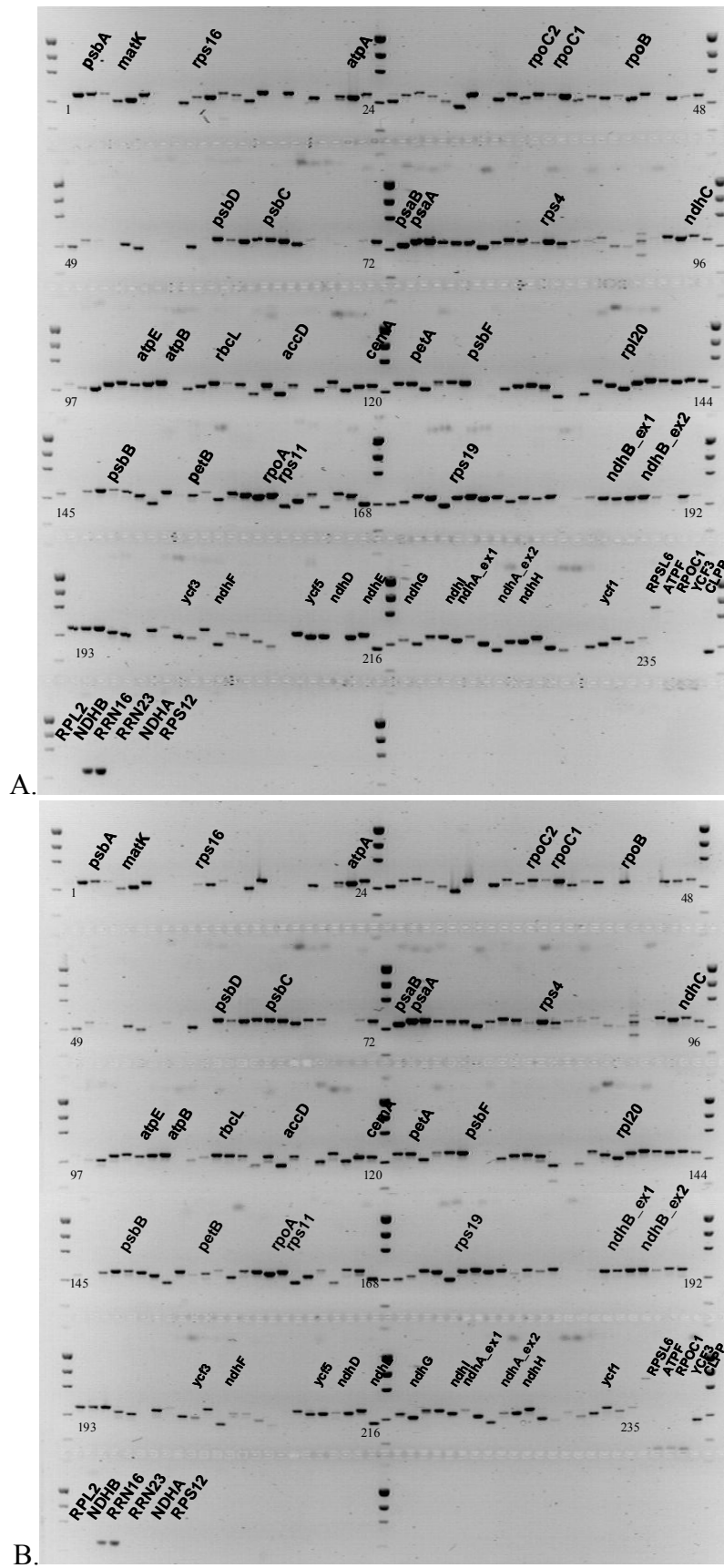
- Allison, L., Simon, L. and Maliga, P.** (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J.*, **15**, 2802-2809.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C. and Ecker, J.R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653-657.
- Aluru, M.R., Yu, F., Fu, A. and Rodermel, S.** (2006) *Arabidopsis* variegation mutants: new insights into chloroplast biogenesis. *J Exp Bot*, **57**, 1871-1881.
- Anthonisen, I.L., Salvador, M.L. and Klein, U.** (2006) Specific sequence elements in the 5' untranslated regions of *rbcL* and *atpB* gene mRNAs stabilize transcripts in the chloroplast of *Chlamidomonas reinhartii*. *RNA*, **7**, 1024-1033.
- Arnon, D.I.** (1949) Cooper enzymes in isolated chloroplasts. *Plant Physiol.*, **24**, 1-15.
- 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. and Goldschmidt-Clermont, M.** (2000) Participation of nuclear genes in chloroplast gene expression. *Biochimie*, **82**, 559-572.
- Barkan, A., Voelker, R., Mendel-Hartvig, J., Johnson, D. and Walker, M.** (1995) Genetic analysis of chloroplast biogenesis in higher plants. *Physiologia Plantarum*, **93**, 163-170.
- Barkan, A., Walker, M., Nolasco, M. and Johnson, D.** (1994) A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.*, **13**, 3170-3181.
- Bechtold, N., Ellis, J. and Pelletier, G.** (1993) In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C.R. Acad. Sci. (Paris)*, **316**, 1194-1199.
- Bentolila, S., Alfonso, A.A. and Hanson, M.R.** (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA*, **99**, 10887-10892.
- Bollenbach, T.J., Tatman, D.A. and Stern, D.B.** (2003) CSP41a, a multifunctional RNA-binding protein, initiates mRNA turnover in tobacco chloroplasts. *Plant J*, **36**, 842-852.
- Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J.** (1996) Engineered GFP as a vital reporter in plants. *Current Biology*, **6**, 325-330.
- Christensen, A.C., Lyznik, A., Mohammed, S., Elowsky, C.G., Elo, A., Yule, R. and Mackenzie, S.A.** (2005) Dual-domain, dual-targeting organellar protein presequences in *Arabidopsis* can use non-AUG start codons. *Plant Cell*, **17**, 2805-2816.
- Church, G.M. and Gilbert, W.** (1984) Genomic sequencing. *Proc Natl Acad Sci USA*, **81**, 1991-1995.
- Davis, B.H.** (1976) Tetraterpens or Carotenoids. In *Chemistry and biochemistry of plant pigments* (Goodwin, T.W., ed). London: Academic Press Inc., pp. 486-505.
- Deng, X. and Gruissem, W.** (1988) Constitutive transcription and regulation of gene expression in non-photosynthetic plastids of higher plants. *EMBO J*, **7**, 3301-3308.
- Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Cattolico, L., Falentin, C., Giancola, S., Renard, M., Budar, F., Small, I., Caboche, M., Delourme, R. and Bendahmane, A.** (2003) Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep*, **4**, 588-594.
- 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.
- Drager, R.G., Higgs, D.C., Kindle, K.L. and Stern, D.B.** (1999) 5' to 3' exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. *Plant J*, **19**, 521-531.
- Emanuelsson, O., Nielsen, H. and von Heijne, G.** (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, **8**, 978-984.
- 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.

- 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.
- Gutierrez-Marcos, J.F., Dal Pra, M., Giulini, A., Costa, L.M., Gavazzi, G., Cordelier, S., Sellam, O., Tatout, C., Paul, W., Perez, P., Dickinson, H.G. and Consonni, G. (2007) Empty pericarp4 Encodes a Mitochondrion-Targeted Pentatricopeptide Repeat Protein Necessary for Seed Development and Plant Growth in Maize. *Plant Cell*, in press.
- Gutiérrez-Nava, M.L., Gillmor, C.S., Jiménez, L.F., Guevara-García, A. and León, P.** (2004) *CHLOROPLAST BIOGENESIS* genes act cell and noncell autonomously in early chloroplast development. *Plant Physiol*, **135**, 471-482.
- 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.
- Herrin, D.L. and Nickelsen, J.** (2004) Chloroplast RNA processing and stability. *Photosynth Res*, **82**, 301-314.
- Jenkins, B.D., Kulhanek, D.J. and Barkan, A.** (1997) Nuclear mutations that block group II RNA splicing in maize chloroplast reveal several intron classes with distinct requirements for splicing factors. *Plant Cell*, **9**, 283-296.
- Kim, J. and Mayfield, S.P.** (2002) The active site of the thioredoxin-like domain of chloroplast protein disulfide isomerase, RB60, catalyzes the redox-regulated binding of chloroplast poly(A)-binding protein, RB47, to the 5' untranslated region of *psbA* mRNA. *Plant Cell Physiol*, **43**, 1238-1243.
- Kim, J. and Mullet, J.E.** (2003) A mechanism for light-induced translation of the *rbcl* mRNA encoding the large subunit of ribulose-1,5-bisphosphate carboxylase in barley chloroplasts. *Plant Cell Physiol*, **44**, 491-499.
- Klein, R.R., Klein, P.E., Mullet, J.E., Minx, P., Rooney, W.L. and Schertz, K.F.** (2005) Fertility restorer locus *Rf1* [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor Appl Genet*, **111**, 994-1012.
- Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S. and Imamura, J.** (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosen radish. *Plant J*, **34**, 407-415.
- Kotera, E., Tasaka, M. and Shikanai, T.** (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, **433**, 326-330.
- Krause, K., Maier, R.M., Kofer, W., Krupinska, K. and Herrmann, R.G.** (2000) Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome. *Mol Gen Genet*, **263**, 1022-1030.
- Lennartz, K., Bossmann, S., Westhoff, P., Bechtold, N. and Meierhoff, K.** (2006) HCF153, a novel nuclear-encoded factor necessary during a post-translational step in biogenesis of the cytochrome *bf* complex. *Plant J*, **45**, 101-112.
- Lezhneva, L. and Meurer, J.** (2004) The nuclear factor HCF145 affects chloroplast *psaA-psaB-rps14* transcript abundance in *Arabidopsis thaliana*. *Plant J*, **38**, 740-753.
- Lopez-Juez, E. and Pyke, K.A.** (2005) Plastids unleashed: their development and their integration in plant development. *Int J Dev Biol*, **49**, 557-577.

- 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.
- Mancebo, R., Zhou, X., Shillinglaw, W., Henzel, W. and Macdonald, P.M.** (2001) BSF binds specifically to the bicoid mRNA 3' untranslated region and contributes to stabilization of bicoid mRNA. *Mol Cell Biol*, **21**, 3462-3471.
- Mansfield, S.G. and Briarty, L.** (1991) Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. *Can J Bot*, **69**, 461-476.
- 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.
- Millar, A.H., Whelan, J. and Small, I.** (2006) Recent surprises in protein targeting to mitochondria and plastids. *Curr Opin Plant Biol*, **9**, 610-615.
- Murakami, S., Kuehnle, K. and Stern, D.B.** (2005) A spontaneous tRNA suppressor of a mutation in the *Chlamydomonas reinhardtii* nuclear MCD1 gene required for stability of the chloroplast petD mRNA. *Nucleic Acids Res*, **33**, 3372-3380.
- 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.
- Nakamura, T., Schuster, G., Sugiura, M. and Sugita, M.** (2004) Chloroplast RNA-binding and pentatricopeptide repeat proteins. *Biochem Soc Trans*, **32**, 571-574.
- Nickelsen, J.** (2003) Chloroplast RNA-binding proteins. *Curr Genet*, **43**, 392-399.
- Oelmüller, R.** (1989) Photooxidative destruction of chloroplasts and its effect on nuclear gene expression and extraplastidic enzyme levels. *Photochem Photobiol*, **49**, 229-239.
- 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.
- 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.
- Rivals, E., Bruyere, C., Toffano-Nioche, C. and Lecharny, A.** (2006) Formation of the Arabidopsis pentatricopeptide repeat family. *Plant Physiol*, **141**, 825-839.
- Rochaix, J.D.** (1996) Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii*. *Plant Mol. Biol.*, **32**, 327-341.
- Rymarquis, L.A., Higgs, D.C. and Stern, D.B.** (2006) Nuclear suppressors define three factors that participate in both 5' and 3' end processing of mRNAs in *Chlamydomonas* chloroplasts. *Plant J*, **46**, 448-461.
- Salvador, M.L., Suay, L., Anthonisen, I.L. and Klein, U.** (2004) Changes in the 5'-untranslated region of the rbcL gene accelerate transcript degradation more than 50-fold in the chloroplast of *Chlamydomonas reinhardtii*. *Curr Genet*, **45**, 176-182.
- Schmitz-Linneweber, C., Williams-Carrier, R. and Barkan, A.** (2005) RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell*, **17**, 2791-2804.

- 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.
- Serino, G. and Maliga, P.** (1998) RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiol*, **117**, 1165-1170.
- Small, I., Peeters, N., Legeai, F. and Lurin, C.** (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, **4**, 1581-1590.
- 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.
- Stern, D.B. and Gruissem, W.** (1987) Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell*, **51**, 1145-1157.
- Stern, D.B., Hanson, M.R. and Barkan, A.** (2004) Genetics and genomics of chloroplast biogenesis: maize as a model system. *Trends Plant Sci*, **9**, 293-301.
- Stockel, J., Bennewitz, S., Hein, P. and Oelmuller, R.** (2006) The evolutionarily conserved tetratricopeptide repeat protein pale yellow green7 is required for photosystem I accumulation in Arabidopsis and copurifies with the complex. *Plant Physiol*, **141**, 870-878.
- Voelker, R. and Barkan, A.** (1995) Two nuclear mutations disrupt distinct pathways for targeting proteins to the chloroplast thylacoid. *EMBO J.*, **14**, 3905-3914.
- Vothknecht, U.C. and Westhoff, P.** (2001) Biogenesis and origin of thylakoid membranes. *Biochem. Biophys. Acta*, **1541**, 91-101.
- Walter, M., Kilian, J. and Kudla, J.** (2002) PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. *EMBO J.*, **21**, 6905-6914.
- Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y. and Liu, Y.G.** (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell*, **18**, 676-687.
- Williams, P.M. and Barkan, A.** (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J*, **36**, 675-686.
- Yamazaki, H., Tasaka, M. and Shikanai, T.** (2004) PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in Arabidopsis. *Plant J*, **38**, 152-163.
- Ye, G.N., Daniell, H. and Sanford, J.C.** (1990) Optimization of delivery of foreign DNA into higher-plant chloroplasts. *Plant Mol Biol*, **15**, 809-819.
- Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W.** (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol*, **136**, 2621-2632.
- Zou, Z., Eibl, C. and Koop, H.U.** (2003) The stem-loop region of the tobacco psbA 5'UTR is an important determinant of mRNA stability and translation efficiency. *Mol Genet Genomics*, **269**, 340-349.



**Figure 18 : Transcriptomes chloroplastiques A. du mutant *clb19* et B. de l'écotype *Col-0* en faible lumière**

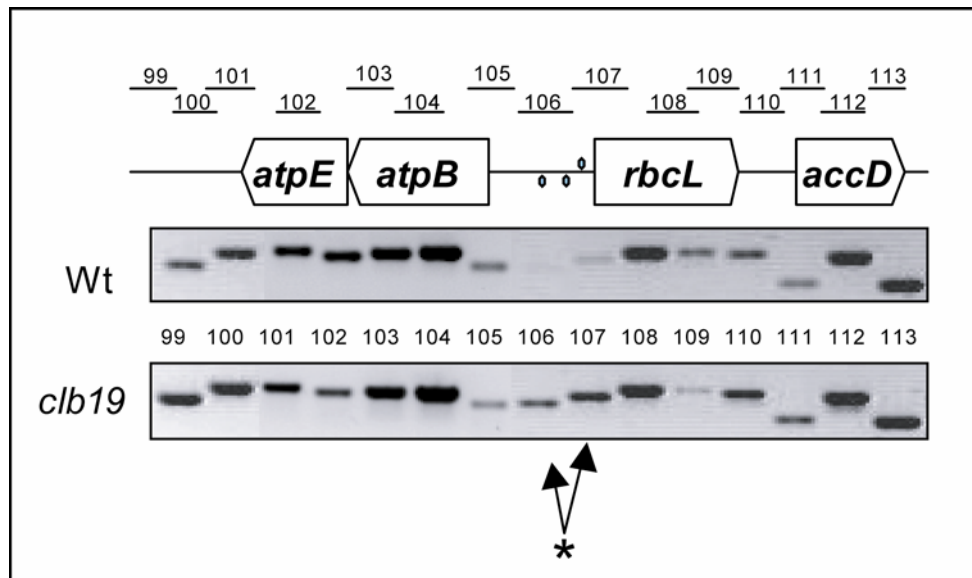
Lors de cette expérience les couples d'amorces 105 et 146 étaient défectueux (de nouveaux couples ont été commandés par la suite).

### II.1.5.2 Analyse du transcriptome chloroplastique de *clb19*

Pour comprendre le rôle de CLB19 dans la biogenèse des chloroplastes, nous avons effectué une analyse du transcriptome chloroplastique. Cette analyse a été effectuée sur des plantules après 15 et 30 jours de culture *in vitro* à basse luminosité. (Figure 18 a et b).

L'analyse a donné des résultats cohérents avec ceux de Patricia León pour la région comportant les gènes *atpB*, *rbcL* et *accD*. La région couverte par les couples cp106 et cp107 apparaît nettement plus présente chez le mutant que chez le sauvage. D'un point de vue général, le mutant semble présenter une légère accumulation des transcrits. Par exemple, à l'instar des couples cp106 et cp107, les couples cp43 et cp44 correspondant aux parties 5' et 3' de *rpoB* amplifient un fragment plus fréquemment chez les mutants *clb19* que chez le sauvage. L'opéron *rpl16-rps3-rpl22-rps19-rpl2-rpl23* pourrait également être affecté. D'après cette analyse, il semblerait que la région non codante située entre les gènes *atpB* et *rbcL* soit surreprésentée dans le mutant *clb19*. La Figure 19 montre la région comportant les gènes *atpE*, *atpB*, *rbcL* et *accD*. La séquence comprise entre les gènes *atpB* et *rbcL* comporte plusieurs promoteurs. Il est difficile de distinguer si ces amplifications sont le reflet de la présence d'un ARN plus long en amont de *rbcL* ou plutôt d'*atpB*.

Il a donc été nécessaire d'utiliser une autre méthode pour affiner ces résultats. La méthode de circularisation des ARN a été employée sans succès à plusieurs reprises. Notre choix s'est donc porté sur l'utilisation de la RACE PCR (Rapid Amplification of cDNA ends).



**Figure 19 : Amplification des transcrits dans la région de *atpB-rbcL-accD* chez *clb19* et chez le sauvage**

L'organisation du génome dans cette région et la position des fragments d'amplification sont indiquées dans la partie supérieure de la figure. La position des promoteurs est indiquée par un diamant, deux promoteurs pour *atpB* et un pour *rbcL* (Isono *et al.* 1997; Swiatecka-Hagenbruch *et al.* 2007).

La méthode de RACE-PCR permet d'identifier, en théorie, les extrémités 5' et 3' d'un transcrit. La Figure 20 (reprise de l'article) présente les résultats obtenus pour les transcrits de *rbcL*. L'extrémité 5' du transcrit de *rbcL* semble se situer nettement plus en amont du gène dans le mutant *clb19* que dans le sauvage. Les deux fragments ont été partiellement séquencés. Le fragment court correspond à un transcrit initié en -179. Il résulte très certainement d'une transcription initiée au niveau du promoteur de la PEP décrit -179 (Hanaoka *et al.* 2003). Le fragment long n'a été séquencé que sur 728 paires de bases (dans le sens 3'→ 5'). Il s'aligne bien à la séquence génomique connue. Des expériences ont été menées sur les transcrits d'*atpB* et d'*accD* mais n'ont pas permis pour le moment d'isoler des fins de transcrits claires.

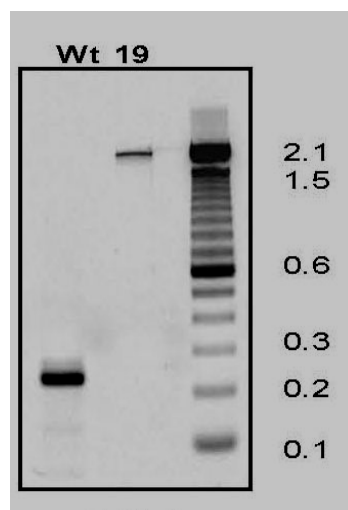


Figure 20 : RACE PCR sur le transcrit de *rbcL*

Le phénotype moléculaire des mutants *clb19* illustre la difficulté d'identifier dans certains cas le défaut primaire. La taille et la quantité de transcrits chloroplastiques ont été analysées par différentes techniques. Pour être certains d'avoir envisagé toutes les cibles possibles de CLB19, il restait à vérifier le niveau d'édition des différents transcrits du mutant. De nombreuses méthodes d'analyse de sites d'édition sont décrites dans la littérature, nous avons recherché les défauts par séquençage dans un premier temps. Cette méthode ne permet pas de réellement quantifier le taux d'édition d'un transcrit, mais elle est parfaitement adaptée pour effectuer le crible d'un grand nombre de sites en même temps. Les résultats pour les sites d'édition chloroplastiques connus sont indiqués dans le Tableau 7.

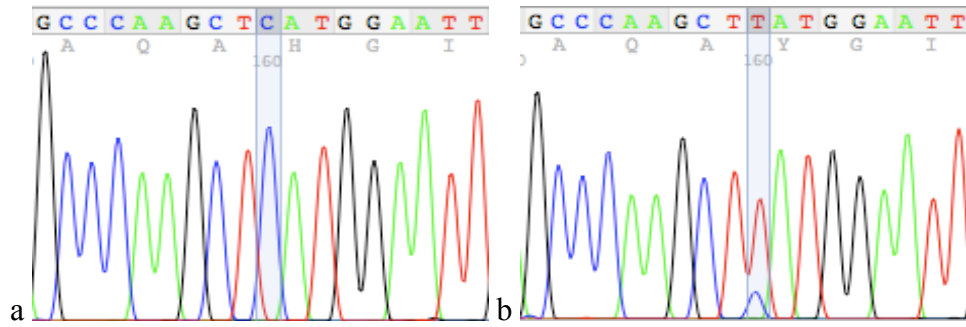


gène	site	codon modifié	codon	Change	<i>clb19</i>	Mutants PPR	Col-0
<i>accD</i>	1	265	uCg	S > L	36	93	
<i>atpF</i>	1	31	cCa	P > L	-	57	
<i>clpP1</i>	1	187	Cau	H > Y	0	66	75
<i>matK</i>	2	236	Cau	H > Y	88	64	91
	1	50	uCa	S > L	83	81	88
	2	156	cCa	P > L	95	69	
	3	196	Cau	H > Y	100	73	
	7	249	uCu	S > F	100	75	
<i>ndhB</i>	8	277	uCa	S > L	85	94	
	9	279	uCa	S > L	64	91	
	11	291	uCa	S > L	77	88	
	12	419	Cau	H > Y	72	78	
	10	494	cCa	P > L	73	85	
	1	1	aCg	T > M	16	26	33
	2	128	uCa	S > L	80	-	
<i>ndhD</i>	4	225	uCa	S > L	100	88	
	3	293	uCa	S > L	82	97	
	5	296	cCc	P > L	86	75	
<i>ndhF</i>	2	97	uCa	S > L	78	76	100
<i>ndhG</i>	1	17	uCc	S > F	-	38	
<i>petL</i>	2	2	cCu	P > L	37	97	
<i>psbE</i>	1	72	Ccu	P > S	95	69	97
<i>psbF</i>	1	26	uCu	S > F	99	70	100
	1	113	uCu	S > F	89	85	83
<i>rpoB</i>	3	184	uCa	S > L	96	87	83
	7	811	uCa	S > L	88	74	69
<i>rps14</i>	1	27	uCa	S > L	88	86	
	2	50	cCa	P > L	90	74	

Tableau 7 : Editotype des mutants *clb19*

Résultats exprimés en pourcentage de forme « T » par rapport à « C » (estimation obtenue par comparaison des pics C et T sur les chromatogrammes). Pour comparaison, la valeur observée pour l'écotype Col-0 et la moyenne des valeurs obtenues pour 8 autres mutants de protéines PPR cultivés dans les mêmes conditions que les mutants *clb19* ont été reportées dans le tableau. Le taux d'édition des sites AtpF-1 et NdhG-1 n'ont pu être déterminés pour le moment.

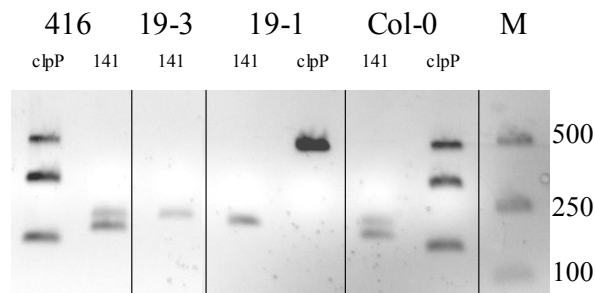
On observe que 3 sites semblent particulièrement affectés dans les mutants *clb19*, ces sites sont situés dans les gènes *clpP1*, *accD* et *petL*. Le transcrit *accD* est également touché au niveau quantitatif puisqu'il est surexprimé dans *clb19*. Il est possible que la diminution de l'édition de ce site dans *clb19* soit due à une limitation de la quantité du facteur reconnaissant le site en raison de la surexpression du transcrit. Le site ClpP1-1 apparaît comme étant non édité, le chromatogramme (Figure 21) ne montre en effet aucun pic correspondant à une trace d'uridine dans l'ARN.



**Figure 21 : Analyse de l'édition du transcrit de clpP1 par séquençage**

Extraits des chromatogrammes obtenus après séquençage de fragments de PCR amplifiant la région couvrant le site ClpP1-1, le site d'édition est surligné en bleu, en (a) *clb19* (b) Col-0.

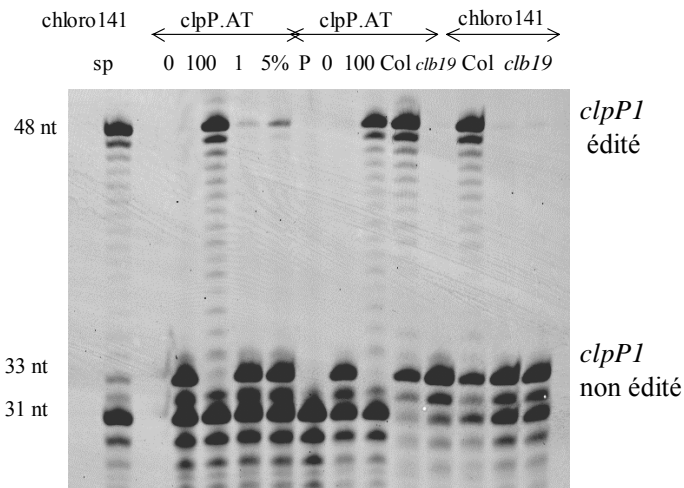
Dans le génome chloroplastique, l'édition de ClpP1-1 crée un site *HindIII*. Différents fragments de PCR couvrant ce site ont été digérés par *HindIII* chez deux mutants *clb19-1* et *clb19-3*, ainsi que chez deux témoins (Figure 22) Cette expérience a confirmé une absence totale d'édition de ClpP1-1.



**Figure 22 : Analyse de ClpP1-1 par digestion enzymatique**

Les fragments obtenus par digestion *HindIII* des produits de PCR couvrant le site ClpP1-1 sont analysés sur gel d'agarose 2%. Le site ClpP1-1 est couvert par deux couples de primers ClpP: utilisé également pour le séquençage, et Chloro141, provenant de la collection transcriptome. Deux témoins différents ont été utilisés, Col-0 et 416, un mutant affectant une autre protéine PPR (At3g16890) présentant un retard de croissance marqué.

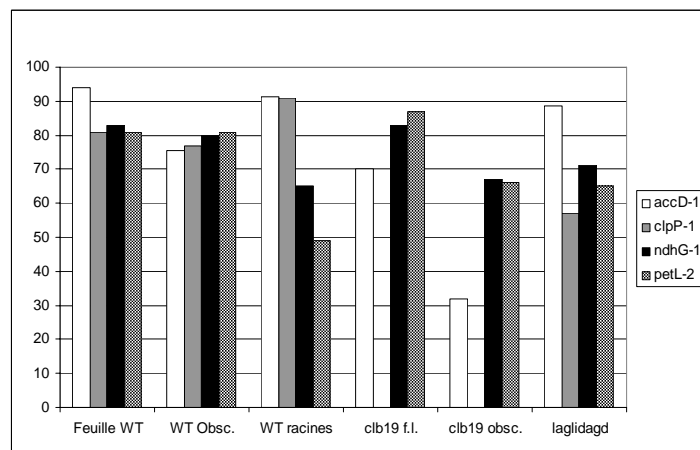
Pour définitivement statuer sur le niveau d'édition général chez *clb19*, Anne-Laure Chateigner-Boutin a effectué une quantification par extension d'amorce empoisonnée ("Poisoned Primer Extension", PPE). Cette méthode permet théoriquement de détecter une variation de quelques pourcents du taux d'édition. La Figure 22 présente les résultats bruts de cette méthode pour le site ClpP1-1, confirmant que ce site n'est pas du tout édité dans les plantes *clb19*.



**Figure 23 : Analyse par Poisoned Primer Extension (PPE) de l'édition de *clpP1-1***

Une première PCR réalisée avec des amorces entourant le site d'édition est réalisée. Elle sert de matrice pour une réaction d'extension à partir d'une amorce marquée à la fluorescence et choisie, celle-ci, pour s'hybrider près du site d'édition (au niveau du dernier C en 5' du site) en présence de 3 dNTP et d'un dideoxynucleotide (ddCTP). L'extension par une ADN polymérase est stoppée par incorporation du ddCTP, au niveau du site d'édition si la molécule n'est pas éditée ou au C suivant si la molécule est éditée. Les produits d'extension de tailles différentes sont séparés sur gel d'acrylamide. L'efficacité d'édition d'un site correspond au pourcentage de molécules éditées et est calculée en faisant le ratio du signal correspondant aux molécules éditées sur le signal total (AL Chateigner Boutin, ARC CoE PEB, Perth, AUSTRALIE).

Par cette méthode Anne-Laure a également analysé les autres sites d'édition semblant être affectés dans le mutant, Figure 23. PetL-2 n'est pas touché contrairement à ce que semblait indiquer le séquençage. AccD-1 présente une diminution d'édition chez les mutants cultivés en faible luminosité, accentuée lorsque le mutant est cultivé à l'obscurité. Les expériences de *northern blot* ont montré que le transcrit d'*accD* était accumulé chez *clb19*, il est donc possible que l'observation de la diminution d'édition soit due à un effet de saturation du facteur d'édition d'AccD-1.



**Figure 24 : Taux d'édition des sites ClpP1-1, AccD-1, NdhG-1 et PetL-2**

Le taux d'édition des sites AccD-1, ClpP1-1, NdhG-1 et PetL-2 ont été déterminés par PPE et indiqués dans ce graphique. Les mesures ont été faites à partir d'ARN de feuilles de plantes sauvages (feuilles en conditions normales, plantules cultivées à l'obscurité, racines cultivées à l'obscurité, du mutant *clb19-3* (plantules cultivées à l'obscurité ou en faible luminosité) et du mutant *ppr\_laglidagd* (cf. troisième partie des résultats).

### II.1.5.3 Interaction physique de CLB19 avec des transcrits chloroplastiques

Les analyses des transcrits chloroplastiques du mutant *clb19* ont montré de nombreux phénotypes moléculaires. Ceux-ci ne permettent pas de conclure sur les cibles potentielles de CLB19. Notre modèle d'action des PPR implique une interaction directe entre la protéine et sa cible, nous avons donc recherché si CLB19 présentait une affinité particulière pour un transcrit chloroplastique.

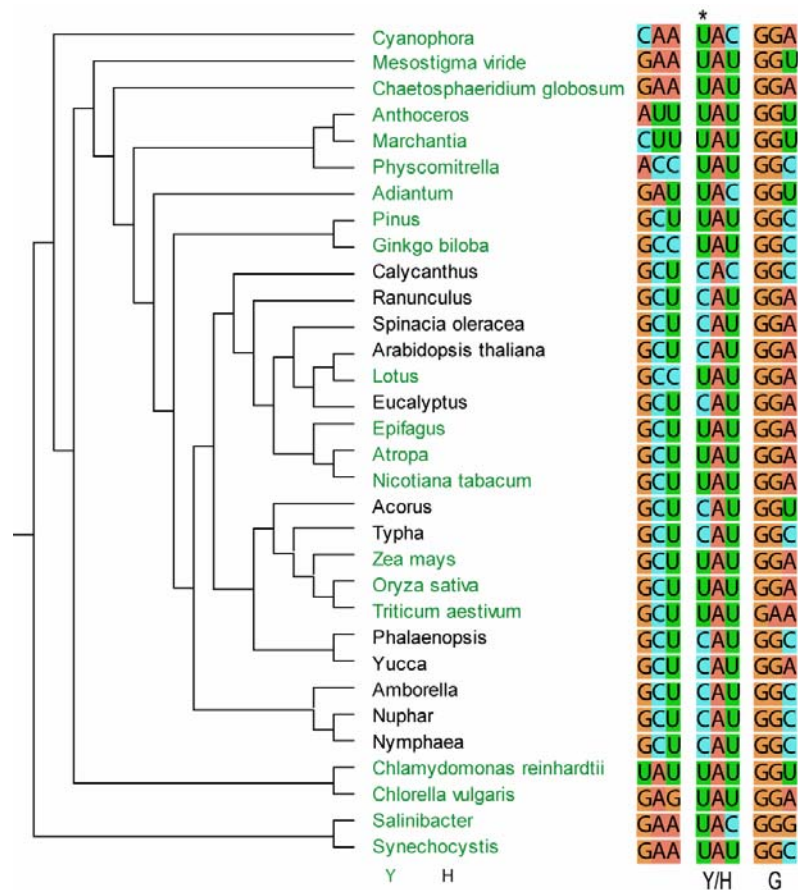
Deux approches sont menées en parallèle. Patricia León a commencé une recherche par retard sur gel. Christian Schmitz-Linneweber, puis Guillaume Barbier (deux post-doctorants du laboratoire) ont d'autre part essayé par RIP-CHIP à l'URGV.

Cette méthode, développée dans le laboratoire d'Alice Barkan à l'université de l'Oregon; permet l'identification des cibles ARN de protéines d'intérêt (Schmitz-Linneweber *et al.* 2005). La protéine étudiée est immunoprécipitée en conditions non-dénaturantes. Les ARN présents dans le culot précipité et ceux restant dans le surnageant sont marqués chimiquement, puis hybridés sur une puce à ADN. A l'URGV, nous avons utilisé une puce constituée des fragments d'amplification obtenus avec la collection d'amorces du crible présenté dans cette thèse. Elle présente donc une couverture complète du génome chloroplastique. Le résultat de l'hybridation sur puces des deux échantillons est analysé pour rechercher l'enrichissement d'ARN dans la fraction immunoprécipitée avec la protéine.

Plusieurs expériences ont placé le transcrit de *clpP* parmi les fragments les plus enrichis par l'immunoprécipitation. Mais plusieurs expériences témoins doivent être effectuées, et le traitement statistique des résultats doit être validé.

### II.1.5.4 Conservation du site ClpP1-1

Les séquences de *clpP1* de 32 organismes photosynthétiques et 2 bactéries ont été alignées pour observer la conservation de séquence du codon affecté par l'édition. Seul quatre codons différents ont été observés à cette position, UAC, UAU, CAC et CAU. Les séquences protéiques de toutes les ClpP disponibles dans la banque de séquence du NCBI ont également été alignées. A la position concernée par l'édition, seulement trois résidus sont présents, tyrosine, phénylalanine et histidine (Figure 25).



**Figure 25 : Conservation de la séquence correspondant au résidu tyrosine 187 de *clpP1* d'*Arabidopsis***

Les séquences génomiques de *clpP1* ont été extraites des séquences de 32 génomes chloroplastiques et de 2 génomes bactériens. Les séquences des ARN primaires ont été déduites, la séquence correspondant aux codons 186, 187 et 188 chez *Arabidopsis* a été alignée à ses homologues dans les 33 autres organismes. L'arbre représente schématiquement les liens phylogénétiques entre les organismes, il a été déterminé en utilisant la base de données taxonomiques du NCBI (Wheeler *et al.* 2000).

Chez *Arabidopsis* le codon CAU (histidine) est édité en première position pour donner UAU (Tyrosine), chez les autres organismes possédant le codon CAC ou CAU nous soupçonnons que la première position soit éditée pour donner les codons UAC ou UAU codant pour une tyrosine. Nous l'avons effectivement montré dans le cas de l'épinard (Figure 23), chez l'orchidée cette édition a été montrée récemment (Zeng *et al.* 2007). Si cela se confirme chez les autres organismes, la tyrosine à cette position serait conservée chez toutes les ClpP.

## II.1.6 Discussion

Le mutant *clb19* illustre les difficultés à analyser certains mutants présentant des phénotypes sévères. *Clb19* présente les phénotypes macroscopiques typiques d'un défaut dans la biogenèse des chloroplastes: plante *albinos* ou vert pâle, défaut de croissance en conditions normales, susceptibilité à la lumière, complémentation partielle des phénotypes par apport de

saccharose et réduction de l'intensité lumineuse. Pourtant les allèles étudiés sont bel et bien affectés dans un unique gène, *At1g05750*, codant pour une protéine PPR. Quel(s) est(sont) le(s) rôle(s) de CLB19? L'analyse des mutants n'a pas permis de trancher sur cette question, mais a apporté plusieurs indices.

#### II.1.6.1 CLB19 agit-elle sur le transcrit de *rbcL*? uniquement?

Les mutants *clb19* présentent un fort phénotype moléculaire pour plusieurs transcrits chloroplastiques, et en particulier le transcrit de *rbcL*. L'analyse par RACE-PCR du mutant *clb19* révèle effectivement la présence de transcrits plus longs que la normale chez le mutant, mais ces résultats ne semblent pas correspondre à un défaut de maturation. Le transcrit de *rbcL* est créé par la PEP à partir d'un promoteur situé 179 nucléotides avant l'AUG du gène (Isono *et al.* 1997). La forme 5' mature du transcrit est raccourcie d'une centaine de nucléotide (Hanaoka *et al.* 2003), cette forme courte du transcrit semble être la forme traduite (Reinbothe *et al.* 1993). Chez *Chlamydomonas*, il a également été montré qu'il existait des structures "tige-boucle" à l'extrémité 5' du transcrit de *rbcL*, et que celles-ci étaient impliquées dans la stabilisation du transcrit (Suay *et al.* 2005). Le fait que la forme courte soit la forme traduite pourrait expliquer la forte diminution de la quantité de la protéine RBCL.

Dans les mutants, les deux formes courtes (-179, et ~-70) semblent apparaître au cours du développement, expliquant une certaine réversion du phénotype macroscopique, il est probable que les transcrits plus longs soient le produit de transcription "illégitime" en amont du promoteur normal de *rbcL*. Il reste à élucider s'il s'agit d'une initiation unique, et de sous produits de dégradation, ou d'initiations multiples. Le fait que d'autres transcrits (*rpoB*, *rpl16-rps3-rpl22-rps19-rpl2-rpl23*) apparaissent comme légèrement affectés semblerait indiquer que dans les mutants *clb19* de nouveaux transcrits, absents dans les plantes sauvages, soient initiés en plusieurs points du génome. Ces transcrits ne sont pas observés dans le mutant *clb6*, indiquant ainsi qu'il ne s'agit pas d'un effet dû à l'état physiologique du mutant mais plutôt d'une conséquence de la disruption du gène *CLB19*. Pour le moment, il est difficile de comprendre pourquoi l'absence de protéines CLB19 provoque ce dysfonctionnement au niveau des transcrits *rbcL*. Les différentes hypothèses en fonction de ce qui est connu sur les protéines PPR sont : i) CLB19 est un facteur de maturation des ARN en 5' et/ou 3', ii) CLB19 est requis pour la transcription chloroplastique, iii) CLB19 est un facteur de maturation d'un gène de ménage chloroplastique dont le rôle affecte la transcription chloroplastique comme indiqué dans les point i) et ii). Le défaut d'édition de ClpP1-1 et les autres défauts moléculaires observés seraient plutôt en faveur de l'hypothèse iii).

### II.1.6.2 Edition de ClpP1-1

Le second phénotype moléculaire observé est l'abolition de l'édition de ClpP1-1. ClpP1 est une sous-unité catalytique du complexe protéolytique ClpP/R, faisant partie de la famille des Clp-protéases.

#### II.1.6.2.1 Les Clp-protéases

Les chloroplastes de plantes possèdent différents types de protéases. Les Clp-protéases sont des protéases de type serine, ATP dépendante. Ces protéases sont trouvées également chez les bactéries. Le séquençage de *Synechocystis PCC6803* a montré l'existence de 3 gènes *clpP* et un gène *clpR*. Cela pourrait indiquer que l'endosymbionte originel comportait déjà plusieurs gènes *clpP*, certains ayant pu être transférés ou perdus. Arabidopsis possède 5 gènes *clpP* et 4 gènes *clpR*, tous adressés vers le chloroplaste ou les mitochondries (Adam *et al.* 2001).

Le complexe protéolytique ClpP/R est constitué de deux anneaux heptamériques : l'anneau 1 est constitué d'une sous-unité ClpP1 et de six sous-unités ClpR1, R2, R3 ou R4. L'anneau 2 est constitué de sept sous-unités ClpP3, P4, P5 ou P6 (Sjogren *et al.* 2006). Ces protéases sont caractérisées par la présence d'une tétrade catalytique portée par les sous-unités ClpP. Le rôle des ClpR n'est pas réellement connu : bien qu'elles soient homologues aux ClpP, elles présentent des variations au niveau de la tétrade catalytique qui semblent altérer l'activité protéase (Figure 26). Il est donc à noter que la protéine ClpP1 n'est présente qu'en une seule copie au sein du complexe et qu'elle est la seule protéine à avoir une activité protéase au sein de l'anneau 1. La famille des protéines Clp chez Arabidopsis comporte également des homologues d'HSP100 : ClpB1 à B4, ClpC1 et C2, ClpD, ClpX1 à X3. Deux autres protéines ClpS1 et S2 sont spécifiques des plantes, elles présentent une similarité avec la partie N-terminale des HSP100, et sont trouvées associées au complexe (Peltier *et al.* 2004). Ces protéines Clp de type « chaperone » présenteraient au complexe ClpP/R les peptides devant être dégradés.

La liste des substrats des Clp-protéases n'a pas encore été établie. Elles sont généralement connues pour prendre une part importante dans le contrôle de la qualité de synthèse des protéines (Adam and Clarke 2002; Hengge and Bukau 2003). Chez *Chlamydomonas reinhardtii*, la dégradation du cytochrome *b6f* et de sous-unités du photosystème II est dépendante du complexe ClpP/R dans certaines conditions (Majeran *et al.* 2000; Majeran *et al.* 2001).

ClpP1 est la seule sous-unité du complexe ClpP/R codée par le génome chloroplastique.

Par ailleurs, il a été montré qu'elle est essentielle au développement (Huang *et al.* 1994; Shikanai *et al.* 2001). Cependant, certaines cultures cellulaires non-photosynthétiques de maïs peuvent perdre ce gène, sans incidence sur la survie des cellules. Il semblerait donc que l'aspect "essentiel" de ce gène soit intimement liés à la nécessité d'effectuer la photosynthèse (Cahoon *et al.* 2003).

#### **II.1.6.2.2 Conservation de la tyrosine 187**

Le mode de fonctionnement des protéases ClpP/R implique l'intervention de nombreuses protéines organisées en complexes et super-complexes, la structure tridimensionnelle des composants est donc essentielle. La combinaison du fait que ClpP1 soit encore codée par le génome chloroplastique et qu'elle ne soit présente qu'en une seule copie dans le complexe ClpP/R pourrait indiquer un rôle majeur au niveau de la régulation de l'assemblage du complexe ClpP/R.

Les protéines ClpP/R présentent une nette identité de séquence avec leurs homologues chez *E. coli*, la structure tridimensionnelle de cette dernière ayant été déterminée par cristallographie (Szyk and Maurizi 2006). La Figure 26 présente un alignement des séquences protéiques des ClpP/R d'*Arabidopsis* avec la séquence de 1TYF (la protéine d'*E. coli*) et deux protéines présentant d'importantes analogies structurelles avec 1TYF. Une flèche rose indique l'acide aminé modifié par l'édition de ClpP1-1.

La conservation de cet acide aminé est un indice fort pour indiquer un rôle essentiel au niveau de la protéine ClpP1 ou du complexe en entier. Cette tyrosine se situe à proximité d'un aspartate faisant partie de la tétrade catalytique et à la fin d'une structure en hélice alpha très conservée (Figure 26). Néanmoins, il est difficile de prévoir l'impact de la non édition et donc de la conservation de l'histidine dans la séquence protéique.

La tyrosine187 est située à proximité d'un acide aspartique faisant partie de la tétrade catalytique, il se pourrait que la présence d'une histidine à cette position ait une influence sur la reconnaissance ou la dégradation des substrats des complexes ClpP/R. Dans ce cas il devrait pouvoir être possible d'observer chez *clb19* une variation dans les produits de dégradation du complexe ClpP/R.

Pour tester l'effet de la non-édition de ClpP1-1, Benoît Derrien (IBPC, Paris) a modifié le gène *clpP1* de *Chlamydomonas* codant naturellement pour une tyrosine. Le gène endogène a été remplacé par une version modifiée codant pour une histidine. Les mutants observés ne présentent ni phénotype majeur ni altération de la composition du complexe ClpP/R. Chez *Chlamydomonas*, le complexe ClpP/R est nettement plus lourd que chez les plantes



supérieures, 540 kDa contre 350, cette différence est due à la présence d'une forme alternative de ClpP1. Deux protéines sont produites à partir du gène *clpP1*, l'une de 59 kDa (ClpP1<sub>H</sub>) et l'autre 25 kDa (ClpP1<sub>L</sub>) (Majeran *et al.* 2005). Il se peut que le rôle structural de la tyrosine ait été modifié chez *Chlamydomonas*, expliquant la différence de phénotype entre les deux espèces.

### Conclusion

CLB19 est une protéine PPR appartenant à la sous-famille E+, localisée dans les chloroplastes. Trois allèles mutants (insertion de T-DNA) indépendants présentent les mêmes phénotypes macroscopiques et moléculaires, deux de ces mutants ont été croisés entre eux et ont montré être alléliques (*clb19-1* et *clb19-2*). Les phénotypes observés sont donc la conséquence de la disruption de *clb19*. Les mutants *clb19* présentent de nombreux phénotypes moléculaires, notamment la présence d'au moins un ARN couvrant les ORF de *atpB*, *rbcL* et *accD* et l'abolition de l'édition du site d'édition ClpP1-1. Les expériences menées jusqu'à présent ne nous permettent pas de statuer si la mutation est pléiotropique, ou s'il existe une relation de causalité entre les différents phénotypes. La caractérisation des propriétés de reconnaissance de l'ARN, si CLB19 est réellement une protéine se liant à l'ARN, permettra de déterminer qu'elle est la fonction primaire de CLB19.

Le rôle du complexe ClpP/R chez les chloroplastes peut intervenir au niveau du contrôle qualité des protéines chloroplastiques. Si ce complexe est défaillant, il devrait être possible d'observer l'accumulation de ses substrats dans les mutants. D'un point de vue fonctionnel, cela impliquerait que dans les chloroplastes, des protéines mal repliées ou des complexes mal assemblés ne seraient plus éliminés. Le génome chloroplastique est transcrit par au moins deux polymérases, dont une produite *in organello*. Les transcrits observés dans le mutant pourraient être le fruit de polymérases défaillantes, initiant la transcription à d'autre point que les promoteurs ordinaires.

Les premières analyses protéomiques du mutant *clb19* n'ont pas permis de tirer de conclusion sur la présence, et le cas échéant l'action, des complexes ClpP/R. De nouvelles analyses du protéome chloroplastique de *clb19* par BN/PAGE permettront d'établir d'une part si les sous-unités du complexe sont présentes, et d'autre part s'il est possible d'observer l'accumulation de protéines qui ne peuvent pas être dégradées par le complexe ClpP/R dans les mutants.

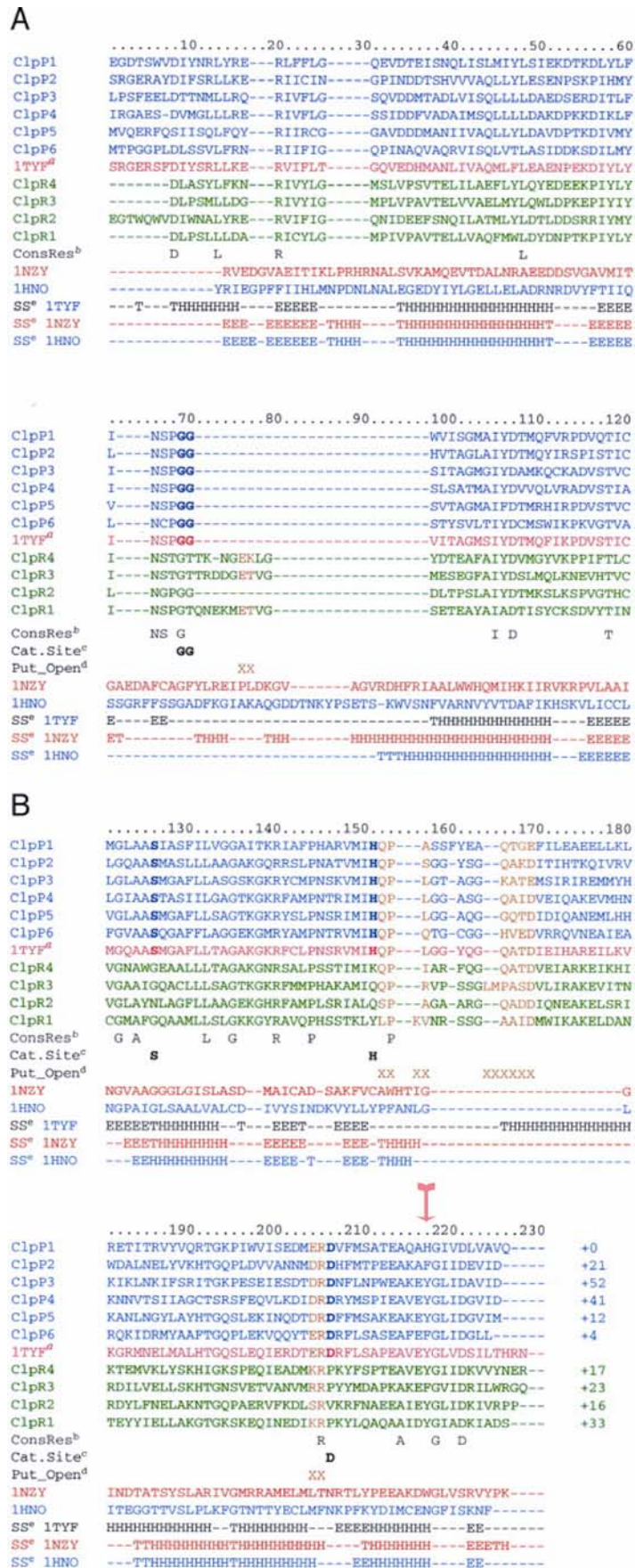


Figure 26 : Alignement des séquences protéiques des protéines ClpP/R avec les modèles 1TYF, 1NZY et 1HNO

## II.2 Identification de défauts moléculaires chez des mutants PPR à l'aide du crible

La collection de primers a été utilisée par Andéol Falcon de Longevialle (ARC CoE PEB, Perth, AUSTRALIE) pour analyser deux nouveaux mutants de protéines PPR. Dans les deux cas, ce crible a permis d'identifier le défaut moléculaire associé à la mutation.

### II.2.1 Etude du mutant PPR43<sup>1</sup> (Andéol Falcon de Longevialle *et al*, en préparation)

La protéine PPR43 est codée par le gène At1g74900, c'est une PPR de type "P" contenant 11 motifs PPR. Elle est prédite pour être localisée au niveau des mitochondries. Les défauts associés à ce gène ont été identifiés dans le cadre du projet Agrikola (Hilson *et al.* 2004). Le projet Agrikola a consisté en la réalisation d'une collection de vecteurs binaires permettant l'extinction par "RNAi" ciblant la plupart des gènes d'*Arabidopsis*. Plusieurs études tests ont été menées dans les différents laboratoires du consortium. Dans notre équipe, les plantes "RNAi" analysées ciblaient des gènes de la famille PPR. Les transformants portant la construction "RNAi" dirigée contre le gène At1g74900 ont manifesté un retard de croissance marqué corrélé avec la présence de la structure en épingle à cheveux produite par le vecteur RNAi. Une lignée provenant de la collection SAIL (Sail\_867F07) portant une insertion d'un ADN-T à l'extrémité 3' du gène At1g74900 a été identifiée et analysée. Cette lignée semblait présenter un phénotype "léthal" à l'état homozygote, 25% de la descendance d'une plante hétérozygote ne germant pas en terre. Une observation plus détaillée a montré qu'il existait deux populations de graines différentes. Environ 25% des graines issues d'une plante hétérozygote sont plus sombres et plus petites que des graines de type sauvage. Ces graines ont été mises à germer sur milieu ½ MS *in vitro*. Dans ces conditions, quelques graines parviennent à germer, et au bout de quelques semaines, quelques plantes naines ont pu être obtenues. Ces plantes ont produit des graines, mais ces dernières n'ont germé dans aucune des conditions testées. Le transcriptome mitochondrial de ces mutants (RNAi et T-DNA), dénommés *ppr43-RNAi* et *ppr43* a été analysé par la suite.

#### II.2.1.1 Analyse du mutant *ppr43*

Le crible de défauts moléculaires sur *ppr43* a révélé une absence totale de la forme mature

---

<sup>1</sup> Dénomination temporaire ne correspondant pas à la nomenclature en place pour l'identification des gènes et mutants chez *Arabidopsis*, le symbole PPR étant réservé pour "Peronospora Parasitica Resistant".

du transcrit de *nad1* au niveau mitochondrial, et l'absence de tout autre défaut. Le gène *nad1* code pour une sous-unité intramembranaire de la NADH : ubiquinone oxidoreductase (complexe I), *a priori* essentielle à l'assemblage du complexe en entier. Le gène est morcelé en 5 exons situés sur trois transcrits indépendants. 3 épissages en "trans" et 1 en "cis" sont donc nécessaires à l'assemblage du transcrit mature. Le mutant *ppr43* est spécifiquement altéré au niveau de l'épissage en trans réunissant les exons 1 et 2. L'analyse de *ppr43-RNAi* a montré une diminution de cet épissage de *nad1* chez les plantes très affectées au niveau macroscopique.

L'absence du complexe I dans le mutant *ppr43* a été montrée par la suite par BN/PAGE. Un test d'activité enzymatique a été effectué et a confirmé l'absence de complexe I fonctionnel.

### II.2.2 PPR\_LAGLIDADG<sup>2</sup> (Andéol Falcon de Longevialle *et al*, en préparation)

Le second mutant caractérisé par Andéol Falcon de Longevialle est le fruit d'une approche gène candidat. La protéine PPR\_LAGLIDADG est codée par le gène At2g15820, cette PPR comporte 6 motifs PPR (4P et 2S) et deux motifs LAGLIDADG. Ce motif est caractéristique des endonucléases de "homing" trouvées chez les intéines et les introns mobiles de groupe I et II, et les introns archéens (Stoddard 2005). Le motif LALIDADG joue trois rôles importants : il permet la i) structure tridimensionnelle du site actif, ii) l'homodimérisation des endonucléases comportant un seul motif et le repliement des endonucléases en comportant deux et iii) l'activité catalytique, par fixation de cations. D'autre part, ce motif peut également agir comme ARN maturase pour l'épissage de certains introns (Ho *et al.* 1997; Bassi *et al.* 2002; Bassi and Weeks 2003). Les prédictions *in silico* prévoient une localisation chloroplastique de la protéine PPR\_LAGLIDADG.

Le mutant *ppr\_laglidag* présente un phénotype *albinos* fort *in vitro* et une incapacité à pousser en terre. Le phénotype peut être partiellement complétement en culture *in vitro* à faible luminosité. Le transcriptome chloroplastique a été analysé et un défaut d'épissage du transcrit de *ycf3* a été observé, et là encore aucun autre défaut. Ycf3 est une protéine interagissant avec PsaA et PsaD pour permettre l'assemblage du photosystème I (Boudreau *et al.* 1997; Naver *et al.* 2001). La forme mature du transcrit de *ycf3* est obtenue après 2 épissages en "cis". Le mutant *ppr\_laglidag* est spécifiquement affecté dans l'épissage de l'intron 2.

---

<sup>2</sup> Dénomination temporaire

### II.2.3 Discussion

Le crible RT-PCR a permis d'identifier la cible moléculaire de deux nouvelles protéines PPR. Ces deux protéines PPR sont impliquées dans l'épissage d'introns. La première, PPR43, est essentielle au trans-épissage entre les exons 1 et 2 de *nad1* dans la mitochondrie. PPR43 ne présente que des motifs P, ce qui impliquerait une absence d'activité catalytique. Il est fort probable que PPR43 soit un facteur de spécificité recrutant un ou plusieurs partenaires et permettant ainsi le bon épissage de *nad1*. Il reste à déterminer la séquence cible de PPR43 ce qui nous aidera à comprendre à quelle étape de l'épissage elle intervient. Réciproquement, l'identification de partenaires protéiques portant une activité catalytique pourrait nous orienter vers la région reconnue par PPR43.

La seconde protéine PPR, PPR\_LAGLIDADG, est requise pour l'épissage en "cis" de l'intron 2 de *ycf3* dans les chloroplastes. Le rôle précis de PPR\_LAGLIDADG reste à déterminer; mais l'hypothèse la plus simple serait que la protéine fonctionne en dimère, les motifs PPR reconnaissant l'ARN de *ycf3*, et le motif LAGLIDADG agissant comme ARN maturase.

Le crible RT-PCR a permis d'identifier rapidement la cible moléculaire de ces deux protéines PPR, le fait qu'elles soient impliquées dans l'épissage a facilité l'identification. Néanmoins le crible a également permis de montrer en une expérience unique qu'aucun autre transcrit de l'organite concerné n'était affecté.

## II.3 Analyse des génomes cytoplasmiques

La nature de la collection d'amorces permet également d'effectuer une cartographie à basse résolution des génomes chloroplastiques et mitochondriaux. En effet, la couverture complète du génome chloroplastique ainsi que le chevauchement des zones amplifiées par chaque couple d'amorces permettent d'obtenir une vue d'ensemble des génomes cytoplasmiques. En théorie, l'utilisation de la collection de primers permet facilement d'identifier les zones des deux génomes présentant des réarrangements.

Ces résultats m'ont conduit à proposer un sujet de stage de Master M2 (Biologie Cellulaire et Physiologie, Université d'Evry) sur la comparaison des génomes mitochondriaux dans différents écotypes d'*Arabidopsis*. La partie suivante présente les résultats obtenus pendant ce stage par Chantal Chauvet. Les résultats obtenus dans le cadre d'une collaboration avec Vincent Zaegel et José Gualberto (IBPC, Strasbourg, Fr) sont également présentés.

### II.3.1 Comparaison des génomes mitochondriaux de différents écotypes d'*Arabidopsis thaliana*

En 2004, Heather McKhann *et al* ont analysé la diversité génétique (nucléaire) et morphologique de différentes accessions d'*Arabidopsis thaliana*. Cette analyse a permis d'établir plusieurs "Core collections" représentant au mieux ces variations (McKhann *et al.*). Il existe 6 collections imbriquées comportant 8, 16, 24, 32, 40 et 48 accessions. Les écotypes constituant la "core collection" 24, couvrent 96% de la variabilité génétique nucléaire et quasiment toute la variabilité morphologique (McKhann *et al.*).

La "core collection" 48 étant disponible au laboratoire nous avons entrepris d'analyser par PCR les génomes mitochondriaux de ces accessions, ne sachant pas *a priori* si la variabilité du génome mitochondrial était comparable à celle du génome nucléaire. Le Tableau 8 présente les accessions utilisées, et leur niveau d'analyse.

Core collection	Nom de l'écotype	Numéro d'identification de Versailles (AV)	ADN collecté	PCR1	PCR2
	Col-0	186 AV	oui	oui	oui
	Ler-1	213 AV	oui	oui	oui
	Ws	244 AV	oui	oui	oui
40	Kondara	190 AV	oui	oui	oui
48	Pi-0	040 AV	oui	oui	oui
8	Jea	025 AV	oui	oui	oui
8	Cvi-0	166 AV	oui	oui	oui
8	Oy-0	224 AV	oui	oui	oui
8	Shahdara	236 AV	oui	oui	oui
16	Ge-0	101 AV	oui	oui	non
40	Ri-0	160 AV	oui	oui	non
8	Bur-0	172 AV	oui	oui	non
16	Mt-0	094 AV	oui	non	non
24	Bl-1	042 AV	oui	non	non
24	Kn-0	070 AV	oui	non	non
24	Stw-0	092 AV	oui	non	non
32	Ta-0	056 AV	oui	non	non
32	Rubezhnoe-1	231 AV	oui	non	non
40	Sp-o	053 AV	oui	non	non
40	Sav-0	235 AV	oui	non	non
40	Yo-0	250 AV	oui	non	non
48	Pa-1	050 AV	oui	non	non
48	N6	262 AV	oui	non	non
8	Blh-1	180 AV	oui	non	non

**Tableau 8 : liste des accessions utilisés dans cette étude**

Les accessions sont classées selon le fait que 2 analyses PCR aient pu être réalisées, puis 1 PCR, et si nous avons juste pu collecter un échantillon d'ADN. La "core collection" d'appartenance est indiquée dans la première colonne. Les "core collections" couvrent respectivement, 84, 92, 96, 98, 100 et 100% de la variabilité génétique étudiée lors de leur constitution. Les accessions n'ayant pas pu être collectées ne sont pas montrées ici mais en annexe.

Parmi les douze écotypes qui ont été analysés au moins une fois par PCR, seulement trois absences d'amplification se situant dans la partie codante d'un gène ont été observées. Toutes les autres différences observées affectent des couples de primers ne ciblant pas une phase codante, ou chevauchant l'extrémité 5' ou 3' d'un gène.

Lors de la mise au point de la collection de primers des amplifications avaient été effectuées en utilisant comme matrice les ADN génomiques des écotypes Columbia (Col-0) et Wassilewskija (Ws). Au niveau mitochondrial, deux régions semblaient absentes de l'écotype Ws: une région d'au moins 1,6kb dans la région 3' non traduite d'*atp9*, et une région d'au moins 2kb dans la partie 5' non traduite d'*atp1*. La comparaison avec les analyses menées sur les ADNc de mutants montre que dans un fond génétique Col-0 une partie de ces régions donne lieu à une transcription, ce qui n'est pas le cas dans un fond génétique WS.

Ces deux zones varient également beaucoup dans les autres écotypes. Les accessions Ler, Shakhara, Ge-0 et Jea présentent le même profil d'amplification que Col-0, les accessions Pi-

0, Cvi-0, Bur-0 et Oy-0 présentent, elles, le même profil que celui observé chez Ws. Ces résultats seront discutés de nouveau au regard des résultats exposés dans l'article Zaegel *et al.* L'analyse complète des génomes mitochondriaux des écotypes est disponible en annexe.

Ces études préliminaires nous ont montré que l'approche pouvait donner des indications intéressantes sur la structure des génomes, elle ne permet cependant que de cibler une zone présentant une variation. Il est difficile de différencier un événement d'inversion d'une délétion : dans les cas d'évènements courts (moins de 500pb) ces deux phénomènes mèneront à un défaut d'amplification; pour des tailles plus grandes, une inversion mènera à deux défauts d'amplification séparés par des amplifications réussies. L'utilisation de logiciels d'analyse de photos de gels permet d'obtenir une estimation de la taille du fragment de PCR. Cette valeur n'est pas absolue, mais peut aider à distinguer des variations de taille de plus de 10% de la taille du fragment.

Un protocole d'analyse plus précis a été mis en place et utilisé pour obtenir une partie des résultats présentés dans l'article Zaegel *et al.* 2007 (disponible en annexe). Les aspects relatifs à ces expériences sont décrits ci-dessous.

### II.3.2 Analyse des génomes mitochondriaux de mutants *osb1*

Les génomes mitochondriaux de cinq plantes mutantes (4 *osb1-1*, 1 *osb1-2*) ont été analysés par PCR et comparés à une plante Col-0. Les résultats pour Col-0 et un mutant sont présentés dans la Figure 27.

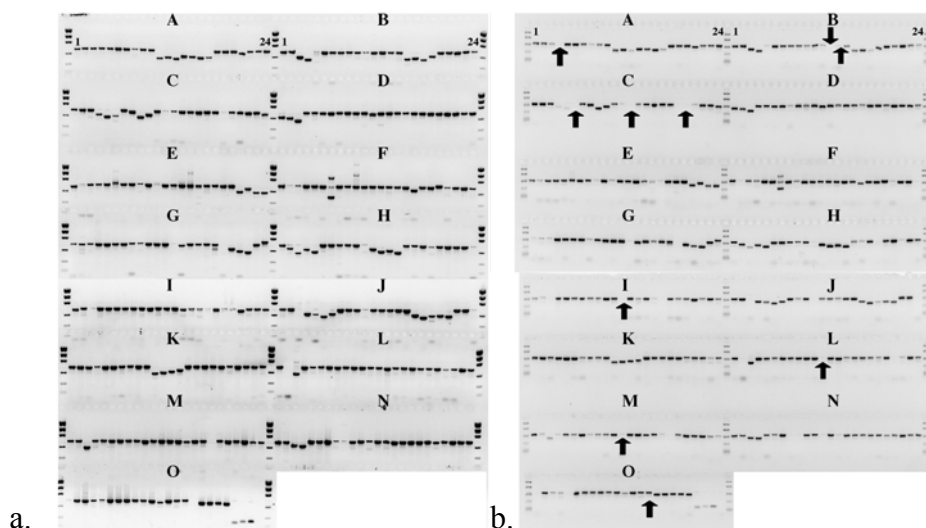


Figure 27 : Analyse par PCR du mutant *osb1-1* et de l'écotype Col-0

Photos des gels d'agarose sur lesquels ont été déposés les produits d'amplification obtenu avec les 359 paires de primers couvrant le génome mitochondrial, a. Col-0, b. *osb1-1*. Les absences d'amplification dans le mutant sont indiquées par les flèches noires. Les lettres A à O représentent les 15 groupes de 24 paires de primers, A : Mito1 à 24, B : Mito25 à 48 ....



Les gels d'agarose ont été analysés à l'aide du logiciel Genetools© (Syngene™, GB). Cette analyse a permis de déterminer la taille de chaque fragment d'amplification. Aucune variation de taille n'a pu être mise en évidence. Cependant, plusieurs absences d'amplification ont été observées parmi les 5 plantes, dans les régions des gènes *nad5*, *ccmFC*, *rpl5*, *atp1*, *cox1* et *cox2*. Des expériences de *Southern blot* ont confirmé l'existence de réarrangements dans ces régions du génome de *osb1-1* (figure 7 de l'article).

### II.3.3 Discussion

Ces deux analyses ont permis de montrer une utilisation alternative de la collection d'amorces. Cette application n'a pas de lien direct avec la fonction des protéines PPR. Cependant, elle a permis, en multipliant le nombre d'expériences, d'affiner notre connaissance sur le comportement des amorces au niveau technique (certains couples sont nettement moins efficaces que d'autre). Ces données nous permettent de mieux analyser les résultats obtenus sur matrice ADNc. D'autre part, elle a permis de mettre en évidence des différences au niveau génomique entre les écotypes Col-0 et Ws qui sont les deux fonds génétiques des mutants d'insertion.

L'analyse fine de la région 3' d'*atp9* réalisée à Strasbourg dans le cadre de l'analyse d'*osb1* a mis en évidence la présence de zones répétées pouvant donner lieu à des recombinaisons homologues. Les tests effectués sur différents écotypes suggèrent que ces régions aient été altérées, et peut-être aient perdu la capacité à recombiner.

### **Chapitre III Conclusion / Discussion**

En introduction à cette discussion, je dresserai le bilan des connaissances sur les protéines PPR grâce aux études publiées dans la littérature.

Les protéines PPR présentes dans la littérature avant le mois d'avril 2006, ont fait l'objet d'une revue publiée en janvier 2007. Cette revue avait pour but de dresser le tableau du rôle des protéines PPR dans les mitochondries de plantes supérieures. La partie suivante présente cette revue, puis une mise à jour sur les protéines PPR caractérisées depuis sa rédaction.

## REVIEW

# The multifarious roles of PPR proteins in plant mitochondrial gene expression

Charles Andrés<sup>a</sup>, Claire Lurin<sup>a</sup> and Ian D. Small<sup>b,\*</sup>

<sup>a</sup>Unité de Recherche en Génomique Végétale, CP 5708, 91057 EVRY cedex, France

<sup>b</sup>ARC Centre of Excellence in Plant Energy Biology, Molecular and Chemical Sciences Building (M310), University of Western Australia, 35 Stirling Highway, Crawley, Perth 6009, WA, Australia

## Correspondence

\*Corresponding author,  
e-mail: iansmall@cylleene.uwa.edu.au

Received 6 April 2006; revised  
15 May 2006

doi: 10.1111/j.1399-3054.2006.00766.x

With a few rare exceptions, genes encoding pentatricopeptide (PPR) proteins are present in all sequenced eukaryotic genomes but absent from prokaryotic and archaeal genomes. The family has greatly expanded in plants, to more than 400 genes in each species. So far, the evidence indicates that PPR proteins are generally involved in regulation of organelle genome expression, in other words they are eukaryotic proteins selected for the control of genomes of prokaryotic origin. PPR proteins are localised in both plastids and mitochondria, and appear to have similar roles in both cases. They have been implicated in almost all stages of gene expression, including messenger RNA (mRNA) transcription, splicing, processing, editing, translation and stability. The most probable hypothesis for explaining these diverse roles is that PPR proteins are sequence-specific RNA-binding adaptors capable of directing effector enzymes to defined sites on mRNAs. Much of the recent interest in the role of PPR proteins in mitochondria has been driven by the discovery that most cytoplasmic male sterility systems comprise fertility restorer genes that are members of this fascinating family.

## Introduction

The plant mitochondrial proteome can be estimated to be 2000–3000 gene products (Millar et al. 2005). However, the mitochondrial genome itself only codes for a tiny fraction of these proteins (30–40 in plants, Adams et al. 2002), including subunits of complexes involved in oxidative phosphorylation, cytochrome *c* biogenesis and ribosomes. All other mitochondrial proteins are nuclearly encoded, in particular most of the components of the transcription and translation apparatus. Mitochondrial gene expression is reliant on multiple posttranscriptional steps and any regulation of expression is likely to occur at this level (for review see Binder and Brennicke 2003). Very few of the proteins involved in these posttranscriptional processes are known, but in recent years pentatricopeptide (PPR) proteins have emerged as likely candidates in many

facets of organelle gene expression. The PPR family is a eukaryote-specific protein family particularly prevalent in higher plants. Mutants defective for individual PPR proteins have been known since 1995 with the analysis of the yeast *Pet309* gene, but the extent and coherence of the family was only recognised after the completion of the *Arabidopsis* genome sequencing project (Aubourg et al. 2000, Small and Peeters 2000). This review concentrates on what is known about PPR proteins in plant mitochondria, but will discuss results obtained on chloroplasts or mitochondria from other organisms when these suggest ways to fill in gaps in our knowledge (Table 1).

## The PPR family

PPR proteins are named after their eponymous 35 amino acid motifs (Small and Peeters 2000). PPR motifs are

*Abbreviations* – BSF, Bicoid mRNA stability factor; GFP, green fluorescent protein; LSFC, Leigh syndrome, French Canadian variant; mRNA, messenger RNA; NAD(P)H, nicotinamide adenine dinucleotide phosphate reduced; ORF, open reading frame; PPR, pentatricopeptide; QTL, quantitative trait locus; Rf factors, fertility restorers; TPR, tetratricopeptide; UTR, untranslated region.

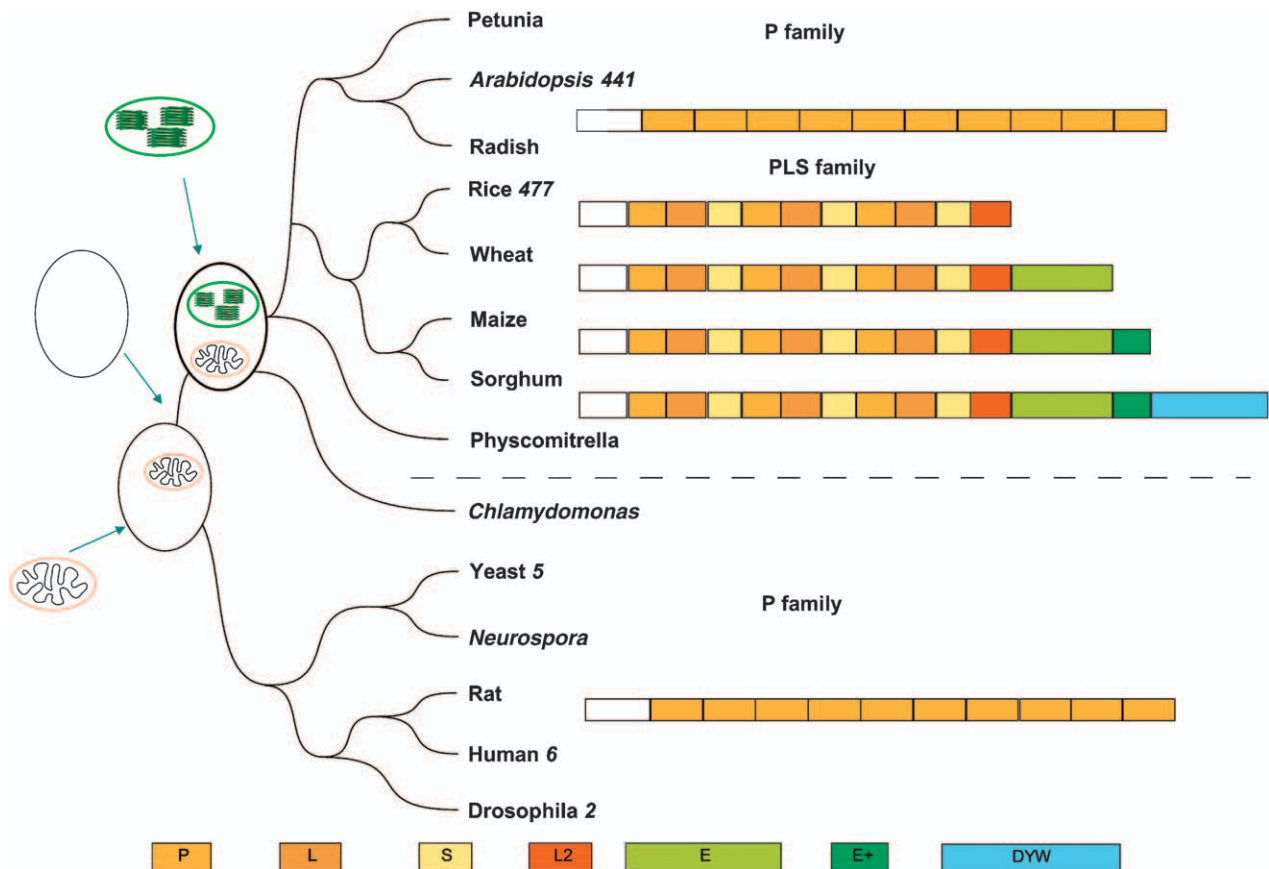
**Table 1.** A table summarising the pentatricopeptide (PPR) proteins described in this review. PPR proteins are involved in an impressively diverse array of molecular and physiological processes. Cp, chloroplast; Mt, mitochondria; CMS, cytoplasmic male sterility.

Gene	Organism	Physiological process	Molecular process	Target	Subclass	Localisation	Reference
<i>PGR3</i>	<i>Arabidopsis</i>	Photosynthesis	RNA stabilisation, translation	<i>petL operon</i>	P	Cp	Yamazaki et al. 2004
<i>CRR2</i>	<i>Arabidopsis</i>	Photosynthesis	RNA cleavage	<i>ndhB</i>	DYW	Cp	Hashimoto et al. 2003
<i>Hcf152</i>	<i>Arabidopsis</i>	Photosynthesis	Splicing, processing	<i>petB</i>	P	Cp	Meierhoff et al. 2003
<i>CRR4</i>	<i>Arabidopsis</i>	Photosynthesis	Editing	<i>ndhD</i>	E+	Cp	Kotera et al. 2005
<i>PTAC2</i>	<i>Arabidopsis</i>	Plastid biogenesis	Transcription		P	Cp	Pfalz et al. 2005
<i>MCA1</i>	<i>Chlamydomonas</i>	Photosynthesis	RNA stabilisation	<i>petA</i>	P	Cp	Lown et al. 2001
<i>PPR2</i>	Maize	Plastid biogenesis	Translation?		P	Cp	Williams and Barkan 2003
<i>CRP1</i>	Maize	Photosynthesis	Translation	<i>petA, psaC</i>	P	Cp	Fisk et al. 1999, Schmitz-Linneweber et al. 2005
<i>P67</i>	Radish	Unknown	RNA binding		P	Cp	Ikeda and Gray, 1999
<i>OsPPR1</i>	Rice	Plastid biogenesis	Unknown		E	Cp	Gothandam et al. 2005
<i>Bsf</i>	<i>Drosophila</i>	Unknown	RNA stabilisation	<i>bicoid 3'</i>	P	Cytoplasm	Mancebo et al. 2001
<i>LRPPRC, hLRP130</i>	Human	Respiration	RNA stabilisation	<i>coxI, coxIII</i>	P	Mt	Mili and Pinol-Roma 2003, Mootha et al. 2003, Xu et al. 2004
<i>Cya5</i>	<i>Neurospora</i>	Respiration	Translation	<i>coxI</i>	P	Mt	Coffin et al. 1997
<i>Rf</i>	Petunia	CMS	RNA degradation		P	Mt	Bentolila et al. 2002
<i>Rfo</i>	Radish	CMS	Translation	<i>orf138</i>	P	Mt	Brown et al. 2003, Desloire et al. 2003
<i>Rfk1</i>	Radish	CMS	Translation	<i>orf125</i>	P	Mt	Koizuka et al. 2003
<i>Rf-1, Rf1a</i>	Rice	CMS	RNA cleavage	<i>atp6-orf79</i>	P	Mt	Akagi et al. 2004, Komori et al. 2004, Wang et al. 2006
<i>Rf1b</i>	Rice	CMS	RNA degradation	<i>atp6-orf79</i>	P	Mt	Wang et al. 2006
<i>Rf1</i>	Sorghum	CMS	Unknown		E	Mt	Klein et al. 2005
<i>P63</i>	Wheat	Unknown	Transcription	<i>coxII</i>	P	Mt	Lahmy et al. 2000
<i>Aep3</i>	Yeast	Respiration	RNA stabilisation, processing	<i>atp6, atp8</i>	P	Mt	Ellis et al. 2004
<i>Pet309</i>	Yeast	Respiration	RNA stabilisation, translation	<i>coxI</i>	P	Mt	Krause et al. 2004, Manthey and McEwen 1995, Manthey et al. 1998
<i>GRP23</i>	<i>Arabidopsis</i>		Transcription		P	Nuclear	Ding et al. 2006

characteristically found as tandem arrays, from two to as many as 26 motifs per protein, with an average in plants of 12 (Lurin et al. 2004). The PPR motif is not one, but in fact three closely related motifs: the canonical PPR motif (P motif), common to all eukaryotes, and two variants specific to plants, the PPR-like S motif (for short) and the PPR-like L motif (for long). *Arabidopsis* has at least 442 PPR proteins, 242 of them with only P motifs, whereas the others are predominantly comprised of repeats of P-L-S triplets PLS subfamily (Lurin et al. 2004). The latter subfamily was previously termed the *Arabidopsis* plant combinatorial and modular protein family (Aubourg et al. 2000). The plant-specific PPR proteins usually contain additional C-terminal domains, respectively, E, E+ and DYW motifs found in 191, 145 and 87 *Arabidopsis* proteins (Lurin et al. 2004). These domains are like

‘Russian dolls’; an E motif always follows the final PPR motif, an E+ motif always follows an E motif and a DYW motif always follows an E+ motif (Fig. 1). Whereas PPR, E and E+ motifs are highly degenerate and can be difficult to recognise, DYW motifs contain highly conserved regions including some invariant amino acids. It has been suggested that this indicates that this domain may have catalytic activity (Aubourg et al. 2000, Lurin et al. 2004).

Almost half of the plant PPR proteins are predicted to be targeted to mitochondria and one-quarter to plastids (Lurin et al. 2004), and given the uncertainties in gene models and targeting predictions, these numbers could be underestimated. Exclusively organellar targeting has been confirmed by the limited experimental testing done so far (Lurin et al. 2004). Only one plant PPR protein has been shown so far to localise outside mitochondria or plastids



**Fig. 1.** The pentatricopeptide (PPR) family. Typical structures of proteins from each of the principal subfamilies and subgroups are shown. The structures are purely indicative, and the number and even order of repeats can vary in individual proteins. P family PPR proteins are found in all eukaryotes with mitochondrial genomes. PLS family PPR proteins are confined to land plants. The numbers of different PPR proteins encoded in each species are given when known (i.e. when the genome has been fully or almost fully sequenced).

(Ding et al. 2006). It is too early to say whether this recent example is the forerunner of a sizeable group of non-organellar PPR proteins or whether it will prove to be an extremely rare exception, but the bioinformatics predictions leave open the possibility of tens and maybe as many as a 100 PPR proteins with primary roles in the nucleus or cytosol. PPR proteins from other organisms are without exception mitochondrial proteins, although some may also have roles in other compartments (Mili and Pinol-Roma 2003). *Giardia*, which lacks true mitochondria and a mitochondrial genome, also lacks PPR proteins (McArthur et al. 2000).

PPR motifs appear to be related to tetratricopeptide (TPR) motifs (Small and Peeters 2000) known to mediate protein–protein interactions (Blatch and Lassle 1999). Like TPR motifs, PPR motifs are predicted to consist of a pair of antiparallel  $\alpha$  helices, and tandem PPR motifs probably form a helical binding surface. Characteristic differences between TPR and PPR motifs suggested that

PPR proteins might bind to RNA sequences (Small and Peeters 2000), a prediction that has been confirmed in several cases (Lahmy et al. 2000, Lurin et al. 2004, Mancebo et al. 2001, Meierhoff et al. 2003, Mili and Pinol-Roma 2003, Nakamura et al. 2003). PPR–RNA complexes can be highly specific, especially in vivo (e.g. the maize chloroplast). Chloroplast RNA processing 1 (CRP1) protein binds specifically to *petA* and *psaC* transcripts (Schmitz-Linneweber et al. 2005). It is not clear whether other RNA-binding proteins contribute to the specificity and strength of these interactions in vivo, as although PPR proteins are generally in complexes of relatively large size, their partner proteins have yet to be identified (Nakamura et al. 2004).

Much of our knowledge of the roles of PPR proteins comes from genetic evidence in a variety of organisms. Mutant phenotypes have revealed a wide range of defects in all stages of RNA metabolism, indicating surprisingly diverse functions for even very similar PPR proteins.

## Transcription

In most mitochondria, transcription is effected by phage T7-type RNA polymerases imported from the cytosol. Transcription initiates at more or less well-defined promoters and often gives rise to long polycistronic transcripts. In plants, multiple degenerate promoters are used and the specificity of transcription initiation appears to be highly relaxed (Holec et al. 2006, Kühn et al. 2005). The necessity for accessory transcription factors to achieve efficient and specific initiation on promoters is controversial and may vary between organisms. A PPR protein has been identified as implicated in transcription in wheat mitochondria (Ikeda and Gray 1999); p63 copurified with transcriptionally active fractions of mitochondrial proteins, improved yield of *coxII* transcripts in in vitro assays and was characterised as a DNA-binding protein, with affinity for the core promoter of *coxII*. A related PPR protein (pTAC2) has been shown to be associated with transcriptionally active chromosomes from plastids (Pfalz et al. 2005). More recently, the *Arabidopsis* protein GRP23 has been described as a nuclear PPR protein, which interacts with the nuclear RNA polymerase II (Ding et al. 2006). Again, the proposition was made that the PPR protein could be implicated in recognition of DNA *cis*-regulatory elements. At this point, it is worth highlighting that the mammalian mitochondrial RNA polymerases contain PPR motifs and that yeast mitochondrial RNA polymerase associates with protein complexes, including PPR proteins (Shadel 2004). Clearly an association between RNA polymerases (and thus transcription) and PPR proteins is a widespread and recurring theme. However, given the data on other PPR proteins, it is more likely that polymerase-associated PPRs are binding and stabilizing nascent RNA chains rather than the DNA substrate. As Shadel (2004) suggests, PPR proteins are likely to be involved in coupling transcription to downstream processing and translation.

## RNA stabilisation, degradation, processing

One of the simplest ways to control gene expression in mitochondria is by acting on the quantity of particular messenger RNAs (mRNAs), either by promoting RNA stability or by activating RNA degradation. Many PPR proteins have been described to be involved in these processes. In *Drosophila melanogaster*, Bicoid mRNA stability factor (BSF) is a PPR protein that binds to the 3' untranslated region of *bicoid* mRNA, and promotes the stabilisation of this mRNA (Mancebo et al. 2001). Given that BSF contains a strongly predicted mitochondrial targeting sequence and BSF tagged with GFP localizes to what look like mitochondria (Mancebo et al. 2001), it is likely that this PPR protein also has a mitochondrial

function. A human homologue of BSF, leucine-rich PPR-motif containing, LRPPRC (also known as LRP130) has been shown to bind to both nuclear and mitochondrial mRNA in vivo (Mili and Pinol-Roma 2003). Mutations in LRPPRC give rise to Leigh syndrome, French Canadian variant (LSFC) (Mootha et al. 2003). In LSFC patients, synthesis of COXI and COXIII subunits are specifically decreased, and the level of *cox1* and *cox3* transcripts is very much lower than usual. This indicates that LRPPRC acts as an mRNA stabilisation factor (Xu et al. 2004). In *Chlamydomonas*, MCA1 is a PPR protein essential for the accumulation of *petA* transcript in chloroplasts (Lown et al. 2001). It is not clear to what extent the effects of mutations described above are direct effects on mRNA stabilisation. Often a lack of processing or translation can lead to similar destabilisation of mRNAs. A typical example is the *Saccharomyces cerevisiae* Aep3 protein, a PPR protein involved in the processing of polycistronic mRNAs containing *atp6* and *atp8* (Ellis et al. 2004). In the absence of Aep3, the precursor RNA is not processed correctly and the level of *atp8/6* mRNAs drastically reduced. Similar entwined effects on processing and RNA accumulation have been seen in chloroplast PPR mutants; chlororespiratory reduction 2 (CRR2) is essential for the cleavage of *rps7-ndhB* RNAs (Hashimoto et al. 2003), proton gradient regulation 3, PGR3 is essential for the accumulation of *petL-petG-psaI* tricistronic transcripts (Yamazaki et al. 2004), and high chlorophyll fluorescence (HCF152) is required for the processing and accumulation of *psbB-psbT-psbH-petB-petD* RNAs (Meierhoff et al. 2003).

## The case of cytoplasmic male sterility (CMS) and fertility restoration

Cytoplasmic male sterility is a maternally inherited trait leading to pollen abortion and that thus converts normally hermaphroditic plants into females (reviewed in Budar et al. 2003). It can be selected for in some natural populations, which tend towards an equilibrium of male-fertile and male-sterile plants. In the vast majority of cases, the sterility-inducing factor is a novel open reading frame (ORF) in the mitochondrial genome that codes for a hydrophobic protein capable in some way of altering mitochondrial function leading to pollen abortion. Fertility can be restored by nuclearly encoded factors (known as fertility restorers or Rf factors) that suppress the effects of the mitochondrial ORF or its expression. Mitochondrial CMS ORFs and nuclear Rfs appear to be species-specific, but for virtually all the CMS/Rf systems studied so far, the Rf factor is a PPR gene. The single exception is the *Rf2* gene from maize, encoding an aldehyde dehydrogenase (Cui et al. 1996) and which probably restores fertility by an entirely different mechanism.

The first of the PPR restorers to be cloned was *Rf* from petunia, a gene needed to suppress expression of the mitochondrial CMS gene *pcf* (Bentolila et al. 2002). As in most of the CMS systems studied to date, the exact mode of action of *Rf* is not clear. The Petunia CMS-associated fused gene (PCF) protein is much reduced in *Rf* plants and *pcf* transcripts show differences in processing patterns and abundance. As discussed in the previous section, this could result from *Rf* action in RNA processing, RNA stability or translation. Whatever the mode of action of *Rf*, it appears to be highly specific to the *pcf* transcript whereas very similar petunia PPR genes have no effect. Almost exactly comparable CMS/*Rf* systems have been subsequently characterised in several plant species. Radish *Rfo* is essential to suppress the expression of mitochondrial *orf138* (Brown et al. 2003, Desloire et al. 2003), and the almost identical radish *Rfk1* suppresses expression of *orf125* (Koizuka et al. 2003).

The most intensively studied CMS/restorer systems in recent times are undoubtedly those found in rice, given the considerable potential agricultural interest for CMS in this species. Rice *Rf* genes have been cloned by four different groups (Akagi et al. 2004, Kazama and Toriyama 2003, Komori et al. 2004, Wang et al. 2006). The male sterility in this case is the because of the product of *orf79*, just downstream of one of the two copies of the *atp6* gene. The product of this ORF is predicted to be a transmembrane protein, and causes male sterility when expressed in transgenic rice plants (Wang et al. 2006). In fertility-restored plants, mRNAs containing *atp6* and *orf79* are drastically reduced. The four groups have shown that the *Rf-1* locus contains a cluster of related PPR genes. Depending on the plant material being studied, only one or two of these related genes encodes the fertility restoration factor. *Rf-1a* promotes cleavage of the *atp6-orf79* transcript, whereas *Rf-1b* promotes its degradation. Either can restore fertility, and when both are present, *Rf-1a* is epistatic over *Rf-1b*.

All of these *Rf* genes encode classical PPR proteins containing between 11 and 17 P motifs, but a recently cloned restorer gene from sorghum shows a quite different structure (Klein et al. 2005). The sorghum *Rf1* gene encodes an E-class PPR with no relation to the *Rf* genes described above. The mode of action of sorghum *Rf1* is unknown.

## Editing

Amongst the different posttranscriptional processes, editing is one of the least understood. In plant mitochondria, editing occurs with high frequencies, as there are 441 sites in *Arabidopsis* (Giege and Brennicke 1999) and 491 in rice (Notsu et al. 2002). It consists of the replacement of a cytosine originally encoded at the

DNA level by a uracil in the mature RNA. Although the numbers of editing sites in plastids and mitochondria are very different (plastids contain at least 10 times fewer sites), most of the evidence implies that the mechanisms in the two organelles are extremely similar. In vitro editing systems suggest that the process proceeds by deamination of the target cytosine (Hirose and Sugiura 2001, Yu and Schuster 1995), and have permitted the identification of the region of mRNA essential for efficient edition (Takenaka et al. 2004). In plant organelles, this corresponds to the 20 nucleotides before the editing site, but often less than 10 after the editing site. As little obvious conservation of sequence is observed in these regions, it is likely that numerous *trans*-factors are necessary to specify editing sites. Recently, a PPR protein was identified as a good candidate specificity factor. CRR4 is essential for the edition of the cytosine present in the ACG initiator codon of the plastid *ndhD* transcript (Kotera et al. 2005). CRR4 is an E-class protein, with 12 PPR-related motifs. CRR4 is the first editing factor ever identified in plants, but is unlikely to encode the enzymatic activity responsible for the base alteration (Kotera et al. 2005, Shikanai 2006). RNA editing of the type found in plant organelles appears to be lacking from green algae and in fact the appearance of editing in the plant lineage approximately corresponds to the expansion of the PPR family in plants, which predates the appearance of mosses (Hattori et al. 2004). The fact that the first editing factor to be identified in plants is a member of the plant-specific PLS subfamily makes it attractive to imagine that many of the other members of this subfamily may also be involved in editing. Bentolila et al. (2005) have recently mapped a quantitative trait locus (QTL) implicated in RNA editing in *Arabidopsis* mitochondria. The region spanning the QTL contains two PPR genes belonging to the DYW class, but more proof is needed before concluding that they encode editing factors.

## Translation

The final posttranscriptional step in the process of gene expression is translation, and once more PPR proteins have been found to be involved. Historically, the first two PPR proteins to be described, Pet309p from yeast and its orthologue from *Neurospora* CYA5 (Coffin et al. 1997, Manthey and McEwen 1995, Manthey et al. 1998), have been since shown to be factors essential for translation of COX1 in mitochondria. Pet309p has been shown genetically to associate with the 5' untranslated leader of *cox1* mRNA (although biochemical evidence of this is lacking) and in addition is part of a protein complex attached to the mitochondrial inner membrane (Naithani et al. 2003). The same complex includes translational activators of other mRNAs (Krause et al. 2004, Naithani et al. 2003).

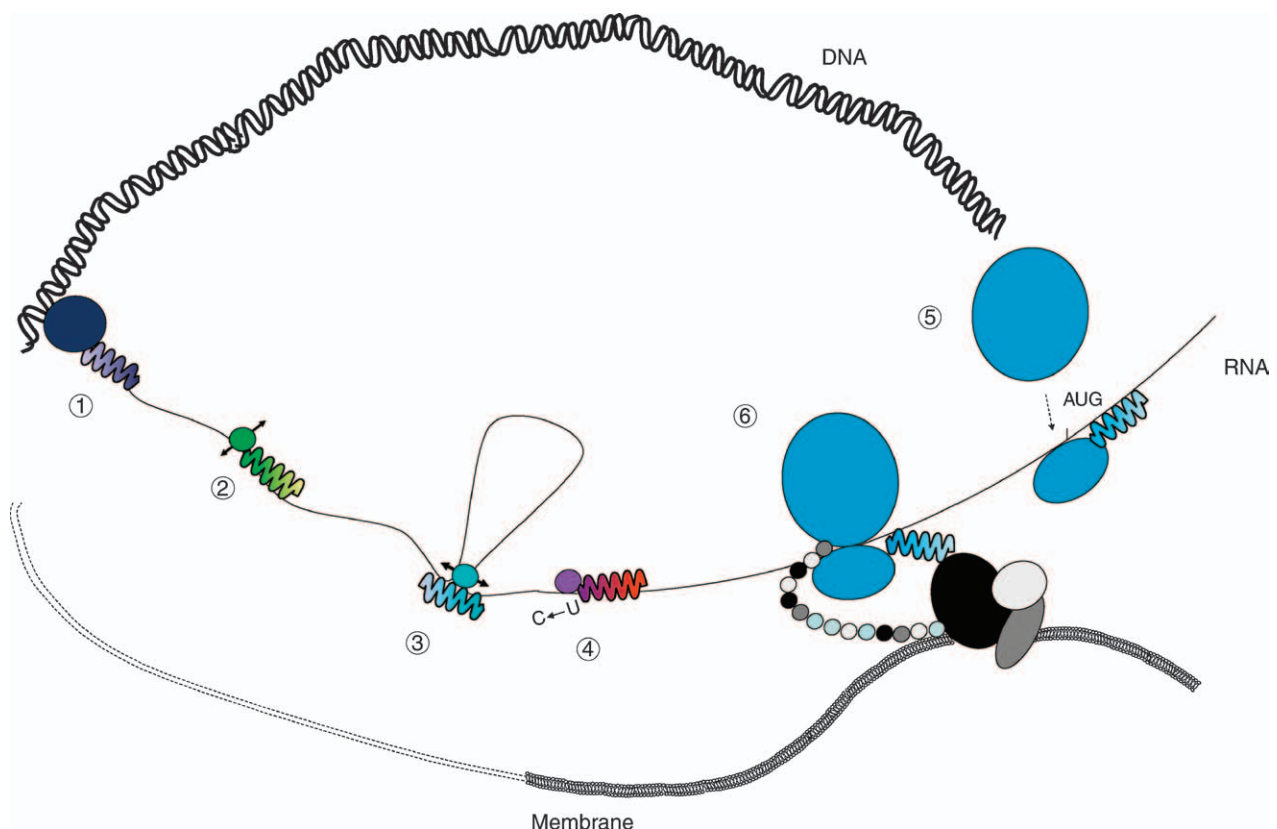
PPR proteins have been found to be associated with the small subunit of the mitochondrial ribosome in both yeast and mammals (Gavin et al. 2002, Koc and Spremulli 2003). Pet309p also associates with Nam1p, which in turn is known to bind to the mitochondrial RNA polymerase (Naithani et al. 2003). Thus is created a link between transcription and translation, and the general belief now is that transcription, translation and assembly into respiratory complexes are all linked at the inner mitochondrial membrane (Krause et al. 2004, Shadel 2004). PPR proteins such as Pet309p and Aep3p undoubtedly play a big role in creating and maintaining these links (Fig. 2).

In plants, our knowledge is much less advanced, but it seems highly likely that many PPR proteins have similar functions in plant organelles. Maize *ppr2* mutants are defective for plastid translation (Williams and Barkan 2003), but the molecular role of PPR2 is unknown and it may be required for expression of a component of the translation apparatus rather than play a role in translation

itself. More complete evidence is available for another maize PPR protein, CRP1. *Crp1* mutants are unable to translate *petA* and *psaC* mRNAs leading to defects in cytochrome *b<sub>6</sub>f* and the photosystem I core complexes (Fisk et al. 1999). CRP1 is associated with the 5' UTRs of *psaC* and *petA* mRNAs (Schmitz-Linneweber et al. 2005) and thus resembles Pet309p in many respects. Many of the PPR proteins described earlier in this review probably play similar roles but the data are lacking to show that they directly affect translation or bind to regions of their target mRNAs likely to be implicated in translational control.

### Concluding remarks and perspectives

PPR proteins are involved in many (at first sight) unrelated aspects of RNA processing within organelles. They can also be divided into multiple different subfamilies and subclasses depending on their motif structure. However, one would be mistaken to believe that there is a rigid



**Fig. 2.** Multifarious roles for pentatricopeptide (PPR) proteins in mitochondrial gene expression. PPR proteins have been found associated with every stage of gene expression imaginable between transcription and translation. 1: PPR proteins have been found associated with the mitochondrial RNA polymerase; 2: PPR proteins have been implicated in RNA cleavage; 3: PPR proteins have been implicated in splicing; 4: PPR proteins have been implicated in editing; 5: PPR proteins are strongly thought to play a role in translation initiation; 6: PPR proteins are thought to be associated with ribosomes and to in some cases to tether the translation machinery to the mitochondrial inner membrane to facilitate insertion of newly synthesized polypeptides into the correct complex.



relation between the structure of the PPR protein and the process it is involved in; for the moment the evidence suggests that closely similar PPR proteins often appear to have quite different physiological and even molecular functions. The easiest way to explain this seems to be that they share a single primary function—recognition of specific RNA sites. Binding of a PPR protein to its RNA target may have widely different and even antagonistic effects in different cases depending on the position of the target site in the RNA and the nature of other proteins bound to, or competing for, the same site. For example CRP1 is required for translation of its target transcripts, whereas the very similar PPR proteins Rfo and Rfk probably block translation of their target transcripts. In addition, as depicted in Fig. 2, transcription, posttranscriptional processing and translation are undoubtedly linked to some extent within organelles and thus multiple ‘knock-on’ effects are to be expected. The physiological role of each PPR is almost entirely dependent on the nature of the gene product encoded by its target transcript and thus entirely unpredictable at the current time.

Over the next few years, research is likely to concentrate on two areas in particular; (1) the identification of the RNA targets of PPR proteins and (2) the identification of the other proteins present in these RNA–protein complexes. The former will be facilitated by the new RIP-chip (RNA immunoprecipitation and chip hybridisation) approach pioneered by Schmitz-Linne-weber (2005), which shows considerable promise for mapping PPR binding sites. Similar immunoprecipitation techniques coupled with proteomics should go some way towards solving the second problem too. Identifying the other proteins complexed with PPR proteins should help characterise the effector enzymes (nucleases, deaminases) involved in different stages of posttranscriptional processing. Finally, an open question that needs to be solved is whether any of these PPR proteins actually regulate the expression of their target genes in response to cellular requirements, or are they just an unavoidable requirement? Many PPR proteins are absolutely essential for plant embryogenesis and development (Cushing et al. 2005, Lurin et al. 2004) so cannot be lost, but it is tempting to imagine that they play more subtle roles in optimising expression of organellar genes, for example by being implicated in control by epistasy of synthesis (Wostrikoff et al. 2004). Genetic approaches are probably best for detecting and unravelling functions in signalling and feedback control, but they will be difficult to carry out successfully when the processes affected in the mutants are so important at the whole plant level. The best hope will be to concentrate on non-essential components of the organellar photosynthetic or respiratory machinery, for example the plastid nicotinamide adenine dinucleotide

phosphate reduced (NAD(P)H) dehydrogenase (Hashimoto et al. 2003, Kotera et al. 2005).

## References

- Adams KL, Qiu Y, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci USA* 99: 9905–9912
- Akagi H, Nakamura A, Yokozeki-Misono Y, Inagaki A, Takahashi H, Mori K, Fujimura T (2004) Positional cloning of the rice *rf-1* gene, a restorer of bt-type cytoplasmic male sterility that encodes a mitochondria-targeting ppr protein. *Theor Appl Genet* 108: 1449–1457
- Aubourg S, Boudet N, Kreis M, 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
- Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA* 99: 10887–10892
- Bentolila S, Chateigner-Boutin A, Hanson MR (2005) Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in *Arabidopsis*. *Plant Physiol* 139: 2006–2016
- Binder S, Brennicke A (2003) Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci* 358: 181–188; discussion 188–189
- Blatch GL, Lassle M (1999) The tetratricopeptide repeat: a structural motif mediating protein–protein interactions. *Bioessays* 21: 932–939
- Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS (2003) The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J* 35: 262–272
- Budar F, Touzet P, De Paepe R (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* 117: 3–16
- Coffin JW, Dhillon R, Ritzel RG, Nargang FE (1997) The *Neurospora crassa cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* Pet309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded *CoxI* protein. *Curr Genet* 32: 273–280
- Cui X, Wise RP, Schnable PS (1996) The *Rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* 272: 1334–1336

- Cushing DA, Forsthoefel NR, Gestaut DR, Vernon DM (2005) Arabidopsis *emb175* and other PPR knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* 221: 424–436
- Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small I, Caboche M, Delourme R, Bendahmane A (2003) Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep* 4: 588–594
- Ding Y, Liu N, Tang Z, Liu J, Yang W (2006) Arabidopsis glutamine-rich protein 23 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
- Ellis TP, Helfenbein KG, Tzagoloff A, Dieckmann CL (2004) Aep3p stabilizes the mitochondrial bicistronic mRNA encoding subunits 6 and 8 of the H<sup>+</sup>-translocating ATP synthase of *Saccharomyces cerevisiae*. *J Biol Chem* 279: 15728–15733
- Fisk DG, Walker MB, 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
- Gavin A, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A, Cruciat C, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier M, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141–147
- Giege P, Brennicke A (1999) RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci USA* 96: 15324–15329
- Hashimoto M, Endo T, Peltier G, Tasaka M, Shikanai T (2003) A nucleus-encoded factor, *CRR2*, is essential for the expression of chloroplast *ndhB* in Arabidopsis. *Plant J* 36: 541–549
- Hattori M, Hasebe M, Sugita M (2004) Identification and characterization of cDNAs encoding pentatricopeptide repeat proteins in the basal land plant, the moss *Physcomitrella patens*. *Gene* 343: 305–311
- Hirose T, 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
- Holec S, Lange H, Kuhn K, Alioua M, Borner T, Gagliardi D (2006) Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol Cell Biol* 26: 2869–2876
- Ikeda TM, Gray MW (1999) Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol Cell Biol* 19: 8113–8122
- Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. *FEBS Lett* 544: 99–102
- Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF (2005) Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor Appl Genet* 111: 994–1012
- Koc EC, Spemullil LL (2003) RNA-binding proteins of mammalian mitochondria. *Mitochondrion* 2: 277–291
- Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosenia radish. *Plant J* 34: 407–415
- Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N (2004) Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant J* 37: 315–325
- Kotera E, Tasaka M, Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433: 326–330
- Krause K, Lopes de Souza R, Roberts DGW, Dieckmann CL (2004) The mitochondrial message-specific mRNA protectors Cbp1 and Pet309 are associated in a high-molecular weight complex. *Mol Biol Cell* 15: 2674–2683
- Kühn K, Weihe A, Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. *Nucleic Acids Res* 33: 337–346
- Lahmy S, Barneche F, Derancourt J, Filipowicz W, Delseny M, Echeverria M (2000) A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett* 480: 255–260
- Lown FJ, Watson AT, Purton S (2001) Chlamydomonas nuclear mutants that fail to assemble respiratory or photosynthetic electron transfer complexes. *Biochem Soc Trans* 29: 452–455
- 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, Mireau H, Peeters N, Renou J, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089–2103
- Mancebo R, Zhou X, Shillinglaw W, Henzel W, Macdonald PM (2001) BSF binds specifically to the bicoid mRNA

- 3' untranslated region and contributes to stabilization of *bicoid* mRNA. *Mol Cell Biol* 21: 3462–3471
- Manthey GM, McEwen JE (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
- Manthey GM, Przybyla-Zawislak BD, McEwen JE (1998) The *Saccharomyces cerevisiae* Pet309 protein is embedded in the mitochondrial inner membrane. *Eur J Biochem* 255: 156–161
- McArthur AG, Morrison HG, Nixon JE, Passamaneck NQ, Kim U, Hinkle G, Crocker MK, Holder ME, Farr R, Reich CI, Olsen GE, Aley SB, Adam RD, Gillin FD, Sogin ML (2000) The giardia genome project database. *FEMS Microbiol Lett* 189: 271–273
- Meierhoff K, Felder S, Nakamura T, Bechtold N, 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
- Mili S, Pinol-Roma S (2003) Lrp130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. *Mol Cell Biol* 23: 4972–4982
- Millar AH, Heazlewood JL, Kristensen BK, Braun H, Moller IM (2005) The plant mitochondrial proteome. *Trends Plant Sci* 10: 36–43
- Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, Delmonte T, Villeneuve A, Sladek R, Xu F, Mitchell GA, Morin C, Mann M, Hudson TJ, Robinson B, Rioux JD, Lander ES (2003) Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci USA* 100: 605–610
- Naithani S, Saracco SA, Butler CA, Fox TD (2003) Interactions among *cox1*, *cox2*, and *cox3* mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of *Saccharomyces cerevisiae*. *Mol Biol Cell* 14: 324–333
- Nakamura T, Meierhoff K, Westhoff P, 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
- Nakamura T, Schuster G, Sugiura M, Sugita M (2004) Chloroplast RNA-binding and pentatricopeptide repeat proteins. *Biochem Soc Trans* 32: 571–574
- Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* 268: 434–445
- Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmüller R (2005) Pta2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18: 176–197
- Schmitz-Linneweber C, Williams-Carrier R, Barkan A (2005) RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell* 17: 2791–2804
- Shadel GS (2004) Coupling the mitochondrial transcription machinery to human disease. *Trends Genet* 20: 513–519
- Shikanai T (2006) RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci*
- Small ID, Peeters N (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25: 46–47
- Takenaka M, Neuwirt J, Brennicke A (2004) Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res* 32: 4137–4144
- Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu Y (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 18: 676–687
- Williams PM, Barkan A (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J* 36: 675–686
- Wostrikoff K, Girard-Bascou J, Wollman F, Choquet Y (2004) Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of *Chlamydomonas*. *EMBO J* 23: 2696–2705
- Xu F, Morin C, Mitchell G, Ackerley C, Robinson BH (2004) The role of the LRPPC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of *cox* (cytochrome c oxidase) I and *cox III* mRNA. *Biochem J* 382: 331–336
- Yamazaki H, Tasaka M, Shikanai T (2004) PPR motifs of the nucleus-encoded factor, *PGR3*, function in selective and distinct steps of chloroplast gene expression in Arabidopsis. *Plant J* 38: 152–163
- Yu W, 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

### III.2 Protéines PPR publiées depuis Juin 2006

Chez le maïs, deux nouvelles protéines PPR ont été analysées: EMP4 (emptypericarp4) et PPR4. La première est une protéine PPR contenant neuf motifs PPR-P et un motif PPR-S, elle est localisée dans la mitochondrie et présente une forte identité de séquence avec deux protéines PPR d'*Arabidopsis* et une de riz. L'analyse de mutants touchés dans ce gène par "puces" et *northern blot* a montré une diminution de la quantité de certains transcrits mitochondriaux codant pour des protéines ribosomales et d'un transcrit codant pour une protéine dont la fonction n'est pas totalement élucidée (*mttB*) (Gutierrez-Marcos *et al.* 2007).

Une approche de RIP-CHIP, a permis de mettre en évidence que, chez le maïs, CRP1 était associée à la partie 5' non traduite de deux ARN chloroplastiques, *petA* et *psaC*. Pour mieux comprendre le rôle des protéines PPR, Christian Schmitz-Linneweber a entrepris la caractérisation de plusieurs mutants de protéines PPR chez le maïs, c'est dans ce cadre qu'il a mis en évidence que le produit du gène *PPR4* était impliqué dans le trans-épissage de l'ARN de *rps12* dans le chloroplaste de maïs. La méthode du RIP-CHIP a permis de mettre en évidence une association entre PPR4 et l'intron 1 de *rps12* (Schmitz-Linneweber *et al.* 2006). PPR4 est une protéine PPR particulière car elle comporte 16 motifs PPR-P et un motif RRM N-terminal.

CRR4 est une protéine PPR appartenant à la sous-famille E qui a été montrée comme impliquée dans le contrôle de l'édition du transcrit de *ndhD* dans le chloroplaste d'*Arabidopsis* (Kotera *et al.* 2005). Une analyse des caractéristiques de cette protéine a été entreprise par Okuda *et al.* et a permis de mettre en évidence que CRR4 se fixait spécifiquement sur la région couvrant le site d'édition (25 bases en amont et 10 bases en aval), et que l'état de ce site n'influaient pas sur la fixation de CRR4 sur sa cible (Okuda *et al.* 2006).

Très récemment, Kobayashi *et al.* ont identifié une nouvelle protéine PPR de la sous-famille DYW, celle-ci pourrait jouer un rôle dans la régulation de la biosynthèse des isoprénoides chez *Arabidopsis* (Kobayashi *et al.* 2007). Les isoprénoides représentent un des plus grands groupes de métabolites, ils sont essentiels à de nombreux processus comme la photosynthèse, la respiration ou encore la défense contre les pathogènes. Chez les plantes tous les isoprénoides dérivent du même précurseur, l'isopentenyl diphosphate (IPP), produit à partir de deux voies différentes, la voie de l'acide mévalonique (MVA) dans le cytoplasme et la voie du 2-C-méthyl-D-érythritol 4-phosphate (MEP) dans le chloroplaste. Le mutant *loi1* a la particularité de présenter un seuil de sensibilité élevé à certaines drogues bloquant ces deux voies. Cette résistance accrue semble conférée par une accumulation de la HMGR (3-

Hydroxy-3-methylglutaryl coenzyme A réductase), une enzyme de la voie MVA. La protéine LO11 étant localisée dans les mitochondries, une action directe dans une des voies de synthèse des isoprénoïdes est peu probable; mais les auteurs suggèrent un défaut affectant l'expression de constituants de la chaîne respiratoire, donc indirectement l'ubiquinone qui est isoprénylée (Kobayashi *et al.* 2007). Le mutant serait donc affecté dans une voie rétrograde mitochondrie-noyau régulant l'expression de *HMGR* au niveau post-transcriptionnel.

Chez la mousse *Physcomitrella patens*, une protéine PPR est requise pour l'épissage du transcrit de *clpP1* et le clivage de la partie non codante entre le 3' de *clpP1* et le 5' de *rps12*. Le mutant présente une diminution de la quantité de protéines PsbA (D1), PsbO, AtpB, RbcL, ainsi que de RBCS. Par contre les protéines codées par le noyau LHCII et cp28 ne sont pas affectées (Hattori *et al.* 2007).

### III.2.1 Le cas particulier des Kinétoplastides

Les kinétoplastides sont des eucaryotes unicellulaires bactériophages ou parasites, le représentant le plus connu est *Trypanosoma brucei*, responsable de la maladie du sommeil. Ce groupe est considéré comme l'un des plus anciens parmi les eucaryotes, ayant divergé très tôt après l'apparition des mitochondries. Ces dernières possèdent quelques particularités spécifiques, comme l'organisation du génome mitochondrial en kinétoplastes, de large réseaux de molécules d'ADN circulaire entrelacées, et l'édition de son ARN par insertion ou délétion de bases (Uridine) (Lukes *et al.* 2005). Comme tous les eucaryotes, les kinétoplastides possèdent des protéines PPR, au nombre de 23 chez *Trypanosoma brucei* (Mingler *et al.* 2006). Le contrôle post-transcriptionnel de l'expression des gènes mitochondriaux est un processus extrêmement complexe chez les kinétoplastides, il est largement impliqué dans les étapes de contrôle du cycle de développement. Au regard de ce qui est connu sur les protéines PPR, il est légitime de suspecter un rôle à ce niveau. Pour vérifier cette hypothèse, Mingler *et al.* ont entrepris l'analyse de TbPPR1 une protéine PPR de *Trypanosoma brucei*. L'observation de trypanosomes transfectés par une construction RNAi anti-TbPPR1 montre une diminution de la quantité de certains ARN produits à partir des maxi-cercles, ainsi qu'une perte de la longue queue poly-A dans certains cas (Mingler *et al.* 2006). Une autre information importante sur les rôles possibles des PPR nous est donnée par les kinétoplastides. Maslov *et al.* ont identifié 3 protéines PPR dans un complexe ribonucléique inhabituel dénommé "45S SSU\*". Ce ribosome particulier ne comporte pas l'ARN 12S habituel, et une grande partie des protéines présentes en sont spécifiques. Parmi elles, trois protéines PPR. Leur rôle dans ce complexe est loin d'être identifié, mais il est probable

qu'elles soient impliquées dans la reconnaissance des ARN à traduire (Maslov *et al.* 2007).

Le tableau 9 présente toutes les protéines PPR caractérisées à ce jour dans la littérature, c'est à dire pour lesquelles la fonction physiologique et/ou moléculaire a été identifiée.

Gène	Organisme	processus physiologique	processus moléculaire	cible	PPR	Loc.	ref
<b>Rf</b>	Pétunia	CMS	dégradation de l'ARN	-	P	Mt	(Bentolila <i>et al.</i> 2002)
<b>Rfk1</b>	radis	CMS	Traduction	<i>orf125</i>	P	Mt	(Imai <i>et al.</i> 2003).
<b>Rfo</b>	radis	CMS	Traduction	<i>orf138</i>	P	Mt	(Brown <i>et al.</i> 2003; Desloire <i>et al.</i> 2003)
<b>Rflb</b>	Riz	CMS	dégradation de l'ARN	<i>atp6-orf79</i>	P	Mt	(Wang <i>et al.</i> 2006)
<b>Rf-1, Rfla</b>	Riz L.	CMS	Clivage de l'ARN	<i>atp6-orf79</i>	P	Mt	(Akagi <i>et al.</i> 2004; Komori <i>et al.</i> 2004; Wang <i>et al.</i> 2006)
<b>Rfl</b>	Sorgho	CMS	-	-	E	Mt	(Klein <i>et al.</i> 2005)
<b>GRP23</b>	Arabidopsis	Embryogenèse	Transcription	-	P	N	(Ding <i>et al.</i> 2006)
<b>LO11</b>	Arabidopsis	Synthèse des isoprénoïdes	-	-	DYW	Mt	(Kobayashi <i>et al.</i> 2007)
<b>CRR2</b>	Arabidopsis	Photosynthèse	Clivage de l'ARN	<i>ndhB</i>	DYW	Cp	(Hashimoto <i>et al.</i> 2003)
<b>CRR4</b>	Arabidopsis	Photosynthèse	Edition	<i>ndhD</i>	E	Cp	(Kotera <i>et al.</i> 2005)
<b>PGR3</b>	Arabidopsis	Photosynthèse	Stabilisation, traduction de l'ARN	<i>Opéron petL</i>	P	Cp	(Yamazaki <i>et al.</i> 2004)
<b>HCF152</b>	Arabidopsis	Photosynthèse	Epissage, maturation	<i>petB</i>	P	Cp	(Meierhoff <i>et al.</i> 2003)
<b>MCA1</b>	Chlamydomonas	Photosynthèse	Stabilisation de l'ARN	<i>petA</i>	P	Cp	(Lown <i>et al.</i> 2001)
<b>CRP1</b>	Mais	Photosynthèse	Traduction	<i>petA, psaC</i>	P	Cp	(Fisk <i>et al.</i> 1999; Schmitz-Linneweber <i>et al.</i> 2005).
<b>PPR4</b>	Mais	Biogènes des ribosomes	Epissage	<i>rps12</i>	P	Cp	(Schmitz-Linneweber <i>et al.</i> 2006).
<b>PPR531-11</b>	<i>Physcomitrella patens</i>	Biogenèse des plastes	Epissage, maturation	<i>clpP</i>	P	Cp	(Hattori <i>et al.</i> 2007).
<b>OsPPR1</b>	Riz	Biogenèse des plastes	-	-	E	Cp	(Gothandam <i>et al.</i> 2005).
<b>LRPPRC, hLRP130</b>	Humain	Respiration	Stabilisation de l'ARN	<i>coxI coxIII</i>	P	Mt	(Liu and McKeehan 2002; Tsuchiya <i>et al.</i> 2002; Mili and Pinol-Roma 2003; Mootha <i>et al.</i> 2003; Tsuchiya <i>et al.</i> 2004)
<b>Cya5</b>	<i>Neurospora crassa</i>	Respiration	Traduction	<i>coxI</i>	P	Mt	(Coffin <i>et al.</i> 1997)
<b>Pet309</b>	Levure	Respiration	Stabilisation, traduction de l'ARN	<i>coxI</i>	P	Mt	(Manthey and McEwen 1995; Manthey <i>et al.</i> 1998; Krause <i>et al.</i> 2004)
<b>Aep3</b>	Levure	Respiration	Stabilisation, maturation de l'ARN	<i>atp6 atp8</i>	P	Mt	(Ellis <i>et al.</i> 2004)
<b>EMP4</b>	Mais	Maturation des graines	-	-	P	Mt	(Gutierrez-Marcos <i>et al.</i> 2007).
<b>At1g53330</b>	Arabidopsis	-	-	-	P	Mt	(Kocabek <i>et al.</i> 2006).
<b>BSF</b>	Drosophile	-	Stabilisation de l'ARN	3' de BICOID	P	C	(Mancebo <i>et al.</i> 2001)
<b>P67</b>	radis	-	Fixation à l'ARN	-	P	Cp	(Lahmy <i>et al.</i> 2000)
<b>GUN1</b>	Arabidopsis	Signaux rétrogrades	-	-	P	Cp	(Koussevitzky <i>et al.</i> 2007).
<b>TbPPR1</b>	<i>Trypanosoma Brucei</i>	-	Transcription	-	P	Mt	(Mingler <i>et al.</i> 2006).
<b>P63</b>	blé	-	Transcription	<i>coxII</i>	P	Mt	(Ikeda and Gray 1999).

**Tableau 9 : Protéines PPR décrites dans la littérature au 1er Avril 2007**

Toutes les protéines PPR caractérisées dans la littérature sont présentées dans ce tableau. La localisation sub-cellulaire indiquée peut être le résultat de prédictions ou d'expériences de localisation, Cp : Chloroplaste, Mt : Mitochondrie, C : Cytoplasme, N : Noyau. Les protéines ayant été seulement montrées comme essentielles ou exprimées dans certains tissus ne sont pas indiquées.

### **III.3 A la recherche des défauts moléculaires présents chez les mutants de protéines PPR d'Arabidopsis.**

Le but de cette thèse était d'avancer dans la compréhension du rôle de la famille PPR par l'analyse de mutants de gènes de cette famille.

Deux étapes étaient nécessaires afin d'atteindre cet objectif. D'une part, constituer un échantillon représentatif de mutants de PPR chez Arabidopsis. D'autre part, développer une méthode de caractérisation des mutants collectés.

#### **III.3.1 Première étape: développer un outil de détection**

Une des étapes de cette thèse était l'établissement d'une méthode de crible de défauts de maturation des ARN des organites d'Arabidopsis. La première ambition de ce crible était de pouvoir observer tous les processus subits par les ARN entre leur production et leur dégradation. C'est-à-dire, la maturation des extrémités 5' et 3', le contrôle de la stabilité, l'édition, l'épissage ou encore le clivage des ARN. La seconde ambition de ce crible était d'être suffisamment "léger" à mettre en œuvre, pour pouvoir être utilisé sur un très grand nombre de mutants.

##### **III.3.1.1 La méthode vu par le manipulateur**

L'objectif était de développer un crible "haut-débit" de recherche de défauts de maturation des ARNm cytoplasmiques. Pour satisfaire à ce "cahier des charges", il fallait que le crible ne soit pas trop coûteux, qu'il soit reproductible et réalisable en un laps de temps relativement court. Le coût du crible reste très compétitif, chaque analyse ne consommant qu'une quarantaine d'unités de Taq polymérase classique et 4 µl d'amorces à une concentration de 0,5µM. La partie biologie moléculaire est très rapide à effectuer, et peut prendre moins de 24H entre la collecte des échantillons et l'obtention de la photographie des 2 gels (chloroplaste + mitochondrie, sauvage+mutant, etc.). La faible quantité d'ARN requise pour réaliser ces expériences (3µg suffisent pour effectuer 2 fois l'expérience complète, Cp+Mt) permet également d'effectuer le crible sur très peu de tissus. Ceci permet de le réaliser relativement tôt dans le développement, ou sur des mutants très affectés (utilisation de plantules). Les différentes expériences menées, par plusieurs opérateurs, ont indiqué une forte reproductibilité de la méthode.

Il y a tout de même "un petit bémol" à l'ergonomie du crible : la distribution des amorces dans les plaques PCR 384 puits. En effet, cette phase ne peut être effectuée par un robot, les



volumes étant trop faibles pour qu'il n'y ait pas de phénomènes d'évaporation pendant le processus de distribution. Elle doit être réalisée à la pipette monocanal, puits par puits. De manière assez surprenante, les personnes m'ayant aidé à préparer quelques plaques ne m'ont plus jamais proposé leur aide!

### III.3.1.2 Le crible vu par l'analyste

Si la réalisation des RT-PCR est rapide, leur analyse est beaucoup plus longue. Nous avons envisagé de mettre en place un système d'analyse informatique semi-automatique des photos de gels. Cela n'a pu aboutir pour différentes raisons. Il n'existait pas, et n'existe toujours pas, de logiciels de ce type. Ian Small a adapté un logiciel existant pour aider à l'analyse de gels de même taille dans les projets Agrikola et Orféome. En réalité cette version du logiciel ne satisfait pas à l'analyse des produits de PCR obtenus avec le crible (profils d'amplification complexes, possibilité de variation de taille, caractéristique semi-quantitative de la méthode). Pour répondre à nos attentes, il aurait nécessité de nombreux développements.

Le logiciel Genetools© utilisé dans le cadre de l'analyse des mutants *osb1* permet d'effectuer une analyse que l'on peut qualifier "d'assistée par ordinateur". Il présente la possibilité d'analyser plusieurs lignes par gels, chaque ligne comportant un grand nombre de puits (51 dans notre cas). Il permet aussi de "désigner" des bandes qui n'auraient pas été détectées automatiquement. Par contre, l'impossibilité de fixer des paramètres par défaut, tels que le mode de calcul de la taille ou de la quantité, est un frein important à son utilisation à "haut-débit", voir même "moyen-débit".

En fait, la complexité et la diversité des phénotypes observables fait qu'il est difficile de programmer un logiciel d'analyse totalement automatique. En effet, l'apparition de fragments de PCR est aisément décelable par une analyse automatique, ce qui est le cas avec un défaut d'épissage par exemple, mais cela ne représente qu'une partie des phénotypes possibles. Beaucoup de mutants vont affecter la maturation en 5' ou 3' d'un transcrit. Dans ce cas, le phénotype moléculaire consistera en une variation de ratio entre la partie codante et les parties non codantes. Le développement d'un tel outil ou l'amélioration de logiciels commerciaux sera tout de même nécessaire pour tirer le meilleur parti de ce crible. J'ai rencontré l'ingénieur de la société Ozyme (distributeur du logiciel Genetools), avec qui j'ai fait remonter au niveau des concepteurs du logiciel un certain nombre d'améliorations à apporter à ce logiciel. Ces modifications devraient améliorer son ergonomie dans le cadre d'une utilisation à "haut-débit".

### III.3.1.3 Sensibilité du crible

L'approche RT-PCR est en elle-même limitée au niveau de l'appréhension des variations quantitatives. J'ai constitué la collection de primers en tenant compte de ce point. Deux critères dans le choix des amorces devaient nous permettre de dépasser cette limite, i) les caractéristiques thermodynamiques des amorces, ii) leurs positions. Les amorces ont été choisies en respectant au maximum le cahier des charges d'amorces utilisables en PCR quantitative. Cet aspect est développé par Etienne Delannoy à Perth. Les amorces ont été dessinées de manière à ce que chaque gène soit couvert par : i) un couple d'amorces s'hybridant uniquement dans la phase codante, ii) des couples d'amorces assurant une couverture d'au moins 2000pb en amont et en aval du gène. Les structures de plusieurs transcrits mitochondriaux et chloroplastiques sont connues, cela m'a permis de confronter les résultats obtenus en RT-PCR avec les résultats théoriques. Nous observons systématiquement des amplifications plus faibles dans les régions 5' et 3' non traduites, et une amplification forte au cœur du gène.

Cette observation est pour moi une indication forte de la capacité intrinsèque du crible à détecter d'autres défauts de maturation que les seuls défauts d'épissage. La possibilité d'avoir accès à un certain niveau de quantification des produits de PCR semble être un point essentiel pour la réussite du crible. Pour cela, je pense que le crible devra être désormais effectué en incluant deux phases de PCR, avec deux nombres de cycles différents, afin d'appréhender de façon "grossière" l'abondance des transcrits.

Le mutant CRR2 pourra être utilisé pour valider ce changement. Ce mutant présente un défaut de clivage du transcrit de *ndhB* dans le chloroplaste. Ce défaut est observable par le crible, mais a nécessité plusieurs répétitions avec des dilutions de matrice d'ADNc différentes. La manifestation de ce défaut dans le crible est une amplification systématique du couple cp190 (couvrant le site de clivage) chez les mutants, alors que ce couple disparaît lors des dilutions d'ADNc de plantes sauvages. Ce mutant constitue un cas de figure certainement fréquent où une trace de la molécule non maturée est toujours présente chez le témoin de type "sauvage".

### III.3.1.4 Bilan de l'outil

L'analyse de mutants pour lesquels nous n'avions pas *a priori* sur le processus mis en jeu, n'a permis d'identifier pour l'heure que des défauts d'épissage. Est-ce le seul défaut identifiable par ce crible? Ou bien, un grand nombre de protéines PPR sont-ils des facteurs d'épissage?

Les données sur les mutants de protéines PPR publiées dans la littérature montrent que ce n'est pas le cas, les protéines PPR sont clairement impliquées dans de nombreux processus de maturation des ARN.

Est-il possible d'identifier d'autres défauts que ceux d'épissage? Tout d'abord il est évident que ces défauts sont les plus "simples" à identifier par ce crible, aucune analyse quantitative n'est requise pour identifier ce type de défaut. Il suffit d'une seule expérience pour détecter l'absence d'amplification de l'un des couples spécifiques des gènes contenant des introns. Chacune des amorces de ces couples est située à une extrémité du gène, les produits d'amplification pouvant être obtenus avec ces couples ne présentent donc pas d'ambiguïté.

L'efficacité de ce crible n'a pas encore été établi par l'identification d'un défaut de maturation, autre qu'un défaut d'épissage. Cependant, le travail effectué sur les mutants pour lesquels nous connaissons le défaut, ou avons de fortes présomptions, devrait aider à améliorer la sensibilité de ce crible.

D'autre part, les observations faites par le crible RT-PCR sont toujours cohérentes avec les structures connues des ARN cytoplasmiques (extrémités 5' ou 3', points de clivage). Ceci nous indique que le crible est capable de dresser la cartographie des transcriptomes chloroplastiques et mitochondriaux. Il reste à déterminer les meilleures conditions pour identifier les variations dans ces transcriptomes.

### **III.3.2 Seconde étape: collecter des mutants**

La collecte de lignées d'insertion pour des gènes de protéines PPR a permis d'isoler 24 lignées d'insertion, correspondant à 22 gènes différents. Parmi eux, deux ont été publiés par d'autres équipes au cours de ma thèse (*CRR2*, *PGR3*), un est décrit dans cette thèse (*CLB19*).

La recherche de défauts moléculaires a été initiée pour 9 gènes, mais elle reste à être approfondie. Le crible RT-PCR a participé à l'identification des défauts moléculaires de trois mutants. Deux, présentent un défaut d'épissage (*PPR43* et *PPR\_LAGLIDADG*), le troisième un profil complexe de défauts moléculaires (*CLB19*).

Au niveau de la collecte de l'échantillon de mutants plusieurs problèmes se sont présentés, parmi eux deux principaux.

La liste primaire représentait un bon échantillonnage de la famille PPR. Cependant, sur les 19 gènes pour lesquels nous avons isolés des plantes homozygotes, six codent pour des protéines sans localisation sub-cellulaire claire, et ces six gènes font partie des neuf par lesquels j'ai commencé les analyses. Les mutants associés présentent des phénotypes forts, nanisme et/ou fertilité extrêmement réduite. *A posteriori*, ces gènes ne représentaient pas les

meilleurs candidats à analyser, puisqu'il est probable que les produits de ces gènes ne soient pas transportés dans les organites.

Un deuxième biais se situe au niveau de l'organisation du plan de recherche. J'ai fait l'erreur de vouloir mener de front le développement du crible (d'un point de vue technique), et l'isolement de nombreuses lignées homozygotes pour l'insertion. J'ai systématiquement essayé de mener l'analyse d'une vingtaine de lignées simultanément, représentant à peu près 200 plantes. Cela est loin d'être infaisable, mais aurait nécessité une organisation différente, comme par exemple effectuer des semis décalés sur plusieurs semaines. L'erreur suivante, la plus gênante pour mon ego, a été de répéter plusieurs fois la précédente!

### III.4 Et les PPR dans tout cela?

Les protéines PPR forment une famille de protéines spécifiques aux eucaryotes. L'évolution de cette famille s'est déroulée de manière particulière chez les eucaryotes photosynthétiques terrestres. Alors que le nombre de protéines PPR oscille autour de la dizaine chez les métazoaires ou les algues vertes, les plantes supérieures en comportent plusieurs centaines. De la même manière, la structure de ces protéines a également divergé : lorsque les PPR humaines ou de levure ne présentent que quelques motifs tous apparentés, les PPR de plantes peuvent posséder une vingtaine de motifs PPR répartis en plusieurs sous-familles. Dernier point de divergence, l'apparition de trois motifs C-terminaux, non homologues aux motifs PPRs mais spécifiques de la famille PPR.

L'analyse des mutants touchés dans des gènes codant pour des protéines PPR permet de définir deux règles sur le rôle des protéines PPR, i) leur action se déroule dans les organites des eucaryotes, ii) cette action implique une interaction avec les ARN messagers produits dans ces organites. Chez les plantes supérieures, les protéines PPR sont divisées en plusieurs sous-familles en fonction de leurs structures. Il serait très tentant d'essayer d'assigner une fonction moléculaire à chaque sous-famille, mais les évidences obtenues de l'analyse de mutants montrent pour le moment qu'il n'existe pas de règle de ce type.

Le sous-groupe DYW illustre ce fait. Ce sous-groupe présente deux caractéristiques qui en font le parfait candidat pour contenir les enzymes responsables de l'édition dans les organites de plantes. Premièrement, la séquence du domaine DYW contient des résidus invariants rappelant le site actif de cytidine-déaminases. Deuxièmement, la distribution phylogénétique de ce domaine chez les chlorophytes est strictement corrélée à la présence d'édition dans les organites (V. Salone, communication personnelle). Pourtant, actuellement, la seule protéine PPR caractérisée appartenant à cette sous-famille est impliquée dans la coupure endonucléolytique du transcrit de *ndhB* dans le chloroplaste d'*Arabidopsis* (Hashimoto *et al.* 2003). Cela implique-t-il que le motif DYW ne soit pas le facteur d'édition attendu? Non, le rôle de CRR2 dans la maturation de *ndhB* n'est établi pour le moment que d'un point de vue génétique, on ne peut exclure qu'il existe une étape d'édition intermédiaire. Pour cela, nous avons vérifié l'édition du transcrit complet dans ce mutant et aucun défaut n'a été observé. Pour autant, il serait hasardeux d'écarter le modèle concernant le motif DYW sur les résultats concernant une seule protéine, par contre il devra être adapté, et tenir compte d'une probable bi-fonctionnalité des protéines portant ce motif. Il a été montré que le tabac avait conservé la capacité à éditer certains sites bien que son propre génome chloroplastique ne nécessite plus

cette édition (Tillich *et al.* 2006). Cette information tend à montrer que les facteurs responsables de l'édition dans les chloroplastes des plantes subissent une pression de sélection forte, les conservant même après la perte de leur cible. L'explication la plus simple à cette observation est que ces facteurs soient requis pour un ou plusieurs autres sites d'édition, ou une toute autre fonction. Dans ce cas, la disparition du site à éditer ne correspondrait qu'à une perte partielle des cibles du facteur d'édition, et ne diminuerait en rien la pression de sélection exercée sur ce facteur.

Pour le moment, deux protéines PPR sont connues pour être impliquées dans l'édition de transcrits des organites, CRR4 et CLB19, décrits dans cette thèse. Ces deux protéines appartiennent respectivement aux sous classes E et E<sup>+</sup>.

Ces deux motifs ne présentent pas d'homologie notable avec quelque autre motif protéique. La présence des motifs E, E et DYW dans les mêmes protéines pourrait indiquer qu'ils ont la capacité à interagir ensemble. Chez Arabidopsis, il existe un gène ne comprenant que le motif DYW, celui-ci pourrait être recruté par des protéines PPR E ou E<sup>+</sup> pour exécuter l'édition des ARN.

Gène	processus moléculaire	PPR
CRR2	Clivage de l'ARN	DYW
CRR4	Edition	E
CLB19	Edition	E+
PPR43	Epissage	P
PPR_LAGLIDADG	Epissage	P
Rf-1, Rf1a	Clivage de l'ARN	P
Rf1b	Dégradation de l'ARN	P
PPR4	Epissage	P
PPR531-11	Epissage, maturation	P
HCF152	Epissage, maturation	P
BSF	Stabilisation de l'ARN	P
LRPPRC	Stabilisation de l'ARN	P
MCA1	Stabilisation de l'ARN	P
Aep3	Stabilisation, maturation de l'ARN	P
Pet309	Stabilisation, traduction de l'ARN	P
PGR3	Stabilisation, traduction de l'ARN	P
Cya5	Traduction	P
Rfk1	Traduction	P
Rfo	Traduction	P
CRP1	Traduction	P
P63	Transcription	P

**Tableau 10 : Lien entre les sous-famille PPR et les processus moléculaires**

Dans la littérature 25 protéines PPR ont un rôle moléculaire identifié. La nature de ce

processus et la sous-famille de la protéine PPR sont indiquées dans le Tableau 10. Ce tableau illustre la variété de fonctions moléculaires sous le contrôle d'une protéine PPR, puisque tous les processus de maturation d'ARNm sont représentés.

A l'heure actuelle, 22 mutants de protéines PPR de la sous classe P ont été caractérisés, toutes les possibilités de maturations des ARN des organites ont été observées, exceptée l'édition. Une telle diversité d'action indique que le rôle principal de ces protéines PPR se situe au niveau de la reconnaissance de l'ARN. La différence entre le nombre de protéines de la famille P et des autres familles s'explique en partie par sa distribution chez tous les eucaryotes (5 protéines caractérisés) et son implication dans les phénomènes de CMS (4 protéines caractérisées).

L'analyse des protéines PPR-P caractérisées permet de proposer deux modèles d'action différents, i) les protéines PPR se fixent à l'ARN et empêchent la fixation d'autres facteurs, ii) les protéines PPR guident les enzymes nécessaires sur leurs sites d'action. La première hypothèse permet d'expliquer la diversité de processus sous le contrôle de protéines PPR et peut correspondre aux cas où une protéine PPR est montrée comme impliquée dans la régulation de la traduction, ou de la dégradation d'un ARN messager. Par contre ce modèle impliquerait que la disruption de ces protéines lève une inhibition, les phénotypes observés devraient alors correspondre à un "gain de fonction" comme la dégradation ou la coupure systématique d'un transcrit. La seconde hypothèse correspond pour le moment à toutes les observations faites chez les mutants. Mais certaines de ces protéines ne comportent visiblement aucun autre motif protéique que les PPR-P, cela impliquerait donc que les motifs PPR-P soient également impliqués dans le recrutement de partenaires protéiques. Dans les deux cas, il est nécessaire d'envisager que les motifs PPR en eux-mêmes soient capables d'effectuer la reconnaissance d'une séquence ARN, et d'interagir avec d'autres protéines.

La solution pour identifier le ou les rôles des motifs PPR viendra peut-être d'approches biochimiques. Les évidences génétiques prouvent l'implication des protéines PPR dans les mécanismes de maturation des ARN, mais sont limitées pour permettre l'élaboration d'un modèle d'action des protéines PPR. L'identification des séquences reconnues par les protéines PPR, ainsi que des résidus dans les motifs PPR impliqués dans l'interaction, devient nécessaire pour aller plus loin dans la compréhension de ces protéines. Pour l'heure, trois protéines PPR ont été montrées se fixant à l'ARN, et leurs cibles identifiées. L'une d'elle, PPR4, comporte également un motif RRM (RNA Recognition Motif). Pour le moment les rôles respectifs des deux types de motifs, PPR et RRM, n'est pas élucidé (Schmitz-Linneweber *et al.* 2006).

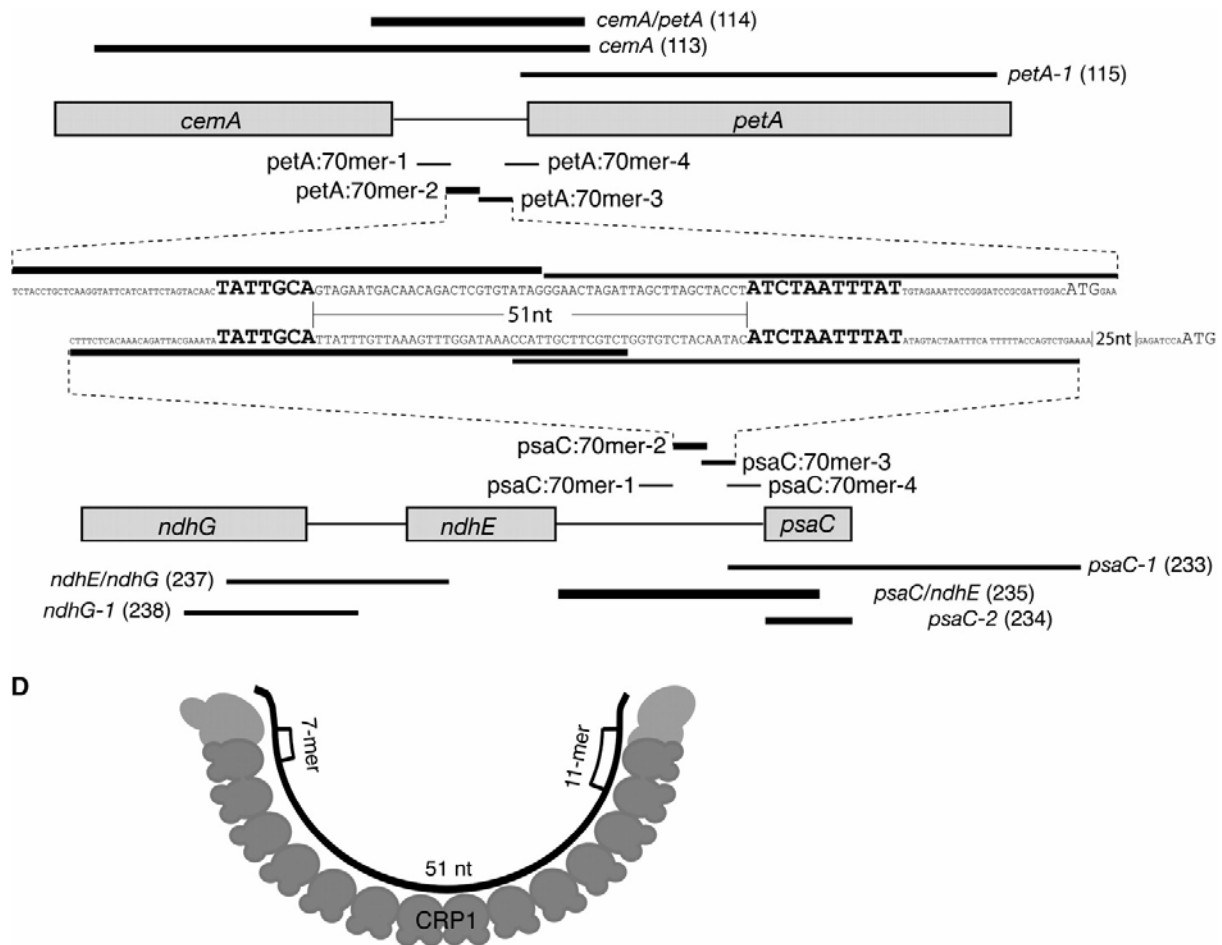


Figure 28 : Modèle de fixation de CRP1 à ses deux cibles ARN

La méthode de RIP-CHIP a permis d'identifier deux cibles pour la protéine CRP1, les transcrits contenant *petA* et *psaC*. Ces deux régions comportent une structure similaire : 2 séquences conservées (7 et 11 nucléotides) espacées de 51 nucléotides. La partie inférieure de la figure présente un modèle de reconnaissance dans lequel les 14 motifs PPR de CRP1 seraient en contact avec les 2 structures de 68 nucléotides analogues entre *psaC* et *petA* (Schmitz-Linneweber *et al.* 2005).

L'analyse de CRP1 chez le maïs a permis de présenter un modèle de reconnaissance de l'ARN illustré dans la Figure 28. Ce modèle implique la reconnaissance de deux séquences au niveau de l'ARN espacées de 51 nucléotides (Schmitz-Linneweber *et al.* 2005). Le modèle de reconnaissance proposé (partie inférieure de la Figure 28) inclut l'association des 14 motifs PPR de CRP1 avec une séquence de 68 nucléotides. Néanmoins, l'association de CRP1 avec toute ou partie de ces séquences reste à démontrer. D'autre part, il existe une certaine incompatibilité entre la structure tertiaire supposée de CRP1 et la taille de la séquence à reconnaître (voir ci-dessous).

CRP1 semble reconnaître 2 courtes séquences espacées de 51 nucléotides, ce qui n'est pas du tout le cas de CRR4 chez *Arabidopsis*. Chez cette dernière, il a été démontré que les 25 nucléotides précédant le site d'édition ciblé, et les 10 nucléotides suivants, étaient reconnus



par la protéine CRR4 (Okuda *et al.* 2006). CRR4 comporte également 14 motifs PPR, mais à la différence de CRP1, ceux-ci appartiennent à la famille PLS, plus un motif E terminal.

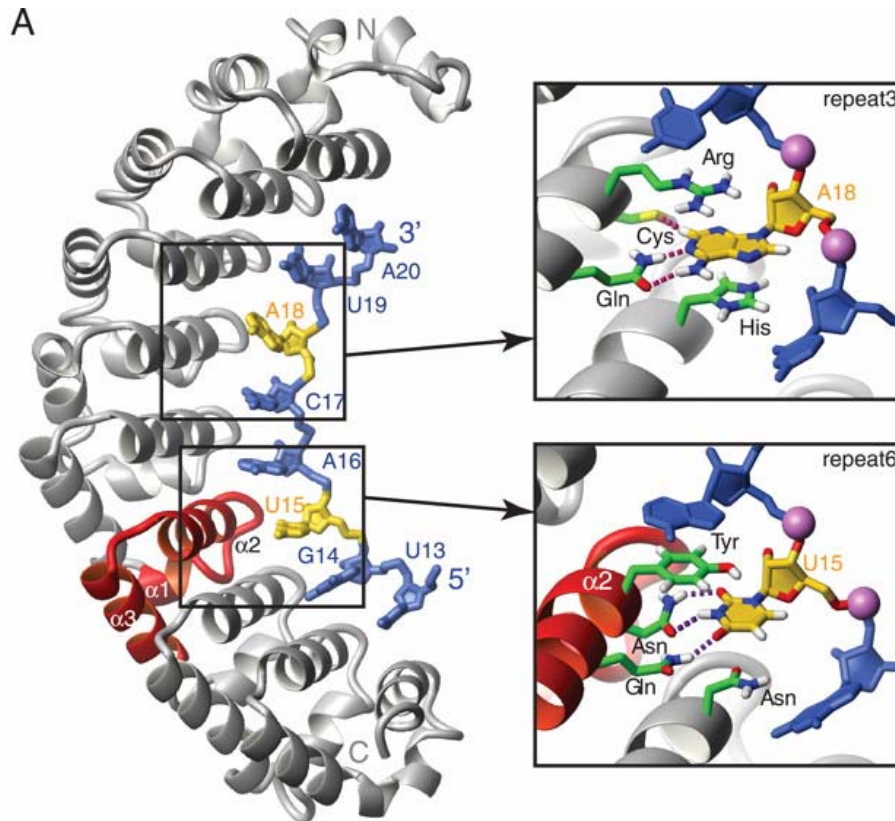
Ces deux observations ne permettent pas pour le moment de proposer un modèle unique de reconnaissance des cibles ARN par les motifs PPR, mais révèlent deux points importants. La cible ARN reconnue peut être bipartite (CRP1) et couvrir le site où une activité catalytique doit être exercée. De nombreuses protéines PPR ne comportent que quelques motifs PPR, si le modèle proposé pour CRP1 se révèle vrai, cela pourrait indiquer que ces protéines doivent fonctionner en dimère (-homo ou -hétéro) pour permettre la reconnaissance des deux séquences distantes.

L'idée que plusieurs protéines PPR puissent agir en homo ou hétéro dimères est assez séduisante, elle pourrait expliquer la variété de processus contrôlés par des protéines PPR, et la difficulté de discerner un schéma protéine-fonction.

Le cas de CRR4 pourrait signifier que les protéines PPR "présentent" la base à éditer, ou le site de clivage, à un facteur catalytique. Les sites d'édition se trouveraient alors "entre" la protéine PPR assurant la fonction de facteur de spécificité et un facteur portant l'activité enzymatique.

La résolution de la structure tridimensionnelle de protéines PPR associées à leur cibles devrait être la seule approche permettant de définir le mode de reconnaissance de ces protéines. De ce point de vue, il est intéressant de prendre en compte la résolution de la structure tridimensionnelle de l'interaction du domaine Pumilio1 humain avec une cible ARN de 8 nucléotides (Figure 29)(Wang *et al.* 2002).

Si les motifs PPR reconnaissent les molécules d'ARN d'une manière semblable à Pumilio, il est probable que le modèle proposé pour CRP1 soit erroné. La reconnaissance de séquence de 8 nucléotides se ferait par environ 10 motifs PPR.



**Figure 29 : Pumilio1 humain reconnaissant une séquence d'ARN**

La protéine est indiquée en gris et la molécule d'ARN en bleu et jaune. Le sixième domaine de Pumilio est indiqué en rouge, les phosphates sont indiqués en violet, code PDB : 1M8Y (Auweter *et al.* 2006).

### III.4.1 Quel rôle pour les protéines PPR au niveau de l'organisme?

L'analyse des génomes complets présents dans les banques de données montre que les protéines PPR sont présentes chez tous les eucaryotes possédant des mitochondries, et absentes des génomes procaryotes. A l'exception de GRP23 chez *Arabidopsis*, toutes les protéines PPR caractérisées sont actives dans les organites. Il semble donc que les protéines PPR soient une famille de protéines typiquement eucaryotes, dédiées au contrôle des deux organites d'origine bactérienne des cellules eucaryotes.

Les génomes des procaryotes à l'origine des plastes et des mitochondries ont été largement réduits depuis l'endosymbiose originelle. Cependant, ces génomes ont été conservés dans tous les organites (à l'exception des hydrogénosomes et des mitosomes (Embley *et al.* 2003)). La réduction des génomes des organites est-elle un processus en cours qui aboutira à la disparition de celui-ci? Ceci est peu probable, le schéma de diminution des génomes n'apparaît pas être totalement aléatoire, tout du moins, certains gènes résistent encore et toujours à l'appel du noyau! Dans les chloroplastes par exemple, les gènes *clpP1*, *rbcL* et *accD* n'ont jamais été transférés au noyau (*accD* a été purement et simplement perdu chez les

graminées). L'hydrophobicité et les différences de code génétique sont fréquemment citées comme les deux forces empêchant le transfert de certains gènes vers le noyau. Ceci est certainement vrai depuis l'origine des eucaryotes, mais il semble que des raisons d'ordre plus fonctionnel agissent également. Par exemple, quelle force peut maintenir le gène *clpP1* dans le chloroplaste alors que plusieurs paralogues sont situés dans le noyau, et leurs produits importés dans le chloroplaste? Une des raisons pourrait être qu'au cours de l'évolution les cellules eucaryotes ont sélectionné un certain nombre de processus de coordination entre le noyau et les organites. Le système de CES illustre parfaitement cette notion. La synthèse des sous-unités des complexes du chloroplaste se fait en fonction de l'état d'assemblage de ces complexes. La machinerie de synthèse des protéines chloroplastiques s'autorégule donc en fonction de la disponibilité des partenaires protéiques codés par le noyau.

Les protéines PPR possèdent plusieurs caractéristiques qui pourraient en faire des candidats pour être des composants de systèmes CES indirects. Les protéines PPR caractérisées interviennent toutes dans la maturation d'ARN des organites, ces maturations sont nécessaires à la synthèse des protéines codées. Bien sûr, pour démontrer cette hypothèse, il reste à prouver qu'une protéine PPR puisse interagir avec les protéines en question.

## Chapitre IV Matériels et méthodes

### IV.1 Matériel biologique

#### IV.1.1 Lignées d'insertions utilisées

AGI	Accession	Ecotype	Sous-famille	localisation	
				TargetP	Predotar
At1g05750	Salk_123752	Col-0	E+	Cp	Cp
	Salk_104250	Col-0			
	pde247-1	Col-0			
	clb19-1	Col-0			
At1g06710	Flag_055B12	Ws	P	Mt	Mt
At1g08070	Salk_034229	Col-0	DYW	Cp	Cp
	Salk_027812	Col-0			
	Salk_044609	Col-0			
	Salk_111305	Col-0			
	Salk_131850	Col-0			
At1g09410	Salk_000083	Col-0	DYW	Mt	Mt
	GABI_104E05	Col-0			
	Salk_067026	Col-0			
At1g14470 0	Flag_076C12	Ws	PLS	S	null
	Salk_071266	Col-0			
At1g15510	Salk_112251	Col-0	DYW	Cp	Cp
	Salk_070819	Col-0			
	Salk_112258	Col-0			
	Salk_112265	Col-0			
	Salk_142130	Col-0			
At1g18900	Salk_013683	Col-0	P	Mt	Mt
	Salk_040448	Col-0			
At1g20230	Salk_033409	Col-0	DYW	null	null
	Flag_214CO1	ws			
At1g22830	Flag_070G09	ws	E	null	Mt
At1g51965	Flag_119G07	ws	P	Mt	Mt
At1g59720	Salk_012256	Col-0	DYW	Cp	null
	Salk_012455	Col-0			
	Salk_115133	Col-0			
At1g61870	SALK_037390	Col-0	P	Mt	Mt
	SALK_108746	Col-0			
	SALK_139562	Col-0			
At1g68930	Salk_148973	Col-0	DYW	Mt	null
At1g74900	SAIL_867_F07		P	Mt	Mt
	SAIL_867_G07				
	SAIL_710_C03				
At1g80550	Flag_044A06	ws	P	Mt	Mt

AGI	Accession	Ecotype	Sous-famille	localisation	
				TargetP	Predotar
At1g80550	Salk_024306	Col-0	P	Mt	Mt
At2g01860	Flag_036D03	ws	P	Cp	Mt
At2g15820	sail_713_d08		P	Mt	Mt
At2g15820	Salk_110002		P	Mt	Mt
At2g20540	Flag_047B08	ws	E+	null	null
At2g29760	Salk_092402	Col-0	DYW	Cp	Cp
At2g33760	Salk_008044	Col-0	DYW	null	Cp
At2g33760	Salk_008240	Col-0	DYW	null	Cp
At2g40720	Flag_187E04	ws	E+	null	null
At2g41720	Flag_380H05	ws	P	null	Cp
At2g41720	Flag_633F06	ws	P	null	Cp
At3g02010	Salk_119701	Col-0	DYW	Mt	null
At3g08820	Salk_023916	Col-0	DYW	null	Cp
At3g08820	Salk_064348	Col-0	DYW	null	Cp
At3g08820	Salk_064349	Col-0	DYW	null	Cp
At3g08820	Salk_147615	Col-0	DYW	null	Cp
At3g13150	Flag_105E08	ws	P	Cp	Mt
At3g16890	salk_071712	Col-0	P	Cp	Mt
At3g13880	SALK_040554	col	E+	Mt	Mt
At3g22150	Flag_371D05	ws	E+	Cp	Cp
At3g22150	Salk_041683	Col-0	E+	Cp	Cp
At3g22150	Salk_040626	Col-0	E+	Cp	Cp
At3g22690	Salk_018914	Col-0	DYW	Cp	Cp
At3g22690	Salk_073746	Col-0	DYW	Cp	Cp
At3g22690	Salk_123515	Col-0	DYW	Cp	Cp
At3g22690	Salk_144420	Col-0	DYW	Cp	Cp
At3g25970	Flag_038A08	ws	E+	S	null
At3g25970	Salk_127777	Col-0			
At3g25970	Salk_122927	Col-0			
At3g25970	Salk_122936	Col-0			
At3g46790	Salk_030786	Col-0	DYW	Cp	null
	GABI_425F06	Col-0			
	Salk_046131	Col-0			
At3g49140	Salk_015469	Col-0	DYW	Mt	Mt
At3g49740	Salk_073095	Col-0	E	Cp	null
	Salk_086074	Col-0			
	Salk_086075	Col-0			
At3g62890	Salk_044324	Col-0	DYW	Cp	null
At3g63370	Salk_008652	Col-0	DYW	Cp	Cp
	Flag_592C05	ws			
	Salk_102445	Col-0			
	Salk_025267	Col-0			
At4g14820	Salk_043173	Col-0	DYW	null	Cp
	Salk_076959	Col-0			

AGI	Accession	Ecotype	Sous-famille	localisation	
				TargetP	Predotar
	Salk_096272	Col-0			
	Salk_125349	Col-0			
	Salk_128858	Col-0			
	Salk_136027	Col-0			
At4g18750	Salk_139995	Col-0	DYW	Cp	Cp
At4g31850	Flag_088D06	ws	P	Cp	Cp
At4g31850	Salk_091893	Col-0			
At4g33170	Salk_005468	Col-0	DYW	Cp	null
	Salk_043048	Col-0			
	Salk_068978	Col-0			
	Salk_068987	Col-0			
At4g34830	Flag_568C09	ws	P	Cp	Cp
At4g35130	Salk_060152	Col-0	DYW	Cp	Cp
	Salk_064949	Col-0			
	Salk_118555	Col-0			
At5g15340	Salk_033610	Col-0	DYW	Mt	Mt
At5g21222	Salk_008247	Col-0	DYW	null	null
	Salk_064877	Col-0			
At5g27270	Flag_026C02	ws	P	Cp	null
At5g39680	Salk_006056	Col-0	DYW	null	null
At5g40410	Salk_004882	Col-0	DYW	null	null
At5g44230	Salk_045515	Col-0	DYW	Cp	Cp
	Salk_082500	Col-0			
	Salk_082502	Col-0			
	Salk_085176	Col-0			
	Salk_131636	Col-0			
	Salk_136636	Col-0			
At5g46460	Salk_033891	Col-0	DYW	Mt	Mt
At5g46580	Salk_072319	Col-0	P	Cp	Cp
At5g48910	Salk_039948	Col-0	DYW	Cp	null
	Salk_039953	Col-0			
	Salk_039955	Col-0			
	Salk_039956	Col-0			
	Salk_039963	Col-0			
At5g50390	Salk_064582	Col-0	DYW	Cp	Cp
	Salk_076606	Col-0			
At5g52630	Salk_113095	Col-0	DYW	null	null
At5g52850	Salk_048941	Col-0	DYW	Cp	Cp
	Salk_085242	Col-0			
At5g55740	Salk_090410	Col-0	E+	Cp	Cp
At5g59200	Salk_029916	Col-0	E+	Cp	Cp
	Salk_060533	Col-0			
	Salk_096638	Col-0			
	Salk_096653	Col-0			

AGI	Accession	Ecotype	Sous-famille	localisation	
				TargetP	Predotar
	Salk_111721	Col-0			
	Salk_139702	Col-0			
At5g62370	Flag_129B09	ws	P	Cp	Mt
At5g65570	Flag_155A12	ws	DYW	Mt	null

Tableau 11 : Liste des lignées d'insertion collectées

## IV.1.2 Méthode de culture d'Arabidopsis

### IV.1.2.1 Culture en terre

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

### IV.1.2.2 Culture *in vitro*

En condition *in vitro*, les plantes ont été cultivées sur un milieu Murashige et Skoog appelé ½ MS. Pour 1L : 2.2g de MS (Sigma Aldrich™ M5519 Murashige and Skoog Basal Medium), 1 à 3% de saccharose et 8% d'agar type E pour les cultures en milieu solide (A6674 sigma Aldrich©). Le pH doit être ajusté à 5,8 par ajout de d'hydroxyde de potassium (KOH). Les conditions de culture de la chambre sont 25°C, 8 heures d'obscurité, 16h de luminosité.

## IV.2 Méthodes

### IV.2.1 Protocoles relatifs à l'ADN

#### IV.2.1.1 Extraction rapide d'ADN génomique :

Ce protocole est utilisé en plaque 96 pour le génotypage d'un grand nombre de plantes en parallèle.

##### Tampon de Lyse :

	Concentration finale
SDS	0,5 %
NaCl	250 M
EDTA	25 mM
Tris-HCl pH : 7,5	200 mM

1. Collecter et broyer les tissus dans l'azote liquide, en plaque 96

2. Ajouter 400µl de Tampon de lyse
3. Vortexer
4. Incuber 15 minutes à 50°C
5. Centrifuger 15 minutes à 1500 G et 4°C
6. Récupérer 150 µl de surnageant propre dans une nouvelle plaque
7. Ajouter 150µl d'isopropanol
8. Vortexer
9. Incuber 20 minutes à température ambiante
10. Centrifuger 30 minutes à 1500 G à 4°C
11. Eliminer le surnageant, rincer les culots avec 125µl d'éthanol à 70%
12. Centrifuger 15 minutes à 1500 G à 4°C
13. Eliminer le surnageant et sécher les culots
14. Reprendre les culots dans 125µl de TE (10mM, 1mM)

#### **IV.2.1.2 Extraction fine d'ADN génomique, méthode dite "au CTAB"**

##### Tampon d'extraction :

	Concentration finale
CTAB	2 %
NaCl	1,4 M
EDTA	20 mM
Tris-HCl pH 8	100 mM
β-mercaptoéthanol <sup>3</sup>	0,2 %
RNase <sup>4</sup>	0,1 mg/ml

1. Récolter et peser les tissus
2. Préchauffer le volume adéquat de Tampon d'extraction à 65°C, sans la RNase
3. Réduire les tissus en poudre fine au mortier dans l'azote liquide
4. Ajouter le volume adéquat de tampon d'extraction (1 ml pour 100 mg de tissus)
5. Agiter pendant 1 heure à 65°C
6. Centrifuger le mélange pendant 1 min à 14000 G afin de faire tomber tout le

---

<sup>3</sup> A ajouter au dernier moment

<sup>4</sup> A ajouter au dernier moment



liquide

7. Ajouter 1 volume de chloroforme par volume de tampon d'extraction
8. Centrifuger le mélange pendant 20 min à 14000 G et 4°C
9. Transférer la phase aqueuse dans un nouveau tube
10. Ajouter 1/10<sup>e</sup> de volume d'acétate de sodium et 8/10<sup>e</sup> de volume d'isopropanol
11. Centrifuger 40 min à 14000 G et 4°C
12. Eliminer l'alcool
13. Faire un lavage à l'éthanol 70%
14. Eliminer l'alcool, sécher sous hotte ou à l'étuve
15. Reprendre le culot dans de l'eau ou du TE, 1/10<sup>e</sup> du volume de tampon d'extraction utilisé

#### **IV.2.1.3 Electrophorèse**

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

Rouge de crésol, solution mère concentrée 10 fois :

- Rouge de crésol 2 mg/ml
- Saccharose 600 mg/ml

TBE 1X, pH8 :

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

#### **IV.2.1.4 Amplification d'ADN par PCR**

Les amplifications d'ADN ont été effectuées en utilisant la Taq DNA polymérase (M0267L) de chez New England Biolabs, elle est distribuée à une concentration de 5 unités par microlitre.

##### ***IV.2.1.4.1 PCR dite "standard"***

PCR utilisée pour le génotypage, ou l'amplification de fragments spécifiques en petit nombre, volume total de 10 à 50 µl.

- Matrice ADN 25-75ng d'ADN génomique
- Rouge de crésol 1X
- Thermopol Buffer 1X
- dNTP 0,2 mM chacun
- Taq polymérase 1 à 5 unités
- Oligonucléotides 1 $\mu$ M

#### **IV.2.1.4.2 PCR dite "384 puits"**

Pour l'analyse du transcriptome mitochondrial ou chloroplastique, il est nécessaire d'effectuer la phase de PCR en plaques de 384 puits. Le mélange d'amorces est préalablement distribué dans chaque puits. Avant d'effectuer la PCR "384 puits" une PCR standard est réalisée avec 4 couples d'amorces tests (CHLORO44, CHLORO23, MITO041, MITO352). Ces 4 couples ciblent la partie codante de 2 gènes chloroplastiques (*rpoB* et *atpA*) et 2 gènes mitochondriaux (*cox2* et *cox1*), dont le profil d'expression est relativement bien connu. Ainsi nous pouvons déterminer la dilution de cDNA adéquate à utiliser par la suite dans la PCR "384 puits".

Le mélange, par puits, de la réaction de PCR est :

- primers 0.5 $\mu$ M 4 $\mu$ l
- cDNA dilué 1 $\mu$ l
- Taq polymérase 1 unité
- Rouge de crésol 1X
- Thermopol Buffer 1X
- dNTP 1 $\mu$ M chacun
- H<sub>2</sub>O qsp 6 $\mu$ l

#### **IV.2.1.5 Test d'édition par profil de restriction**

De nombreux transcrits mitochondriaux ou chloroplastiques sont modifiés par édition d'une cytidine vers une uridine, dans certains cas ce changement peut être observé par utilisation d'une enzyme de restriction car un site de restriction peut être soit créé soit détruit.

Après une PCR classique dans un volume de 20 $\mu$ l, 5 $\mu$ l de produits PCR sont visualisés sous UV sur gel d'agarose 1.7%, 5  $\mu$ l sont conservés pour contrôle de digestion. Les 10 $\mu$ l restant sont précipités par ajout de 40 $\mu$ l d'éthanol 96°, 2  $\mu$ l d'Acétate de sodium (3M) et 1  $\mu$ l de Glycogène Azur (20 mg/ml). Ce mélange est laissé une heure à -20°C, puis centrifugé 30

min à 4°C et à 14000G. Les culots sont rincés deux fois à l'éthanol 70%. Après séchage à l'air libre les culots sont repris dans 10 µl d'eau, 5µl mélange de digestion sont ensuite ajoutés (0.5µl d'enzyme de restriction, 1.5 de tampon adéquat, 3µl d'eau). La digestion est effectuée à la température adéquate pendant 1H en bain marie liquide.

Les produits de digestion sont ensuite visualisés sous UV sur gel d'agarose 1.7%.

#### **IV.2.1.6 Purification d'acides nucléiques**

Les produits de PCR destinés à être séquencés ont été concentrés et dessalés sur membrane à très faible adsorption en cellulose régénérée, Microcon® YM-100 (Millipore).

Les produits de PCR sont déposés sur la membrane puis centrifugés à 14000G, la membrane est ensuite transférée et disposée à l'envers dans un nouveau tube après ajout de 10µl d'eau ultrapure, les produits de PCR purifiés sont récupérés par centrifugation à 14000G.

#### **IV.2.1.7 Séquençage d'ADN**

Le séquençage des produits de PCR a été effectué par une entreprise externe, Genome Express puis Macrogen à partir de janvier 2006.

### **IV.2.2 Protocoles relatifs aux ARN**

#### **IV.2.2.1 Extraction d'ARN total, méthode dite "au TRIzol®"**

Méthode d'extraction d'ARN en une seule étape d'après (Chomczynski and Sacchi 1987)).

1. Récolter les tissus, les peser et les plonger dans l'azote liquide
2. Réduire les tissus en fine poudre au mortier et dans l'azote liquide
3. Reprendre dans 1ml de TRIzol® par 100mg de tissus
4. Incuber 5 minutes à température ambiante
5. Centrifuger 10 minutes à maximum 12000 G à température ambiante pour éliminer les déchets
6. Transférer le surnageant dans un nouveau tube, sans toucher le culot
7. Ajouter 0,2 volume de chloroforme et vortexer
8. Incuber 5 minutes à température ambiante
9. Centrifuger 15 minutes à un maximum de 12000 G à température ambiante
10. Récupérer la phase aqueuse<sup>5</sup>

---

<sup>5</sup> A cette étape, si la phase aqueuse est encore colorée en rose, ou l'interface trop épaisse, renouveler l'étape chloroforme plusieurs fois, en rajoutant de l'eau traitée au DEPC si besoin est.

11. Ajouter un volume d'isopropanol (0,5 volume de TRIzol<sup>®</sup> initial)
12. Incuber 10 à 30 minutes à température ambiante
13. Centrifuger 15 minutes à maximum 12000 G à 4°C
14. Récupérer le culot, puis le laver dans l'éthanol 70%, eau DEPC 30%
15. Centrifuger 15 minutes à maximum 12000 G
16. Eliminer l'éthanol
17. Sécher le culot sous hotte de préférence
18. Reprendre le culot dans de l'eau DEPC à 65°C, généralement 1/10ème à 1/20ème du volume de TRIzol<sup>®</sup> utilisé, compléter par de l'eau DEPC si le culot n'est pas totalement dissous.

Les ARN totaux sont quantifiés au spectrophotomètre par mesure de l'absorption à 260 nm, ensuite visualisés sur gel d'agarose 2 %, tampon TBE. La migration se fait dans une cuve préalablement rincée au peroxyde d'hydrogène 10% pour diminuer les risques de dégradation par des RNAses (Sambrook *et al.* 1989).

#### **IV.2.2.2 Elimination de l'ADN contaminant des extraits d'ARN totaux**

Les ARN totaux sont traités à la DNaseI (DNA-free<sup>™</sup> DNase Treatment and Removal Reagents AM1906) suivant les instructions du fabricant (Ambion<sup>®</sup>) :

100µl d'ARN totaux sont traités pendant 1 heure à 37°C par 4 unités de DNaseI (ajout de 2µl de DNaseI, 1µl de RNase OUT et 11,4 µl de tampon de réaction spécifique de la DNase). La DNase est ensuite éliminée par ajout de 0,1 volume de tampon d'inactivation. Après deux minutes d'incubation à température ambiante, le mélange est centrifugé 1 minute à 10000G, les microbilles contenues dans le tampon d'inactivation fixent la DNase et les cations bivalents. Les ARN totaux (dans le surnageant) sont ensuite transférés dans un tube propre en prenant garde à ne pas toucher au culot contenant les billes, puis visualisés sur gel et quantifiés au spectrophotomètre.

#### **IV.2.2.3 Synthèse de premier brin d'ADN complémentaire d'ARN d'organites.**

Les ARN produits à partir des génomes cytoplasmiques n'étant pas polyadénylés de manière homogène, l'initiation de la synthèse de premier brin se fait à l'aide d'hexanucléotides aléatoires.

La synthèse d'ADN complémentaire a été effectuée en utilisant la reverse transcriptase SuperscriptII<sup>™</sup> (11904-018, Invitrogen<sup>™</sup>).

1. Le mélange suivant est incubé 5 minutes à 65°C :
  - 3 µg d'ARN<sup>6</sup>
  - 1 µl de dNTP (10mM)
  - 100ng d'hexanucléotides
  - Eau DEPC, qsp 12µl
2. Le mélange est plongé dans la glace
3. Ajout de
  - 4µl de tampon de réaction 5X
  - 2µl de DTT (0,1 M)
  - 1µl de RNaseOUT™ (10777-019 Invitrogen™)
  - 1µl de SuperscriptII™ (200 unités)
4. Incubation 10 minutes à 25°C
5. Incubation 60 minutes à 42°C
6. Incubation 15 minutes à 75°C
7. Le mélange est plongé dans la glace

#### IV.2.2.4 5' et 3' Race

Les expériences d'identification des extrémités 5' et 3' des transcrits ont été effectuées avec le Kit SMART™ RACE cDNA Amplification Kit de Clontech™, en suivant exactement les instructions du fabricant :

##### IV.2.2.4.1 Synthèse du premier brin d'ADNc

1. Dans deux tubes de type eppendorf de 1.5 ml mélanger :

<b>5'-RACE cDNA</b>	<b>3'-RACE cDNA</b>
• 1–3 µl d'ARN	1–3 µl d'ARN
• 1 µl 5'-CDS primer	1 µl 3'-CDS amorce A
• 1 µl SMART II A oligo	
2. Compléter à 5µl avec de l'eau stérile.
3. Mélanger et centrifuger brièvement.
4. Incuber à 70°C pendant 2 min; puis 2 min dans la glace.
5. Centrifuger brièvement pour faire descendre tout le liquide au fond du tube.
6. Ajouter dans chaque tube :
  - 2 µl 5X First-Strand buffer
  - 1 µl DTT (20 mM)

---

<sup>6</sup> Ou 10µl maximum si la concentration est trop faible

- 1 µl dNTP Mix (10 mM)
  - 1 µl PowerScript Reverse Transcriptase
7. Mélanger et centrifuger brièvement.
  8. Incuber à 42°C par 1H30 dans un bain marie sec fermé
  9. Dilution du mélange dans un tampon Tricine-EDTA
    - 20 µl si moins de 200 ng d'ARN totaux ont été utilisés
    - 100 µl si plus de 200 ng d'ARN totaux ont été utilisés
  10. Chauffer les tubes à 72°C pour 10 min

#### ***IV.2.2.4.2 "Rapid Amplification of cDNA Ends" (RACE)***

1. Préparer le master mix (quantités par gène et par type d'extrémité 5' ou 3') :
  - 34,5 µl d'eau ultra pure
  - 5 µl 10X Advantage 2 PCR Buffer
  - 1 µl dNTP Mix (10 mM)
  - 1 µl 50X Advantage 2 Polymerase Mix
  - 41,5 µl Volume final
2. Mélanger vigoureusement, sans former de bulles, centrifuger brièvement.
3. Chaque réaction de PCR est préparée ainsi :
  - 2,5 µl d'ADNc préparé précédemment
  - 5 µl d'amorces universelles UPM
  - 1 µl d'amorce spécifique au gène (10 mM)
  - 41,5 µl de master mix
4. Utiliser le programme de PCR suivant
  - 25 cycles :

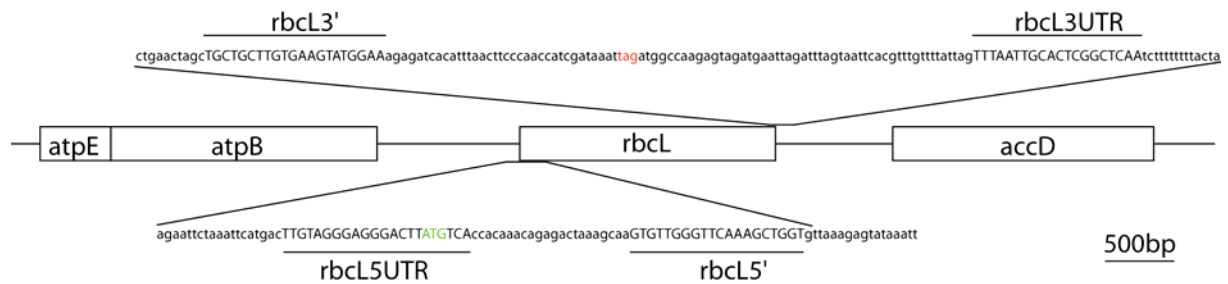
94°C	30 sec
68°C	30 sec
72°C	3 min
5. Visualisation sur gel d'agarose 2% des produits de PCR

#### ***IV.2.2.4.3 Etapes supplémentaires***

Dans le cas d'une amplification insuffisante après le programme PCR, les échantillons peuvent être remis pour 5 cycles supplémentaires. Une "Nested PCR" peut également être effectuée. Cette seconde PCR a pour intérêt d'éliminer une partie des faux positifs pouvant être obtenus après la première PCR. Ces PCR sont effectuées en utilisant une nouvelle amorce spécifique du gène, placée en amont de l'amorce précédent pour l'extrémité 5', et en aval pour l'extrémité 3'. Le protocole précédent est repris en utilisant, comme matrice 5µl d'une dilution au 50<sup>ème</sup> du produit de RACE (dilution dans le tampon Tricine-EDTA), et en remplaçant les amorces universelles UPM par les amorces universelles NUP (reconnaissent uniquement

l'adaptateur utilisé lors de la synthèse du premier brin).

#### IV.2.2.4.4 Amorces gène spécifique



**Figure 30 : Amorces pour l'analyse par RACE PCR du transcrit *rbcL***

Les zones d'hybridation des amorces utilisées pour effectuer l'analyse par RACE PCR du transcrit *rbcL* sont indiquées en majuscule, le codon d'initiation de la traduction est indiqué en vert, le codon stop en rouge.

- *rbcL5UTR*                    TGACATAAGTCCCTCCCTACAA
- *rbcL3UTR*                    TTTAATTGCACTCGGCTCAA
- *clpP3UTR*                    AAAAATAGGAGCGATTTTCATGC
- *clpP5UTR*                    GCAGAATGAGACGGGATAACA
- *rbcL5'*                        ACCAGCTTTGAACCCAACAC
- *rbcL3'*                        TGCTGCTTGTGAAGTATGGAA
- *clpPex3\_5'*                   AACGAACTAGCGGGTTGATG
- *clpPex3\_3'*                   GAAGACATGGAACGGGATGT

#### IV.2.2.4.5 Amorces généraux et Tampons

##### SMART II™ A Oligonucleotide

5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3'

##### 3'-RACE CDS Primer A<sup>7</sup>

5'-AAGCAGTGGTATCAACGCAGAGTAC(T)30V N-3'

##### 5'-RACE CDS Primer A<sup>7</sup>

5'-(T)25V N-3'

##### Universal Primer A Mix (UPM)

- Long (0,4 μM) :

5'-ctaatacagactcactatagggcCTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'

- Short (2 μM) :

5'-ctaatacagactcactatagggcCTAATACGACTCACTATAGGGC-3'

##### Nested Universal Primer A (NUP)

5'-AAGCAGTGGTATCAACGCAGAGT-3'

##### 5X First-Strand Buffer

- 250 mM Tris-HCl (pH 8,3)

<sup>7</sup> (N = A, C, G, ou T; V = A, G, ou C)

- 375 mM KCl
- 30 mM MgCl<sub>2</sub>

Tricine-EDTA Buffer

- 10 mM Tricine-KOH (pH 8,5)
- 1 mM EDTA

## IV.2.3 Protocoles relatifs aux protéines

### IV.2.3.1 Extraction de protéines

#### IV.2.3.1.1 Protéines totales

L'extraction de protéines totales se fait sur tissus frais, broyés dans l'azote liquide. Environ 100mg de tissus sont incubés 30 minutes dans un tampon de broyage, en vortexant deux à trois fois pendant cette période. Le mélange est ensuite centrifugé 10 minutes à 12000 G. Les protéines totales sont situées dans le surnageant.

Tampon d'extraction protéines totales :

	Concentration finale
Tris-HCl à pH 7,4	50 mM
β-mercapto-éthanol	1 mM
EDTA	1 mM
PMSF	1 mM
SDS	1 %

#### IV.2.3.1.2 Enrichissement en protéines chloroplastiques

Pour une analyse plus précise du contenu en protéines chloroplastiques, il est intéressant d'effectuer un enrichissement en chloroplastes; pour cela les échantillons sont broyés sur glace dans le tampon d'extraction, pour faciliter le broyage du sable est ajouté dans le mortier (Sigma Aldrich™274739). L'enrichissement est obtenu par plusieurs centrifugations successives en tube Corex dans une centrifugeuse réfrigérée à 4°C de type Sorvall. Il est important de laisser la fonction freinage doux active pour assurer une bonne séparation. Les échantillons sont centrifugés 5min à 200G, puis 5min à 600G, puis 5min à 800G, et enfin 5min à 3500G, à chaque fois le surnageant est transféré dans un nouveau tube avant la centrifugation suivante, à l'issue de l'étape à 3500G le culot est repris dans le tampon de rinçage.



Tampon d'extraction Protéines chloroplastiques :

	Concentration finale
Sucrose	0.45 M
EGTA	1.5mM
PVP40	0.6%
MOPS/KOH	15 mM

Tampon de rinçage

	Concentration finale
Sucrose	0.3 M
EGTA	1 mM
MOPS/KOH	10 mM

**IV.2.3.2 Dosage des protéines**

La quantité de protéines extraites est mesurée par Test de Bradford (Bradford 1976), en utilisant le réactif Coomassie Plus de Pierce Biotechnology™.

**IV.2.3.3 Observation sur gel dénaturant**

Les protéines sont séparées sur gel 7cm, de type SDS-PAGE 10% :

Composition d'un gel SDS-PAGE

	Gel de migration	Gel de concentration
30% Acryl/bis	1,1ml	0,2ml
Tris <sup>8</sup>	0,8ml	0,2ml
SDS 10%	43µl	20µl
APS 10%	13µl	10µl
TEMED 6,6 M	3,3µl	6,5µl
Eau	1,9 ml	1,5ml

Les échantillons sont mélangés à un tampon de dénaturation/charge (tampon concentré 5X: 15% β-mercaptoéthanol, 15% SDS, 1,5% bleu de bromophénol et 50% glycérol) puis dénaturés à 65°C pendant 5 minutes. La migration se fait à ampérage constant compris entre 10 et 30 mA et dans un tampon de migration type Laemmli (Glycine 1,92M, Tris base 0,25M, SDS 1%) (Laemmli 1970).

La coloration est faite par la méthode dite de Neuhoff, une incubation sur la nuit dans une solution de coloration (80 ml de solution mère et 20ml de méthanol), renouvelée le matin pour deux heures supplémentaires avant un rinçage à l'eau d'une heure.

Solution mère, coloration de Neuhoff

---

<sup>8</sup> solutions mères : Séparation : 2M pH 8,9, Concentration 1,5M pH 6,8

	Quantité dans 500ml
Acide orthophosphorique	10%
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	50g
Bleu de Coomassie G250	0,5g

#### IV.2.3.4 Western blot

Après séparation sur gel SDS-PAGE, les protéines sont transférées sur une membrane PVDF par électro-transfert avec le Mini Trans-Blot Cell de Bio-Rad™ dans un tampon (Glycine 0,1 M, Tris base 768 mM). La membrane est ensuite saturée par incubation pendant une heure dans une solution de TBST contenant 7,5% de lait lyophilisé. Après lavage dans du TBST, la première incubation avec l'anticorps primaire anti-rubisco (AgriSera, AS01-017) est réalisée sur la nuit à température ambiante. Après deux lavages dans du TBST, la seconde incubation avec cette fois-ci l'anticorps secondaire (Anti IgG de poulet, couplé à la peroxydase de raifort (HRP), Sigma A9792) est réalisée pendant une heure à température ambiante. Pour finir, la membrane est lavée trois fois dans du TBST, puis deux fois dans un tampon Tris-Mg acétate (50mM-10mM, pH 9,5).

Le Western blot est révélé en utilisant le Kit de réaction ECL Western Blotting de Amersham Biosciences® (RPN2106/8/9, RPN2209, RPN2134), et visualisé grâce à un GeneGnome de Syngene®.

#### IV.2.4 Oligonucléotides utilisés lors de cette thèse

##### IV.2.4.1 Génotypage de lignées d'insertion

**Tableau 12 : liste des amorces utilisées pour le génotypage des lignées d'insertion**

Les amorces spécifiques des différents vecteurs sont indiquées en gras

NOM	SEQUENCE	ID	SEQUENCE
<b>GABI_LB</b>	GGACGTGAATGTAG	96638_RP	GAGATTCATGCCAGGTTTTGA
<b>GABI_RB</b>	CCAAAGATGGACCCCCACCCAC	111305_LP	TTTTGATTAACTATGCTTCTTGAA
<b>LB1</b>	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTC	111305_RP	AATGTCGAGTTGGGCGAATCA
<b>LB2</b>	GCTTCCTATTATATCTTCCCAAATTACCAATACA	111721_LP	CCTCCATGGCTACATGCGTTT
<b>LB3</b>	TAGCATCTGAATTTTCATAACCAATCTCGATACAC	111721_RP	TCCATCGATTACGCCAAGAT
<b>SALK_LBA1</b>	TGGTTCACGTAGTGGGCCATCG	112251_LP	CGAATGGGAAGCTGGAGGAAG
<b>LB_6313R</b>	TCAAACAGGATTTTCGCCTGCT	112251_RP	TGTCTTTGCGCTCCATTCTGC
<b>LB_BAR1</b>	CAACCCTCAACTGGAAACGGGCCGGA	118555_LP	TCAAGGAATCTCTTGCTGCG
<b>SALK_LBB1</b>	GCGTGGACCGCTTGCTGCAACT	118555_RP	CGGTATGATCCAGAGGAACATCG
<b>TAG_3</b>	CTGATACCAGACGTTGCCGCATAA	12256_LP	TCCAAACACCCACAAGACCCA
<b>TAG_5</b>	CTACAAATTGCCTTTTCTTATCGAC	12256_RP	GGGTTCATGCTACCGTTTCTACG
<b>TAIL2</b>	TCGTTAAACTGCCTGGCACA	123515_LP	GCTGCATCCACCTCACCATTCT
		123515_RP	GCATCTTCTGGGTTATGCAACG
123655_LP	CAGCAAGAGGACTTTTCAGGTTCA	125349_LP	AATGGCTTTTGGGCTTTGCT
123655_RP	TGGCGTTTCGGTTTATGGATG	125349_RP	TTGCAGCCAAACGTATCTCTCG

NOM	SEQUENCE	ID	SEQUENCE
090410_LP	TTCAAGCTCATTTCACAAGCA	128858_LP	ACGAAGGAGGTCCGGTGAGTT
090410_RP	TCGAGGAAATGCCATTCAAGC	128858_RP	TTGAGCATTATGGATGCATGG
123752_RP	TTGCGGTCAACCAATTTGAGC	131636_LP	CCCTGGAAACCTGAACCATCC
123752_LP	GCTGTCTTAAACGCTTGCACA	131636_RP	TGTGTGGACCATCATATTTCCAAAA
092402_LP	GATGGATCCCCAAGTTGACC	136027_LP	CATCCACCAAAAACGCTACCACA
092402_RP	TGCTGACTTAACAGCCATCCCA	136027_RP	TGCTCTGGGATCAAACCTGATG
104250_LP	AACCGGATCAAACCGGATTCT	136636_LP	TCAACCGGATTTGAGCTCTGTG
104250_RP	CATGTGCATTTCCATTAGCAGCA	136636_RP	CGAAAAAGTTTGATCCTCTCGCA
041683_RP	GCGACAACATTCTTCCTACGCA	139702_LP	TCCATCAGCAGAACGACCAGA
041683_LP	ACATAGGATGAGTTTGGGCTTC	139702_RP	CTCCAGCAAGCAACACAAGCA
024306_LP	TTTCCCGCTCTTGCACATGAT	139995_LP	ATGGCGATGTTGGTGACGAAT
024306_RP	TTTTCACCTCTGGATGGCTCAA	139995_RP	CGAGAATCAGCACAACCAGCA
067026_RP	CGTCAGCCCACCTACCTTGAG	142130_LP	GCAAAATCCCAAATGGTGAGACA
067026_LP	TGGACTTCCATGCTTATGGGC	142130_RP	AAAAGGATGATCAACAACACTGGA
122927_LP	TTTTGAATAACACCCAAACAAAAC	144420_LP	TCCCAACTCTCAAACCCATTCC
122927_RP	TCGCCACTAAATCAGGCTTCG	144420_RP	CGGAGTTGTGAATTGGGTACTCG
040626_LP	TTTCCGGGTACACGCAGAATG	147615_LP	CCTTCCCTTCCAACCTCGATCC
040626_RP	GCCAAACCGTCGATTACATCA	147615_RP	ATGCCAAGTGTGGGAAGATGG
071266_RP	CAAAATCCCGAATCTCCCAGC	18914_LP	AGTCTTGCAAGAGCGCAGGGAT
071266_LP	CATGCAAAAACAAACTAACGGTGA	18914_RP	TGCTTGGTCGAGCAGGATTGT
091893_RP	CCATCAACCCTCAAAGCTTCA	23916_LP	GAATCACTTGGCCATGATCCG
091893_LP	TGTTCTTCTCTTTTTCAATTCTGG	23916_RP	TGCAAGCTTAGGAGCCCTTGA
013683_LP	TACGAAGATTCAGGTGGGGGC	27812_LP	TGTGTGTAACCACCGATCAAGG
013683_RP	AGGTCGCAGACCAGCATGAAG	27812_RP	CTCCGCATTATCCACGCTCAG
046131_LP	TCGAATTTGAGGGCACAATGAA	29916_LP	TCGTTGCTTACTCAATCCAGCC
046131_RP	TGATCTGTTAGGTCGGGCGAA	29916_RP	TGCTCGGATTTAGGTGCGTTG
040448_RP	ACTGAAACCGCTTGGCACAAA	127777_LP	GTCTTTTGAGCCACCCAAGCC
040448_LP	GGTTCATGCTTGGAAACCGTCA	127777_RP	CACCCATGACCGAAATCTAGAAAA
33610_LP	TCTAGCGCAGTGCAAGCACAC	39948_LP	CACGAGCATCTTTCATAAAACCACA
33610_RP	TCTGAGAAAAGCGAGGCATCAGA	39948_RP	TGCTTGCCTGACTCTCATGCT
119701_LP	CGAGGCATCCCTCCACTTCTT	43048_LP	TCCTAGCGAAACCGAATTTTCTCA
119701_RP	TGCTGTCTCTCAGCAAGGCT	43048_RP	CGGAAAATGTACACACGCACG
113095_LP	TCCAATGCTTATGCAGCCGAT	43173_LP	CATCCACCAAAAACGCTACCACA
113095_RP	GCCATCGCAACGAACTTTGAT	43173_RP	TGCTCTGGGATCAAACCTGATG
115133_LP	CGACCCGCATTTGCAATACAT	44324_LP	GACATTTGTGGGCATTCTCGG
115133_RP	GTCTTGCTTGCCATCCGGTTT	44324_RP	TCTCTGCAAGAGCAAGACCCG
148973_RP	TAACAGCTCTTCAGCACGCC	45515_LP	TGAGCTCTGTGCCATACGACG
148973_LP	TCGACAGAAATCCCCAACCAA	45515_RP	CGAAAAGTTTGATCCTCTCGCA
71712_LP	GCTTCTTAGCACGCCGTGAA	48941_LP	TCGAACAGTTTACGTGCATTCCA
71712_RP	GCTGTGATTGATGCGTTGGTG	48941_RP	CATCATGCAACTTTGGTTTTCCC
12455_LP	CATTTCCGGTAAAGCATAAAGGC	5468_LP	CGCAGAGTGATTTTATTGGGA
12455_RP	TGTAACAGCTTGGAGACCAATTATG	5468_RP	AGGCGCCGATTAATAAACTT
25267_LP	GAGCTGTCTCACAAGGTCGGC	60533_LP	CACCAATTCCCCAGATTTCCC
25267_RP	GGCCAGCAGCGATCATATCAG	60533_RP	TGGATCGGTTACAGTGGCACA
8652_LP	TCGAAAACACGAAACCAATG	64348_LP	AGCGGCCACATCGTGATTA
8652_RP	AATGGTGTAGCTGTTGGGGGC	64348_RP	CGCACTCTGTCCATCTTGCT
592C05_RP	TCTTGGTGGGACACATGCAAA	60152_LP	TCTCCGCATAACTCACTCACCA
592C05_LP	ATGCAGGTCTGCTTGACGAGG	60152_RP	TTTCCGCGGGAAAAATGAAGAT
371D05_LP	GAGAAGAAAACAGAGCCAATAAAGG	64582_LP	ACCCATCAAAGCGTTCCATGA
371D05_RP	TGACTCAATATCACCAAGCTCAGCA	64582_RP	TTTGAGATTCTGGAGATTCGCTG

NOM	SEQUENCE	ID	SEQUENCE
34168_LP	AAAGCGAACTCCAGAGCAAA	64877_LP	AAGCTTTGCCATCGCATTGTG
34168_RP	CAACCACACTCGGACGATAA	64877_RP	AGCCCTTGGGAGGTAAAGGGA
064124_LP	TGGGCCATGTTAGAGGCAACT	64949_LP	AATTTCCACACGAGCAGCGAC
064124_RP	TCAGGTCTAAGGTTTCCCGCC	64949_RP	TGGAACAGCATCATCATGGCA
CRI6_LP	CTCTGCTCTGCGAGTTTGTG	68978_LP	TTCCGCTCTTTCTTTCCAGCC
CRI6_RP	GCACAATCGAGCTTCCTAGC	68978_RP	CTCTTGTTTGGCTGATGCGCT
72319_LP	AACGATGGCGCTGCATATGAT	70819_LP	CTTAAATGCTGGGTCATGGCG
72319_RP	TATGGCTTGTTTGGAGCGAGC	70819_RP	CAAATCAACCACGCACGCATA
142675_LP	CGCTCTGCATCTCATCCAACA	73095_LP	AACGACGGCGTTTTCTGCTTA
142675_RP	TGGCGAAGAAGAGAAGAAGCAGA	73095_RP	TTGAAACCAGCCAAACCATCAA
380H05_LP	CCAAGCCAACTCAGATCCACTG	73746_LP	GCTGCATCCACCTCACCATTC
380H05_RP	GGGTGCCTATGCTGTTTCATGG	73746_RP	GCATCTTCTGGGTTATGCAACG
568C09_LP	TGAAGAATGCCAAAAGCCTCG	76606_LP	CCAGATTTACTGATCTGGGTATCA
568C09_RP	TCAAATTCTGGAGTGAAGCCA	76606_RP	GAGATAGCTGCGAGCCAACCA
633F06_LP	TCTGCACATAACAAGCTGGATCCTAT	76959_LP	AATCGGTGACGTGCGAGTGTT
633F06_RP	ATCCGTACTGCCATCGCTTT	76959_RP	AGCTGGAGCCTGATCATGACG
DDU146_LP	AGAAGGGTCTCAGCGAGGAT	8247_LP	TGGATGGAGTTGGTGAGGTGC
DDU146_RP	AATCAAGCCACTGCATTTC	8247_RP	TCGATCGACACCAATCTTTGA
96272_LP	GCCGAAATCATTGTGGTCCAA	82500_LP	GGTCTTCTTCAACAATGCATGGAC
96272_RP	CTCCATGCTCACATCCTCCGT	82500_RP	TGCATGAAACTGACGACCCAA
713_D08_LP	GTACACAAACGCTGAGAAGG	85176_LP	TCGGGACTTGTGACCAAGGA
713_D08_RP	TTTTACGGCTAATTCCACAG	85176_RP	TGATTCAGATCCGTGGCGTTT
112013_RP	TTCACTTGGTACACGTCCTTG	85242_LP	TCTCTGCGTAAATCCCATTG
112013_LP	TTTCAATTTTAGTTTCGAACATGAAC	85242_RP	TATGCCAGCTCGAGGAAGGTG
131077_LP	TGCTTGTTCTCGTGTTCCTGC	86074_LP	TGATGCTCGGAAGCCAAACTC
131077_RP	GCTCCAGATGATAGAGTCCGGC	86074_RP	CGCTGCGTTTTGTGTTTTTACA
144661_LP	TGCTTGTTCTCGTGTTCCTGC	261G09_LP	CACGGCCTCTTGTAAATGTTTTG
144661_RP	GCTCCAGATGATAGAGTCCGGC	261G09_RP	GCAACACCTCCTTCTCCACCC
131079_LP	TGCTTGTTCTCGTGTTCCTGC		
131079_RP	GCTCCAGATGATAGAGTCCGGC		

#### IV.2.4.2 Analyse de l'éditotype chloroplastique

Pour l'établissement de l'éditotype chloroplastique, une collection d'amorces dérivée de celle définie par (Tillich *et al.* 2005) a été établie.

Oligo	Sequence (5'→3')	Target(s)	length
accD.AT.for	TTCATTTGTAGTGAAAGCGG		
accD.AT.rev	TTTCGCCTACTACGGATCCC	<i>accD-1</i>	468
AT7atpA	GTAATACGACTCACTATAGGGAAATTACATTAGTAGGAATA		
atpHF6	CGTATATGAGCCTCTTCAA	<i>atpA-2</i>	584
ATPHfor2	ATAAGAGGAGATTGTATGAAA		
AT7atpF	GTAATACGACTCACTATAGGGCTAATGGTACGTAAATGTAAGTC	<i>atpF-1</i>	522
clpP.AT.for	GTAATGATCCATCAACCCGC		
clpP.AT.2rev	TGAACCGCTACAAGATCAAC	<i>clpP-1</i>	220
matK.AT.for	CGTTACCGGGTAAAAGATGC		
matK.AT.rev	AGCGGCGTATCCTTTGTTGC	<i>matK-2</i>	367
ndhBfor2	TCATGATCTGGCATGTACAG		
ndhBedIII	ATTTCTTGAAGCTCAATCTCTCCCCGGAT	<i>ndhB-1,2</i>	642
nb11	TTCATGCTTGTGGAGTAATAGC	<i>ndhB-</i>	
P12	GGTCTAATGAGGCTACTATG	3,4,5,6,7,8,9,11	437
ndhBrev2	CTAAAAAAGGCTATCCTGAGC		
AndhB	GTCGTTGCTTTTCTTTCTG	<i>ndhB-10,12</i>	742
ndhDstart5'	GGTCCAAGTGATCTTGTC		
ndhD.AT.3rev	ACAGTTGAATTAATGGATCG		458
iz	AGGAATTAATTCTAACTCCC	<i>ndhD-1,2</i>	210
ndhD.AT.for	CAAGCCTAATTCTATCATAACTCG		
ndhD.AT.rev	AAGTTTATATGGTTCGAACG	<i>ndhD-3,4,5</i>	520
ndhF.AT.for	AAAACCTTCGCCGCATGTGG		
ndhF.AT.rev	GCATTTCGCTGCAATAGGTCG	<i>ndhF-2</i>	420
ndhGforM	TAGAATGGATTTGCCTGGAC		
ndhG.AT.rev	AGAATTATTGAAATGAGTTC	<i>ndhG-1</i>	491
petL.AT.for	AAATTTGGTAATTAACACGG		
petL.AT.rev	ATTTCAATTGAACTTAGGG	<i>petL-1</i>	260
psbF.AT.rev	CGTTGGATGAACTGCATGTC	<i>psbE-1</i>	
psbE.AT.for	ACAGGAGAACGTTCTTTTGC	<i>psbF-1</i>	364
rpoB.AT.for	GAAAACCAGTAGGAATATGC		
rpoB.AT.rev	GTCTCCAATTAATATTTCCGGCG	<i>rpoB-1,3</i>	544
rpoB.AT.3for	GAGGTGGGTTTCAGAAAAAGG		
rpoB.AT.3rev	TATCTGTCCTACATTCATGCG	<i>rpoB-7</i>	234
rps14.AT.for	TTATAGGGAGAAGAAGAGGC		
rps14.AT.rev	TACCAGCTTGATCTTGTTGC	<i>rps14-1,2</i>	281

## Chapitre V Références

- Abdelnoor, R. V., *et al.* (2003). "Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to muts." PNAS **100**(10): 5968-5973.
- Adam, Z., *et al.* (2001). "Chloroplast and mitochondrial proteases in arabidopsis. A proposed nomenclature." Plant Physiol **125**(4): 1912-8.
- Adam, Z. and Clarke, A. K. (2002). "Cutting edge of chloroplast proteolysis." Trends in Plant Science **7**(10): 451-456.
- Adams, K. L., *et al.* (2002). "Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts." Plant Cell **14**(4): 931-943.
- Aguettaz, P., *et al.* (1987). "Relations between the plastid gene dosage and the levels of 16s rna and rbcL gene transcripts during amyloplast to chloroplast change in mixotrophic spinach cell suspensions." Plant Molecular Biology **V8**(2): 169-177.
- Akagi, H., *et al.* (2004). "Positional cloning of the rice rf-1 gene, a restorer of bt-type cytoplasmic male sterility that encodes a mitochondria-targeting ppr protein." Theor Appl Genet **108**(8): 1449-57.
- Allison, L. A. (2000). "The role of sigma factors in plastid transcription." Biochimie **82**(6-7): 537-548.
- Allison, L. A., *et al.* (1996). "Deletion of rproB reveals a second distinct transcription system in plastids of higher plants." Embo J **15**(11): 2802-9.
- Archibald, J. M., *et al.* (2003). "Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga bigeloviella natans." PNAS **100**(13): 7678-7683.
- Aubourg, S., *et al.* (2000). "In arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants." Plant Mol Biol **42**(4): 603-13.
- Aubrey, D. N. J. d. G. (2005). "Forces maintaining organellar genomes: Is any as strong as genetic code disparity or hydrophobicity?" BioEssays **27**(4): 436-446.
- Austin, J. R., 2nd, *et al.* (2006). "Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes." Plant Cell **18**(7): 1693-703.
- Auweter, S. D., *et al.* (2006). "Sequence-specific binding of single-stranded rna: Is there a code for recognition?" Nucl. Acids Res. **34**(17): 4943-4959.
- Azevedo, J., *et al.* (2006). "Sub-plastidial localization of two different phage-type rna polymerases in spinach chloroplasts." Nucl. Acids Res. **34**(2): 436-444.
- Baba, K., *et al.* (2004). "Organellar gene transcription and early seedling development are affected in the rpoT2 mutant of arabidopsis." Plant J **38**(1): 38-48.
- Backert, S. (2000). "Strand switching during rolling circle replication of plasmid-like DNA circles in the mitochondria of the higher plant chenopodium album (l.)." Plasmid **43**(2): 166-170.
- Backert, S. and Borner, T. (2000). "Phage t4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant chenopodium album (l.)." Current Genetics **V37**(5): 304-314.
- Backert, S., *et al.* (1997). "The mystery of the rings: Structure and replication of mitochondrial genomes from higher plants." Trends in Plant Science **2**(12): 477-483.

- Barkan, A. and Goldschmidt-Clermont, M. (2000). "Participation of nuclear genes in chloroplast gene expression." Biochimie **82**(6-7): 559-72.
- Barkan, A., *et al.* (2007). "The crm domain: An rna binding module derived from an ancient ribosome-associated protein." Rna **13**(1): 55-64.
- Bassi, G. S., *et al.* (2002). "Recruitment of intron-encoded and co-opted proteins in splicing of the bi3 group i intron rna." Proc Natl Acad Sci U S A **99**(1): 128-33.
- Bassi, G. S. and Weeks, K. M. (2003). "Kinetic and thermodynamic framework for assembly of the six-component bi3 group i intron ribonucleoprotein catalyst." Biochemistry **42**(33): 9980-8.
- Bendich, A. J. (1987). "Why do chloroplasts and mitochondria contain so many copies of their genome?" Bioessays **6**(6): 279-82.
- Bendich, A. J. (2004). "Circular chloroplast chromosomes: The grand illusion." Plant Cell **16**(7): 1661-1666.
- Bentolila, S., *et al.* (2002). "A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants." Proc Natl Acad Sci U S A **99**(16): 10887-92.
- Blatch, G. L. and Lassle, M. (1999). "The tetratricopeptide repeat: A structural motif mediating protein-protein interactions." Bioessays **21**(11): 932-9.
- Bollenbach, T. J., *et al.* (2005). "Rnr1, a 3'-5' exoribonuclease belonging to the rnr superfamily, catalyzes 3' maturation of chloroplast ribosomal rnas in arabidopsis thaliana." Nucleic Acids Res **33**(8): 2751-63.
- Borukhov, S. and Nudler, E. (2003). "Rna polymerase holoenzyme: Structure, function and biological implications." Current Opinion in Microbiology **6**(2): 93-100.
- Borukhov, S. and Severinov, K. (2002). "Role of the rna polymerase sigma subunit in transcription initiation." Research in Microbiology **153**(9): 557-562.
- Boudreau, E., *et al.* (1997). "The chloroplast ycf3 and ycf4 open reading frames of chlamydomonas reinhardtii are required for the accumulation of the photosystem i complex." Embo J **16**(20): 6095-104.
- Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem **72**: 248-54.
- Brown, G. G., *et al.* (2003). "The radish rfo restorer gene of ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats." Plant J **35**(2): 262-72.
- Cahoon, A. B., *et al.* (2003). "The plastid clpp gene may not be essential for plant cell viability." Plant Cell Physiol **44**(1): 93-5.
- Carlsson, J., *et al.* (2007). "Microarray analysis reveals altered expression of a large number of nuclear genes in developing cytoplasmic male sterile brassica napus flowers." Plant J **49**(3): 452-62.
- Cavalier-Smith, T. (2006). "Rooting the tree of life by transition analyses." Biology Direct **1**(1): 19.
- Chomczynski, P. and Sacchi, N. (1987). "Single-step method of rna isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." Anal Biochem **162**(1): 156-9.
- Choquet, Y., *et al.* (1998). "Translation of cytochrome f is autoregulated through the 5' untranslated region of peta mrna in chlamydomonas chloroplasts." Proc Natl Acad Sci U S A **95**(8): 4380-5.
- Choquet, Y. and Vallon, O. (2000). "Synthesis, assembly and degradation of thylakoid membrane proteins." Biochimie **82**(6-7): 615-34.
- Choquet, Y. and Wollman, F.-A. (2002). "Translational regulations as specific traits of

- chloroplast gene expression." FEBS Letters **529**(1): 39-42.
- Choquet, Y., *et al.* (2001). "Assembly-controlled regulation of chloroplast gene translation." Biochem Soc Trans **29**(Pt 4): 421-6.
- Choquet, Y., *et al.* (2003). "Cytochrome f translation in chlamydomonas chloroplast is autoregulated by its carboxyl-terminal domain." Plant Cell **15**(6): 1443-54.
- Clifton, S. W., *et al.* (2004). "Sequence and comparative analysis of the maize nb mitochondrial genome." Plant Physiol **136**(3): 3486-503.
- Coffin, J. W., *et al.* (1997). "The neurospora crassa cya-5 nuclear gene encodes a protein with a region of homology to the saccharomyces cerevisiae pet309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded coxi protein." Curr Genet **32**(4): 273-80.
- de Koning, A. P. and Keeling, P. J. (2006). "The complete plastid genome sequence of the parasitic green alga helicosporidium sp. Is highly reduced and structured." BMC Biol **4**: 12.
- Del Campo, E. M., *et al.* (2002). "Post-transcriptional control of chloroplast gene expression. Accumulation of stable psac mrna is due to downstream rna cleavages in the ndhd gene." J Biol Chem **277**(39): 36457-64.
- Desloire, S., *et al.* (2003). "Identification of the fertility restoration locus, rfo, in radish, as a member of the pentatricopeptide-repeat protein family." EMBO Rep **4**(6): 588-94.
- Dietrich, A., *et al.* (1992). "Nuclear-encoded transfer rnas in plant mitochondria." Annu Rev Cell Biol **8**: 115-31.
- Ding, Y. H., *et al.* (2006). "Arabidopsis glutamine-rich protein23 is essential for early embryogenesis and encodes a novel nuclear ppr motif protein that interacts with rna polymerase ii subunit iii." Plant Cell **18**(4): 815-30.
- Dombrowski, S., *et al.* (1997). "3'-inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators." Embo J **16**(16): 5069-76.
- Drescher, A., *et al.* (2000). "The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes." Plant J **22**(2): 97-104.
- Dreyfus, M. and Regnier, P. (2002). "The poly(a) tail of mRNAs: Bodyguard in eukaryotes, scavenger in bacteria." Cell **111**(5): 611-613.
- Ellis, T. P., *et al.* (2004). "Aep3p stabilizes the mitochondrial bicistronic mrna encoding subunits 6 and 8 of the h<sup>+</sup>-translocating atp synthase of saccharomyces cerevisiae." J. Biol. Chem. **279**(16): 15728-15733.
- Emanuel, C., *et al.* (2006). "Development- and tissue-specific expression of the rpot gene family of arabidopsis encoding mitochondrial and plastid rna polymerases." Planta **V223**(5): 998-1009.
- Embley, T. M., *et al.* (2003). "Mitochondria and hydrogenosomes are two forms of the same fundamental organelle." Philosophical Transactions of the Royal Society B: Biological Sciences **358**(1429): 191-203.
- Embley, T. M. and Martin, W. (2006). "Eukaryotic evolution, changes and challenges." **440**(7084): 623-630.
- Escoubas, J. M., *et al.* (1995). "Light intensity regulation of cab gene transcription is signaled by the redox state of the plastoquinone pool." Proc Natl Acad Sci U S A **92**(22): 10237-41.
- Falkowski, P., *et al.* (2003). Phytoplankton and their role in primary, new, and export production. Ocean biogeochemistry: The role of the ocean carbon cycle in global change. M. J. R. Fasham. Berlin, Springer: 99-121.
- Fargo, D. C., *et al.* (1998). "Shine-dalgarno-like sequences are not required for translation of chloroplast mRNAs in chlamydomonas reinhardtii chloroplasts or



- in escherichia coli." Mol Gen Genet **257**(3): 271-82.
- Faure, E. and Casanova, J.-P. (2006). "Comparison of chaetognath mitochondrial genomes and phylogenetical implications." Mitochondrion **6**(5): 258-262.
- Fauron, C., *et al.* (1995). "The maize mitochondrial genome: Dynamic, yet functional." Trends Genet **11**(6): 228-35.
- Fernandez-Silva, P., *et al.* (2003). "Replication and transcription of mammalian mitochondrial DNA." Exp Physiol **88**(1): 41-56.
- Fernandez-Silva, P., *et al.* (1997). "The human mitochondrial transcription termination factor (mterf) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions." Embo J **16**(5): 1066-79.
- Fey, J., *et al.* (2002). "Role of editing in plant mitochondrial transfer rnas." Gene **286**(1): 21-4.
- Fisk, D. G., *et al.* (1999). "Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression." Embo J **18**(9): 2621-30.
- Freyer, R., *et al.* (1997). "Occurrence of plastid rna editing in all major lineages of land plants." PNAS **94**(12): 6285-6290.
- Fujii, S., *et al.* (2007). "Retrograde regulation of nuclear gene expression in cw-cms of rice." Plant Mol Biol **63**(3): 405-17.
- Fukui, K., *et al.* (2007). "Nuclease activity of the muts homologue muts2 from thermus thermophilus is confined to the smr domain." Nucleic Acids Res **35**(3): 850-60.
- Gagliardi, D., *et al.* (2004). "Messenger rna stability in mitochondria: Different means to an end." Trends in Genetics **20**(6): 260-267.
- Gallois, J.-L., *et al.* (2001). "The arabidopsis chloroplast ribosomal protein l21 is encoded by a nuclear gene of mitochondrial origin." Gene **274**(1-2): 179-185.
- Giege, P. and Brennicke, A. (1999). "Rna editing in arabidopsis mitochondria effects 441 c to u changes in orfs." Proc Natl Acad Sci U S A **96**(26): 15324-9.
- Giege, P. and Brennicke, A. (2001). "From gene to protein in higher plant mitochondria." C R Acad Sci III **324**(3): 209-17.
- Goldschmidt-Clermont, M., *et al.* (1990). "Trans-splicing mutants of chlamydomonas reinhardtii." Mol Gen Genet **223**(3): 417-25.
- Gothandam, K. M., *et al.* (2005). "Osppr1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis." Plant Mol Biol **58**(3): 421-33.
- Gualberto, J. M., *et al.* (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." Nucl. Acids Res. **18**(13): 3771-3776.
- Gutierrez-Marcos, J. F., *et al.* (2007). "Empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize." Plant Cell.
- Hanaoka, M., *et al.* (2003). "Molecular genetic analysis of chloroplast gene promoters dependent on sig2, a nucleus-encoded sigma factor for the plastid-encoded rna polymerase, in arabidopsis thaliana." Nucl. Acids Res. **31**(24): 7090-7098.
- Hanic-Joyce, P. and Gray, M. (1990). "Processing of transfer rna precursors in a wheat mitochondrial extract." J. Biol. Chem. **265**(23): 13782-13791.
- Hashimoto, M., *et al.* (2003). "A nucleus-encoded factor, crr2, is essential for the expression of chloroplast ndhb in arabidopsis." Plant J **36**(4): 541-9.
- Hattori, M., *et al.* (2007). "A pentatricopeptide repeat protein is required for rna processing of clpp pre-mrna in moss chloroplasts." J Biol Chem.
- Hedtke, B., *et al.* (1997). "Mitochondrial and chloroplast phage-type rna polymerases

- in arabidopsis." Science **277**(5327): 809-811.
- Hedtke, B., *et al.* (2000). "One rna polymerase serving two genomes." EMBO Rep **1**(5): 435-40.
- Hengge, R. and Bukau, B. (2003). "Proteolysis in prokaryotes: Protein quality control and regulatory principles." Molecular Microbiology **49**(6): 1451-1462.
- Herrmann, R. G., *et al.* (2003). "Eukaryotic genome evolution: Rearrangement and coevolution of compartmentalized genetic information." Philosophical Transactions of the Royal Society B: Biological Sciences **358**(1429): 87-97.
- Hilson, P., *et al.* (2004). "Versatile gene-specific sequence tags for arabidopsis functional genomics: Transcript profiling and reverse genetics applications." Genome Res **14**(10B): 2176-89.
- Hirose, T., *et al.* (1998). "Translation of tobacco chloroplast rps14 mrna depends on a shine-dalgarno-like sequence in the 5'-untranslated region but not on internal rna editing in the coding region." FEBS Lett **430**(3): 257-60.
- Hirose, T. and Sugiura, M. (1997). "Both rna editing and rna cleavage are required for translation of tobacco chloroplast ndhd mrna: A possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes." Embo J **16**(22): 6804-11.
- Hirose, T. and Sugiura, M. (2004). "Functional shine-dalgarno-like sequences for translational initiation of chloroplast mRNAs." Plant Cell Physiol **45**(1): 114-7.
- Hirose, T. and Sugiura, M. (2004). "Multiple elements required for translation of plastid atpb mrna lacking the shine-dalgarno sequence." Nucl. Acids Res. **32**(11): 3503-3510.
- Ho, Y., *et al.* (1997). "A protein encoded by a group i intron in aspergillus nidulans directly assists rna splicing and is a DNA endonuclease." Proc Natl Acad Sci U S A **94**(17): 8994-9.
- Hoffer, P. H. and Christopher, D. A. (1997). "Structure and blue-light-responsive transcription of a chloroplast psbd promoter from arabidopsis thaliana." Plant Physiol **115**(1): 213-22.
- Holec, S., *et al.* (2006). "Relaxed transcription in arabidopsis mitochondria is counterbalanced by rna stability control mediated by polyadenylation and polynucleotide phosphorylase." Mol. Cell. Biol. **26**(7): 2869-2876.
- Huang, C., *et al.* (1994). "The chlamydomonas chloroplast clpp gene contains translated large insertion sequences and is essential for cell growth." Mol Gen Genet **244**(2): 151-9.
- Huber, A. H., *et al.* (1997). "Three-dimensional structure of the armadillo repeat region of beta-catenin." Cell **90**(5): 871-82.
- Ikeda, T. M. and Gray, M. W. (1999). "Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria." Mol. Cell. Biol. **19**(12): 8113-8122.
- Imai, R., *et al.* (2003). "Delimitation of the fertility restorer locus rfk1 to a 43-kb contig in kosena radish (raphanus sativus l.)." Mol Genet Genomics **269**(3): 388-94.
- Ingman, M., *et al.* (2000). "Mitochondrial genome variation and the origin of modern humans." Nature **408**(6813): 708-13.
- Isono, K., *et al.* (1997). "Evidence for transcriptional regulation of plastid photosynthesis genes in arabidopsis thaliana roots." Plant Physiol. **114**(2): 623-630.
- Jenkins, B. D. and Barkan, A. (2001). "Recruitment of a peptidyl-trna hydrolase as a facilitator of group ii intron splicing in chloroplasts." Embo J **20**(4): 872-9.
- Jenkins, B. D., *et al.* (1997). "Nuclear mutations that block group ii rna splicing in maize chloroplasts reveal several intron classes with distinct requirements for

- splicing factors." Plant Cell **9**(3): 283-96.
- Kabeya, Y. and Sato, N. (2005). "Unique translation initiation at the second aug codon determines mitochondrial localization of the phage-type rna polymerases in the moss *Physcomitrella patens*." Plant Physiol **138**(1): 369-82.
- Kajander, O. A., *et al.* (2000). "Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states." Hum. Mol. Genet. **9**(19): 2821-2835.
- Kanamaru, K. and Tanaka, K. (2004). "Roles of chloroplast rna polymerase sigma factors in chloroplast development and stress response in higher plants." Biosci Biotechnol Biochem **68**(11): 2215-23.
- Kessler, F., *et al.* (1999). "Identification of proteins associated with plastoglobules isolated from pea (*Pisum sativum* L.) chloroplasts." Planta **208**(1): 107-13.
- Kessler, F. and Schnell, D. J. (2006). "The function and diversity of plastid protein import pathways: A multilane gtpase highway into plastids." Traffic **7**(3): 248-257.
- Klein, R. R., *et al.* (2005). "Fertility restorer locus rf1 of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12." Theor Appl Genet.
- Kmieć, B., *et al.* (2006). "Heteroplasmy as a common state of mitochondrial genetic information in plants and animals." Current Genetics **50**(3): 149-159.
- Kobayashi, K., *et al.* (2007). "Lovastatin insensitive 1, a novel pentatricopeptide repeat protein, is a potential regulatory factor of isoprenoid biosynthesis in arabidopsis." Plant Cell Physiol **48**(2): 322-31.
- Kocabek, T., *et al.* (2006). "Isolation and characterization of a novel semi-lethal arabidopsis thaliana mutant of gene for pentatricopeptide (ppr) repeat-containing protein." Genetica **128**(1-3): 395-407.
- Kolodner, R. D. and Tewari, K. K. (1975). "Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism." Nature **256**(5520): 708-11.
- Komori, T., *et al.* (2004). "Map-based cloning of a fertility restorer gene, rf-1, in rice (*Oryza sativa* L.)." Plant J **37**(3): 315-25.
- Korpelainen, H. (2004). "The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes." Naturwissenschaften **V91**(11): 505-518.
- Kotera, E., *et al.* (2005). "A pentatricopeptide repeat protein is essential for rna editing in chloroplasts." Nature **433**(7023): 326-30.
- Koussevitzky, S., *et al.* (2007). "Multiple signals from damaged chloroplasts converge on a common pathway to regulate nuclear gene expression." Science.
- Krause, K., *et al.* (2004). "The mitochondrial message-specific mRNA protectors cbp1 and pet309 are associated in a high-molecular weight complex." Mol Biol Cell **15**(6): 2674-83.
- Krause, K., *et al.* (2000). "Disruption of plastid-encoded rna polymerase genes in tobacco: Expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome." Mol Gen Genet **263**(6): 1022-30.
- Kubo, T., *et al.* (2000). "The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for trnA(cys)(gca)." Nucleic Acids Res **28**(13): 2571-6.
- Kühn, K., *et al.* (2005). "Multiple promoters are a common feature of mitochondrial genes in arabidopsis." Nucl. Acids Res. **33**(1): 337-346.
- Kunzmann, A., *et al.* (1998). "5' end maturation and rna editing have to precede trnA

- 3' processing in plant mitochondria." Proc Natl Acad Sci U S A **95**(1): 108-13.
- Kuzmin, E. V., *et al.* (2005). "A mitochondrial mutator system in maize." Plant Physiol. **137**(2): 779-789.
- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage t4." Nature **227**(5259): 680-5.
- Lahmy, S., *et al.* (2000). "A chloroplastic rna-binding protein is a new member of the ppr family." FEBS Lett **480**(2-3): 255-60.
- Lang, B. F., *et al.* (1997). "An ancestral mitochondrial DNA resembling a eubacterial genome in miniature." **387**(6632): 493-497.
- Legen, J., *et al.* (2002). "Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and pep-deficient transcription machineries." Plant J **31**(2): 171-88.
- Levinger, L., *et al.* (2004). "Mitochondrial trna 3' end metabolism and human disease." Nucl. Acids Res. **32**(18): 5430-5441.
- Lezhneva, L. and Meurer, J. (2004). "The nuclear factor hcf145 affects chloroplast psaa-psab-rps14 transcript abundance in arabidopsis thaliana." Plant J **38**(5): 740-53.
- Liere, K. and Link, G. (1995). "Rna-binding activity of the matk protein encoded by the chloroplast trnK intron from mustard (sinapis alba l.)." Nucleic Acids Res **23**(6): 917-21.
- Linder, T., *et al.* (2005). "A family of putative transcription termination factors shared amongst metazoans and plants." Current Genetics **V48**(4): 265-269.
- Lister, R., *et al.* (2005). "Protein import into mitochondria: Origins and functions today (review)." Molecular Membrane Biology **22**(1): 87 - 100.
- Liu, L. and McKeehan, W. L. (2002). "Sequence analysis of lrpprc and its sec1 domain interaction partners suggests roles in cytoskeletal organization, vesicular trafficking, nucleocytoplasmic shuttling, and chromosome activity." Genomics **79**(1): 124-36.
- Liu, S., *et al.* (2005). "Domain graph of arabidopsis proteome by comparative analysis." J Proteome Res **4**(2): 435-44.
- Lopez-Juez, E. (2007). "Plastid biogenesis, between light and shadows." J. Exp. Bot. **58**(1): 11-26.
- Lown, F. J., *et al.* (2001). "Chlamydomonas nuclear mutants that fail to assemble respiratory or photosynthetic electron transfer complexes." Biochem Soc Trans **29**(Pt 4): 452-5.
- Lukes, J., *et al.* (2005). "Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates." Current Genetics **V48**(5): 277-299.
- Lunn, J. E. (2007). "Compartmentation in plant metabolism." J Exp Bot **58**(1): 35-47.
- Lupold, D. S., *et al.* (1999). "The maize mitochondrial cox2 gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units." J Biol Chem **274**(6): 3897-903.
- Lurin, C., *et al.* (2004). "Genome-wide analysis of arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis." Plant Cell **16**(8): 2089-103.
- Lutz, K. A. and Maliga, P. (2001). "Lack of conservation of editing sites in mrnas that encode subunits of the nad(p)h dehydrogenase complex in plastids and mitochondria of arabidopsis thaliana." Curr Genet **40**(3): 214-9.
- Magee, A. M. and Kavanagh, T. A. (2002). "Plastid genes transcribed by the nucleus-encoded plastid rna polymerase show increased transcript accumulation in transgenic plants expressing a chloroplast-localized phage t7 rna polymerase." J Exp Bot **53**(379): 2341-9.

- Maier, R. M., *et al.* (1995). "Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing." J Mol Biol **251**(5): 614-28.
- Majeran, W., *et al.* (2005). "The chloroplast clpp complex in chlamydomonas reinhardtii contains an unusual high molecular mass subunit with a large apical domain." FEBS Journal **272**(21): 5558-5571.
- Majeran, W., *et al.* (2001). "The light sensitivity of atp synthase mutants of chlamydomonas reinhardtii." Plant Physiol. **126**(1): 421-433.
- Majeran, W., *et al.* (2000). "Evidence for a role of clpp in the degradation of the chloroplast cytochrome b6f complex." Plant Cell **12**(1): 137-150.
- Malek, O., *et al.* (1996). "Rna editing in bryophytes and a molecular phylogeny of land plants." Embo J **15**(6): 1403-11.
- Malik, H. S. and Henikoff, S. (2000). "Dual recognition-incision enzymes might be involved in mismatch repair and meiosis." Trends Biochem Sci **25**(9): 414-8.
- Mancebo, R., *et al.* (2001). "Bsf binds specifically to the bicoid mrna 3' untranslated region and contributes to stabilization of bicoid mrna." Mol. Cell. Biol. **21**(10): 3462-3471.
- 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**(16): 4031-43.
- Manthey, G. M., *et al.* (1998). "The saccharomyces cerevisiae pet309 protein is embedded in the mitochondrial inner membrane." Eur J Biochem **255**(1): 156-61.
- Marchfelder, A., *et al.* (1996). "Rna editing is required for efficient excision of trna(phe) from precursors in plant mitochondria." J Biol Chem **271**(4): 1898-903.
- Marchfelder, A., *et al.* (1990). "In vitro processing of mitochondrial and plastid derived trna precursors in a plant mitochondrial extract." Nucl. Acids Res. **18**(6): 1401-1406.
- Maslov, D. A., *et al.* (2007). "Proteomics and electron microscopic characterization of the unusual mitochondrial ribosome-related 45s complex in leishmania tarentolae." Mol Biochem Parasitol.
- McFadden, G., *et al.* (1994). "Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga  
10.1073/pnas.91.9.3690." PNAS **91**(9): 3690-3694.
- McKhann, H. I., *et al.* (2004). "Nested core collections maximizing genetic diversity in arabidopsis thaliana." The Plant Journal **38**(1): 193-202.
- Meierhoff, K., *et al.* (2003). "Hcf152, an arabidopsis rna binding pentatricopeptide repeat protein involved in the processing of chloroplast psbb-psbt-psbh-petb-petd rnas." Plant Cell **15**(6): 1480-1495.
- Meinke, D. and Koornneef, M. (1997). "Community standards for arabidopsis genetics." The Plant Journal **12**(2): 247-253.
- Meurer, J., *et al.* (1996). "A nuclear mutant of arabidopsis with impaired stability on distinct transcripts of the plastid psbb, psbd/c, ndhh, and ndhc operons." Plant Cell **8**(7): 1193-207.
- Mili, S. and Pinol-Roma, S. (2003). "Lrp130, a pentatricopeptide motif protein with a noncanonical rna-binding domain, is bound in vivo to mitochondrial and nuclear rnas." Mol Cell Biol **23**(14): 4972-82.
- Minai, L., *et al.* (2006). "Chloroplast biogenesis of photosystem ii cores involves a series of assembly-controlled steps that regulate translation." Plant Cell **18**(1):

159-75.

- Mingler, M. K., *et al.* (2006). "Identification of pentatricopeptide repeat proteins in trypanosoma brucei." Mol Biochem Parasitol **150**(1): 37-45.
- Miyagishima, S.-y. (2005). "Origin and evolution of the chloroplast division machinery." Journal of Plant Research **118**(5): 295-306.
- Mohr, G. and Lambowitz, A. M. (2003). "Putative proteins related to group ii intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants  
10.1093/nar/gkg153." Nucl. Acids Res. **31**(2): 647-652.
- Mootha, V. K., *et al.* (2003). "Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics." Proc Natl Acad Sci U S A **100**(2): 605-10.
- Moreira, D. and Philippe, H. (1999). "Smr: A bacterial and eukaryotic homologue of the c-terminal region of the muts2 family." Trends Biochem Sci **24**(8): 298-300.
- Muschner, V. C., *et al.* (2006). "Differential organellar inheritance in passiflora's (passifloraceae) subgenera." Genetica **V128**(1): 449-453.
- Nakagawa, N. and Sakurai, N. (2006). "A mutation in at-nmat1a, which encodes a nuclear gene having high similarity to group ii intron maturase, causes impaired splicing of mitochondrial nad4 transcript and altered carbon metabolism in arabidopsis thaliana." Plant Cell Physiol. **47**(6): 772-783.
- Naver, H., *et al.* (2001). "Functional studies of ycf3: Its role in assembly of photosystem i and interactions with some of its subunits." Plant Cell **13**(12): 2731-2745.
- Notsu, Y., *et al.* (2002). "The complete sequence of the rice (oryza sativa l.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants." Mol Genet Genomics **268**(4): 434-45.
- Nott, A., *et al.* (2006). "Plastid-to-nucleus retrograde signaling." Annu Rev Plant Biol **57**: 739-59.
- Oelmüller, R. and Mohr, H. (1986). "Photooxidative destruction of chloroplasts and its consequences for expression of nuclear genes." Planta **V167**(1): 106-113.
- Okuda, K., *et al.* (2006). "A pentatricopeptide repeat protein is a site recognition factor in chloroplast rna editing." J Biol Chem **281**(49): 37661-7.
- Oldenburg, D. J. and Bendich, A. J. (2004). "Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms." Journal of Molecular Biology **335**(4): 953-970.
- Olof Karlberg, B. C., Charles G. Kurland, Siv G. E. Andersson, (2000). "The dual origin of the yeast mitochondrial proteome." Yeast **17**(3): 170-187.
- Orphan, V. J., *et al.* (2002). "Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments." Proc Natl Acad Sci U S A **99**(11): 7663-8.
- Ostersetzer, O., *et al.* (2005). "Crs1, a chloroplast group ii intron splicing factor, promotes intron folding through specific interactions with two intron domains." Plant Cell **17**(1): 241-55.
- Ostheimer, G. J., *et al.* (2003). "Group ii intron splicing factors derived by diversification of an ancient rna-binding domain." Embo J **22**(15): 3919-29.
- Peltier, J.-B., *et al.* (2004). "Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications." J. Biol. Chem. **279**(6): 4768-4781.
- Perrin, R., *et al.* (2004). "Atmtmpnase is required for multiple aspects of the 18s rna metabolism in arabidopsis thaliana mitochondria." Nucl. Acids Res. **32**(17): 5174-5182.

- Perrin, R., *et al.* (2004). "Two exoribonucleases act sequentially to process mature 3'-ends of atp9 mnas in arabidopsis mitochondria." J. Biol. Chem. **279**(24): 25440-25446.
- Pombert, J. F., *et al.* (2006). "The complete chloroplast DNA sequence of the green alga oltmannsiellopsis viridis reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes." BMC Biol **4**: 3.
- Qiu, Y. L. and Palmer, J. D. (2004). "Many independent origins of trans splicing of a plant mitochondrial group ii intron." J Mol Evol **59**(1): 80-9.
- Raghavendra, A. S. and Padmasree, K. (2003). "Beneficial interactions of mitochondrial metabolism with photosynthetic carbon assimilation." Trends in Plant Science **8**(11): 546-553.
- Reinbothe, S., *et al.* (1993). "A methyl jasmonate-induced shift in the length of the 5' untranslated region impairs translation of the plastid rbcl transcript in barley." Embo J **12**(4): 1505-12.
- Rhoads, D. M. and Subbaiah, C. C. (2007). "Mitochondrial retrograde regulation in plants." Mitochondrion.
- Richly, E. and Leister, D. (2004). "An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of arabidopsis and rice." Gene **329**: 11-16.
- Rivals, E., *et al.* (2006). "Formation of the arabidopsis pentatricopeptide repeat family." Plant Physiol **141**(3): 825-39.
- Rotte, C., *et al.* (2000). "Origins of hydrogenosomes and mitochondria: Commentary." Current Opinion in Microbiology **3**(5): 481-486.
- Ruf, S., *et al.* (1997). "Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem i-related gene." J Cell Biol **139**(1): 95-102.
- Rumpho, M. E., *et al.* (2001). "Mollusc/algal chloroplast symbiosis: How can isolated chloroplasts continue to function for months in the cytosol of a sea slug in the absence of an algal nucleus?" Zoology **104**(3-4): 303-312.
- Sambrook, J., *et al.* (1989). Molecular cloning: A laboratory manual Cold Spring Harbor, NY, Cold Spring Harbor Press.
- Schiffer, S., *et al.* (2002). "Assigning a function to a conserved group of proteins: The trna 3'-processing enzymes." Embo J **21**(11): 2769-77.
- Schmitz-Linneweber, C., *et al.* (2002). "The plastid chromosome of atropa belladonna and its comparison with that of nicotiana tabacum: The role of rna editing in generating divergence in the process of plant speciation." Mol Biol Evol **19**(9): 1602-1612.
- Schmitz-Linneweber, C., *et al.* (2005). "Rna immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5' region of mnas whose translation it activates." Plant Cell **17**(10): 2791-804.
- Schmitz-Linneweber, C., *et al.* (2006). "A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mrna." Plant Cell **18**(10): 2650-63.
- Schonfeld, C., *et al.* (2004). "The nucleus-encoded protein moc1 is essential for mitochondrial light acclimation in chlamydomonas reinhardtii." J Biol Chem **279**(48): 50366-74.
- Schürer, H., *et al.* (2001). "This is the end: Processing, editing and repair at the trna 3-terminus." Biol Chem. **382**(8): 1147-1156.
- Sheahan, M. B., *et al.* (2005). "Mitochondria as a connected population: Ensuring continuity of the mitochondrial genome during plant cell dedifferentiation

- through massive mitochondrial fusion." The Plant Journal **44**(5): 744-755.
- Shiina, T., *et al.* (2005). Plastid rna polymerases, promoters, and transcription regulators in higher plants  
international review of cytology. A survey of cell biology. K. W. Jeon, Academic Press: 1-68.
- Shikanai, T., *et al.* (2001). "The chloroplast clpp gene, encoding a proteolytic subunit of atp-dependent protease, is indispensable for chloroplast development in tobacco." Plant Cell Physiol **42**(3): 264-73.
- Siegfried, S., *et al.* (1999). "Plant mitochondrial rna editing." Journal of Molecular Evolution **48**(3): 303-312.
- Sjogren, L. L. E., *et al.* (2006). "Structural and functional insights into the chloroplast atp-dependent clp protease in arabidopsis  
10.1105/tpc.106.044594." Plant Cell **18**(10): 2635-2649.
- Small, I., *et al.* (1989). "Evolution of plant mitochondrial genomes via substoichiometric intermediates." Cell **58**(1): 69-76.
- Small, I. D. and Peeters, N. (2000). "The ppr motif - a tpr-related motif prevalent in plant organellar proteins." Trends Biochem Sci **25**(2): 46-7.
- Steinmuller, K., *et al.* (1989). "Characterization of the ndhc-psbg-orf157/159 operon of maize plastid DNA and of the cyanobacterium *synechocystis* sp. Pcc6803." Mol Gen Genet **216**(1): 60-9.
- Stoddard, B. L. (2005). "Homing endonuclease structure and function." Q Rev Biophys **38**(1): 49-95.
- Streatfield, S. J., *et al.* (1999). "The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression." Plant Cell **11**(9): 1609-22.
- Suay, L., *et al.* (2005). "Specific roles of 5' rna secondary structures in stabilizing transcripts in chloroplasts  
10.1093/nar/gki760." Nucl. Acids Res. **33**(15): 4754-4761.
- Swiatecka-Hagenbruch, M., *et al.* (2007). "High diversity of plastidial promoters in arabidopsis thaliana." Mol Genet Genomics.
- Szyk, A. and Maurizi, M. R. (2006). "Crystal structure at 1.9 a of e. Coli clpp with a peptide covalently bound at the active site." Journal of Structural Biology AAA+ Proteins **156**(1): 165-174.
- Teixeira, R. T., *et al.* (2005). "Expression levels of meristem identity and homeotic genes are modified by nuclear-mitochondrial interactions in alloplasmic male-sterile lines of brassica napus." Plant J **42**(5): 731-42.
- Tillich, M., *et al.* (2005). "Editing of plastid rna in arabidopsis thaliana ecotypes." Plant J **43**(5): 708-15.
- Tillich, M., *et al.* (2006). "Maintenance of plastid rna editing activities independently of their target sites." EMBO Rep **7**(3): 308-13.
- Timmis, J. N., *et al.* (2004). "Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes." Nature Reviews Genetics **5**(2): 123-135.
- Tsuchiya, N., *et al.* (2004). "Lrp130, a single-stranded DNA/rna-binding protein, localizes at the outer nuclear and endoplasmic reticulum membrane, and interacts with mrna in vivo." Biochem Biophys Res Commun **317**(3): 736-43.
- Tsuchiya, N., *et al.* (2002). "Lrp130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite pc-1." Eur J Biochem **269**(12): 2927-33.
- Tsudzuki, T., *et al.* (2001). "Comparative analysis of rna editing sites in higher plant chloroplasts." J Mol Evol **53**(4-5): 327-32.
- Tullberg, A., *et al.* (2000). "Photosynthetic electron flow regulates transcription of the



- psab gene in pea (*Pisum sativum* L.) chloroplasts through the redox state of the plastoquinone pool." Plant Cell Physiol **41**(9): 1045-54.
- Unsold, M., *et al.* (1997). "The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides." Nat Genet **15**(1): 57-61.
- Vaishnav, S. and Striepen, B. (2006). "The cell biology of secondary endosymbiosis--how parasites build, divide and segregate the apicoplast." Mol Microbiol **61**(6): 1380-7.
- van der Giezen, M., *et al.* (2005). Mitochondrion-derived organelles in protists and fungi  
international review of cytology. A survey of cell biology. K. W. Jeon, Academic Press: 175-225.
- Wahleithner, J., *et al.* (1990). "A sequence encoding a maturase-related protein in a group II intron of a plant mitochondrial *nad1* gene." PNAS **87**(2): 548-552.
- Waller, R. F. and McFadden, G. I. (2005). "The apicoplast: A review of the derived plastid of apicomplexan parasites." Curr Issues Mol Biol **7**(1): 57-79.
- Walter, M., *et al.* (2002). "Pnpase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts." Embo J **21**(24): 6905-14.
- Wang, X., *et al.* (2002). "Modular recognition of RNA by a human Pumilio-homology domain." Cell **110**(4): 501-12.
- Wang, Y., *et al.* (2003). "Comparison of plastid DNA replication in different cells and tissues of the rice plant." Plant Molecular Biology **52**(4): 905-913.
- Wang, Z., *et al.* (2006). "Cytoplasmic male sterility of rice with *boro2* cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing." Plant Cell **18**(3): 676-87.
- Wheeler, D. L., *et al.* (2000). "Database resources of the National Center for Biotechnology Information." Nucl. Acids Res. **28**(1): 10-14.
- Williams, M. A., *et al.* (2000). "Addition of non-genomically encoded nucleotides to the 3'-terminus of maize mitochondrial mRNAs: Truncated rps12 mRNAs frequently terminate with CCA." Nucleic Acids Res **28**(22): 4444-51.
- Wostrikoff, K., *et al.* (2001). "Tca1, a single nuclear-encoded translational activator specific for *petA* mRNA in *Chlamydomonas reinhardtii* chloroplast." Genetics **159**(1): 119-32.
- Wostrikoff, K., *et al.* (2004). "Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of *Chlamydomonas*." Embo J **23**(13): 2696-705.
- Wostrikoff, K. and Stern, D. (2007). "Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts." PNAS.
- Yamaguchi, K. and Subramanian, A. R. (2000). "The plastid ribosomal proteins. Identification of all the proteins in the 50S subunit of an organelle ribosome (chloroplast)." J Biol Chem **275**(37): 28466-82.
- Yamaguchi, K. and Subramanian, A. R. (2003). "Proteomic identification of all plastid-specific ribosomal proteins in higher plant chloroplast 30S ribosomal subunit." Eur J Biochem **270**(2): 190-205.
- Yamaguchi, K., *et al.* (2000). "The plastid ribosomal proteins. Identification of all the proteins in the 30S subunit of an organelle ribosome (chloroplast)." J Biol Chem **275**(37): 28455-65.
- Yamazaki, H., *et al.* (2004). "PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*." Plant J **38**(1): 152-63.

- Yang, D. H., *et al.* (2001). "The redox state of the plastoquinone pool controls the level of the light-harvesting chlorophyll a/b binding protein complex ii (Lhc ii) during photoacclimation." Photosynth Res **68**(2): 163-74.
- Yesodi, V., *et al.* (1995). "Involvement of two different urf-s related mitochondrial sequences in the molecular evolution of the cms-specific s-pcf locus in petunia." Mol Gen Genet **248**(5): 540-6.
- Yoshinaga, K., *et al.* (1996). "Extensive rna editing of u to c in addition to c to u substitution in the rbcl transcripts of hornwort chloroplasts and the origin of rna editing in green plants." Nucl. Acids Res. **24**(6): 1008-1014.
- Zaegel, V., *et al.* (2006). "The plant-specific ssdna binding protein osb1 is involved in the stoichiometric transmission of mitochondrial DNA in arabidopsis." Plant Cell **18**(12): 3548-63.
- Zeng, W.-H., *et al.* (2007). "Identification of rna editing sites in chloroplast transcripts of phalaenopsis aphrodite and comparative analysis with those of other seed plants." Plant Cell Physiol. **48**(2): 362-368.

## Chapitre VI Annexes

### VI.1 Données supplémentaires de l'article 2, Andrés C., Falcon de Longevialle A., Lurin C. and Small ID. Rapid low-resolution transcript mapping for identifying mutants affecting organelle gene expression in *Arabidopsis thaliana*

**Tableau 13 : Tableau S1 amorces chloroplastiques**

Ce tableau répertorie, la position dans la plaque 384, la position du primer forward sur le génome, le nom du couple, la taille du produit de PCR, et les séquences des amorces

plaque	génom	nom	PCR	FORWARD	REVERSE
A1	7	CHLORO1	491	GAACGACGGGAATTGAACC	TTGGGCTGATATTATTAACCGTG
A2	850	CHLORO2	485	TGTGAAATGGGTGCATAAGAA	GATCCCTACCTTATTGACCGC
A3	1263	CHLORO3	491	TCAATATCTACTGGAGGAGCAGC	ATCCGACTAGTTCCGGGTTC
A4	1700	CHLORO4	478	TTGAATACGATCCGTAGTGGG	GGGAAGACCAAGTCTTTCTTTA
A5	2523	CHLORO5	407	TTTCTAGCATTGACTACGTACCAC	TGTATGTGAATACGAATCCATCTTT
A6	3321	CHLORO6	429	GGTTCACATCCTTAGTGGGAA	AACGAACAGAAACCTATTAAGATAGAG
A7	3686	CHLORO7	476	AGATACCTCATAAACAGGTAGACTGC	TGCTTCTGATCGGTACAAT
A8	4090	CHLORO8	488	TTCACCTGACCCAAATTGGAA	TCATTGTATGAACATAATGAACCAAGA
A9	4541	CHLORO9	460	AAAGCTTAACCTCTTGGTTCATTAG	TCTACTCTTCTTTCACTTCCATCA
A10	4966	CHLORO10	392	AAATAAATATGATGGAAGTAAAGAGAA	TTCTCGAGCCGTACGAGG
A11	5335	CHLORO11	445	TTTGGCCTCGTACGGCT	TTTGTATAGAAAGAACAGAAATCATAAAACA
A12	5724	CHLORO12	444	AAAGGGCAGCAACAAGCC	CGATGTGGTAGAAAAGCAACG
A13	6114	CHLORO13	483	CCACAGCAGATAATGTCTTCA	CAATCTTCTATTGCCTTTGTACACC
A14	6551	CHLORO14	461	TTTAAACAGAAAGAGAAAGATGGTG	AAATTGGTGTCTTGTGGATAATTT
A15	6982	CHLORO15	412	AAATTAATATCCACAAGAACACCA	TCTAATTAGGTTCTTTCATTAGGGAA
A16	7316	CHLORO16	492	GGTAGCCATACTAAATCTGGATCA	GGGTTGGAATCCCTCTCTTT
A17	7800	CHLORO17	510pb	TCGAACCTCGGTACGATTA	TTTGAGTTCGCGCCTTTTT
A18	8239	CHLORO18	484pb	GAACCTTCTCAGCAATCTAGTTACATT	AATCAAACCGAGAGACCCTTT
A19	8630	CHLORO19	429	CAATAATTGCATTGTAGCGGG	CAATATAATATGTCGATGTCTTTAGTAAACC
A20	9019	CHLORO20	430	GAGAAGCTTGGTTACTAAAGACATCG	GGCAATTGTGAATTTAGAGCG
A21	9380	CHLORO21	400	AAACCCTTAGCCTTCCAAGC	AAGCATTGAGAAATTTCTTTCTTATTCT
A22	9749	CHLORO22	452	AAATAGAATAAGAAAGAAATCTGAATGC	TGAAACAATCCCAATCAGCC
A23	10596	CHLORO23	449	CCTGGATAAGCTTCACGACC	AGACGTTCTGTATATGAGCCTCTTC
A24	11404	CHLORO24	479	ATTGCTCAATACGTTACAGG	GAAGAAGTTCGCTGAAGGAGC
B1	11812	CHLORO25	423	TCCGTTTCTACGTTACGCAAG	AAAGAACAACCTTTGCTGACAATTA
B2	12177	CHLORO26	481	TTCGGAGGATTCTTCTGACC	GTGTGTGTGAGTTGTTCAATTTCAA
B3	12624	CHLORO27	507	CAGCCTATTCTTGAATGAACAAC	GATTGCTTTCTTCTCTCC
B4	13111	CHLORO28	464	GGGAAGGAAGAAAGCGAATC	TTACGTCCTAAGTCATTGGATGAT
B5	13433	CHLORO29	420	TCCAGGTCCAATAGAAGCAAG	TTTAGATTAGATAGGGATTGATTAGATA
B6	13816	CHLORO30	377	GGAAGAAGTCTAATCGAATCCC	CGGAAATATATTAGCCGATGAA
B7	14120	CHLORO31	499	AGGTATAGGAACCACTAAAGTACTAAAGA	CCTATTAGGTTCCGAGTCTAGC
B8	14561	CHLORO32	483	GCCGTCAGTTGGAATTGTTT	GGGCCGTTCTAGCTATATACAAA
B9	15021	CHLORO33	446	TTTGTATATAGCTAGAAGCGCCC	GGCTGCAATACGGGCTC
B10	15399	CHLORO34	500	ACCAATTTGTTAATACATGCCG	TTGTTTGTGTAGTCAATTTGCTT
B11	15852	CHLORO35	459	TTTATTAGTGTAAACAATACGAAAGCA	GCGTGCTTTAGAAGAAGCAA
B12	17948	CHLORO36	496	GATCAACCCGCCGACTT	TTCAGAATTGAATGGAACCTTAGG
B13	20006	CHLORO37	492	TCGTTTAAATAGCAGTTCCATCTATCA	TAGTCTTTGTGGCTTCGGT
B14	20855	CHLORO38	487	GGATGTAACAATAAGTGCCTCC	GGGAATTAGTAATGTGTCAGGAAA
B15	21666	CHLORO39	436	TTGCAAACTAAATCGGCT	CACTAAGAATTTAAATAGAATCAAAGAAG

Ecole doctorale "des génomes aux organismes"

plaque	génom	nom	PCR	FORWARD	REVERSE
B16	22023	CHLORO40	475	TCAAGTTCCTCAGTCAAACCA	AAAGGCGATTGGTCTATGGTC
B17	22440	CHLORO41	464	TTCACGAGGTACGCCTATCA	TTGGACCTATAAAGAGTGGAAATTTG
B18	22844	CHLORO42	478	TTTCATCTCCGATCACTCGAT	TTTACAAGAGATGCTTACTTATAAACTCGA
B19	24487	CHLORO43	455	GTTGATACATAATTAATGGAATGCTTAGAG	AGCACGGATACGGATGCTTT
B20	25302	CHLORO44	490	GCTCCGGAGATAGTTCCCTT	TTCTAGTTCTATCATCAGCTATGGG
B21	25720	CHLORO45	492	AAATTCAGGGTAGCAAACATTC	AAATTGAAGATATAGATCACGAAATTGA
B22	26140	CHLORO46	455	TCAGAGTTCTACCAATTGATATGTT	AAACTCCGTAAGTTCCAGCG
B23	26538	CHLORO47	456	AAGAAATAGATCCAAATGGAACAGA	TCCCTTGAAGAGAACATTTCC
B24	26924	CHLORO48	471	TTGTGGTACAGTTCTATTTGGGA	CTTACCACCTCGGCCATGC
C1	27377	CHLORO49	408	GCATGGCCGAGTGGTAAG	AAATGAAAGGCAATTCTTAATCTAATC
C2	27716	CHLORO50	471	CTCATGAAAGTCTCGGAATGC	AAGTATTTCTAGAGTCCACTTCTCCC
C3	28110	CHLORO51	460	TGGGCTGCTTAATGGTAGTTT	AGCCATTATGGTTATTGTAATGTATAAAGA
C4	28531	CHLORO52	437	TCACTTGTATCTTTATACATTACAATAACCA	CTCCGATATTGTTATTCATGATATTTCT
C5	28890	CHLORO53	436	TCGCTTGTAACTTACTCAGATGAA	TTTACCGGACTCGTCTTTG
C6	29286	CHLORO54	398	AACAAAGAAAGGGAATGTGACAA	TTAACTAGGGTAGAGAAACATATGTAATGA
C7	29608	CHLORO55	457	CCCATACAGATCTCTTTCTGAAGT	TCAATTGATTTGATAATAATGTAAGGATT
C8	30004	CHLORO56	467	CACHTTGTAGTACAGCAAGACGG	GGGTAGGGTACTACGAAAGGAA
C9	30398	CHLORO57	500	CATCGACCCAGGAAGAATCTA	GAGGCAATCTATCAAGGAAAGTAAA
C10	30876	CHLORO58	422	ACTTTCCTTGATAGATTGCCTCA	ACAATAGACGCATAGTGGTTAGTAGC
C11	31233	CHLORO59	473	TGAAGTCAATCAAAATCTCATTTT	TTTGGATTTGTCGATTTCTAATCT
C12	31631	CHLORO60	500	GTCGTCTCTTGAATCGCCA	AGTTTCTTCTCACATACACCTTCG
C13	32068	CHLORO61	486	TCCATATTGATGCATCCGAG	TTTCATTTAATTGGTTTCGAAGAT
C14	32489	CHLORO62	488	TCCGTTTACCTAAGTAAGGACCA	GAATGCGCTAAACTATTAGCAGG
C15	32904	CHLORO63	499	CCAGTTCCTATTTAGAAGTTGC	TTTGACCATCACCATCTTC
C16	33334	CHLORO64	497	GGCTCTGCTATGCGCTATTC	GGCCACCAAGCGAA
C17	33769	CHLORO65	487	TTAATGGAACHTTAGCTTTAGCTGG	GCCTTGAATACTAGAAGAAAGGCA
C18	34217	CHLORO66	453	TTTGTAGGTGTAGGTGCCTTTT	AAAGTAAATGCTTGAGCTTGAGAAG
C19	35024	CHLORO67	491	TTTCTATTCTGGGTCATTT	AGACCTATTAAGAAAGTACACCC
C20	35441	CHLORO68	493	AGAACATAGCCCTATGAGTTAATACGA	GATAAGAGAATTAAGGATACCCACCA
C21	35938	CHLORO69	499	GAATTCATTCGTTGCAGATCC	ACGAGATACACTCTTTGGTTTGG
C22	36414	CHLORO70	492	CCAAACCAAAGAGTGTATCTCG	TTCTATGATCGATGAGGCC
C23	36700	CHLORO71	482	AAAGTTGCGGAGACAGGATTT	TCATTTGATTCGTCATCCTC
C24	37117	CHLORO72	477	TCCATTTCTCACTTAGCGACG	TTCTTAATTTCTGGCGTGG
D1	38422	CHLORO73	450	AGCTCCTGTCATGATGAATCC	CCCGACTCGAGTAGTCATTTATTT
D2	39234	CHLORO74	490	CTATATTACCAGGCGCAAGA	CTCAGCCTAGACCTTGAGC
D3	40509	CHLORO75	489	AAACTGTGGAAGCTAGAAATATACA	GTTAACCGATATAGCACATCATCA
D4	41728	CHLORO76	475	TGGTAGTATCAGGTCCCTTAGCTATT	AAATAGGCTACATTTCCGAACTTAAT
D5	42147	CHLORO77	477	AACCTAATCGAGTGTATTCATATTTCA	AGCACAGAATTTGGTTGACGA
D6	42574	CHLORO78	477	AACCTAATCGAGTGTATTCATATTTCA	AGCACAGAATTTGGTTGACGA
D7	43028	CHLORO79	428	AAAGCATGTCCACTGTTATTAGATATTAGA	AGGGCTTTCTACATAAGCATCG
D8	43384	CHLORO80	465	CGATGGATGAAGTTTAGTTATTTGC	TTTCATTATGCACCTAAAGGAGG
D9	43775	CHLORO81	498	AACGAACTCTCCGTTTCAGAAC	TTTACCTATGGTTCCCGTATTACA
D10	44202	CHLORO82	485	CCCTATGCCTCTATTCTGATATT	AGTTCGTACTATACAATTCCTTTGTTG
D11	44652	CHLORO83	453	TTTCTCACAACAAGGAATTGTAT	TTTCATGATCATAGATTGGTAGGGA
D12	45053	CHLORO84	476	GGGAAGAGCAAACAAGGAT	GCTTTGACTATTCCTCAAGCG
D13	45486	CHLORO85	451	GTCCATGATTAACATAATTGGCG	TCTTTGATTCCAATAGTGGAGTGTATT
D14	45877	CHLORO86	462	ACTGAACTAATGATAATGATAAATGACG	TTGATACGATCTAGTAGTTTAGTAATTTG
D15	46269	CHLORO87	464	CTCTGAGCTAAGCGGGCTC	TTTGTATTTGCAATTCATATGGTT
D16	46632	CHLORO88	498	TCGACCGTTGACTATTTCTG	TGAACACAAGGTAGTGAACTCCA
D17	47065	CHLORO89	424	AAGGGATAGGTGCAGAGACTCA	GGGATAGAGGACTTGAACCC
D18	47459	CHLORO90	414	AAATCGTGAAGGTTCAAGTCC	TTTCAGTCTCTACTAATGGTGACA
D19	47798	CHLORO91	458	CCTTCGTAATGGTGCAGATAG	AAATCCTCTGCCAAGAACCA
D20	48185	CHLORO92	484	CTCAGTTGGTAGAGCAGAGGACT	AAATGAGCCAATAGGATGAACATAA
D21	48593	CHLORO93	500	TTGACTAAGTTATGTGCGCGA	TCGATTATCAAGGAATAGAGACTTTACA

plaque	génom	nom	PCR	FORWARD	REVERSE
D22	49027	CHLORO94	487	AAATTACAGCAATAGAATGCCAAT	CTAAACCAGAGGCTGTATAGACG
D23	49439	CHLORO95	498	TTGATCCTTATAGATTTCTCTAGCTATT	TTATGAATCCATTAAGTTCCCA
D24	49864	CHLORO96	462	TCATTTAACGTAGTTGAAATAACGGA	TTTACTAATATCAAGTGCTATTCTGTTT
E1	50277	CHLORO97	461	TCCGGAAATTAGAAATGCCA	AAGATGAGTATGGTAAAGTCTTACTGGA
E2	50707	CHLORO98	426	AAATCCAGTAAGACTTTACCATACTCATC	TTTGATTCTTTATTTCTAACAGGTTACTTT
E3	51051	CHLORO99	431	CAGGACTTATGGTTTATGATATGGAA GAATTCATAACAATACAGAAGAATACTACCA A	AAAGCAAACCTACGTCAAGAGAA
E4	51411	CHLORO100	481	AAACTTGGCGCACGTGTA	AAATGTGTATCACAACAACAAGGG
E5	51819	CHLORO101	493	AAACTTGGCGCACGTGTA	CTCTCAGACGAGCTCGGACA
E6	52266	CHLORO102	470	CAAATCGTATTGAGAGCCTCG	TTACTTAGTGGTAACATCGATGAAG
E7	52659	CHLORO103	487	GTCATTTCTCAATTTACTCTCCATT	GCACCTGCCACCACATT
E8	53077	CHLORO104	499	ACCTTTGGCAGCTAATCCTCT	TGGTGTATCCGTATTTGGTGG
E9	53937	CHLORO105	424	CTCGATTATTTCTAATAATTGCTGT	AACTAGGTTTCAGTTAGTTGAGCTTGA
E10	54332	CHLORO106	435	TTTCAAGCTCAACTAACTGAAACC	TTCTTTCATATGTATAGCGCAACC
E11	54731	CHLORO107	461	AAAGAAGTATTAGTTGCGCTATACA	ATCAAGGCTGGTAAGCCCAT
E12	55552	CHLORO108	491	TTTACCAAAGATGATGAGAATGTGA	ATACCGCGGCTTCGATCT
E13	55974	CHLORO109	486	GGAGTCAACTTTGGGCTTTG	AAAGATTGAGCCGAGTGCAA
E14	56601	CHLORO110	481	TGTTGTGTCCACAAGAAATCCT	TTTCCATAGAGCTTCTGGCCT
E15	56931	CHLORO111	389	TCTTAGCAATTGCATCGATATGTT	TCATCCGATATGAAATTGCG
E16	57671	CHLORO112	477	AAATTACAAGATAAGAATAAGGAATCG	AGAGGTAACATTGATTGGTAGCAT
E17	58453	CHLORO113	390	TGGATGCAATTGTACCACGTA	AGAATCTGATCTAACACACGGGAA
E18	58774	CHLORO114	472	TGCAGATCTATAGAGAGTCTCCATC	AAATGGGTACCTCAATTTAATTTGT
E19	59188	CHLORO115	477	CTTTAGAATTCAGAATAACAATAACAGGT	TTACCTTAGGAGTAATTCATAGAATTGTTTA
E20	59632	CHLORO116	417	TATAAACAATTCTATGAATTACTCCTAAGGT	CATAACCACTACCCACATTTCCA
E21	59976	CHLORO117	500	TATGGGATCGCAGGTCTCTT	AATTTGAGACAAATAAATCTCTT
E22	60451	CHLORO118	423	AAAGGAGTTTATTTGTCTCGAAATTAG	CGCATTGTCTAGTATCCACCA
E23	60801	CHLORO119	463	TGTTGATCTCTCTGTCTGTAA	CAATCATTAGTCCCACCGT
E24	61220	CHLORO120	456	ATGTATCGGATTCATTGCG	AGGTATTTCTAGTTGCATGGTCC
F1	61607	CHLORO121	482	TCGACCTAGCTAACTTTATTGTAGAAA	TTTATTGGGACGATAATTCTGAAA
F2	62019	CHLORO122	487	CGATCGTATTTCCACCGAGA	TGGATCTTGAAGTACTATTTCCGC
F3	62438	CHLORO123	427	CATTAACGAGTAATCCTAATGTGGG	TCAATTTCAATTGGTCAATAACG
F4	62961	CHLORO124	496	GGGAACCATAAATCGGCG	GGGTCCCGTTGAACTGCTA
F5	63511	CHLORO125	500	ACGTCTATAACTCAGCTCATCCAA	CATGGACTAGTGACTACCTACCGT
F6	63939	CHLORO126	498	GGTTTATCGTGGATGAACTGC	AAATACCCTTGGTACTTTATTGACG
F7	64387	CHLORO127	457	CACTGAGATTGCATCTTTGTGAG	TTCTTGTCTGCAAAATTAATCTCT
F8	64815	CHLORO128	363	TTTAAGGAATTAATTTGCAGAAACA	GAGGTATCCTACACGGAGAACG
F9	65151	CHLORO129	428	AATCTGTTCTCCGTGTAGGA	TTGTGAATTACATCAATATAGGTGG
F10	65501	CHLORO130	458	GCCAAGACGATGAATTGAGTT	GAGAGGTACTATCTATTCTAGCTCTTAATC
F11	65904	CHLORO131	464	TCATTGAGATTCATTGGCAATAC	CGAGTCGGGTTAAAGAATGC
F12	66332	CHLORO132	446	AAAGGCTAGAAAGAGGGCATT	AGAACGATCTTATCGATTCCATT
F13	66764	CHLORO133	352	GATAAGATCGTTCTAGTCGACAATATAAA	CCCATTGTTGATTGTATCTCT
F14	67092	CHLORO134	419	TTAGAGATACAATCAACAATCGGG	CGAACATCTTTACCCCTTGGC
F15	67452	CHLORO135	364	AAGATTGAGGTGAATCATAAATCCA	AGAATATCGGATTAGATTTCTTCTATTG
F16	67483	CHLORO136	494	GTTTATCGCCAAGGGTAAA	GGGCAAACGCCCTACGAA
F17	68055	CHLORO137	455	TTATCTAGACGGGTGAATAGAGTGA	GGATTGGAAGAAATCCACGA
F18	68451	CHLORO138	418	CCCAGACTCATTCTCCTTAG	TCAATGACTAGAATTAACCGCG
F19	68795	CHLORO139	472	AAGCTTGATGCAATAAACGAA	GGGATCTCATTGTGAAAGACCT
F20	69196	CHLORO140	498	CTTGTAACAATCACATTCTTTGGC	CAAGACAGCCAATCCGAAAC
F21	69623	CHLORO141	484	ACATGTTCTCGTCTGCTGA	TTTATGAGGCACAAACGGGA
F22	70037	CHLORO142	474	GTGATGGTTTCGCGAAGTTT	TGAAAGAGTAGTATGAGATAAGGAAGAGG
F23	70431	CHLORO143	484	AAGTTTGTGACGCTGAAATGTG	TTTAGCCAAGAGGTTGATACC
F24	70872	CHLORO144	469	AATAAGTTGATTGAGATTTTCGGT	TTGGTTCATTGTCTATATGTGC
G1	71315	CHLORO145	446	TTGCACATATAGGACAAATGAACC	TCGATCGAGATATCCTCTGTTTC
G2	71688	CHLORO146	498	TTTGGCCTTCGCGTTAC	TCCGATAAGTACCAATATGCAATG
G3	72119	CHLORO147	478	ACAACAATAGCTATCTTGGTTCATAGA	CCCAAGAATTGGTTATTCCTAAA

plaque	génom	nom	PCR	FORWARD	REVERSE
G4	72954	CHLORO148	498	CGGGAGGAATAGCCTCTCAT	TAGGCATGCGGCGTACA
G5	73780	CHLORO149	497	TGCTAGAACCCTCTCAGAGATG	AGATCCATTGAGGAACACG
G6	74223	CHLORO150	482	AAATATTGGAGGCTGATTACGTT	AATTCAGTAAATCCATCCAG
G7	74668	CHLORO151	455	AATCTTCTGTCTACTGGATGGAA	AAATACAAAGAAGCGTTTGTCT
G8	75091	CHLORO152	381	TTGGATCAAAGGACAACGC	AATGCGAAAGCTCGTCTTTG
G9	75402	CHLORO153	495	AACACAACATAAAGACGAATGAGC	TGATCGGATTAACCAACCAA
G10	75862	CHLORO154	449	TGACTGAAGCTAATTTGGTTGG	ACCTAGAATCTATATTATTATAAGGGACCA
G11	76240	CHLORO155	462	TGTTAATGCATTTCTAATGATACG	TGTTCCATCCCTTAATCTACTGG
G12	76629	CHLORO156	499	TTTCTATGATCGTACCCGACG	CTCGAGCAAGAATTAACCAAC
G13	77102	CHLORO157	426	TTTAGTTTGGTAATTCTTGCTCG	AAGGGTACGGTTAATAATCCCG
G14	77563	CHLORO158	498	GTGCTCCAGTAGCGACAACC	CCAATTAGAATTGCTCCCA
G15	78259	CHLORO159	491	GGAGTTAAACTTCCATTTGTCCA	GCGATGCGAAGAGCTTACT
G16	78709	CHLORO160	470	CATGTTCTTCTATTTGCGCA	GGAACACCCCTTGCTGCC
G17	79112	CHLORO161	487	GCATACCTTGATCCACTACTGCT	AAATAAGGGCTTCCGTTCTG
G18	79545	CHLORO162	379	TCTGCGGATTAGTCGACAC	CTTAATCACTTCCGAACGGC
G19	79892	CHLORO163	441	AAATCGAACCATGCCGTTT	CGTAAGGAACATCGCGAAA
G20	80262	CHLORO164	500	CTTTCTATTTCGTCTATGTCTTAGGG	ATCCACGGGAATTGAGAC
G21	80739	CHLORO165	372	AATTGTCTCAATTCCTGTTG	TTTACTTCGTCTTTGCATTA
G22	81037	CHLORO166	500	CATTCAAATAGTTTGTGTTGA	TCTTCTCGTGGGAATCGATTT
G23	81464	CHLORO167	499	TAGATGTATCCAAGCGGGT	ATTGAAAGAGTAAATATTCGCGG
G24	81937	CHLORO168	409	CGCCGGCGAATATTTACTCT	CCAATTTGGAAAGTCTATGAGGTC
H1	82273	CHLORO169	402	TTCTGAATGAATCTGTATTGATGCT	AAACTCCTTTGATATAATTGCTATGC
H2	82608	CHLORO170	430	TTTATCTTTATCTAATCCTAAACCAACGA	TTGAATTAAGTGAACAGGCGAA
H3	82971	CHLORO171	494	TTTCCGTCGATACGTCCTG	CCACTTGGTTTCAGACTTGG
H4	83406	CHLORO172	481	GGTTGTGAAACCAAGAGAA	ATGTCTGCCATAAAGCACG
H5	83825	CHLORO173	402	GCTTCTTCATAGAACGTCCA	GAGAAGCTTAATACAAAGGCGG
H6	84157	CHLORO174	486	TGGGTATAATGGTGGATGCC	TTACCTTCTGGAGAGGTCCG
H7	84572	CHLORO175	500	CCAACATTTCCACTTGTCC	GAAGTATCGACGTAATTTATAGAGTCA
H8	84998	CHLORO176	479	CGTTCGTCATCTGGCTTAT	AAATGGGAAATGCCCTACCT
H9	85419	CHLORO177	487	TCCAATTACGTAATCAAATAGTTCAA	AACCGGTTATTCTATTCCACC
H10	85880	CHLORO178	449	AAGAGGTGGAATAGAATAACCCG	AATTCCTACTGGATGCACGC
H11	86259	CHLORO179	491	TTCCATTGATACAGAGCCAATTC	AATACGATATAGAAGGGCCGC
H12	87961	CHLORO180	475	TTTCTCTGACAGATGGTCAGAACT	AACCTCCTCAAGAATTGTCTTTG
H13	92930	CHLORO181	499	AGGTCTTCCAAGAACAAGGC	TCCAATTAATCCCTATTTCTATGG
H14	93363	CHLORO182	476	ACAGGAGAAGGGTTTCCAT	ACAATTAGAATTAGGCTCGGTCA
H15	93789	CHLORO183	474	CCCTATATGATAAGACACATTCCA	GAATAAGTTGCGCAGCGGAT
H16	94206	CHLORO184	486	CCATTCATTAGATAGAGAAGATCGC	TCCCTACGTCGTTACATTCCA
H17	94640	CHLORO185	477	AGGTGATTTCCGCTATTCC	TGACTGGACGAACCAAGAA
H18	95069	CHLORO186	477	TTCTATAATTTGCGATGTGAGGG	TCATTGCTATTACTCAAACAAGCA
H19	95481	CHLORO187	489	TCCGATTTGACCTATGGACG	GTGCTCCTACGGAACCAAGG
H20	95899	CHLORO188	495	GGAATCCTTCATCAGTGGTTG	GAAGGAGTGCAGTTGCTTT
H21	96342	CHLORO189	470	TCTCAGAGATAGATAGAGAGGTAGGAATTT	CATTGAATGTACAGAAATGGCTATAA
H22	96739	CHLORO190	496	AAACATTCCTCTAGAGTAGCTGTTAAT	TTCCGTAATATTCCATTCCAAA
H23	97161	CHLORO191	493	TCATGAATTAGTAAGTGCATCAA	CACCTGCCATTCGTTGGTTA
H24	97591	CHLORO192	486	TGAAAGCCATATTTGACCC	CGTATGTACGGTTTGGAGGG
I1	98017	CHLORO193	489	CCCATATTGTAGGGTGGATCTC	CTAGAACATGTGAAGTGTATGGCT
I2	98430	CHLORO194	494	GTCTCGTGGTAGCCTGCTC	GAGGAGATTGCGGAACGA
I3	98855	CHLORO195	498	ACCTCATACGGCTCCTCGTT	AATCTGATTGAATTTATTTGTAACCC
I4	99310	CHLORO196	440	TTTCTCATTGAGCGAAGGG	CGGTTCCGTTGCTACTGC
I5	99708	CHLORO197	417	GATCCGATCAATTCGCTAAA	TCATAGTAATAGAATACATGTCTACCCG
I6	108302	CHLORO198	457	GGGCTTGATGCTCAGAGGATT	TCTTTGATGCGATCATCGATT
I7	108698	CHLORO199	470	ATTCATATGTTTGTCTCGGG	GGGTTACGAAGGAAACTTCGAG
I8	109098	CHLORO200	496	AGATCTCATAGAGTTCAATCCCG	TGATACTTTCTCTCGGTTCTTCT
I9	109928	CHLORO201	413	TTGTTGGTTGGTTAATTTGGTCA	TGTTGCAACGTTGGAATTCAT

Ecole doctorale "des génomes aux organismes"

plaque	génom	nom	PCR	FORWARD	REVERSE
I10	110271	CHLORO202	403	TGAAATGGAAAGAAATCCGAGT	TTTGTATATAAATGGTCATATAATCGTGG
I11	111371	CHLORO203	495	TCATTAAGAATTCATCTTTCGACC	TCCCATTTCGGCTCTTATACA
I12	112275	CHLORO204		GGTATCCTTGATCATGCGACATA	TCAATATTCATGGATCATCCCTTT
I13	112554	CHLORO205	454	AAAGGAGAAGTCCAACCTCTAGTAAA	TTTGTCTGGAAATCTTTGTTC
I14	112979	CHLORO206	455	AAGAATAGAACAAAGATTTCCACGA	AATTCAGTGGTATAATTGGATGTCA
I15	113402	CHLORO207	391	TGTCTTTGACATCCAATTATACCAC	TGTCTCAAGGTAGTGATGGGAA
I16	113758	CHLORO208	350	GGGATTCGTATTGTCCCATC	AATTAAGTAGTTGGAACCTCAACCC
I17	114073	CHLORO209	458	TTTGCTTTGGGTGAAGTTCC	TTGTCAGCACAATTGAAACGA
I18	114873	CHLORO210	493	GGTTAATGATGCACGTAAGTATGA	AAACCGATCAAAAGCCACAAT
I19	115331	CHLORO211	462	GGTATAAATTCTGCAATTGTGGC	CCCAGAGTTATTTCTTCAATC
I20	116139	CHLORO212	464	AGCGCCAATAAATCCATGAG	GTTGCAACGAACCAACATTA
I21	116530	CHLORO213	469	AATAGTATTTCAAGTGAACAGGATAGGA	TGGACGATCCATTAATCAACT
I22	116953	CHLORO214	473	TCGATCCATTATAATCTTCGGA	GACTGTGTTGGTTGAAGAGATGTG
I23	117350	CHLORO215	496	CATGCCATAAATAAATCGAACG	AAATCAATTCGCATAAACCAATC
I24	117817	CHLORO216	362	AAGGTCGATTGGTTATGCG	TCTAGTTGATGGAATCCGGTG
J1	118107	CHLORO217	456	CATTGAGCAACTCCTTATCAATTT	CTCGATACGTCATGGTACGG
J2	118492	CHLORO218	419	CTTATTAATCTTGCTCTAGAATCTGGTT	ATAATGGATTGCCTGGACC
J3	118846	CHLORO219	493	ACTAGAAGACCGGATCCAGAG	TGGGTCGTTACCAATGTCA
J4	119269	CHLORO220	493	TTGTTTGGAGTGAATCCAAA	TGCTTCTATGATAACCGGG
J5	119694	CHLORO221	465	ACCAATATACCTTGACGCTCG	GTTGCTTCTATCTAAATCTTAATTTCC
J6	120079	CHLORO222	498	GGAATAGAAATATCCAACCGC	GAATTTGAGCAGACAGAATCCA
J7	120535	CHLORO223	435	GGGTCGCAATATACCAATG	GGGCTTAAATAGTAGAAATGATCAGA
J8	120891	CHLORO224	371	ATTTGTCGGTGATAGGAGC	TGTACAAAGGTTAAACGGAAATAA
J9	121212	CHLORO225	460	GGGTAAAGTGATTCTACACTTATGTTTATT	TTCAGTAATTCCTTTAGCAATCA
J10	121616	CHLORO226	463	AAATACCGATATTGAGATCCGC	TGAAATGGCTGATATTATGACG
J11	122409	CHLORO227	498	AATAACGAGCTAATGAATCTCCTTG	AAAGGGCCAGCTATATCAGAGT
J12	122836	CHLORO228	400	GCCGAGCCATAACAGATGAG	TTCAGAAATCTAATTCATGCCAGA
J13	123171	CHLORO229	384	TTCTGTAACTGGTCTTTCATAAGTT	AAATATAGTCATTTCAATGGAAGACAAA
J14	123376	CHLORO230	454	CCAGCAGTCGTTGACGTTT	TGCACAAATGAGGGATTGTT
J15	123808	CHLORO231	412	AACAATCCCTCATATTTGTGCAT	AATTATGATTTCTTTGCCCTGAAA
J16	126215	CHLORO232	446	TTGCTAAGTGAAGAATTTGTCC	AATTTCAAGCATCTCAGAAGGC
J17	128944	CHLORO233	500	ACGGTAGAGCTAGGACAGTTATTG	TTAATAGGGAACCTCAAAGCG
J18	129381	CHLORO234	457	TGAATGGAATTGGGATATGGA	AGGACTATAGAATAAAGGATTATCCCG
J19	129769	CHLORO235	489	CTGTGATATTTGAGAATCACCGTC	GAAGGACCTTCCCTCCG
J20	5213	pl16	97	ACCCACCTTACTAAGATCTCTCCTT	TTTAAACAGATGTGGTAGAAAGCA
J21	11857	atpF	149	TGAATAGCTCCTTACGCAGT	TTAGCAACAATCCAATAAATCTAAGTG
J22	21798	poC1	150	ATTCAAATGAACCTCGTAATCGTAA	TATGGTATTTGAAACGCTTCTAGTTAT
J23	42661	ycf3	144	TCCAATACTCAGCGGCTTG	TTCGGGCATTAGAACGAAAC
J24	70115	clpP	358	GCGGGTTGATGGATCATTAC	TTCGAAGTCTGGAGAAGGA
K1	84623	pl2	172	CGGACCTCTCCAGAAGGTAAT	AAATGGGAAATGCCCTACCT
K2	95582	ndhB	152	CCAGAAGAAGATGCCATTCA	TCATCAATGGACTCCTGACG
K3	101686	n16S	500	GAAAGCACTCTGCTGGGC	GTCGCCAGGGCATAAG
K4	106334	n23S	500	GAAACCGACACAGGTGGTA	CCTTCCACCTAAGCTGCG
K5	120297	ndhA	159	TTGACGCCACAAATCCAT	TTAGGTGGTCTGCGAGCTG
K6	98626	ps12	266	TCTCACACCGGTAATCCT	ATCCGAAACGTCACGAAATC

K9	425	CHLORO258	498	AGGGAAGTTGTGAGCATTACG	AGGAATCTCTGGTACTTTCACTTT
K10	2125	CHLORO259	470	TTTAGAAGCATAATAACTTCTGGGAA	AATTTAGAGTTTCTGGGTTATCTGTCA
K11	2862	CHLORO260	490	AAGATGTCGATCGTAAATGCG	TCGTTTGATTATCCCACTAAGG
K12	10134	CHLORO261	482	TTTGTTCCGGTATAAATGGTCA	GGGTCGTGAAGCTTATCCAG
K13	10974	CHLORO262	485	CCGCGTCTATAGGGATCAT	GGTAACCATTAGAGCCGACG
K14	16245	CHLORO263	922	TTCAAAGATGCTCGAGTTATTCC	ATTAATAATTCGTGGGCC
K15	17103	CHLORO264	493	GGCAAAAACGAATAAAATTTGA	ATTTCTGACACAAGTATTC

Ecole doctorale "des génomes aux organismes"

plaque	génom	nom	PCR	FORWARD	REVERSE
K16	17540	CHLORO265	484	TTTGTCCCAATTCAAGACTAAACA	TGGTACAGAACTATAGTATCATTTGGAGTG
K17	18411	CHLORO266	839	TGACTAGTACCTAAGTTTCCATTCAATTC	TGTCATATCGATTAATAACCTTTGG
K18	19178	CHLORO267	896	AGATGTACTCCGACAAGTAAAGGG	GTACTIONATGGCGGAACGGG
K19	20433	CHLORO268	491	CGATAGGAACCTCTCTTGAAGCA	CCCTACATAGATTAGGCATACAGTCA
K20	21267	CHLORO269	439	GTCCACGGATTCCATTATCAA	GGGTGCTATCCGAGAACAAT
K21	23263	CHLORO270	1300	CAAGTACTTCTTGGCGAGCTCTA	CCTTTATACTGACACTGAGAAGATCG
K22	24872	CHLORO271	498	AGAATTTCTGCGCTATCA	TTTACAAGACCAGCTCGGG
K23	34599	CHLORO272	489	AATACCGCTTACCCTAGTGAGTTT	TTTCAAATCCCGCTGCC
K24	37530	CHLORO273	915	GTTTCATGAGCCACGCTAA	GCAGGATTCATCATGACAGG
L1	38803	CHLORO274	490	CCCAAGAAGGGTTAGAATGG	TTGGTCAACCGGCTGTG
L2	39655	CHLORO275	913	TTCCACCCAGAAGGTAATGG	TCGCGATGCAATCATATCAC
L3	40927	CHLORO276	872	TCCTATACATATGACCCGCGA	AAACTTCTTTCGAGGAATGGG
L4	53526	CHLORO277	477	CATTTCTTCCAGAGTCCGT	CGAGATACTCTTGGTCAAGAAATTA
L5	55121	CHLORO278	492	CTGCGGTAGCTGCTGAATCT	AAGAAACGGTCTCTCCAACG
L6	56433	CHLORO279	406	AGTTTAATTGCACTCGGCTCA	AGAAATATTTGTTCTTCCGTTATGAA
L7	57249	CHLORO280	490	TCTTCTAGTATTCCAATAATGTTGATCTT	TGAGTTATATCGAAATCCTTAGAACTATT
L8	58079	CHLORO281	437	TATGGGATCCGTAGTAGGCG	GCATGGAGCTGAAATAACTCG
L9	62798	CHLORO282	786	CAATAAAAAATTTGAAAAATGAGAAG	GTATTTTCTTTTATGGTTCATAT
L10	67721	CHLORO283	390	TTTAAGAACAGGAATAATGCTAGTATC	AGCAATAGTGATTAATCGTTGTTGTTTTA
L11	72526	CHLORO284	496	TGATCCAATGTGGAGACAAGG	TAAGATGGAATAGGCCCGCT
L12	73404	CHLORO285	444	TCTTTAGAAATAAAGAAGGGCGTG	CCGAATCCACTTGAGCATC
L13	77976	CHLORO286	773	TTGACTGTTATTCAAAGGTCCAA	GCGATGCGAAGAGCTTTACT
L14	86680	CHLORO287	1336	AACAAGCAATCGATATTTACG	TGGACCTCTCAGTAGGATTCG
L15	88131	CHLORO288	1249	CCGTCTCAATTCATCCTATTTC	TGGTGATATTGTTCTCCATTG
L16	89346	CHLORO289	1254	CGAAATCTTGATCAATGGAGG	TAGGCTCGGGAACAAGAG
L17	90538	CHLORO290	1320	TTTCCTTCTTCTTGTGCTGG	GGGAATATGAGTCGAAGCAATAA
L18	91785	CHLORO291	863	TGAACTCTCTCTCCAGGGATTG	AGAGAAATCTCTGATCAACGATAGAA
L19	92501	CHLORO292	497	GGCGCTTTGGTGGGAT	AAAGAGTGGATCTGCGGG
L20	100008	CHLORO293	484	TGTTGATCCGAATGAATCATCTT	AAACCGTACGATTCTCCCG
L21	100424	CHLORO294	498	AGAACATGCCGAGTCAAATCTT	CGCTTCATATTCGCATGGAG
L22	100852	CHLORO295	489	CTCGTGGGATTGACGTGAG	CACTGCTGCCTCCCGTAG
L23	101280	CHLORO296	476	AGAGGATGATCAGCCACACTG	GGGTATCTAATCCCATTCGC
L24	102123	CHLORO297	494	GTGATAAGCCGGAGGAAGGT	GTCGAGAAACGAAAGAAGGC
M1	102558	CHLORO298	479	AAAGAAGGGAGCTACGTCTGA	TTGGATCTACATTGGATCTCACC
M2	102970	CHLORO299	491	GTTTGAAACCAAACCTCTCCTCA	TGTACTIONGCTCTCCAAGTG
M3	103407	CHLORO300	483	ACTTCTATGGTAAGGAAGAGGG	TGACCAACTGACCATCCTACC
M4	103819	CHLORO301	499	CCACTGAAAGACTCTACTGAGACAA	TCTAGAGGATCCAGCTACAGGAG
M5	104274	CHLORO302	470	GGAAAGCTCTCCGTTCTG	CCTTCTCGTCTCTGGGTG
M6	104675	CHLORO303	901	AAAGTTAAGGGCGAGGTTCAA	ACCGAAACAGTGCTTTACCC
M7	105517	CHLORO304	888	CCGAAATGCGTTGAGGC	TTTGCCGAGTTCCTTAGAGAG
M8	106774	CHLORO305	489	GAAGCTTCACTGTTCCCTGG	AACAGCCCAACCCCTTGA
M9	107195	CHLORO306	497	AAGGTTTGGCACCTCGAT	TACCGGTCTGTTAGGATGCC
M10	107626	CHLORO307	479	TCATTACGATAGGTGCAAGTGG	AAAGTAATAAGAATGAGAGGTGTTAAGC
M11	108034	CHLORO308	473	GGGAGGTCTGCGGAA	AGAAGTCACTCCACTCTATTTCG
M12	109519	CHLORO309	480	GGGCCCTCTTATCTCTTCTT	TTTGCTGATCCAGACTAATACCAA
M13	110636	CHLORO310	808	TCTATGTAACCAGATTATATGACCAA	TTAATAGGAACACTCTCTTTGTGG
M14	111794	CHLORO311	497	GGAAGAAGACGAGCTACAAGAAA	GCATGATCAAGGATACCTGAGA
M15	114504	CHLORO312	445	TTTCGGTCTGTTCAATTGTG	AAGAGCTATTGATAAATGATCCGC
M16	115726	CHLORO313	487	AAATCAGGATAAATCCCAATACCA	GGTATTAGTCTATAACGGATCCTGG
M17	122010	CHLORO314	472	TTTAACGATCAACTTCTCCATAA	TCGTTATGAGTCTACGATGAATTT
M18	124156	CHLORO315	897	GAATTCGAAATCTCTACGTGCTT	TTTGATTGGATGGGAATGAA
M19	125001	CHLORO316	1237	AAATTTGTTATACGCTGATTTAGTATTT	TGGACAAATCTTCACTTAGCG
M20	126633	CHLORO317	774	AGTCGAGCCTTCTGAGATGC	GTCGAGTTGGAAGAAATGGAAGG
M21	127340	CHLORO318	1253	TCTTGAACGAATCCAGGCT	TCAGAATTGAGAAATCTATGGCTC



Tableau 14 : Tableau S2 Amorces mitochondriales

plaque	génomé	nom	PCR	FORWARD	REVERSE
A01	18529	MITO001	551	TGTAAGATCTCGATGACCGC	AGATGAATAGTACCAGTGGCGTC
A02	19049	MITO002	559	TCATCCTAGACGCCACTGGT	GTTTCTTCTTGATCGAAATGGG
A03	19562	MITO003	560	TGTCTATTGGGCTTCACGCT	GAGCTTCTCGTTGACGTTT
A04	20098	MITO004	560		TGGGACTCTCTATCTTCTGGG
A05	20622	MITO005	560	CGATCGATTATCTACCCAAGAA	CCCATAGCATTCCATCGTCT
A06	21145	MITO006	560	CGGTACAGTACAATACGAGACGA	AAGTGGATTGTGTGCGAA
A07	21682	MITO007	559		
A08	22204	MITO008	557	TTTATTATTCGTCGCCGGTC	GATCCATAGGTATGTGCTCGC
A09	22711	MITO009	559	CGATTTCCGCAGTCTCTAT	ACAATCAAGGACTGAGCATGG
A10	23235	MITO010	558	CCCTTGGTCTAAACCATGC	TCTGAGTGGATATGTCCAAGTACG
A11	23717	MITO011	353	TTGGGTCATCTCAATGGGTT	CGATCGATATTTGCGGAGTT
A12	23976	MITO012	381	CGAATGCGTGAGTTATACCG	GTAATGCCCAAAGACTCCCA
A13	24337	MITO013	353	TGGGAGTCTTTGGGCATTAC	GAATCATTGATAAAGCCCTCG
A14	24529	MITO014	416	CGCATTGAAAGGAAGCTGT	GAATCCGAAGTGTGCCAAT
A15	24856	MITO015	404	TGTGGAACCTGCAGCATAGG	CGGGAAACCCACAGAAGTAA
A16	25141	MITO016	377	GAGCATTGCCAAACTCACA	GGGAGACGTCTATATCCGA
A17	25431	MITO017	387		
A18	25783	MITO018	557	GAGATGTCTTGAGCAATCGGA	GGGAGAGCGAATGCACAG
A19	26307	MITO019	559		
A20	26813	MITO019	561		
A21	27324	MITO021	502	AGTTTAAAGTCTATCGTTCCGG	TGCTCAATAAAGGGAAGATTCG
A22	27764	MITO022	557	GAAAGATTGCAGCCTTAGCG	TGATCTGAGAAGTGCTAATTCGG
A23	28281	MITO023	558	TTTGAAGTCTATCCGAACCGA	GGAAGGTTGGTCATAAATTTGG
A24	28799	MITO024	560	TGAGTTTCCGTGACTCTCCAA	GGCCCTACTACTACAAACGA
B01	29302	MITO025	547	GGAGAGGCATGAGTGCAGTT	TGAGTCCTCACTGACCTGAGC
B02	29824	MITO026	560	AATCGCTCAGGTCAGTGAGG	ACCGTACGTACCTTTCGGTG
B03	30224	MITO027	415		
B04	30542	MITO028	356	CGTATATCGTCGTAACGCC	AAGAGCCGAACGAGAATGAA
B05	30888	MITO029	560	TTCGGCTCTTGAATCACAT	ACATTGTTGGAGGGAGAGCA
B06	31388	MITO030	559	GACATAACTCGACATCCCGAA	GCCATGAAGGGAATCCAAG
B07	31919	MITO031	559		
B08	32457	MITO032	560	GACACTGTCCAGCCATTCTG	TCACATCTATTGGTCGTCCG
B09	32973	MITO033	559	ATATCTTCTCCGGAAGCCTG	TCTCCATCTAGCTTCTCAGGG
B10	38537	MITO034	556	GAATCAATGAAATGTTGCTTGG	TTCTCATACAAAGCTGCAAACA
B11	39064	MITO035	560	TCTTGATTGTTGCACGTTTG	GGACTAGCCGGGCTATAAGG
B12	39604	MITO036	554	CCTTATAGCCCGCCTAGTCC	TCCAAGCTCATAACAGGCA
B13	40100	MITO037	560	AGAATTCATTCCCACTGTCTGA	AGCAGTACGAGCTGAAAGGC
B14	40639	MITO038	560	AGCCTTTCAGCTCGTACTGC	AATATGAGGGATCCGGAAGG
B15	41179	MITO039	559	CCTTCCGGATCCCTCATATT	ACGAACTTGAACGCGGAG
B16	41537	MITO040	380	TTCCATAGGTCCTTGGTTCCG	ACATAGGCCGACTGTTGAGC
B17	41823	MITO041	398	CTCCAGCCGCTCACTGTAAT	TCCGATGAGCAGTCACTCAC
B18	42201	MITO042	359	GTGAGTGACTGCTCATCGGA	TCTCAAGACGCAGCTACACC

Ecole doctorale "des génomes aux organismes"

B19	42451	MITO043	445	CGCGAACCAAGATCCATAAT	TTCATAATTCAGTGTGGTGGG
B20	42761	MITO044	551	GGGCGTAAGATTCTTCTTCTTC	AAGAATCTGAGGTGGCAGGA
B21	43290	MITO045	560	AATCCTGCCACCTCAGATTTC	GCCTATGCCTCTAACGATGC
B22	43789	MITO046	560		CAGTGGCTTGTGTAGTGGAG
B23	44315	MITO047	560	AATCATGTTCCGGTCTCCAC	TGTCACCATCCAATCTTGGTTAT
B24	49236	MITO048	559	GAAGTTTCCAGCCTCACAGG	TTACGACGTATGAGACCGTTTG
C01	49766	MITO049	556	AAACAGGCAAACGGTCTCAT	ATGCGCGTGACTAAGGCTAT
C02	50272	MITO050	560	GAACGAGCGAAGAAGCAATC	GCCTTGTATAGTATTAGGAGCG
C03	50776	MITO051	560	TCGCTACGTATTCTTCTCACTG	GC GTTGGAAATCGAGAAAAGTT
C04	51043	MITO052	422	TGCGAAGGGACAGAGCTAAT	TTCCAGAACAGGAGGTTGGT
C05	51313	MITO053	375	CGGAACTTTCTCGATTCCAA	AATCACTTCATGCCGACCTC
C06	51712	MITO054	383	CGTTCGTTAGGGACGGTTTA	AATAGAAAAGCAGCGAAGCG
C07	51830	MITO055	558		GGAACGGGCTTCCCAA
C08	52410	MITO056	520	GCCTGCCAGCGGTAGTG	CTATGGCTGTTGGTCACGACT
C09	52856	MITO057	385	GATGCACACTCCTCCATGTG	ATCTTTCTGCGGTGTCTCT
C10	53102	MITO058	451	ATTTCTGCTGTGCTCATTCC	GGCAGCACCGTACTATTGA
C11	53374	MITO059	560	TTGCTCTGACCAAGACTCCA	CCGCCTACTACTCGTGCT
C12	53905	MITO060	560	TCGAGCTACGAGCACGAGTA	ACGAAGGCACGAAATGCTT
C13	54442	MITO061	561	GAGAAAGCATTTCGTGCCTT	TGGCTCCTTTCGTGGATTAT
C14	54968	MITO062	560	GGCCATTGAGAAATTATAATCCA	GGTTGCTTTCAGCCTAGCTT
C15	55509	MITO063	557	AGCTAGGCTGAAAGCAACCC	AAACGAATGAATGCTTTCCC
C16	56005	MITO064	559	AAGCATTTCGTTGTTGTTG	CCACCATTAAATAAGGCTCCC
C17	56526	MITO065	558	ATAATTGAGTGGGCATTGGG	AGGAATCAAAGTGGGAGAACAA
C18	57050	MITO066	559	TTCTCCTATGCTTGTCTCCCA	GATCAAAGGTTACTCCCAGC
C19	57586	MITO067	560	TTGCTGGGAGTAACCCCTTG	TTTCCGTATTTGACCCGGA
C20	58125	MITO068	558	TTTCCGGTCGAAATACGGA	GACCAGCTGCGGAGCTAAC
C21	58654	MITO069	551	ACAAAGAAAGGTTAGCTCCGC	CGAGTCATTATACCTCTCGGG
C22	59176	MITO070	550	CCCTACTCCCAGAGGTATAAA	AATTTGAATAGGATGACTCAGCG
C23	59662	MITO071	560	GCCAAATCAATCAAGGCAA	CGTGACAACCTTGTTCACCC
C24	60003	MITO072	495	ATTTAGGTCTCTGCCCGCTT	TGCATATAACGGAGCAACCA
D01	60345	MITO073	427	GTTCCGGTCCGTTAGCTGGTA	ATTGTCCACGGAGAAACCAC
D02	60694	MITO074	404	ACAAGCTTAGCTAGCGCCAT	AGAATGGCATGGATCGGTAG
D03	61026	MITO075	350	ACCTGCTAATCCGATGTCCA	CTCTTCCAACCTCGTCCCAGA
D04	61239	MITO076	553	TTTGGCGGATTGCTTACTACT	GGTTGGTCCAATCTGAGTGC
D05	61741	MITO077	560	GGAAGAGGAGAGGCGAAGAA	TTGCAGTAAAGGTACAACCACAA
D06	62275	MITO078	551	TGTTTGTGGTTGTACCTTACTGC	GTTTATGAAGCCATCCTCGC
D07	62806	MITO079	556	GCGAGGATGGCTTCATAAAC	AGCATTGGCAAAGCACAGAT
D08	63311	MITO080	560	CACGGAGGAAAGAGAGACCA	TGCATCATCATCTCGTCTG
D09	74625	MITO081	567	CGCTGCGAGATTAGCACTT	CAAAGGGAACAGAAAGAAAT
D10	75169	MITO082	558	ATTTCTTCTTCTGTTCCCTTG	AGACCCGACTTTCTCTTTCAA
D11	75677	MITO083	561	TGAGCTCATAAACTGGCGAA	CCTGTCTAGCAGATCGTAGCG
D12	76206	MITO084	522	AATAGTAGACAGCTACGATCTGC	AAAGACAGGATGTATTCCGACG
D13	76666	MITO085	559	CGTCTCCTCATTATAGCTCTCC	CTCTGGTTTGTGGTTGCAC
D14	77178	MITO086	560	AAACGGAATGTACCGGATTT	GTTTGCACCTCCGTGACTAGC
D15	77689	MITO087	560	TCTCCACCAGATAGACGTAAGC	AATCAAATCGAGGATCAACCC
D16	78228	MITO088	559	GGTTGATCCTCGATTGATT	TTTGGATTGCTCACTTCT

Ecole doctorale "des génomes aux organismes"

D17	78741	MITO089	559	TTGAGAATCTCGATAACTCCTAAA	CGATACTTCAACTAGAGGAGCG
D18	79241	MITO090	559	GTGTGACGCTGGGTAGGG	GGAAGCCGGCTCGAAATA
D19	79731	MITO091	561	CGGTCTGACCCAAACAATAA	CCCATTCTAACCAGTGGAG
D20	80245	MITO092	558	AAACCCTCCAGCATAAGCG	ACGTGCTGCAACCTAGGGAT
D21	80762	MITO093	560	GAGCGAAAGAGCTGCCTG	GACTCAGTGCTCCATACGGG
D22	81276	MITO094	559	TCCGAACGAATTCTGCTTTC	GCCCCGCATAATCAAGG
D23	81817	MITO095	560	GCCTTGATTATGCCGGG	TGAGCACAAGAGAATGAAGAGG
D24	82350	MITO096	560	AAATTCTCTTCACTCTTGTGC	CAAACCTACACGAAAGTTGCC
E01	82851	MITO097	511	CTGTGCATCTCTGTGATGGG	GTGGCCGAGCTAGATCCTTA
E02	83305	MITO098	560	TTTAATAAAGCGAGCGCCAC	AGGGATTAGAGGAATGGGAGA
E03	83833	MITO099	560	CTTGGCTTACCTCTCCCATTC	AGCCACTGTCCCGACCTT
E04	84335	MITO100	560	TAACGGCGAGAAGCTCCTTT	TCCAAAGGGACTAAGCTTGACTA
E05	84836	MITO101	560	TTGATAACTTTGTCTCATGCGG	GATTAAGTGGCTTCTCTCCACT
E06	109006	MITO102	560	TTTACGGCGGAATCAACTA	GGATTTGGAAGTCTACCGAG
E07	109535	MITO103	560	GGGAACCCGTCTCGGTA	AAAGGAAGTCCCTTACACTAGTCC
E08	109990	MITO104	560	TACGTTCCATCCACATCCG	TCCATACGAGCTCCGGC
E09	110566	MITO105	560	TCTATGAGGCCCTGCACAAA	TAGAAGGGCGAGGTTGGGA
E10	111105	MITO106	560	GTTCCAAACCTCGCCCTT	AGTTGTGGTTGGTTGTGACC
E11	111642	MITO107	560	TTGGTCAACAACCAACCACA	TTTCCGATATGCATAGGAATCA
E12	112151	MITO108	558	CCCCTGGACCAATTTGAGA	TGACCGCCATCATATTAGC
E13	112687	MITO109	561	TTGCTAATATGATGGCCGGT	GGAAGTGGCTTAGGAAAGGA
E14	113215	MITO110	560	CGCATCCGATTAGTCTTTTC	GAGTTGAGCTTGATCACTCCT
E15	113718	MITO111	555	CTTTGTCTGCGCTAGAGGAAG	CCCTCTCCTTATCAGTTTCTATG
E16	127032	MITO112	555	GGGCTCGAACCTACAATATAACC	CTTACTTATTGAGACCCGCC
E17	127535	MITO113	556	CTTCAAACCTTGAACAGAGGAGA	TCTATTCTTATCCCAGTGTGCT
E18	128028	MITO114	561	GCTCGGTAGTTGGATCTTTCA	TGGAGTAGACCTACCTTCATCCA
E19	128535	MITO115	557	AGAAAGTAACCAGGATCCGC	GACCCAGCTTCTGGCTCTC
E20	129034	MITO116	560	GGCTATGTGATAGCACCCAAA	AAGAAAGAATGCCAGGTGA
E21	129482	MITO117	396	TGTTAGGCGGTGTGAGACAG	CCCTCCAGACAGCTTCACTC
E22	129723	MITO118	468	CCCAGCATCAAACCTATCAAT	TAAGTCGACGGCCCTACACC
E23	130004	MITO119	371	TGGAGTACTGGGATCAGCA	TTCTTGGCCATGTACAACA
E24	130359	MITO120	368	GTACATGGCCAAGGAAGCAT	CTGCGCACAAACCACTCTAA
F01	130579	MITO121	560	AGAAGCTCATTCTGAACCGC	AAGCCTTAGCATACAATCCATC
F02	131078	MITO122	554	CATGCATGGGTGGATGAAG	CCTTCTGTAAACCGCCACTC
F03	131601	MITO123	552	TTTCATCTAAGGAGTGGCGG	ATCGTGAAACACCATGAGCA
F04	132118	MITO124	560	ACCTCAACATCCTGCTGCTC	CAGTCGAGTTCCTAAAGCCC
F05	132633	MITO125	556	TAGCCTAACAGAAGGCAGCG	TCAGTCCCTCTAGTCGGGTG
F06	133169	MITO126	559	CACCCGACTAGAGGGACTGA	GCAGAGGAGGAAGGGTGG
F07	133676	MITO127	559	ATGTGTGGGACAGAGGGCT	CAAGTGTCTCTGAACCGTG
F08	134182	MITO128	560	AGTGGGAGAGCCGTGTTATG	CITTTGCTTGTGTGCACTGA
F09	134718	MITO129	560	ACTGTCACTGCACAGCAAGC	CCCAGTCAGCCACCAACT
F10	135257	MITO130	560	CCGAGTTGGTGGCTGACT	AGGCCACCACTTCACTTTA
F11	135797	MITO131	558	TAAAGTGAAGTGGTGGCCCT	AGCTTCCCAACTGGTAGACG
F12	136327	MITO132	560	AAAGGAACCGTCTACCAAGTTG	TTTCCCAACCTGTTGCTTCT
F13	136827	MITO133	560	GTAGGCTGGACCAACATCCA	CGAAATGACAGCATCGAAGA
F14	137336	MITO134	560	ATCACCGTGGCCGTGTAT	TGAGGTAGAGTAGACCGATGATCC

Ecole doctorale "des génomes aux organismes"

F15	137856	MITO135	560	CGGCCAAATGACTACAGGAT	CTTCACAGCCACCTATTCA
F16	138396	MITO136	559	TGAATAGGGTGGCTGTGAAG	AGTGACAGGCTAAGACGGC
F17	138903	MITO137	474	AAGAGACCACCGGTATGCAA	CTGCAGCTTTCTAAAGATGATTG
F18	139334	MITO138	560	ATGTTTCTGCCATTCCCAAC	AGTTCACCAAACATAACCCGA
F19	139853	MITO139	558	CCCATTTAGCGATAAGATACTCG	CCTTGCACCTCCCATCTAAT
F20	140367	MITO140	560	CAGTGGCCGTTTAGTCCAT	GTGGGAACCTTGCTTTCTGG
F21	140901	MITO141	560	TCCCAACCAGAAAGCAAAGT	GATCGCATACTTGGTCACCC
F22	141421	MITO142	560	GTGGGACCCTCCATAGCAT	TTCGCAAATAGGTCCGACTC
F23	141927	MITO143	560	CGGGCTATCGAACACAGAGT	CAAGAGGGCGCTTATGCC
F24	142430	MITO144	540	CTATTGGCTGGGCCCTT	GCTCGGTAGTTCCTAGCAG
G01	142911	MITO145	560	TAGCGTCTCCTTCTGATCCT	TGCACATTGTTCTTTCCAGG
G02	143422	MITO146	556	TGGGAAGATCTAGGATAGGCG	AAAGCGCTAACGATCAAGAAA
G03	143957	MITO147	559	TTTCTTGATCGTTAGCGCTTT	ATTGTGCACCACAGATTCCG
G04	144478	MITO148	560	GTGAATATTGCAGACCAGCG	ATCAAACAGCTGACGAAGGG
G05	144988	MITO149	560	CCGTCCTGTAGACTCCCTGT	AAGGCTGAATTCTATGGTGA
G06	145493	MITO150	561	GGTCTCTCATGGCCCAATTA	TCGACGATCCCAAGTCTTT
G07	145993	MITO151	560	CCTACGAGTCGTCCGGC	CCATCGGGCAGGGATAAT
G08	146531	MITO152	560	AATGATTATCCCTGCCCGAT	AAGGGCGTGGACAGAAAAT
G09	147062	MITO153	560	CAACAAGGCCATTCTGTCC	GATCATCATCATGGCCCTTT
G10	147607	MITO154	554	GCCATGATGATGATCCTATGTT	TGACGTGACTCGAGTGAAACT
G11	148140	MITO155	554	AGTTTCACTCGAGTCACGTCAG	GAATCTTTGCTTACCCGTGG
G12	148633	MITO156	564	CAAACCAACGTTGCGAGC	CGGAAAGGGAATAAGAAAGA
G13	149158	MITO157	560	GCTGCGTCAGCAGTTTATTC	AGCATGTAACGGACTTTCCC
G14	152028	MITO158	520	ACCGAGTATGCAGGTTAACACA	AGGAGTTAGCGTCCAGTAGG
G15	152526	MITO159	560	CCCTACTGGAGCGTAAGTCT	CAACAGCGAAAGATACAGAGAAGA
G16	153067	MITO160	554	TCTGTATCTTTCGCTGTTGGC	TGCCACCTACACCTTGATT
G17	153584	MITO161	559	CCGGGCTAAAGGTTTCAAAT	CCGGTCGATAAGCAGCAG
G18	154097	MITO162	560	GCTGGCTAGCTTAGCTCTCG	AGATTCTCGCCACCTCTGT
G19	154421	MITO163	350	TTGGTCAACAACCAACCAA	AAGTGCTCTTGCTCTCCCTG
G20	154743	MITO164	442	AATGAGACCAGGGAGAGCAA	AGACGAAGACAAGGAGGCA
G21	155025	MITO165	388	GATCCTCGAACCTACCACGA	GACCTCGAAAGAGAGAACG
G22	155393	MITO166	371	CGTTCCTCTTTGCGAGGTC	CGCATCTGATCGAAGTTTCA
G23	155641	MITO167	560	AAAGTCGATCGTGCACCTGT	AGGTCCCAGGTCTCCGAGT
G24	156149	MITO168	562	TGGGTGCAGGATTCTTCG	GCTCAGAATGCCTAATGCCT
H01	156691	MITO169	559	AGGCATTAGGCATTCTGAGC	GCCTGCTTCTAATTGATGACTG
H02	157142	MITO170	456	TTAGGAAGAGCCGTACGAGG	ATCAATATCCGAACGGAACG
H03	157537	MITO171	350	TTAATTTCCGATCGGAAACG	CAGGAACAAGAAGAGCGAGA
H04	157725	MITO172	433	TCGACATTTGTTGCAACGTC	TGGAAGAGCAGCTGTGAGA
H05	158016	MITO173	379	ATTTCGTTCAATCCATCGGT	CGCTGGTAAGGAGTCCCTCTG
H06	158264	MITO174	558	GACGAGTGGAAATGAGGGAGA	ACTCTGCTCTTTGGTGCTCG
H07	158769	MITO175	560	GACGGTGAGTTACTGGTTCTTG	GGATAAGTCCGGTGTGAG
H08	159308	MITO176	559	CCTCAACACCGGAGCTTATC	CGCACTACTACTTGGTCGGA
H09	159832	MITO177	560	CGAGAGTGAGATTGATCCGA	TCCATCACAGACCTTCTCTGG
H10	160365	MITO178	558	TGGTGACCAGAGAAGGTCTGT	GGAGCTATCTCGCTCGAGTG
H11	160902	MITO179	560	ACACTCGACGCAGATAGCTC	TTGCATTCTTATCCGATGG
H12	161314	MITO180	449	ATTTCCGCTATGACCAATGC	TGCTTCTAGCACAGGACAC

Ecole doctorale "des génomes aux organismes"

H13	161626	MITO181	397	GTATTGGAATTTAGCGCCA	AGACCAACCCACTGAAATGC
H14	161905	MITO182	416	CCTTCGATGGCTTCCTTATG	GGGTTCCGACCAGAAATGCTA
H15	162234	MITO183	379	GTCGTCGCGTCGTAAGT	CGAGTTTACGTCCATCCCTG
H16	162481	MITO184	559	GTTTGTTCGCGAGAATGGAT	TCTTTCGGGTTTGATCTTTCC
H17	163005	MITO185	560	GGCCAGGATAATTGGGAAAG	AGGTAGCCACTCCCTTCTCAG
H18	163512	MITO186	559	CCGTATGATCGGAAGTCTC	CGATCTGTCTTGAAGTGGTCA
H19	164011	MITO187	531	TTTATACTTTAAGCGGATTGCT	AAATAGCACGTTTCTCTCGCT
H20	164519	MITO188	557	AAAGCGAGATGAACGTGCT	TCCGGTATAGCAGCTCACG
H21	165016	MITO189	558	CTGCTAACATGTGCCAGTC	CTGATCCAACAGAAATCCGC
H22	165520	MITO190	560	AGCTGTAGCCGATATCACA	AGCTCTCTCGAGCTTACGG
H23	166020	MITO191	558	GCCTACCAATCAAAGAGGCA	CACCACGTTGAGAAAGTCCA
H24	166552	MITO192	560	GGAGTTGGACTTTCTCAACG	GCTCTGTCGATCGAGGGTAGTA
I01	167097	MITO193	523	CTCGATCGACAGAGCGGA	TGGAACTTCTCTGCCATTTG
I02	167592	MITO194	560	TCCGATTCAAATGGCAGAG	AGACCTCGGTCGTGCTTCT
I03	168104	MITO195	557	TCATCAATATTCTAATGGGAATGG	TTGATCCAATATTGACCGGG
I04	168614	MITO196	560	CAGGACCTATGGAAGTCGGG	CGGATTAGGAATGCGAAGG
I05	169120	MITO197	560	GGTAATAGTGGCGGGTCTC	TCCTCTCAATGAAATTTGCCA
I06	169624	MITO198	560	TGCATGCATACATCCGTAAG	TAAAGAGATAGCCGCCCAGA
I07	170152	MITO199	555	ATTGGTACTTGATCTGGGCG	TTGTAGTCAGCCAGCGTTTC
I08	170676	MITO200	560	TGGAGTACTGGAACGCTG	TCATTCTAGCCCGAGAGGAA
I09	171190	MITO201	561	GGCGCTTAATGAAGAAGCAG	CTGCCTCTGCAGGATCAAA
I10	186004	MITO202	559	CCACAAGGAAGAATCAGGCT	CCAACAACATCTCGAGAAAGAA
I11	186501	MITO203	560	TCATGATCTAGGTGCCGTTG	GATCTAGCCCTCTCGCTTGA
I12	187028	MITO204	560	TTCCATGTGTCAGTCAAGCG	TCTCGAAGTTGGATCTATGGC
I13	187564	MITO205	560	CCGGCCATAGATCCAACCTT	GGTGGTCTTAAGAATGGTTCG
I14	188102	MITO206	545	CCGAACCATTCTTAAGACCAC	CGAATATGGATGGTAGAAAGATGT
I15	188616	MITO207	560	GCAAATAACATCTTTCTACCATCC	GGGAATCCTCCTTAATAGACGAA
I16	189121	MITO208	555	TGATTCAATTGGCATTGACATAAT	GCCCAGCTGTCGTGTCTATT
I17	189631	MITO209	558	GGCTCAAGAGCCGCTAAAG	GAGCAGATTTGACTCGGCA
I18	190158	MITO210	560	ATCGTTAGGTTGTGCCGAGT	ACCCGCTGTAGGGATTAGT
I19	190658	MITO211	560	GTGAGTCCCATGCGTCAGT	GAAAGATATTAAGCGGGTCCA
I20	191191	MITO212	558	AACCTCTGACCCTTAAT	CGTTTCATTCTGTGCGCTAGT
I21	191713	MITO213	518	CCAAATATAAATACCACTAGCGCA	AGAAGAGACAGACAAACCAGGC
I22	192203	MITO214	560	CGATAAGCCTGGTTTGTCTGT	TCGGCTATGCCTTAACCATC
I23	192734	MITO215	558	GACTAATGGGATGGTTAAGGCA	CCCCACCTCATTCATTAATA
I24	193272	MITO216	544	TTTAATGAATGAGGTGCGGG	ACAGAACAGATGTTGTTCTCGACT
J01	193787	MITO217	560	CACGGAGTCGAGAACAACAT	CGTTACCTTCTCTCCTTGGC
J02	217038	MITO218	559	AATGAGTCCCAACCCGAC	AAACGAGAGGGAAGAACTCG
J03	217546	MITO219	559	AAGAGCTGTTGCTGGAAATCA	GAGAATCGTTGGTTTGACCG
J04	217913	MITO220	438	AGGGATAGAGGAGGAAAGCG	TCCGAGTGAACCCGAAATAG
J05	218215	MITO221	416	CAAAGATAAAGAGCCCGGTG	TACCTCCGATCTCTACCGCA
J06	218611	MITO222	372	TGCGGTAGAGATCGGAGGTA	TGATGCTCCTTCGTGATG
J07	218837	MITO223	436	GCACCTTCACTATTTCCGA	TGGTGAACATGATCTGCCAT
J08	219105	MITO224	559	AAGCTCGAAGACAAGAGAACC	GTCTAAGAAGCTTCGAATACCC
J09	219611	MITO225	560	GACAAGTTTGAACCAACAA	AAAGGAAGCTTCTCCCTCTC
J10	220113	MITO226	559	TTTAGGCAAGGAGGCTA	GGTCATCTCCATGCTTTCC

Ecole doctorale "des génomes aux organismes"

J11	220650	MITO227	560	TTGGAAGCATGGAAGATGA	TTTAGCACAACTTGGCGTC
J12	229041	MITO228	560	TTGGGATTTAGTTTGTATGCC	GGCCTGGCCTATTCCATA
J13	229577	MITO229	560	TTGTCTATGGAATAGGCCAGG	TCGCAATCAAGCACATAACA
J14	230098	MITO230	558	TGTCTTGCACGAGATTCATG	AAGGGTCGTTTCATTAGCCCT
J15	230636	MITO231	559	AGGGCTAATGAACGACCCTT	GAAAGCAGCTTCACATCACAA
J16	231163	MITO232	560	GGCTTCCTCACTTGTATGTG	GGGAAAGAAGACTGGAACCG
J17	231738	MITO233	422	GCGCTCCTATCCTCTCCTTT	ATGAATGGGATTGTGGCACT
J18	232052	MITO234	370	GCTTGTATACGGATCCGA	GCATCTGGCACGAGATGATA
J19	232309	MITO235	401	GATCGGAACTCGCTGATAGG	AGGAGGCCATAGGGAAGTC
J20	232690	MITO236	409	GACTTCCCTATGGCCTCCT	TCCTGTGGTACC GGATGAT
J21	232950	MITO237	393	AACGCTGGTGGTTGTTCTT	GCTGGTCTTGTGGTCAT
J22	233242	MITO238	560	TCCCTCTCTGCTTATGCCA	CTTGCGAGCCCAAGTCAAC
J23	233767	MITO239	559	AGTTCGTTGGTCCGAAGTTG	ATGCAAGACCCACTCCAGAT
J24	234296	MITO240	560	AGAGGCATCGATCTGGAGTG	TCCTGCAATTTGCTCTCTCA
K01	234836	MITO241	559	TGAGAGGACAAATGCAGGA	GCATGTTCCCTGCCACA
K02	235377	MITO242	560	ATGTGGCAGGGAACATGCTA	GCACCATTCCCAAGTGATTC
K03	235916	MITO243	559	TGAATCACTGGGAATGGTG	GTTGCCCTGATGTCTCCAAC
K04	236432	MITO244	557	AGCAATAGTCTTGGCAGTCCA	CTATCCAAACCCATTGTGCGAA
K05	236962	MITO245	559	GACGATTTGACAATGGGTT	ACTGAGCTTGTTCAGGCGAT
K06	237460	MITO246	554	GATCCTGGACGAGCTAACCA	ACCCATCAGCTTTACCAACAG
K07	237992	MITO247	560	GCTGTTGGTAAAGCTGATGG	TTGAAACGAACATCACTCTCCT
K08	238525	MITO248	560	AAAGAGGAGAGTGATGTTTCGTTT	TTTGCCTACTTATTGACCACTGA
K09	239044	MITO249	560	CGTTAGCTGATCTAGGAGTCAGTG	CAGTCGAAGAAAGGTAGCCC
K10	239563	MITO250	560	TAAGCAATCCAGACCCAAT	GGAAACAGCTTCTCTATTCCA
K11	239833	MITO251	364	TCCTGTTGAACGAGTGGATG	CATGTTCTATGCCATTCC
K12	240173	MITO252	381	GATTGGAATGGGCATAGGAA	ATCGCCACGGCTATAAACAC
K13	240485	MITO253	405	TCCGGAAGAGCGAAGATAAA	GGAGTGCCACGCACAAG
K14	240701	MITO254	451	CGCGTAGCTCTGTTCTTTG	AAAGCGATTCCGACGACTAA
K15	241085	MITO255	560	GCGAAGAAGTCGAGACCTAACA	GTTCTACCCGGTGGTATCCG
K16	241610	MITO256	560	CTCCTCCTAGTCTCCGGATA	GTTGGAAGAGTTCCAGCTTT
K17	242136	MITO257	560	GGGAGGCTCCTCTAAAGCTG	GGTGCGTACCCTTGAACCTA
K18	242665	MITO258	559	CTAGGGCGGCTAGGTTT	AGAGTCAATGGGCTGAGCTG
K19	243201	MITO259	558	GATCAGCTCAGCCATTGAC	AAGCATGATTATGCACTAGCAGA
K20	243714	MITO260	558	TTTCTCTCTTGGCACTTATGTC	TCATAGTGTAGATAAGGCACTCCC
K21	254025	MITO261	560	CTTCCGCTTTCAAATCTTCC	GAGGCCGTATGGAGTATAGCC
K22	254547	MITO262	560	CCGGTGCCTTACCACTTG	TCATTATTTCATCGGATACGTC
K23	255060	MITO263	549	TGACATGAAATGGATCACGG	CAGTAGTTCTTAACCAACCGCC
K24	255575	MITO264	559	TTTATAAAGCTGGCGGTTG	GGAATGAATGTTCCACGAGTC
L01	256111	MITO265	560	ACGACTCGTGAACATTCAAT	ATGCGCATGTGGTCGTAGTA
L02	256646	MITO266	563	CCGACTACTACGACCACATGC	GCCACAGCTCGTATTCTTCT
L03	257013	MITO267	376	TGCCGGTTCATTAGAAGGAAG	GCGAACACAGTGGTCTCTGA
L04	257188	MITO268	554	AGAATGAATACGAGCTGTGGC	CACAAGCAAAGGATAGCTATAACG
L05	257713	MITO269	547	TTTCTCGTTATAGCTATCCTTTC	AAATCAAGATTATCCCGACGTG
L06	258232	MITO270	560	ACTGCTCACGTCGGGATAAT	TGTGGATCTATTTCGAACAAGTG
L07	258769	MITO271	559	CACCTGTTGCAATAGATCCACA	CGCCCAATGCTATTATACGA
L08	259307	MITO272	559	TCGTATAAATAGCATTGGGCG	GTGTGCCAAGTTAGGAAGGTT

Ecole doctorale "des génomes aux organismes"

L09	259822	MITO273	560	AAAGCCTTCAATTAGTCGGTC	TGCAGGAACATTCTACGGT
L10	260323	MITO274	560	ACACCTGGCGAATCTTTCAC	GCCTACGAATGTGGTTTCG
L11	260836	MITO275	560	AAACGACTTCTGGCATCACC	TCCCTCTGCGGAGATCATAG
L12	261352	MITO276	560	CTTCGGCCTCTATTCCGG	TTCGTTTGAAGGGCCG
L13	261889	MITO277	556	TTTCTGCGGCCCTTCA	TGCATACACAGAAAAGGCA
L14	262403	MITO278	543	TGGTTAGTTAGTGTGCATAGCTGC	AAACAACCACCTCGGAACAA
L15	262885	MITO279	560	AGAAAGGTTAGCGAAGCCG	TTCAGCCTATTACAGCGG
L16	263412	MITO280	559	TGCTATTACTTCATCCGCTGTG	TCAAGCTACGTAGAACCAGGAG
L17	263937	MITO281	560	AGAGTCACTCCTCTCTGGTTC	CAGCTTGTGTCCGCACTTAG
L18	264474	MITO282	560	TTGCTAACTGCCACACAAG	GGAGGCGCTACTAACTCTG
L19	264975	MITO283	560	TGAAGTTACATGCCTGCTG	TTCCATTACGTAAGCAGTTCG
L20	265480	MITO284	560	CCCGGATGAAGGAATGTAAT	CTGGTGGTTATGGATGGAGAA
L21	266018	MITO285	531	TTTCTCCATCCATAACCACCA	AAGATGCTTCATTGGTTCCC
L22	276037	MITO286	560	TCAGCCTAGTGGAAGAAGACA	TAAGGAGATTCGGGATGGG
L23	276578	MITO287	559	CCCATCCCGAATCTCCTTAG	ATTCTCCACGGTTTAAGGCA
L24	277104	MITO288	560	TTCCACCAATCTTGCCCTTA	TGATCATCGATCTGTAGCGAA
M01	277603	MITO289	560	TTAGAGGGCGAAGGACAAGA	ACTCGGCTAGACCATTATCCAA
M02	278114	MITO290	560	CTTTGACTTGCTTTAGCTTTCCA	TTGATCCGCTTCTCGTCTTCTC
M03	278713	MITO291	386	CATTCCTCTGACGTGCGAAT	CTTCGGTTAGAGCAAAGCCC
M04	278646	MITO292	558	TTCATAAGAGAAGACGAAGACGG	TGTGAAAGAGGTGCTTGCTT
M05	279173	MITO293	560	TCAGTCTCATAAAGCAAGCACC	CCGCGGAGAAATGTACCAA
M06	279694	MITO294	559	GTAGCCATATCTTGGCGGGT	CCTCAGTCGACGTAATGCC
M07	280233	MITO295	560	CGGCATTACGTCGACTGAG	GGCAGGTTTGAAGCAAGGAG
M08	280735	MITO296	560	TGTTATGAAAGGGAGGAAGAGG	AAGTTGTGCTCAATGACTGGAA
M09	300012	MITO297	559	TTACTGCAAGGAAAGAGAAGGAA	TGACCGTATAACCAAGTTCCG
M10	300514	MITO298	559	GACCACCGGTCTATGCCTCT	CTCAATATCGGGCTTATCCG
M11	301039	MITO299	561	GTTTGAGACTTCGCCGATA	GCGATGAGACTTCTCGGTTAG
M12	301539	MITO300	559	AATAAAGCTCCTCAAAGTACC	CCCGGTCGAGATGTAGTAAGT
M13	302044	MITO301	562	CCATTCTCCTCAGTCTCCA	CGCGGAATTAGACCTGCTAT
M14	302563	MITO302	560	GACTGACAGATAAGCCGACG	GCAATAAACTCAAGGGCCA
M15	303091	MITO303	526	CACTCTCAGAGGTGGCCCT	TTTCAAGTGGATGAGATCGG
M16	303591	MITO304	560	ACTCGACCGATCTCATCCAC	CGCCGAAATATAGAATTGG
M17	304107	MITO305	560	CCTGGATCCGTTGTAGATCC	GAATGACGCTATTGTTGCGA
M18	304614	MITO306	560	CAAAGAAAGTTATGAAGAGCCAGG	TCTCATGGATCAACTTATTGCTG
M19	305127	MITO307	560	CCAACCCTCCTAGAAACG	TTAGGAAGATTGGACGAGGG
M20	305666	MITO308	560	ACCCTCGTCCAATCTTCTTAA	AAGAAGAGTGGTTGTGCTTAC
M21	306203	MITO309	559	AGTGAAGCACAACCACCTTCTT	CACGCTGATCAACAACCATT
M22	306741	MITO310	560	CAATGGTTGTTGATCAGCGT	CAAAGCTTCTTCGATAGACA
M23	312566	MITO311	522	CGATTGAAACGTGGAAGAGG	CATCTCATTATGTTGAGCTGGTT
M24	313023	MITO312	530	ACCTTAGAGGTAGAGGCGATGA	CTCGAAGCATTGAGTTGAGAA
N01	313498	MITO313	560	GGCTTAGCTTATAAGTAGGAACGG	AAATGAAAGGATTCTGTGCTAG
N02	314002	MITO314	560	GGACCTCGAGTTCGATAGATTG	CGCAATCGTTACTGTCCATC
N03	314315	MITO315	449	CCCTTGGTCAAAGGAAAGGT	TACATCGGTTTCACTTCGAG
N04	314632	MITO316	433	GGCTTGGATGGTGAAGCAA	ATCTGAAATGCGCGAAACTT
N05	315053	MITO317	559	CGCATTTCAGATGGTGGTAA	GGCTTAGCTACCTCTCCAGTG
N06	315575	MITO318	559	CTGGAACAGAAGCTTCACTGG	ACTTTCGCGAATGCCTTCTA

Ecole doctorale "des génomes aux organismes"

N07	316079	MITO319	560	GGCAGGCTGCTAACTCAGG	AAGCGGTTTGAGTCAAGTGC
N08	316585	MITO320	560	TTGTGACTCTTGCCACAATTAG	GACCCAGCCAGATGATGTCT
N09	317125	MITO321	558	AGACATCATCTGGCTGGGTC	AACTAAGTAGGTCGTTCCGCC
N10	317630	MITO322	541	CGGCATGCCTCAATTTAC	TTTAGAATGGCTCCAGTGGC
N11	318123	MITO323	561	CCGAGCGACCAGACTTAACA	GGACCGCTAGAGATCTTATTTTC
N12	318642	MITO324	560	TCACGCTCTATTTTCGCTAGA	TTTCCTTGTTTCATTTTCGATATTC
N13	319148	MITO325	560	AAGGGCCAAGCAAAGGAT	CATGTGCGGATCAAGTCTCT
N14	319687	MITO326	548	AAAGGACTTGATCCGCACAT	TTCGAACTGAACTACTCATGCC
N15	320208	MITO327	549	GCCTAGGCATGAGTAGTTCAGTT	ATCTCTGAAATGAGGGTGC
N16	320736	MITO328	560	GGCACCCTCATTTCCAGAG	CATGGCAAAGCACTTGTTC
N17	325505	MITO329	560	CAGCAAGCACAAAGGTTTC	GGGACTTTGTCTCCAGCAAT
N18	326041	MITO330	560	GTGGATTGCTGGAGACAAAG	TAAACTTCTGGCACGTGGTC
N19	326565	MITO331	560	TGATCTTTAGCTAGCCGACCAC	CGGCTTGATTGGCTAAGG
N20	327109	MITO332	538	TTAGCCAATCAAGCCGGAC	GAAAGCACCTGTTGGTTGGT
N21	327612	MITO333	560	TTCGACAATCGTGAACCAA	GCCGTATGATGGGCAACTAT
N22	328152	MITO334	559	ATAGTTGCCCATCATACGGC	AGCCTCGAGAAGCACTGTTG
N23	328651	MITO335	551	AACCCTGTAGAGAATGAAGAGGG	AATCAGTGTGGAGGAGAAGCA
N24	329169	MITO336	556	AATGAGCTCGGCTGCTTCT	GGAAGGTTGCCACGAATAAG
O01	329705	MITO337	557	CTTATTCGTGGCAACCTTCC	ACAGCATTCCCTTTCTATTGGC
O02	330240	MITO338	560	GCGCAATAGAAAGGAATGCT	GATTGCCGGAGCCTCTC
O03	330747	MITO339	559	AGGCTGCTATCCCAATAGGC	TCCGAATGCTTCTTCAATCC
O04	331266	MITO340	560	TTCTCTTCGCGAGAACTTCC	GATGAGGTGAGGGAAGTAGG
O05	331770	MITO341	560	GGAAAGCGAAGAACCAATCTT	CCCTCTCCTTCAGTCGAGTT
O06	332269	MITO342	559	GAAGTCACTCGAGCACTTGCT	CATCACTACCTCGTAAGGGCA
O07	332800	MITO343	560	GTCTCAATGCCCTTACGAGG	GAAGCCAAGCCGATACATACA
O08	333313	MITO344	563	GGGTCTTTCCTTCGCCTATC	CGAAGTGGTCATTAGTAAGCCAA
O09	333831	MITO345	560	TTGCCTTGGGTCAATCAGTT	CCTAACATGTAAATGGGTGC
O10	334344	MITO346	560	TTCCAGGCTGAGCACTTCAT	TTGAGCTCGTCAACCACTTC
O11	334876	MITO347	560	CCTTCACCGAAGTGGTTGAC	AGGAACGAACCGAACCTGAC
O12	347809	MITO348	512	TTTGCTTCCACATCCTATATTGA	TTTGTGCAAGGATAGATTGG
O13	348299	MITO349	560	CCCAATCTATCCCTTGACA	GAGGATTGGTCAGAAAGTGAT
O14	348808	MITO350	560	TGCCTAATCCGTAAGTTGTC	AGAGATTGGAATCGGTTTCG
O15	349346	MITO351	560	AACGAAACCGATTCCAATCTC	TGGTACAAAGTCTCCTGCTT
O16	350331	MITO352	478	GCCAGGACTATTCCAGTGAGT	TCCCGGTAAGTGGCAGG
O17	351135	MITO353	492	GGTGCACCTATCAGAATCGG	CAAGGTATGATTTCTGCCA
O18	351556	MITO354	497	CAAGCAGGAAATCCAACCTA	AACACACGGATTGTCAAT
O19	351992	MITO355	490	TTCTTCGCTACCTTTAGTATTCGG	ACTGGGATCCCGAAAGAAGT
O20	361748	MITO356	479	CCCACCTTCTCCAGTATATCA	CAGCTCGTGTGAGATGT
O21	361001	MITO357	116	GACTCCGTCGCGATGTTT	CCGACCTCGATATGTGGAAT
O22	361153	MITO358	132	TTCACGTCTCACCGTAGTGC	TAGCTTCTTCCACCCACT
O23	361286	MITO359	135	GTGGGTGGGAAGGAAGCTA	GGTCTTATTGGCGCATACCA
O24	143444	NAD1	392	CCAACCTCTGGAAGAACA	GAATTATTATAGCGGGTCGGTCTA
P01	328020	NAD2	1209	AGAATCCATGTCTAGGTGATCAA	TGTGGGTTGGCTTGATTA
P02	162132	NAD4	1000	AAAGCGTGCCAATCCCTA	TGCAGTCCGGGAACACTT
P03	20625	NAD5	1800	TCGATTATCTACCCAAGAAGATAGAGA	GGAGCTAGTGTGCTATCTAAGAAT
P04	132182	NAD7	841	GAACGTGCGGAACCACATA	CTTCTGTATAAAGTTCGAAATGGTGA



P05	155627	RPL2	95	AGGATGGAGCGTGCAAAG	CCAATCGCAATTCATCACC
P06	26984	RPS3	150	GACGAAAGCCGAAGGTGA	CGATTTCCGGTAAGACTTGGA
P07	9777	RRN26S	500	TTTCACCGGGTCCATGTC	CCATAGTCGGAGTCTGTTTAT
P08	51765	CCB452	151	GTGGGTCCATGTAAATGATCG	CACATGGAGGAGTGTGCATC
P09	40531	COX2	151	GATTGGATACCCGAGAACCAT	CCTGGTCGTTTAAATCAAATCTCTA
P10	287310	ALT_MITO365	465	GGGAGGGATTGACCTTTCTAA	GCCTAGCGCGAGTCTTATTG
P11	287767	ALT_MITO366	437	CGCTAGGCGATCACGTTT	TCATGCTGAGTCACAGGCAG
P12	288134	ALT_MITO367	499	CACACGTGCAAGTTTCCCT	CGGTCCGAATCTTATCTGAAC
P13	288571	ALT_MITO368	485	AGAATTACCTTATGGCACAGC	CGATCTGCAGCTCAAATGGT
P14	288988	ALT_MITO369	492	CTCACAAGGCGCACATAAG	TTAGATATGTCCTGAAACAATCCG
P15	349926	ALT_MITO453	478	TTCAGTTCAAGAGCCCAAGG	AAACACCCATGTTATTTGCTGT
P16	350750	ALT_MITO455	447	GGTTGTATTAAGTTTCGATCGGT	GTTATGCCGGCGATGATAG

**VI.2 Article 4 : Zaegel, V., Guermann, B., Le Ret, M., Andres, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J. M. and Imbault, P. (2006). The Plant-Specific ssDNA Binding Protein OSB1 Is Involved in the Stoichiometric Transmission of Mitochondrial DNA in Arabidopsis. *Plant Cell* 18, 3548-63.**

# The Plant-Specific ssDNA Binding Protein OSB1 Is Involved in the Stoichiometric Transmission of Mitochondrial DNA in *Arabidopsis*<sup>www</sup>

Vincent Zaegel,<sup>a</sup> Benoît Guermann,<sup>a</sup> Monique Le Ret,<sup>a</sup> Charles Andrés,<sup>b</sup> Denise Meyer,<sup>a</sup> Mathieu Erhardt,<sup>a</sup> Jean Canaday,<sup>a</sup> José M. Gualberto,<sup>a,1</sup> and Patrice Imbault<sup>a</sup>

<sup>a</sup>Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Université Louis Pasteur, 67000 Strasbourg, France

<sup>b</sup>Unité de Recherche en Génomique Végétale, 91057 Evry, France

**Plant mitochondrial genomes exist in a natural state of heteroplasmy, in which substoichiometric levels of alternative mitochondrial DNA (mtDNA) molecules coexist with the main genome. These subgenomes either replicate autonomously or are created by infrequent recombination events. We found that *Arabidopsis thaliana* OSB1 (for Organellar Single-stranded DNA Binding protein1) is required for correct stoichiometric mtDNA transmission. OSB1 is part of a family of plant-specific DNA binding proteins that are characterized by a novel motif that is required for single-stranded DNA binding. The OSB1 protein is targeted to mitochondria, and promoter- $\beta$ -glucuronidase fusion showed that the gene is expressed in budding lateral roots, mature pollen, and the embryo sac of unfertilized ovules. OSB1 T-DNA insertion mutants accumulate mtDNA homologous recombination products and develop phenotypes of leaf variegation and distortion. The mtDNA rearrangements occur in two steps: first, homozygous mutants accumulate subgenomic levels of homologous recombination products; second, in subsequent generations, one of the recombination products becomes predominant. After the second step, the process is no longer reversible by backcrossing. Thus, OSB1 participates in controlling the stoichiometry of alternative mtDNA forms generated by recombination. This regulation could take place in gametophytic tissues to ensure the transmission of a functional mitochondrial genome.**

## INTRODUCTION

In higher plants, mitochondrial genomes are large (367 and 570 kb in *Arabidopsis thaliana* and maize [*Zea mays*], respectively) (Unsold et al., 1997; Clifton et al., 2004), and their complete genetic repertoire can theoretically assemble in a genome-size circular chromosome. However, electrophoresis and electron microscopy studies showed that plant mitochondrial DNA (mtDNA) is a heterogeneous population of circular, linear, and complex double- and single-stranded molecules (Oldenburg and Bendich, 1996; Backert et al., 1997; Backert and Börner, 2000). These structures exist in dynamic equilibrium and probably result from a rolling-circle mechanism of DNA replication, but also from a recombination-dependent mechanism of DNA replication, similar to the replication of phage T4 (Backert and Börner, 2000). In addition, numerous recombination events between large and small repeats result in a multipartite structure of the plant mtDNA (Adams and Daley, 2004). Because the plant mtDNA is rich in repeated sequences, strict control of homologous recombina-

tion (HR) is essential for its stability. If recombination regulation is relaxed, then alternative mtDNA structures generated by HR could be generated during replication. These structures could give rise to rapidly replicating selfish genomes, containing an incomplete set of genes or expressing deleterious chimeric genes. The *Arabidopsis* mtDNA (ecotype C24) contains 22 pairs of 100% identical repeats of >100 bp, but only the two largest ones (6.5 and 4.2 kb) are involved in frequent reciprocal HR (Unsold et al., 1997). In both prokaryotes and eukaryotes, several proteins are known to suppress HR (Pinto et al., 2005), but in plant mitochondria, the mechanisms that regulate HR have not yet been characterized.

Adding to the complexity of plant mitochondria, the mtDNA is essentially heteroplasmic (Kmiec et al., 2006)—that is, different genome forms can coexist and replicate differentially. The ratios of the different types of mtDNA may vary, but usually one mtDNA configuration is prevalent and alternative configurations are present at sublimon levels (Small et al., 1989). Heteroplasmy can also originate from selfish elements, such as those at the origin of cytoplasmic male sterility, an agronomically important trait used by breeders to create high-yielding hybrids. In plants, very rapid changes may occur in the relative proportions of mtDNA variants, a phenomenon called substoichiometric shifting. These changes can occur under natural conditions (Janska et al., 1998), but they can also be induced in cybrids, in specific cell culture conditions, and in certain nuclear backgrounds (Kanazawa et al., 1994; Bellaoui et al., 1998; Kuzmin et al., 2005).

<sup>1</sup> To whom correspondence should be addressed. E-mail jose.gualberto@ibmp-ulp.u-strasbg.fr; fax 33-3-88-61-4442.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: José M. Gualberto (jose.gualberto@ibmp-ulp.u-strasbg.fr) and Patrice Imbault (patrice.imbault@ibmp-ulp.u-strasbg.fr).

<sup>www</sup>Online version contains Web-only data.

www.plantcell.org/cgi/doi/10.1105/tpc.106.042028

For example, in *Phaseolus vulgaris* pvs-orf239, a subgenomic molecule that undergoes substoichiometric shifting is amplified up to 2000-fold (Arrieta-Montiel et al., 2001), leading to cytoplasmic male sterility when the nuclear fertility-restorer *Fr* gene is inactive.

Mechanisms that regulate the stoichiometric transmission of the different mitotypes are still poorly understood. Substoichiometric shifting could result from increased HR, which continuously generates recombination products in somatic tissues, or from the favored replication of one of the mitotypes. It is also possible that increased HR activity creates a pool of sequences that, by strand invasion, prime the asymmetric replication of mtDNA chimeras. Whatever the mechanism regulating mtDNA heteroplasmy, it is expected to be active in rapidly dividing cells and in gametophyte cells. The nuclear control of substoichiometric shifting was shown to be more effective in undifferentiated meristem cells than in vegetative tissues (Arrieta-Montiel et al., 2001), and mtDNA reorganization was thus postulated to occur in transmitting tissues in which mtDNA replication is active. This hypothesis was recently corroborated by the work of Sheahan and colleagues (2005), who showed that massive mitochondrial fusion precedes fission and the dispersion of the organelles throughout the cytoplasm in newly prepared protoplasts. Massive mitochondrial fusion appears to be specific to the cell dedifferentiation process and therefore should facilitate the repackaging of mitochondrial genomes, thus ensuring the retransmission of all subgenomic molecules. In animals, a sharp reduction in mitochondrial genome number (the so-called bottleneck effect) accompanies oogenesis, and it was suggested that this phenomenon is particularly relevant to understanding how differential mitochondrial segregation is achieved during mitotic divisions (Barr et al., 2005). As a corollary to this model, if the control of HR and illegitimate recombination is relaxed, then selfish mitochondrial genomes arising from rearrangements, deletions, and insertions could accumulate and be preferentially transmitted, leading to mitochondrial dysfunction.

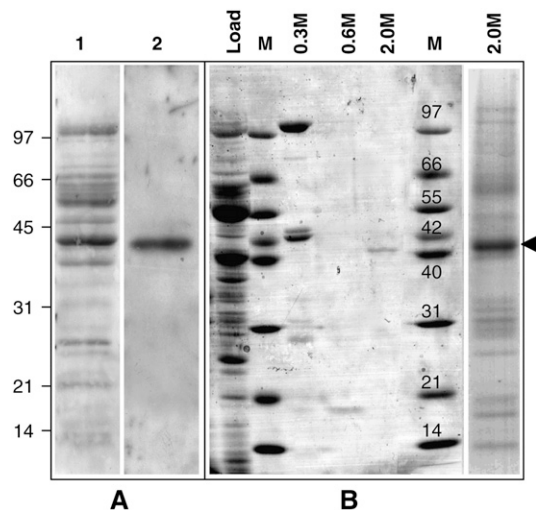
In plants, a component of stoichiometric regulation was identified: the *MSH1* gene (Abdelnoor et al., 2003), which encodes a protein similar to prokaryotic MutS and is responsible for the *chm* mutant phenotype (Martinez-Zapater et al., 1992) in *Arabidopsis*. In bacteria, MutS is involved in mismatch repair and the suppression of ectopic recombination. In yeast mitochondria, MSH1 is an essential gene required for mtDNA stability. However, recent evidence suggests that the mismatch repair function of yeast MSH1 is not essential but that other functions, such as recombination surveillance and heteroduplex rejection, are necessary for homoplasmic mtDNA sorting (Mookerjee and Sia, 2006). The precise mode of action of plant MSH1 in mtDNA substoichiometric shifting is still not known, but it could play a role similar to that of yeast MSH1. Another example of a gene involved in plant mtDNA instability by recombination was found in the maize P2 line (Kuzmin et al., 2005). In this line, a nuclear allele directly controls the amplification of several mitochondrial genome rearrangements. Several rearranged mtDNA forms could result from rare, reciprocal HR and could be generated continuously in vegetative tissues. However, other mtDNA forms could be created de novo by illegitimate recombination and amplified in P2 progeny lines.

Here, we describe the identification of a plant-specific OSB (for Organellar Single-stranded DNA Binding protein) family of mitochondrial and chloroplast proteins, several of which bind single-stranded DNA (ssDNA). Among these proteins, *Arabidopsis* OSB1 was shown to be required for mtDNA stability. In *osb1* T-DNA insertion mutants, the accumulation of mtDNA molecules derived from HR leads to severe morphological phenotypes. OSB1 is expressed primarily in gametophytic cells, in correlation with the need for a nuclear control on gametophytic tissues to ensure the stoichiometric transmission of alternative mitochondrial genomes to the plant progeny.

## RESULTS

### The Family of OSB Proteins Is Characterized by a Plant-Specific Motif

To identify plant mitochondrial proteins involved in mtDNA maintenance, potato (*Solanum tuberosum*) mitochondrial proteins with a high affinity for ssDNA were analyzed by protein/DNA gel blotting. To avoid sequence-specificity bias, a random sequence oligonucleotide was used as a probe. A 40-kD protein was the major mitochondrial ssDNA binding protein (SSB) detected by protein/DNA gel blotting (Figure 1A). A protein of the same size was purified from potato mitochondria by ssDNA-affinity chromatography (Vermel et al., 2002). This protein eluted from the column at high salt concentrations, denoting a high affinity for ssDNA (Figure 1B). It was thus named St OSB1, for



**Figure 1.** Purification of the Potato Mitochondrial OSB1 Protein.

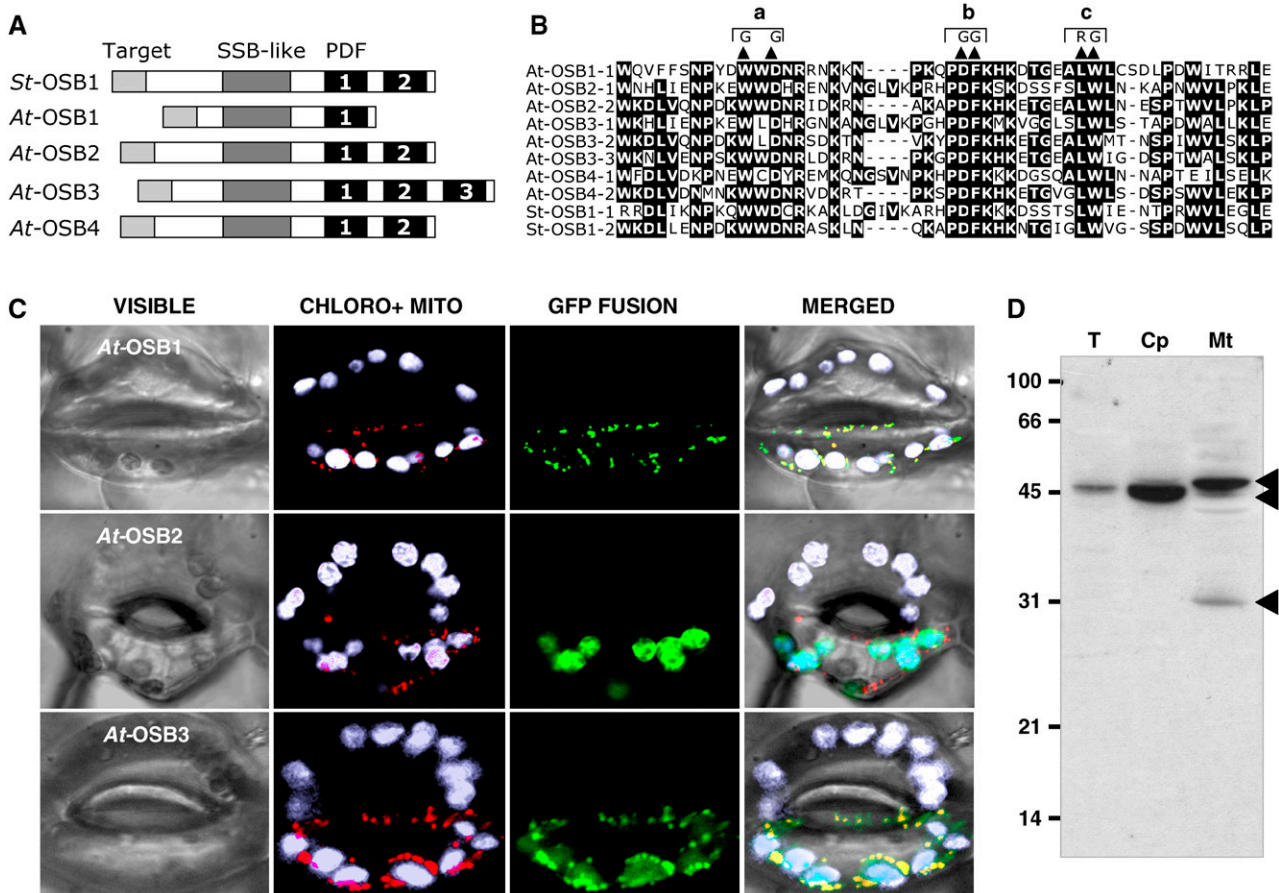
**(A)** Soluble proteins (8  $\mu$ g) from potato mitochondria were fractionated by SDS-PAGE, transferred to polyvinylidene difluoride membranes, and probed for DNA binding. Lane 1, Ponceau S staining of the proteins; lane 2, protein/DNA gel blot of proteins probed with a 5' end-labeled random 37-nucleotide single-strand oligonucleotide.

**(B)** Purification of the 40-kD protein by ssDNA affinity chromatography. Coomassie blue staining of the proteins eluted at 0.3, 0.6, and 2.0 M NaCl. The proteins eluting at 2.0 M were separated (right lane), and the 40-kD protein (arrowhead) was extracted and sequenced.

*Solanum tuberosum* Organellar ssDNA Binding protein1. The complete St OSB1 cDNA was obtained by RT-PCR using degenerate oligonucleotides deduced from tryptic peptide sequences, followed by 5' and 3' rapid amplification of cDNA ends. The cDNA encodes a 43-kD protein predicted to be mitochondrial by Predotar, TargetP, and iSort. Amplification was possible from total potato sprout RNA but not from leaf RNA, suggesting that the gene is expressed preferentially in rapidly dividing tissues.

We found orthologs of the St OSB1 gene in the *Arabidopsis thaliana*, rice (*Oryza sativa*), and *Zea mays* genomes. The highest homology between these sequences (40 to 80% identity) was found in a 50-amino acid C-terminal motif that is present in one to three copies. This motif was named PDF, because these three residues are conserved in all proteins of the family. Two of the three putative OSB genes in the rice genome code for proteins with a single PDF motif, and one codes for a protein with two PDF

motifs. PDF motifs were also found in numerous sequences deduced from ESTs of monocot and dicot plants. In the green alga *Chlamydomonas reinhardtii*, the chloroplast protein RB38 is constituted of four PDF motifs repeated in tandem (Barnes et al., 2004). RB38 is the only PDF motif protein in *Chlamydomonas*, according to the latest release of the *C. reinhardtii* genome (<http://genome.jgi-psf.org/Chlre3/Chlre3.home.html>). However, the PDF motif was not found in proteins of species outside of the plant kingdom. We focused our study on *Arabidopsis*, in which OSBs form a family of four putative proteins encoded by the nuclear genes At1g47720 (At OSB1), At4g20010 (At OSB2), At5g44785 (At OSB3), and At1g31010 (At OSB4). Complete cDNAs were obtained by RT-PCR for OSB1, OSB2, and OSB3. The OSB4 cDNA was not cloned, but the gene is apparently expressed, because there is a corresponding EST sequence. Figure 2A shows that OSB sequences can be divided into three domains: (1) a nonconserved sequence predicted to be



**Figure 2.** Characterization of the OSB Protein Family.

**(A)** Basic structure of OSB proteins from potato (St OSB1) and *Arabidopsis* (At OSB1 to OSB4). The PDF motifs are numbered.

**(B)** Alignment of PDF motifs of the proteins. The double substitution mutations (a, b, and c) analyzed in Figure 3 are indicated.

**(C)** Localization of protein-eGFP fusions in guard cells of *N. benthamiana*. Green, eGFP fluorescence; white, natural fluorescence of chloroplasts; red, fluorescence of the mitochondrial marker.

**(D)** Protein gel blot of protein fractions probed with antibodies raised against OSB1: total fraction (T)-, chloroplast (Cp)-, and mitochondria (Mt)-enriched fractions. Protein fractions were extracted from *Arabidopsis* cells in suspension culture as described (Laloi et al., 2001).

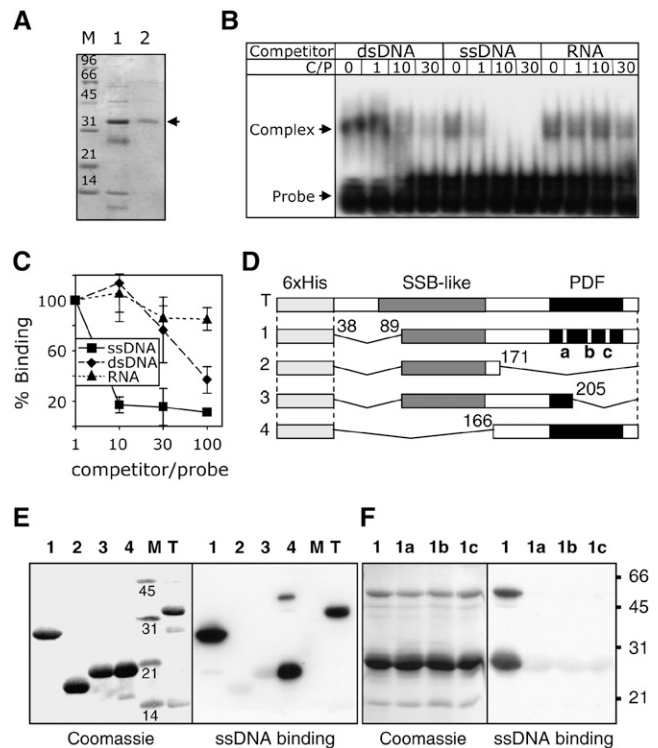
an organellar targeting sequence; (2) a central region that has the signature of the prokaryotic SSB family (PFAM00436); and (3) a C-terminal region whose length depends on the number of PDF motifs: one in At *OSB1*, two in At *OSB2*, At *OSB4*, and St *OSB1*, and three in At *OSB3* (Figure 2B).

### OSB Proteins Are Targeted to Mitochondria and Chloroplasts

OSB proteins are predicted to be organellar from sequence analysis. To localize At *OSB1*, two and three green fluorescent protein (GFP) fusion proteins, the N-terminal regions, up to the conserved SSB domain, were cloned in pCK-GFP3A (Menand et al., 1998) and transiently expressed in epidermal cells of *Nicotiana benthamiana* leaves. The N-terminal sequences of OSB1 and OSB3 target eGFP into mitochondria (Figure 2C). In all cells observed, the OSB3-eGFP fusion is also targeted to the chloroplast. By contrast, the OSB2 fusion protein was detected only in chloroplasts. To confirm these results, antibodies directed against At *OSB1* were generated. Affinity-purified antibodies recognized both At *OSB1* and At *OSB2* recombinant proteins, respectively, expressed in *Escherichia coli* and in planta (see below; data not shown). Three OSB proteins were detected by these specific antibodies in fractions extracted from protoplasts of *Arabidopsis* cultured cells (Figure 2D). In the whole cell extract, a single 45-kD protein is detected. This protein is considerably enriched in the chloroplast fraction. In the mitochondrial extract, two other proteins of 48 and 30 kD are detected. The sizes of these three proteins are compatible with the sizes predicted for mature At *OSB1* (31 kD), At *OSB2* (42 kD), and At *OSB3* (49 kD). These results suggest that OSB2 is targeted to chloroplasts, whereas OSB1 and OSB3 are targeted to mitochondria, in agreement with the localization of the corresponding GFP fusion proteins. The OSB3-GFP fusion protein was also detected in chloroplasts, and a faint band that could correspond to OSB3 was also immunodetected in the chloroplast fraction. However, this could result from mitochondrial contamination of the chloroplast fraction. Therefore, the dual targeting of At *OSB3* should be confirmed by another approach. Given the limitations of the methods used, we cannot completely exclude dual targeting of a small proportion of other At OSB proteins.

### OSB Proteins Preferentially Bind ssDNA

To investigate the nucleic acid binding capacity of *Arabidopsis* OSB proteins, recombinant OSB1 and OSB2 were expressed and purified (Figure 3A; see Supplemental Figure 1 online). The binding of soluble OSB1 protein to a labeled ssDNA probe was tested by electrophoretic mobility shift assay in the presence of cold competitors (double-stranded DNA [dsDNA], ssDNA, or RNA). To avoid sequence-specific effects and stable secondary structures, ACTG (or ACUG) repeats were used as probes. OSB1 binding was competed for by ssDNA, but it was poorly competed for by dsDNA or RNA (Figure 3B). Because it was difficult to obtain large quantities of soluble OSB1, additional binding assays were done on protein/DNA gel blots using renatured OSB1. The protein was purified under denaturing conditions, transferred to



**Figure 3.** ssDNA Binding Activity of At *OSB1*.

(A) Nickel-nitrilotriacetic acid agarose affinity chromatography purification of soluble At *OSB1* expressed in *E. coli*. Lane 1, proteins eluting with 50 mM imidazole; lane 2, purified protein eluting with 150 mM imidazole.

(B) Electrophoretic mobility shift assay of nucleic acid binding. At *OSB1* and  $^{32}\text{P}$ -labeled ssDNA probe were incubated with increasing quantities of cold competitor before electrophoresis. dsDNA, ssDNA, and RNA competitors were of the same size and sequence as the probe. The molar ratio of competitor to probe (C/P) is given. At *OSB1*/ssDNA complexes are shown by arrowheads.

(C) At *OSB1* purified under denaturing conditions was tested on protein/DNA gel blots for binding to labeled ssDNA in the presence of increasing concentrations of cold competitor (ssDNA, dsDNA, or RNA). Results were quantified using a phosphor imager. Error bars indicate the SD of three independent experiments.

(D) Structure of At *OSB1* mutants expressed in *E. coli*. SSB-like, PDF, and deleted regions are indicated. a, b, and c refer to the double substitution mutations described for Figure 2B.

(E) Analysis of ssDNA binding by At *OSB1* mutant proteins. Expressed proteins described for (D) were fractionated by SDS-PAGE, transferred to polyvinylidene difluoride membranes, and probed with labeled ssDNA.

(F) Same as (E) using construct 1 with mutation a, b, or c.

polyvinylidene difluoride membranes, and incubated with labeled ssDNA probe and variable amounts of cold competitor (ssDNA, dsDNA, or RNA). The renatured OSB1 has a much higher affinity for ssDNA than for dsDNA or RNA (Figure 3C), as shown above for the soluble protein. Both OSB1 (Figures 3B and 3C) and OSB2 (see Supplemental Figure 1 online) showed a higher affinity for ssDNA than for dsDNA, and neither bound to RNA. The affinity of OSB2 for ssDNA ( $K_d = 1.9 \pm 0.5$  nM; see

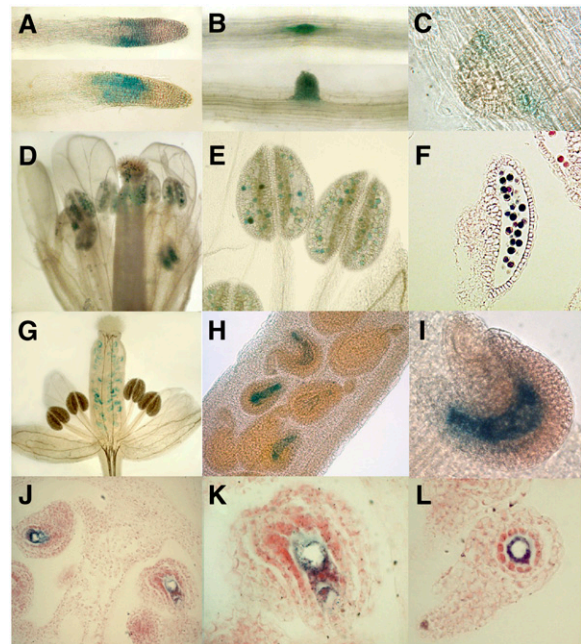
Supplemental Figure 1 online) is comparable to that found for other high-affinity DNA binding proteins.

### The PDF Motif of OSB1 Is Involved in ssDNA Binding

The *Arabidopsis* OSB1 protein has a simple structure consisting of an SSB-like domain followed by a single PDF motif. To define the protein domains involved in ssDNA binding, deletion mutant constructs of OSB1 (Figure 3D) were expressed in *E. coli* and tested on protein/DNA gel blots for ssDNA binding. The results obtained showed that large portions of the protein (mutants 1 and 4) could be deleted without affecting its ability to bind ssDNA (Figure 3E). However, when the PDF motif was partially deleted (mutants 2 and 3), most of the ssDNA binding capacity was lost. To confirm that the PDF motif is directly involved in ssDNA binding, we constructed three additional mutants (a, b, and c; see Figure 2B) starting from deletion mutant 1 (Figure 3D). In each mutant, two amino acids were chosen among the most conserved residues in the PDF motif and were mutated to radically different residues (Figure 2B). Each of the mutations drastically reduced the ssDNA binding capacity of the protein (Figure 3F), thus suggesting that the PDF motif takes part in DNA binding. Additional bands were often observed by protein/DNA gel blots, corresponding to proteins twice the size of the protein monomers (mutants 1 and 4; Figures 3E and 3F). Because both prokaryotic and eukaryotic ssDNA binding proteins are known to form homomultimers or heteromultimers, it is possible that these bands are dimers that do not dissociate during SDS-PAGE.

### At OSB1 Is Expressed in Roots and in Gametes

The *OSB1* transcript is not detectable in total plant RNA by RNA gel blot hybridization. Microarray results, analyzed with GENEVESTIGATOR (<https://www.geneinvestigator.ethz.ch/at/>), showed a higher expression level in cell suspension cultures but did not reveal significant developmental regulation. To investigate expression in different tissues, promoter- $\beta$ -glucuronidase (GUS) fusions were constructed. The sequence between the start codon of *OSB1* and the stop codon of the upstream gene was cloned fused to the GUS gene (ProOSB1:GUS construct) and introduced in *Arabidopsis* by agroinfection. In young seedlings of six independent ProOSB1:GUS plant lines, gene expression was visible primarily in budding secondary roots (Figures 4B and 4C) but was no longer visible in developed lateral roots. In some plantlets, a short region of the root apex was also stained (Figure 4A), corresponding to the root elongation zone. In mature flowers (stage 14 according to Smyth et al. [1990]), anthers were fully colored by GUS staining of the pollen grains (Figures 4D to 4F). In the same flower (Figure 4E), anthers containing stained pollen were observed (from 0 to 100%), indicating that gene expression was independently induced in each maturing pollen grain. In young flowers (stage 11), strong GUS staining was also visible in the embryo sac of unfertilized ovules (Figures 4G to 4I), where the central cell was clearly stained (Figures 4J to 4L). The expression of At *OSB1* thus appears to be temporally and spatially restricted primarily to gametophytic cells. Similarly, promoter-



**Figure 4.** Histochemical Localization of GUS Activity in *Arabidopsis* Plants Transformed with the Intergenic Region Upstream At *OSB1* Gene Fused to the GUS Reporter Gene.

(A) to (C) Roots in an 18-d-old plantlet.

(D) to (F) Expression is detected in pollen grains of mature flowers. Expression in unfertilized ovules is visible in immature flowers.

(G) Whole flower.

(H) and (I) Closer views.

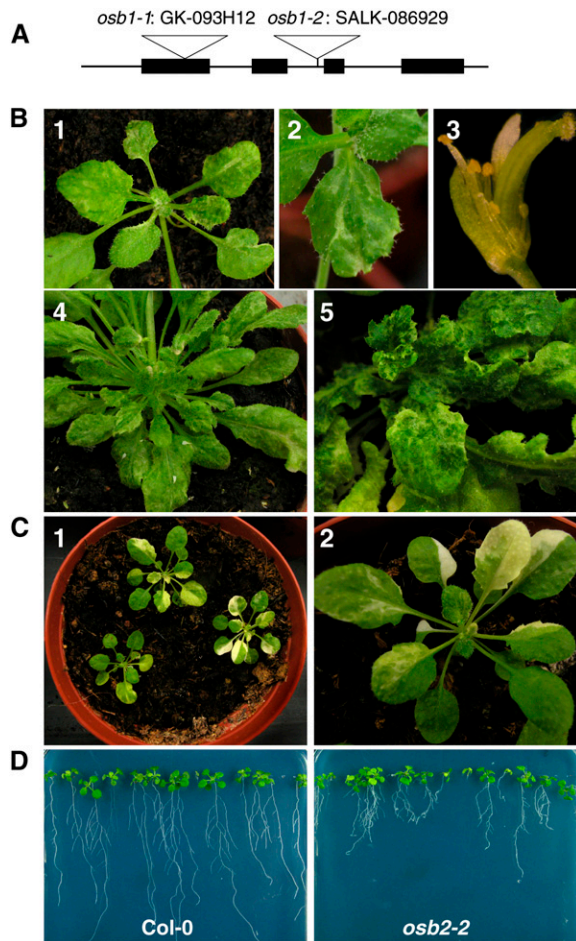
(J) to (L) Sections of unfertilized ovules showing that expression is restricted to the embryo sac.

GUS fusions of the At *OSB2* and At *OSB3* genes showed that the mitochondria-targeted OSB3 protein is expressed primarily in the female gametophyte, but the chloroplast-targeted OSB2 protein is not (data not shown).

The expression pattern we detected is consistent with the idea that mitochondrial OSB proteins are ssDNA binding proteins that could play a major role in mtDNA function during gametogenesis.

### OSB1 T-DNA Insertion Mutants Can Develop Severe Leaf Variegation and Distortion Phenotypes

To identify the possible roles of OSB proteins in organellar genome maintenance, T-DNA insertion mutants were obtained for each of the At *OSB1*, *OSB2*, and *OSB3* genes. Visible phenotypes were observed only for mutants affected in *OSB1* expression. Two mutant alleles in the Columbia (Col-0) background were obtained, *osb1-1* (GK-093H12 from GABI-Kat) and *osb1-2* (SALK\_086929 from the SALK collection). T-DNA insertions were in the first exon (position 173) and in the second intron (position 870) of *OSB1*, respectively (Figure 5A). In the homozygous lines, no *OSB1* mRNA was detected by RT-PCR, suggesting that the



**Figure 5.** Variegated and Distorted Phenotypes of At OSB1 T-DNA Insertion Mutants.

**(A)** Physical map of the At OSB1 gene. The positions of the T-DNA insertions in *osb1-1* and *osb1-2* mutants are indicated.

**(B)** Examples of variegation and distortion phenotypes in leaves and flowers of *osb1-1* (panels 1, 2, and 4) and *osb1-2* (panels 3 and 5) mutants.

**(C)** Examples of variegation in leaves of an *msh1* T-DNA insertion mutant.

**(D)** Comparison of the roots from Col-0 and *osb2-2* plants at 21 d after germination on Murashige and Skoog agar plates.

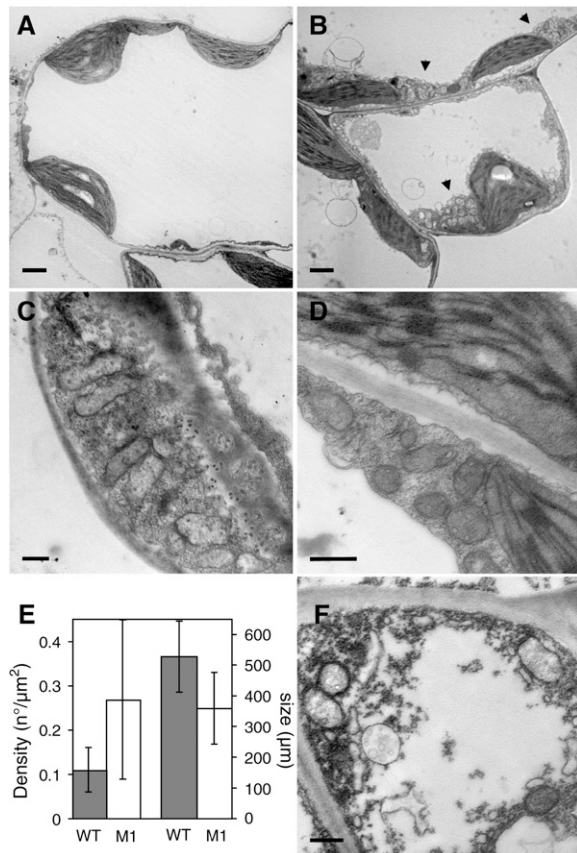
insertion lines are authentic knockout OSB1 mutants. DNA gel blot hybridization of *osb1-1* plants with T-DNA-specific probes indicated that there is a single T-DNA insertion. No particular phenotype was observed in the T3 homozygous plants that were analyzed. However, at generations T4 and T5, several *osb1-1* and *osb1-2* plants showed anatomical and/or developmental defects (Figure 5B): variegation, growth retardation, distortion of leaves and flowers, partial sterility, and production of unviable seeds. The severity of the phenotype varied from plant to plant and became more pronounced in later generations. Retarded growth was most evident in the roots. No major differences were observed between *osb1* plants and wild-type plants at 7 d after germination. However, after 21 d, root growth was delayed in

*osb1* plants and the development of secondary roots was abnormal, giving a bushy appearance to the *osb1* roots (Figure 5D).

The leaf variegation and distortion phenotypes in *osb1* are similar to the phenotypes described for *chm* mutants, which are affected in the expression of the *MSH1* gene. This gene encodes a MutS homolog that is targeted to both mitochondria and chloroplasts and is involved in mtDNA substoichiometric shifting (Martinez-Zapater et al., 1992; Sakamoto et al., 1996; Abdelnoor et al., 2003; Christensen et al., 2005). However, the phenotypes conferred by *osb1* and *msh1* are not identical, because white sectors (attributable to chloroplast deficiency) frequently observed in *msh1* plants (T-DNA insertion line SALK\_046763; Figure 5C) have never been observed in *osb1-1* or *osb1-2* mutant plants. Furthermore, sections of variegated leaf sectors of *osb1-1* plants were observed by transmission electron microscopy (Figure 6) and revealed no abnormal chloroplast morphology, unlike *msh1* plants (Sakamoto et al., 1996). In most sections observed, mitochondria of mesophyll cells also had normal morphology but were smaller. In some cases, abnormal mesophyll cells were also observed (Figure 6F). Clusters of mitochondria were frequently visible in the mutant (Figures 6B to 6D) but not in control plants (Figure 6A). A statistical analysis of mitochondrion size and number showed that mitochondria were  $\sim 30\%$  smaller in *osb1-1*. The reduction in mitochondrion size appeared to be counterbalanced by an increase in density, with  $0.28$  mitochondria/ $\mu\text{m}^2$  in *osb1-1* compared with  $0.11/\mu\text{m}^2$  in wild-type Col-0 (Figure 6E). Variations in mitochondrion number and size may reflect an increase in mitochondrion division necessary to counterbalance a reduction in the respiratory performance of mutant mitochondria (Barr et al., 2005). Mitochondria of low electron density were also often observed (data not shown), suggesting that, in the affected tissues, there is a heterogeneous population of normal and deficient mitochondria.

#### ***Arabidopsis* OSB1 Plays a Role in Regulating mtDNA Recombination**

Because OSB1 binds ssDNA and is localized in gametophytic tissues, OSB1 could be involved in mtDNA maintenance. To investigate whether mtDNA structure was altered in *osb1* mutants, several *osb1* mutant plants were tested by PCR, using primers designed to amplify mtDNA regions of  $\sim 500$  nucleotides corresponding to mitochondrial genes and their flanking sequences (see Methods). In mutant plants, several specific pairs of primers failed to amplify the mtDNA in regions surrounding the *nad5*, *ccmFe*, *rpl5*, *atp1*, *cox1*, and *cox2* genes, whereas these loci could be amplified from Col-0 DNA (see Supplemental Figure 2 online). These results suggested that mtDNA was rearranged or that changes had occurred in gene copy numbers. To test this hypothesis, total DNA from Col-0 wild-type plants and from one plant each of the *osb1-1* and *osb1-2* mutant lines having morphological phenotypes was analyzed by DNA gel blot hybridization using several mtDNA gene-specific probes. The patterns of mtDNA from mutants differed from wild-type mtDNA around several genes, including the *atp9*, *atp1*, *atp6*, *cox2*, and *cob* genes (Figure 7A): stoichiometry was altered, additional fragments were detected, and other fragments were no longer visible. Additional



**Figure 6.** Clustering of Mitochondria in *osb1* Mesophyll Cells.

(A) and (B) Transmission electron microscopy of sections from variegated leaves of *osb1-1* plants. Mesophyll cells from wild-type (A) and *osb1-1* (B) plants, showing clusters of mitochondria in *osb1-1* (arrowheads).

(C) and (D) Details of mesophyll cells from *osb1-1*, showing clusters of mitochondria and the accumulation of mitochondria in regions of cell wall ingrowths.

(E) Increase in the number and decrease in the size of mitochondria in *osb1-1*. The mean surface of mitochondria cross sections was obtained from the analysis of 20 electron microscopy images of wild-type and *osb1-1* (M1) leaves. Error bars indicate SD.

(F) Abnormal mesophyll cell in *osb1-1*.

Bars = 2 μm in (A) and (B) and 500 nm in (C), (D), and (F).

restriction fragments were detected in both *osb1-1* and *osb1-2* plants (*atp9* and *atp1* hybridization). However, some minor bands were absent only in the *osb1-1* plant (*atp6*, *cox2*, and *cob* hybridization). In the wild-type plants, the level of these fragments was low compared with the gene-containing fragments. These fragments do not correspond to the published mtDNA sequence of *Arabidopsis* ecotype C24 and could be rearranged mtDNAs that exist as sublimons in wild-type mitochondria. These results suggest that, in the *osb1* mutant background, there is segregation of plants with rearranged mtDNA, as a result of enhanced recombination activity and/or changes in the relative copy number of substoichiometric molecules.

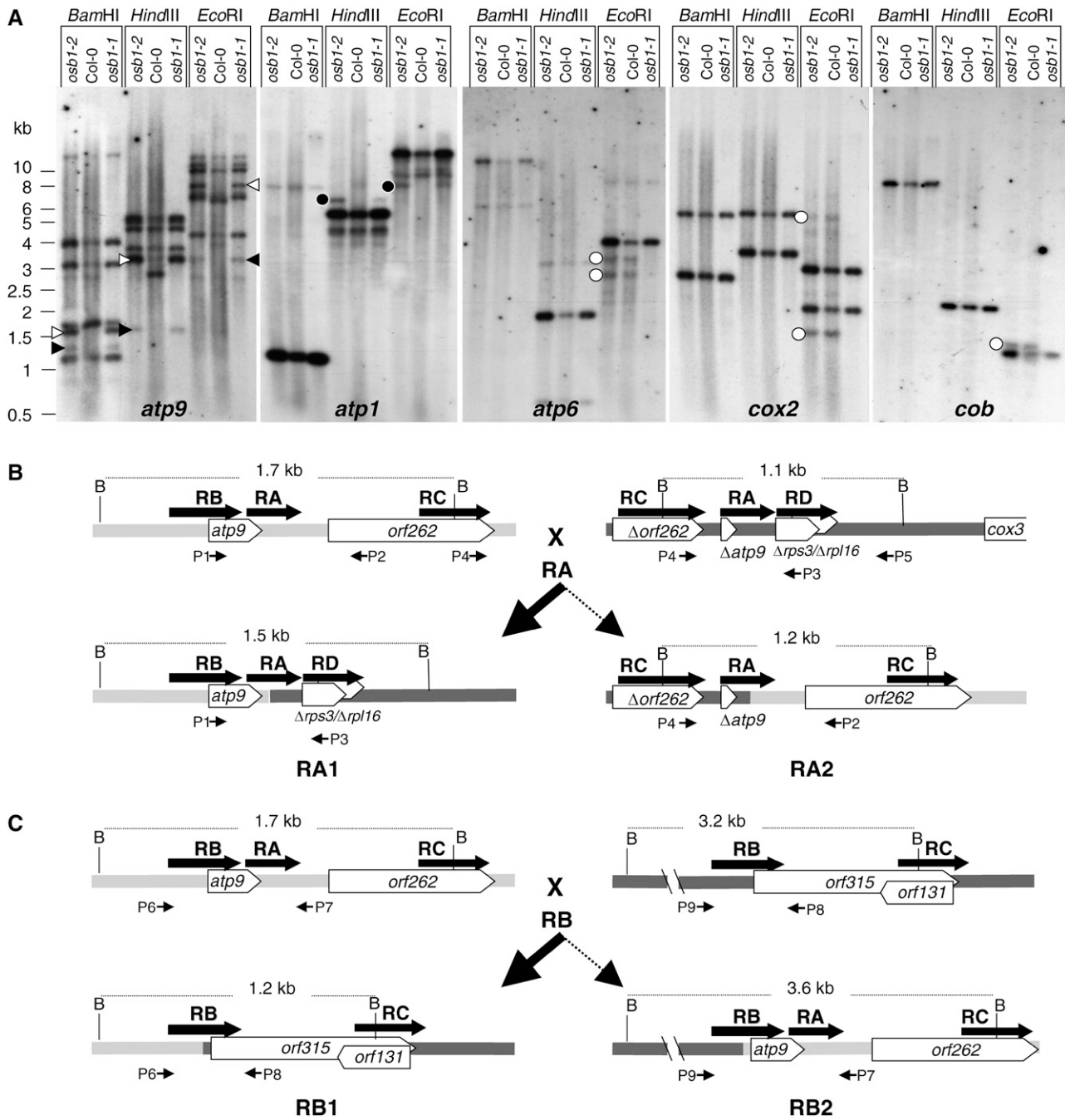
### Substoichiometric Shifting in *osb1* Mutants Results from Asymmetric Amplification of HR Products

To investigate the mechanism that leads to substoichiometric shifting in *osb1* plants, we focused on the *atp9* gene locus, which undergoes substoichiometric shifting in the *msh1* and *msh1*-derived MDL mutants (Sakamoto et al., 1996; Abdelnoor et al., 2003). As shown in Figure 7A, four major *Bam*HI bands hybridize with a 748-bp fragment containing the whole *atp9* gene and the 5' region of *orf262* (fragment amplified using primers P1 and P2 [Figure 7B], positions 278,908 to 279,656 of the mtDNA genome [ecotype C24]) (Unsel et al., 1997). The 1.7-, 3.2-, and 4.0-kb fragments contain the *atp9* gene locus (Figure 7B), a 334-bp repeat (RB) that overlaps the 5' end of *orf315* (Figure 7B; positions 17,635 to 17,968 of the genome sequence), and a 259-bp repeat containing the 5' end of exon e of *nad5* (positions 142,922 to 143,180), respectively. The expected *Hind*III and *Eco*RI fragments were also identified, confirming our interpretation of the hybridization results (Figure 7A).

In Col-0, an additional 1.1-kb *Bam*HI fragment hybridizes to the *atp9* probe, which cannot be accounted for by the C24 mtDNA sequence. We mapped this fragment and showed that it differs in the mitochondrial genomes of ecotypes C24 and Col-0. By PCR and sequence analysis, the 1.1-kb *Bam*HI fragment mapped to the region upstream of *cox3*. Recently, significant differences were reported in the *cox3* region of ecotypes C24, Landsberg *erecta* (*Ler*), and Col-0 (Forner et al., 2005). In Col-0, upstream of *cox3* there are three sequences repeated elsewhere in the mtDNA. We named these sequences repeats RA, RC, and RD (Figure 7B). Repeat RA (248 bp) is constituted of the *atp9* 3' coding region and untranslated region sequences (sequence  $\Delta atp9$  in Figure 7B). RA is the sequence responsible for the hybridization of the 1.1-kb fragment with the *atp9* probe. Repeat RC (407 bp) corresponds to the 3' end of *orf262*, an open reading frame that maps downstream of *atp9* (sequence  $\Delta orf262$  in Figure 7B). Repeat RD (251 bp) overlaps the *rps3* and *rpl16* genes (sequence  $\Delta rps3/\Delta rpl16$  in Figure 7B), containing the 3' end of *rps3* and the 5' end of *rpl16*.

As a result of substoichiometric shifting in the *osb1* mutant lines, two additional *Bam*HI fragments of 1.5 and 1.2 kb hybridizing with the *atp9* probe (Figure 7A) were amplified. According to our PCR and sequence analysis, the 1.5-kb fragment corresponds to one of the products (RA1) obtained from the reciprocal HR between the two copies of the RA repeat (Figure 7B). As a result, the *atp9* gene is relocated upstream of *cox3*. Similarly, the 1.2-kb fragment is the RB1 product resulting from HR between the two copies of the RB repeat (Figure 7C). In this rearrangement, the *atp9-orf262* genes are replaced by the overlapping *orf315-orf131* sequences (Figure 7C). Our interpretation of these hybridization results is supported by the results of the *Hind*III and *Eco*RI digests (Figure 7A). Fragments of the sizes expected for the RA1 and RB1 recombination products undergo the same changes in relative copy number as the *Bam*HI fragments. The expected HR products RA1 and RB1 were amplified by PCR from the *osb1* mutants and sequenced. Interestingly, the reciprocal products RA2 and RB2 resulting from the reciprocal HR processes were not detected by hybridization. This feature is discussed below.





**Figure 7.** Recombination of mtDNA in *osb1* Mutants.

**(A)** DNA gel blots of total flower DNA (3  $\mu$ g/lane) from Col-0 and *osb1-1* and *osb1-2* mutants hybridized with *atp9*, *atp1*, *atp6*, *cox2*, and *cob* probes. Open arrowheads, RA1 recombination product; closed arrowheads, RA2 recombination product; closed circles, fragments that appear only in the mutant lines; open circles, fragments that disappear in one of the mutant plants.

**(B)** and **(C)** Analysis of the recombination process affecting the *atp9* gene locus. RA, RB, RC, and RD indicate repeated sequences. *Bam*HI fragments and their sizes are indicated. Triangles indicate partial gene sequences. P1 to P9 are primers used for PCR amplification.

**(B)** Recombination mediated by repeat RA present in the *atp9* locus and upstream of *cox3*. The structures predicted for the most abundant heteroduplex (RA1) and the unfavored reciprocal heteroduplex (RA2) are shown. The corresponding fragments are indicated by open arrowheads in the *atp9* DNA gel blot.

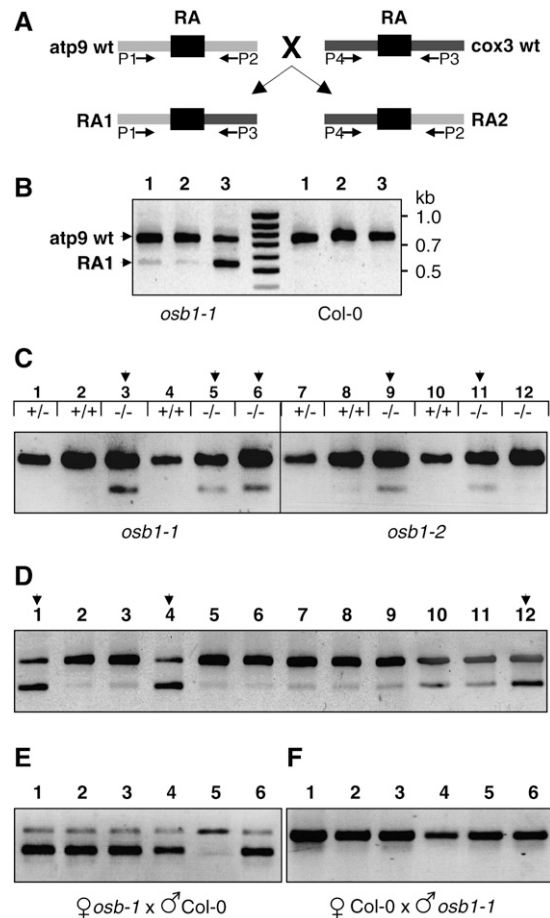
**(C)** Recombination mediated by repeat RB present in the *atp9* and *orf315* mtDNA environments. The structures of the most abundant heteroduplex (RB1) and the unfavored reciprocal heteroduplex (RB2) are shown. The corresponding fragments are indicated by closed arrowheads in the *atp9* DNA gel blot.

In *msh1* mutants, the same *Bam*HI 1.5- and 1.2-kb fragments were associated with substoichiometric shifting in the Col-0 background (Martinez-Zapater et al., 1992; Abdelnoor et al., 2003). However, when an *msh1* mutation was backcrossed into a *Ler* cytoplasm, giving rise to the MDL mutant lines (Sakamoto et al., 1996), different sequences were involved. In *Ler*, a copy of *atp9* maps upstream of *cox3*. The product of the RA-mediated HR amplified by substoichiometric shifting in Col-0 appears to be the predominant wild-type sequence in *Ler* (Forner et al., 2005). The substoichiometric shifting events described by Sakamoto et al. apparently resulted from HR involving the pair of repeats RD (Figure 7B), leading to the relocation of *atp9* upstream of *rp116* and the concomitant deletion of most of *rps3*. In the reciprocal recombination product, *rps3* is found upstream of *cox3* (fragments LMB56 and LMB23, respectively, in Sakamoto et al., 1996). This interpretation is consistent with the sequences reported by those authors and with the sizes of the restriction fragments on DNA gel blots. An important conclusion from this analysis is that substoichiometric shifting in *osb1*, as well as in the *msh1* and *msh1*-derived MDL mutants, involves the products of reciprocal HR between rather large repeats (>200 bp) and not illegitimate recombination between short repeated sequences, as was assumed previously.

#### Analysis of Substoichiometric Shifting at the *atp9* Gene Locus

The distinctive morphological phenotypes of leaf distortion and variegation of *osb1* mutants were observed only at generations T4 and T5, after selfing of homozygous plants. The substoichiometric shifting events described above at the *atp9* gene locus were analyzed in these plants. To determine the effect of the mutation on mtDNA rearrangement, the appearance of recombination products was followed in successive plant generations. We developed a competitive PCR protocol, using three primers (P1 + P2 + P3; Figures 7B and 8A), to determine the proportion of the wild-type *atp9* sequence and the RA1 recombination product. A similar competitive PCR method was used to analyze substoichiometric shifting in *msh1* mutants (Sakamoto et al., 1996; Abdelnoor et al., 2003). It must be emphasized that dramatic differences in amplification of the two fragments were obtained with small differences in the primer ratio. Under these conditions, the *osb1* RA1 sequence was amplified but no product was detected in the wild type (Figure 8B). The RA1 sequence was also amplified in the *msh1* T-DNA insertion mutant using this technique (data not shown). The profiles obtained with *osb1* homozygous mutant DNA (Figure 8B) show that the proportion of the *atp9* wild-type (748 nucleotides) and RA1 (546 nucleotides) products varied according to the plant. In all Col-0 plants, only the *atp9* sequence was detected. However, RA1 and RA2 sequences were present at low levels in wild-type plants, because both could be detected by a conventional two-primer PCR (Figure 9).

Plants issued from mixed seeds of the *osb1-1* insertion mutant obtained from GABI-Kat (line GK-093H12, generation T2) were genotyped and tested (Figure 8C). Fourteen of the 20 plants analyzed were either wild type (Col-0) or heterozygous for the T-DNA insertion. In the heterozygous plants, the *atp9* wild-type



**Figure 8.** Accumulation of the RA1 Recombination Product in *osb1* Mutants Affected by mtDNA Substoichiometric Shifting.

(A) Simplified scheme showing the RA repeats in wild-type Col-0 *atp9* and *cox3* contexts and the resulting recombination products RA1 and RA2. Primers P1 to P4 (Figure 7B) were used to amplify the four mtDNA configurations.

(B) Analysis by optimized competitive, three-primer PCR of three *osb1-1* plants affected by shifting and of three Col-0 plants. The top band (P1 + P2) corresponds to wild-type *atp9* (748 nucleotides), and the smaller fragment (P1 + P3) corresponds to the RA1 recombination product (546 nucleotides).

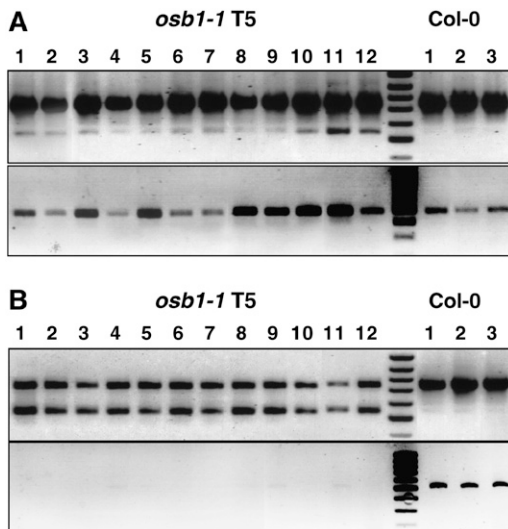
(C) Analysis of RA1 accumulation in the first generation of *osb1* homozygous mutants. The *OSB1* genotype of each plant is indicated. Homozygous plants 3, 5, 9, and 11 have increased levels of RA1 (stage I shift; as described in Results and in Supplemental Table 2 online).

(D) Progeny (generation T3) of homozygous plant 3 from (C): plants 1, 4, and 12 show much higher levels of RA1 than their siblings (stage II shift).

(E) Analysis of plants resulting from the backcrossing, as pollen receptor, of plant 4 from (D).

(F) Analysis of plants resulting from the reciprocal cross.

gene locus was amplified predominantly, whereas the recombination product RA1 was barely visible (Figure 8C; see Supplemental Table 2 online). In Supplemental Table 2 online, this PCR pattern is referred to as shifting stage 0 (plants 1, 2, and 4 in Figure 8C). In pattern stage I, the amount of RA1 is higher, but it is



**Figure 9.** Substoichiometric Shifting Is Accompanied by the Loss of Reciprocal Recombination Products.

**(A)** Detection of the reciprocal recombination product RA2 (bottom panel) in the progeny of a T4 plant that shows partial mtDNA shifting (stage I; top panel).

**(B)** The same as **(A)**, in the progeny of a plant that has a complete mtDNA shift (stage II).

still at a substoichiometric level (see Supplemental Table 2A online). The stage I pattern was found for three of the six homozygous mutant plants analyzed (plants 3, 5, and 6 in Figure 8C). Similar results were obtained for plants issued from the self-fertilization of a T3 *osb1-2* plant that is heterozygous for the T-DNA insertion (line SALK\_086929). Two of the three homozygous plants analyzed (plants 9 and 11 in Figure 8C) showed a significant increase in the RA1 copy number (see Supplemental Table 2A online). Because *osb1-1* and *osb1-2* mutants are both in the Col-0 background, the molecular phenotype is not specific to the plant line used for T-DNA insertion mutagenesis. From these results, we can conclude that the loss of *OSB1* function led to an increase in the copy number of mtDNA molecules that arose from HR. This effect was already detected in the first generation of the mutants.

In *Arabidopsis*, the published mtDNA sequence of ecotype C24 contains >20 pairs of repeats of >100 bp. Only the two largest repeats (6.5 and 4.2 kb) are described as contributing to the multipartite structure of the genome (Unsel et al., 1997). Therefore, we tested whether recombination events could be mediated by short repeats of several hundred base pairs. A 435-bp sequence situated in exon 1 of the *cox2* gene is repeated elsewhere in the genome (at positions 319,174 to 319,608 of the published sequence). A 556-bp sequence covering the first exon of *rps3* and the tRNA-Lys gene is repeated at positions 204,102 to 204,657 in the genome sequence. By PCR, we found an increase in the copy number of sequences that result from reciprocal HR between these repeats (data not shown). As for the RA1 sequence, these differences were observed in the first generation of *osb1* mutants. Hence, the increase in HR products in *osb1*

mutants is not restricted to the *atp9* locus. It is likely that many HR products that are present in wild-type mitochondria at sublimon levels accumulate in *osb1*. The increased pool of HR-derived sequences could then trigger additional HR events.

### The Level of Recombination Products Increases with Successive Generations

The progeny of the five homozygous *osb1* plants described above were analyzed. The results shown in Supplemental Table 2B online demonstrate that the majority of the plants (138 of 194), like their parents, had higher levels of RA1 (stage I). However, 32 plants seemed to revert to wild type stage 0 RA1 levels. Interestingly, several (24) had a new pattern (stage II of stoichiometric shifting) characterized by a level of RA1 PCR product equal to or greater than the level of the PCR product corresponding to unrecombined DNA (plants 1 and 4 in Figure 8D). The segregation of shifted plants differed in the mutant lines: one-quarter of the progeny of homozygous plant 3 was affected, but none of the progeny of homozygous plant 9 showed shifting (see Supplemental Table 2B online). At generation T3, the shifted plants had no visible morphological phenotype.

### After Complete Stoichiometric Shifting, Reversion to the Wild Type Is No Longer Possible

We tested whether the molecular phenotype of shifted mtDNA can be reverted by reintroduction of the wild-type *OSB1* allele. Individual T3 *osb1-1* plants with stage I and stage II mtDNA were selfed and backcrossed with wild-type Col-0. The progeny were then tested for mtDNA shifting by competitive PCR, as described above. Fifteen T4 plants obtained by selfing a stage I T3 plant were analyzed. Nine maintained the same molecular profile, whereas three reverted to the wild-type stage 0 pattern and three others shifted to stage II (see Supplemental Table 2C online). The backcross with wild-type plants revealed that, at generation T3, the shifting process can be reverted: of the 15 F1 daughter plants tested, 13 reverted to stage 0 wild-type patterns and only 1 plant evolved to stage II. As expected, all of the plants analyzed from the reciprocal cross displayed a wild-type profile.

When the same experiment was done with a plant of the same generation (T3) with a stage II shifted pattern, no revertants were found among the 15 T4 plants analyzed (see Supplemental Table 2D online). From the backcross with wild-type Col-0, of 21 F1 daughter plants, 11 conserved the stage II shifted pattern, whereas the others reverted to the wild-type or stage I pattern (Figure 8E). However, the analysis of progeny of the reciprocal cross showed that pollen of the same plant was unable to induce the stage I or stage II pattern in the F1 hybrids obtained from crossing with wild-type plants (Figure 8F). These results establish the non-Mendelian inheritance of the shifted phenotype.

In subsequent generations, reversion from the shifted phenotype was no longer detectable. This was clearly illustrated by results obtained by selfing or backcrossing T4 *osb1* plants with shifted stage II patterns, severe morphological distortion, and reduced fertility. As shown in Supplemental Table 2D online, the T5 plants obtained by selfing had the shifted stage II pattern (except for one plant that had the shifted stage I pattern). Most of

these plants had distorted, variegated leaves. From the backcross with Col-0 (see Supplemental Table 2D online), the majority of the plants (15 of 18 analyzed) maintained shifted mtDNA. One plant had wild-type stage 0 mtDNA. However, further PCR analysis of this plant showed that its mtDNA had undergone numerous recombination events that could be responsible for its spectacular growth delay, leaf distortion, and early death. The reciprocal cross with Col-0 gave F1 plants with wild-type mtDNA, as expected from the maternal inheritance of the phenomenon.

Together, these results suggest that, above a threshold level of alternative structures obtained by infrequent HR, reversion to the original mtDNA structure is infrequent and independent of the *OSB1* allele.

### Shifting of the mtDNA from Stage I to Stage II Is Accompanied by the Loss of Reciprocal Recombination Products

As shown above (Figure 7), *osb1* mutants have clear morphological phenotypes, and recombination products expected from reciprocal HR (RA2 and RB2; Figures 7B and 7C) are not detectable by DNA gel blot analysis. Therefore, we tested for the presence of these molecules using conventional two-primer PCR. Primers P2 + P4 (Figure 7B) were used to specifically amplify reciprocal recombination product RA2. As clearly shown in Figure 9A, plants in the first stage of mtDNA shifting (stage I, as shown by the level of RA1 in the top panel) contained levels of RA2 (bottom panel) comparable to those of wild-type Col-0. The relative copy number of RA2 was even higher in a few T2, T4, and T5 *osb1* plants analyzed (data not shown). On the other hand, the RA2 product was absent, or detected very faintly, in all plants from subsequent generations with a shifted stage II pattern (Figure 9B), even after >35 PCR cycles. Similarly, the recombination product RB2 was not detectable in these plants after amplification using primers P8 + P9 (data not shown). Thus, the transition from partial and reversible mtDNA shifting to irreversible shifting correlates with the loss of reciprocal recombination products. Whether it is a cause or a consequence of the transition is discussed below.

## DISCUSSION

### The OSB Family Is a Novel Plant-Specific Protein Family

We have shown that the *Arabidopsis OSB* genes constitute a small family of organelle-targeted proteins. An OSB protein was initially identified in potato by ssDNA-affinity chromatography (Vermel et al., 2002). Putative OSB sequences were also detected in the *O. sativa* and *Z. mays* genomes, but no OSB sequences were found in nonplant genomes, suggesting that either the OSB proteins evolved to fulfill plant organelle-specific functions or that in nonplant species the same function is accomplished by other proteins.

We showed that GFP fusion proteins are targeted to mitochondria (At OSB1), to chloroplasts (At OSB2), or to both (At OSB3). From our results, we cannot exclude the possibility that a small proportion of At OSB1 and At OSB2 proteins could be dual-

targeted. However, it is unlikely that At OSB1 is dual-targeted, because a protein of the corresponding size was detected only in the mitochondrial protein extract. In addition, *osb1* mutants do not show the phenotypes associated with chloroplast deficiency observed in *msh1* plants, supporting the idea that At OSB1 is targeted only to mitochondria. Although the dual targeting of At OSB3 should be confirmed by an additional approach, it is interesting that dual targeting is also observed for other proteins involved in organellar gene expression, such as aminoacyl-tRNA synthetases, RNA polymerases, DNA gyrases, and DNA polymerases (Hedtke et al., 2000; Wall et al., 2004; Christensen et al., 2005; Duchene et al., 2005). The redundancy of specifically targeted and dual-targeted proteins could reflect a high degree of flexibility in the regulation of organellar gene expression by the nucleus.

### Evolution of OSB Proteins as ssDNA Binding Proteins

The OSB proteins are characterized by the presence of PDF motifs. Our results suggest that, at least for At OSB1, the PDF motif is required for the protein-DNA interaction. It is surprising that the intermediate region of At OSB1 (Figure 2), which is clearly related to the SSB protein family (30 to 50% similarity), is apparently not required for ssDNA binding. However, none of the residues, including Trp and Phe, required for high-affinity binding to ssDNA of *E. coli* SSB are conserved in the SSB-like domains of OSB proteins (Raghunathan et al., 1997). It is also possible that the protein/DNA gel blot analysis underestimates the influence of this domain if, for instance, the SSB domain is involved in protein oligomerization. Nevertheless, the PDF can bind ssDNA and thus constitutes a new ssDNA binding motif. We have shown that At OSB1 and At OSB2 have a preference for ssDNA, but we cannot exclude the possibility that other members of the family evolved to have different binding specificities. Multiple PDF motifs in different OSB proteins could lead to differences in their affinity for nucleic acids.

The phylogenetic origin of the PDF motif remains unknown. Apart from plant OSB proteins, this motif was found only in the RB38 protein of *Chlamydomonas*. RB38 binds the 5' untranslated region of the chloroplast *psbA* mRNA (Barnes et al., 2004) and is composed of a chloroplast-targeting peptide directly followed by four PDF motifs responsible for RNA binding. The apparent different substrate specificities of At OSB1 (ssDNA) and RB38 (RNA) could be attributable to differences in the positions of the predicted  $\alpha$ -helices and  $\beta$ -sheets within the PDF motifs, to nonconserved amino acids, and to sequences surrounding the motif. We can speculate that the PDF motif evolved primarily as an RNA binding domain and that OSB proteins originated in higher plants by the fusion of a bacterium-type SSB and a PDF-containing protein. In this context, the PDF domain could have evolved to preferentially bind ssDNA.

OSB proteins have the common structure of an SSB-like domain followed by a PDF motif(s). Therefore, it is possible that they have overlapping functions, which could explain the absence of visible phenotypes of At OSB2 and At OSB3 T-DNA insertion mutants. In the case of At OSB1, our results show that its deficiency results in mtDNA instability, leading to an unbalanced transmission of alternative mtDNA configurations.

### Accumulation of mtDNA Recombination Products in *osb1* Plants

In *osb1* mutant lines, changes in mtDNA profiles apparently result from reciprocal HR between small repeats, such as the RA and RB repeats, which are a few hundred nucleotides long. In wild-type plants, recombination mediated by these repeats seems infrequent, because the corresponding products usually are present at very low levels and can be detected only by PCR. The very strong HR activity, thought to be responsible for the multipartite structure of plant mitochondrial genomes, involves large repeats of several kilobases in length (Unsel et al., 1997).

The hot spot of HR that we followed as a marker of mtDNA substoichiometric shifting concerns the RA repeat, consisting of *atp9* gene sequences. RA-mediated recombination is a good marker for recombination deregulation in *Arabidopsis* Col-0: the RA1 recombination product is amplified in *osb1* and *msh1* mutants and also in cell suspension cultures (Forner et al., 2005). In other ecotypes, such as *Ler*, the RA1 recombination product is the predominant mtDNA configuration. The repeats surrounding the *atp9* gene are prone to recombination, which could be attributable to the very high transcription activity associated with the *atp9* gene locus. Local melting of the DNA double strand could favor invasion by homologous sequences. However, the RA1 recombination product is not linked to the morphological phenotypes of leaf variegation and distortion, because the phenotype severity is not correlated with the ratio between wild-type and RA recombination products. Systematic analysis by PCR of mtDNA sequences surrounding expressed genes detected changes affecting several mtDNA loci in *osb1* mutants. These changes probably result from additional recombination events and suggest that inactivation of *OSB1* affects the maintenance of the mtDNA genome as a whole. Multiple mtDNA configurations probably segregate from the same original mutant line, explaining the variegated phenotype of the most affected *osb1* plants. A comprehensive analysis of the mtDNA substoichiometric shifting process induced by the *osb1* or *msh1* mutations would require detailed mapping and sequencing of the different mtDNA genomes for several lines segregating from the same mutant plant.

### mtDNA Shifting in *osb1* Lines Is a Two-Step Process

Repression of HR is likely to be essential in ensuring mtDNA stability. The loss of *OSB1* appears to affect nuclear control, which represses HR of the mtDNA. Although our analysis was restricted to the recombination events involving the *atp9* locus, several conclusions can be made concerning the process of mtDNA shifting. (1) In the first mutant generation, there is already a patent increase in the accumulation of HR products. (2) In the next generation, individual lines segregate in which certain rearranged mtDNA configurations obtained by HR are preferentially replicated, compared with the parental sequences (what we called stage II). This process is accompanied by an almost complete disappearance of the reciprocal recombination product, suggesting that this product is no longer produced by reversible reciprocal HR. (3) In the first stages of mtDNA shifting (stage I), the process is reversible if *OSB1* activity is restored. This is probably attributable to the low copy number of HR products,

which are segregated out in the subsequent generations, if their continuous production by HR is repressed. However, after stage II, reversion to wild-type mtDNA no longer occurs, even in the presence of the *osb1* allele. It is possible that the recombination product is now the major mtDNA sequence and can no longer be segregated out during gametogenesis.

The shift from stage I to stage II could result either from the stochastic segregation of the parental mtDNA molecules and HR products or from additional recombination events, leading to the preferential replication of the mtDNA molecules containing recombined sequences. The accumulation of HR products could contribute to the pool of sequences that prime the asymmetric replication of mtDNA chimeras by strand invasion. It is reasonable to assume that, after the stage II mtDNA shift, the accumulation of the predominant recombination product no longer depends on continuous production by HR. This could explain why reciprocal recombination products are no longer detected: if they are not replication-competent, their production in somatic tissues by HR requires the presence of the wild-type parental sequences. Because the wild-type sequences are minor in the shifted mtDNA, they should preferentially undergo HR with the new predominant mtDNA sequence, a process that can no longer generate the reciprocal recombination products.

In cybrid rapeseed (*Brassica napus*) plants, changes in mtDNA stoichiometry are associated with the selective amplification of low copy number molecules produced by recombination (Bellaoui et al., 1998). As for *osb1* mutants, this process involves the amplification of only one of the recombination products, whereas the reciprocal sequence remains at low levels. Therefore, it seems that different causes (cybrid production and the *osb1* mutation) can result in a similar mechanism of mtDNA substoichiometric shifting.

### A Possible Role of *OSB1* as a Repressor of HR Required for mtDNA Genome Stability

ssDNA binding proteins play a central role in HR, both as enhancers (by holding DNA strands open and melting ssDNA secondary structures that block recombinases) and as suppressors (by outcompeting the initial binding of recombinases to ssDNA). Recombination mediators, such as bacterial RecO and eukaryotic Rad52, overcome ssDNA binding protein inhibition by promoting the assembly of recombinases on ssDNA (Gasior et al., 2001). Bacterium-type SSB proteins also exist in plant mitochondria. Like *E. coli* SSB, plant mitochondria SSB promotes the RecA-dependent strand invasion of dsDNA by homologous ssDNA (Edmondson et al., 2005). At *OSB1* is an ssDNA binding protein with no apparent sequence specificity, but its function is apparently not redundant with that of bacterium-type SSB, which is essential for plant viability (our unpublished data). At *OSB1* could function as a suppressor of HR, a function that could explain its role in the mechanism of copy number suppression of subgenomes generated by HR during mtDNA replication.

Such a putative function is in agreement with the preferential expression of *OSB1* in gametophytic tissues, postulated to be the locus where the stoichiometry of the alternative genomes is determined. Models of how mtDNA is maintained with minimal

heteroplasmy, both in animals and in plants, suggest that mitochondria pass through a stringent genetic bottleneck in transmitting tissues (Shoubridge, 2000; Arrieta-Montiel et al., 2001). Thus, genes that directly affect the structure of the transmitted mtDNA are expected to be preferentially expressed in those tissues. According to our model, to replicate a functional, high copy number mitochondrial genome, the role of *At OSB1* in meristematic and gametophytic cells is to reduce the mitochondrial genome complexity resulting from the production of aberrant subgenomes generated by infrequent recombination events between small repeats. In somatic tissues, recombination control is relaxed, and the ratio of normal to subgenomes is dictated by their inherent replication efficiency. In the absence of *At OSB1*, the predominance of these subgenomes during mtDNA replication in transmitting cells results in the preferential replication of selfish subgenomes and in the nonreversible imbalance observed in shifted plants.

We have identified *At OSB1* as a component of the system regulating mtDNA stoichiometry. Our results suggest that *At OSB1* represses the production of recombination products considered illegitimate and thus takes part in the nuclear control that prevents mtDNA instability. Determining the roles of proteins such as OSB1 in plant development will shed light on the mechanism of substoichiometric shifting in the mitochondrial genome and its contribution to plant evolution.

## METHODS

### Identification, Purification, and Cloning of *OSB* Genes

For protein/DNA gel blot analysis of soluble mitochondrial proteins, potato (*Solanum tuberosum*) mitochondria were purified as described (Neuburger et al., 1982) and lysed by freeze-thaw (three cycles). Soluble proteins (3 mg) were recovered by centrifugation, adjusted to 250 mM NaCl, and run through a 1.5-mL DEAE anion-exchange column (DE52; Whatmann) to remove contaminating nucleic acids. The eluted proteins were fractionated by SDS-PAGE (8  $\mu$ g/lane) and probed for ssDNA binding with an oligonucleotide containing a 25-nucleotide random sequence (5'-GGCCACTAGTCGGATCCC[N]<sub>25</sub>GGGTCGACGAATTCGACG-3'). The *St OSB* protein was purified by ssDNA-affinity purification as described previously (Vermel et al., 2002), starting from 200 mg of potato mitochondrial protein. The purified protein was excised from the gel, and tryptic peptides were N-terminal sequenced on an Applied Biosystems 473A apparatus. From two peptide sequences (Pep1, 5'-IEIYDEADVSSWPKPEI-3'; Pep2, 5'-NFLDENDNQHSY-3'), two sets of degenerated primers were derived that allowed RT-PCR amplification of the *St OSB1* cDNA from total potato germ RNA. The sequence of the 450-bp amplified fragment was used to derive new primers and amplify the complete *St OSB1* cDNA by 5' and 3' rapid amplification of cDNA ends (5' RACE system; Invitrogen).

*At OSB* cDNAs were amplified from total *Arabidopsis thaliana* cDNA primed with oligo(dT) using primers complementary to the 5' and 3' untranslated region sequences of genes *At1g47720*, *At4g20010*, and *At5g44785*. For *in vivo* intracellular localization, the sequences encoding the first 43, 85, and 68 amino acids of *At OSB1*, *At OSB2*, and *At OSB3*, respectively, were cloned in plasmid pCK-GFP3A to express protein-eGFP fusions under the control of a 35S promoter (Menand et al., 1998). The resulting plasmids were transfected in *Nicotiana benthamiana* leaves by bombardment (Sanford et al., 1993), and images were obtained at 24 h after transfection with a Zeiss LSM510 confocal microscope. The pCK-mRFP plasmid (Vermel et al., 2002) was cotransfected as an internal mitochondrial marker. Chloroplasts were visualized by the natural fluo-

rescence of chlorophyll. Organelle-targeting predictions were determined with the programs Predotar, TargetP, and iPSort (genoplante-info.infobiogen.fr/tools/predotar/; www.cbs.dtu.dk/services/TargetP/; and biocaml.org/ipsort/iPSORT/, respectively).

### Plants, DNA, and RNA

*Arabidopsis* plants, ecotype Col-0, were transformed by the floral dip method (Clough and Bent, 1998). T1 and T2 plants were sown on agar plates containing Murashige and Skoog salts, 0.5% sucrose, and 50  $\mu$ g/mL kanamycin. After 2 to 3 weeks, kanamycin-resistant plants were transferred to soil and grown in the greenhouse. T-DNA insertion mutant lines were obtained from the GABI-Kat and SALK collections. Insertion sites were determined by PCR using gene- and T-DNA-specific primers and mapped to positions 17,565,117 to 17,565,126 (in *Arabidopsis* chromosome 1) with a 9-nucleotide deletion for mutant GK-93H12 and to positions 175,645,593 to 175,645,568 with a 25-nucleotide deletion for mutant SALK-086929, where insertion is composed of at least two T-DNA molecules head to head. Mutant seeds were genotyped by PCR, and homozygote plants were self-propagated. Plants were grown at 23°C under a 16-h-light/8-h-dark photoperiod. Genomic DNA was extracted with DNAzol (Invitrogen) and total RNA was extracted with TRIzol (Invitrogen) as described (Chomczynski and Sacchi, 1987).

### Protein Expression

The *At OSB1* cDNA sequence encoding the predicted mature protein (between amino acids 38 and 261) was cloned between *Hind*III and *Bam*HI sites in expression vector pRSETc (Invitrogen) (construct T in Figure 3). Deletion mutant constructs 1 and 4 (Figure 3D) were obtained by PCR from construct T, using primers taken at the borders of the deleted region (residues 38 to 89 for construct 1 and residues 38 to 166 for construct 4) followed by clone religation. Similar deletion constructs 2 and 3 were obtained starting from construct 1 and resulting in the removal of residues 171 to 262 (construct 2) and residues 205 to 261 (construct 3). Recombinant proteins were expressed in *Escherichia coli* strain BL21(DE3) (Novagen). The cells were grown at 37°C in Luria-Bertani medium containing carbenicillin (50  $\mu$ g/mL) and 0.2% glucose up to  $A_{600}$  of 0.6, and protein expression was then induced with 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside. After 2 h at 37°C, cells were pelleted and vortexed in the presence of buffer A (10 mM Tris-HCl, pH 8.0, 100 mM NaH<sub>2</sub>PO<sub>4</sub>, and 8 M urea). Debris were removed by centrifugation, and the soluble fraction was incubated with nickel-nitrilotriacetic acid agarose beads (Qiagen) for 2 h at room temperature. After washing with buffer A containing 50 mM imidazole, the expressed proteins were eluted with 150 mM imidazole.

The *At OSB2* cDNA sequence encoding the predicted mature protein (between amino acids 69 and 372) was cloned in the pBin+ vector (van Engelen et al., 1995) fused to a C-terminal tag sequence comprising the FLAG peptide and the calmodulin binding peptide under the control of the cauliflower mosaic virus 35S promoter and terminator sequences, as described (Perrin et al., 2004). These constructs were used to express the protein in *N. benthamiana* 3-week-old leaves by agroinfiltration, as described (Voinnet et al., 2003). After 4 d, infiltrated leaves were ground in liquid nitrogen and the tagged protein was purified by tandem affinity purification, as described previously (Perrin et al., 2004).

### Antibody Production

Purified *At OSB1* (amino acids 38 to 261) produced in *E. coli* using vector pRSETc was transferred to nitrocellulose membranes after SDS-PAGE. The protein was stained with Ponceau red, cut, dried for 12 h under vacuum, and dissolved in DMSO. A rabbit antiserum was prepared, and the antibodies were purified against the protein antigen immobilized on

CNBr-activated Sepharose. For protein gel blot analysis, the purified antibodies were used at a concentration of 1:5000.

### DNA Binding Assays

Electrophoretic mobility shift assays were performed as described in Promega Technical Bulletin 110. Briefly, purified protein (1 to 10 ng) and 5' end radiolabeled oligonucleotide (0.1 pmol) purified on polyacrylamide gels were incubated for 20 min on ice in 10% glycerol, 30 mM Tris-HCl, pH 8.0, 3 mM  $\beta$ -mercaptoethanol, 10 mM MgCl<sub>2</sub>, 100  $\mu$ g/mL BSA, 50 mM NaCl, the appropriate concentration of competitor, and a mixture of protease inhibitors (Complete-EDTA; Roche Applied Science). DNA-protein complexes were resolved on 8% polyacrylamide gels in 1 $\times$  Tris-Borate-EDTA buffer at 4°C. The gels were then vacuum-dried and revealed using a BAS 1000 phosphor imager (Fujix). Quantification of the signals was done with MacBas version 2.2 software (Fuji Photo Film). Routinely, the (ACTG)<sub>8</sub> 32-mer oligonucleotide was used as an ssDNA probe. The dsDNA probe was prepared by annealing with the complementary oligonucleotide, and the RNA probe was prepared by *in vitro* transcription using a primer of the same sequence containing a 5' T7 promoter sequence. All probes were gel-purified before use. Protein/DNA gel blot analysis was done as described by Ghosh et al. (1994) for protein/RNA gel blot analysis. Briefly, purified proteins were fractionated by SDS-PAGE, transferred to polyvinylidene difluoride membranes (Immobilon P; Millipore), and renatured by four 30-min incubations at 4°C in 100 mM Tris-HCl, pH 7.5, 0.1% Nonidet P-40, and 100 mM NaCl. The membrane was blocked for 10 min at room temperature in binding buffer (10 mM Tris-HCl, pH 7.5, 5 mM Mg-acetate, 2 mM DTT, 50 mM NaCl, and 0.01% Triton X-100) containing 5% BSA, and then incubated for 10 min at room temperature in binding buffer added with 100  $\mu$ g of poly(dCdI). Radiolabeled probe was added, and after 2 h of incubation, membranes were washed four times for 10 min each in binding buffer before autoradiography.

### Promoter-GUS Fusion Analysis

The 5' intergenic region of *At OSB1* (1074 nucleotides between the stop codon of the upstream *At1g47730* gene and the initiation codon of *At1g47720*) was cloned in the binary vector pBI101.2 (Clontech) upstream of the GUS gene. Transgenic *Arabidopsis* plants were produced, and tissues from promoter-GUS fusion plants issued from six independent lines were stained with 5-bromo-4-chloro-3-indolyl- $\beta$ -glucuronidic acid (Biosynth) as described (Jefferson et al., 1987). For observation of thin sections, tissues were fixed in 3.7% formaldehyde, 50% ethanol, and 5% acetic acid, stained using eosine (0.002% final concentration), and embedded in Paraplast wax using conventional techniques.

### Transmission Electron Microscopy

Leaf tissue samples were fixed overnight in 3% glutaraldehyde, then treated for 2 h with 10% (w/v) picric acid and for 2 h with 2% uranyl acetate, and stained with 0.1% (v/v) osmium tetroxide in 150 mM phosphate buffer, pH 7.2. Samples were dehydrated through an ethanol series and infiltrated with EPON812 medium-grade resin (Polysciences). Polymerization was done for 48 h at 60°C. Ultrathin sections (90  $\mu$ m) were cut using an Ultracut E microtome (Reichert) and collected on grids coated with formvar (Electron Microscopy Sciences). Samples were visualized with a Hitachi H-600 electron microscope operating at 75 kV. The mitochondrial surface was measured with ImageJ software (<http://rsb.info.nih.gov/ij/>).

### PCR Screening of mtDNA Structural Modifications

The DNA integrity of the *Arabidopsis* mitochondrial genome was checked by PCR using the primer collection described by C. Andrés (unpublished

data; see Supplemental Table 1 online). PCR was performed on 384-well plates in a final volume of 10  $\mu$ L using *Taq* DNA polymerase (M0267S; New England Biolabs). PCR conditions were as described in the product manual except that primer concentrations were 0.1  $\mu$ M and that Cresol Red (0.2 mg/mL) and sucrose (60 mg/mL) were added. Genomic DNA concentration was 25 to 75 ng per reaction. The Thermocycler (Bio-Rad) was set with the following program: 5 min at 94°C, then 35 cycles of 20 s at 94°C, 20 s at 52°C, and 60 s at 72°C, and a final step of 5 min at 72°C. PCR products were analyzed on 2.0% agarose gels and visualized by ethidium bromide staining under UV light. Gel images of wild-type and mutant samples were processed using Genetools software (Syngene).

### Accession Numbers

Sequence data for the genes mentioned in this article can be found in the GenBank/EMBL data libraries under the following accession numbers: AY942642 (*St OSB1*), At1g47720 (*At OSB1*), At4g20010 (*At OSB2*), At5g44785 (*At OSB3*), At1g31010 (*At OSB4*), At3g24320 (*MSH1*), At4g11060 (mitochondrial SSB), and NC\_001284 (*Arabidopsis* ecotype C24 mitochondrion complete genome).

### Supplemental Data

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

**Supplemental Methods.** Analysis of *At OSB2* Binding Activities.

**Supplemental Figure 1.** ssDNA Binding Activity of *At OSB2*.

**Supplemental Figure 2.** PCR Analysis of Col-0 and *osb1* Mutant mtDNA.

**Supplemental Table 1.** Mitochondrial Genomic Positions of PCR Products.

**Supplemental Table 2.** Analysis of Shifting in the Progeny of *osb1* Plants.

**Supplemental Table 3.** Oligonucleotides Used.

### ACKNOWLEDGMENTS

We thank Philippe Hamann, Malek Alioua, and Leo Baettig for technical assistance and the gardeners of the Institut de Biologie Moléculaire des Plantes for excellent plant care. We are grateful to Dominique Gagliardi and Philippe Giege for helpful discussions and critical reading of the manuscript. We also thank the ABRC, GABI-Kat, and the Salk Institute Genomic Analysis Laboratory for providing the T-DNA insertion mutant lines. V.Z. was supported by a PhD fellowship of the French government. This work was supported by the Centre National de la Recherche Scientifique and by the French plant genomic program Génoplante.

Received February 18, 2006; revised November 4, 2006; accepted November 20, 2006; published December 22, 2006.

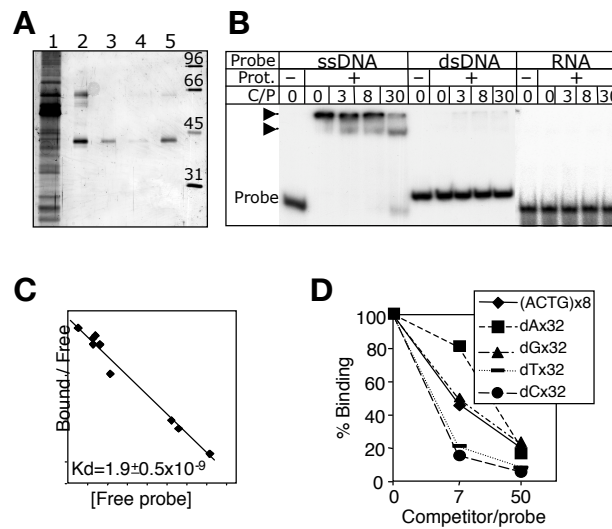
### REFERENCES

- Abdelnoor, R.V., Yule, R., Elo, A., Christensen, A.C., Meyer-Gauen, G., and Mackenzie, S.A. (2003). Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc. Natl. Acad. Sci. USA* **100**, 5968–5973.
- Adams, K.L., and Daley, D. (2004). Plant mitochondrial genome evolution and gene transfer to the nucleus. In *Plant Mitochondria: From Genome*

- to Function, D.A. Day, A.H. Millar, and J. Whelan, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 107–120.
- Arrieta-Montiel, M., Lyznik, A., Woloszynska, M., Janska, H., Tohme, J., and Mackenzie, S.** (2001). Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* **158**, 851–864.
- Backert, S., and Börner, T.** (2000). Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.). *Curr. Genet.* **37**, 304–314.
- Backert, S., Nielsen, B.L., and Börner, T.** (1997). The mystery of the rings: Structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci.* **2**, 477–483.
- Barnes, D., Cohen, A., Bruick, R.K., Kantardjieff, K., Fowler, S., Efuet, E., and Mayfield, S.P.** (2004). Identification and characterization of a novel RNA binding protein that associates with the 5'-untranslated region of the chloroplast psbA mRNA. *Biochemistry* **43**, 8541–8550.
- Barr, C.M., Neiman, M., and Taylor, D.R.** (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol.* **168**, 39–50.
- Bellaoui, M., Martin-Canadell, A., Pelletier, G., and Budar, F.** (1998). Low-copy-number molecules are produced by recombination, actively maintained and can be amplified in the mitochondrial genome of *Brassicaceae*: Relationship to reversion of the male sterile phenotype in some cybrids. *Mol. Gen. Genet.* **257**, 177–185.
- Chomczynski, P., and Sacchi, N.** (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.
- Christensen, A.C., Lyznik, A., Mohammed, S., Elowsky, C.G., Elo, A., Yule, R., and Mackenzie, S.A.** (2005). Dual-domain, dual-targeting organellar protein presequences in *Arabidopsis* can use non-AUG start codons. *Plant Cell* **17**, 2805–2816.
- Clifton, S.W., et al.** (2004). Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* **136**, 3486–3503.
- 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.
- Duchene, A.M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N.M., Zaepfel, M., Marechal-Drouard, L., and Small, I.D.** (2005). Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **102**, 16484–16489.
- Edmondson, A.C., Song, D., Alvarez, L.A., Wall, M.K., Almond, D., McClellan, D.A., Maxwell, A., and Nielsen, B.L.** (2005). Characterization of a mitochondrially targeted single-stranded DNA-binding protein in *Arabidopsis thaliana*. *Mol. Genet. Genomics* **273**, 115–122.
- Förner, J., Weber, B., Wietholter, C., Meyer, R.C., and Binder, S.** (2005). Distant sequences determine 5' end formation of cox3 transcripts in *Arabidopsis thaliana* ecotype C24. *Nucleic Acids Res.* **33**, 4673–4682.
- Gasior, S.L., Olivares, H., Ear, U., Hari, D.M., Weichselbaum, R., and Bishop, D.K.** (2001). Assembly of RecA-like recombinases: Distinct roles for mediator proteins in mitosis and meiosis. *Proc. Natl. Acad. Sci. USA* **98**, 8411–8418.
- Ghosh, A., Ghosh, T., Ghosh, S., Das, S., and Adhya, S.** (1994). Interaction of small ribosomal and transfer RNAs with a protein from *Leishmania donovani*. *Nucleic Acids Res.* **22**, 1663–1669.
- Hedtke, B., Börner, T., and Weihe, A.** (2000). One RNA polymerase serving two genomes. *EMBO Rep.* **1**, 435–440.
- Janska, H., Sarria, R., Woloszynska, M., Arrieta-Montiel, M., and Mackenzie, S.A.** (1998). Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell* **10**, 1163–1180.
- Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W.** (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.
- Kanazawa, A., Tsutsumi, N., and Hirai, A.** (1994). Reversible changes in the composition of the population of mtDNAs during dedifferentiation and regeneration in tobacco. *Genetics* **138**, 865–870.
- Kmiec, B., Woloszynska, M., and Janska, H.** (2006). Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr. Genet.* **50**, 149–159.
- Kuzmin, E.V., Duvick, D.N., and Newton, K.J.** (2005). A mitochondrial mutator system in maize. *Plant Physiol.* **137**, 779–789.
- Laloi, C., Rayapuram, N., Chartier, Y., Grienenberger, J.M., Bonnard, G., and Meyer, Y.** (2001). Identification and characterization of a mitochondrial thioredoxin system in plants. *Proc. Natl. Acad. Sci. USA* **98**, 14144–14149.
- Martinez-Zapater, J.M., Gil, P., Capel, J., and Somerville, C.R.** (1992). Mutations at the *Arabidopsis* CHM locus promote rearrangements of the mitochondrial genome. *Plant Cell* **4**, 889–899.
- Menand, B., Marechal-Drouard, L., Sakamoto, W., Dietrich, A., and Wintz, H.** (1998). A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl-tRNA synthetase in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **95**, 11014–11019.
- Mookerjee, S.A., and Sia, E.A.** (2006). Overlapping contributions of Msh1p and putative recombination proteins Cce1p, Din7p, and Mhr1p in large-scale recombination and genome sorting events in the mitochondrial genome of *Saccharomyces cerevisiae*. *Mutat. Res.* **595**, 91–106.
- Neuburger, M., Journet, E.P., Bligny, R., Carde, J.P., and Douce, R.** (1982). Purification of plant mitochondria by isopycnic centrifugation in density gradients of Percoll. *Arch. Biochem. Biophys.* **217**, 312–323.
- Oldenburg, D.J., and Bendich, A.J.** (1996). Size and structure of replicating mitochondrial DNA in cultured tobacco cells. *Plant Cell* **8**, 447–461.
- Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M., and Gagliardi, D.** (2004). Two exoribonucleases act sequentially to process mature 3' ends of atp9 mRNAs in *Arabidopsis* mitochondria. *J. Biol. Chem.* **279**, 25440–25446.
- Pinto, A.V., Mathieu, A., Marsin, S., Veaute, X., Ielpi, L., Labigne, A., and Radicella, J.P.** (2005). Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. *Mol. Cell* **17**, 113–120.
- Ragunathan, S., Ricard, C.S., Lohman, T.M., and Waksman, G.** (1997). Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-Å resolution. *Proc. Natl. Acad. Sci. USA* **94**, 6652–6657.
- Sakamoto, W., Kondo, H., Murata, M., and Motoyoshi, F.** (1996). Altered mitochondrial gene expression in a maternal distorted leaf mutant of *Arabidopsis* induced by chloroplast mutator. *Plant Cell* **8**, 1377–1390.
- Sanford, J.C., Smith, F.D., and Russell, J.A.** (1993). Optimizing the biolistic process for different biological applications. *Methods Enzymol.* **217**, 483–509.
- Sheahan, M.B., McCurdy, D.W., and Rose, R.J.** (2005). Mitochondria as a connected population: Ensuring continuity of the mitochondrial genome during plant cell dedifferentiation through massive mitochondrial fusion. *Plant J.* **44**, 744–755.
- Shoubridge, E.A.** (2000). Mitochondrial DNA segregation in the developing embryo. *Hum. Reprod.* **15** (suppl. 2): 229–234.
- Small, I.D., Suffolk, R., and Leaver, C.J.** (1989). Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* **58**, 69–76.



- Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M.** (1990). Early flower development in *Arabidopsis*. *Plant Cell* **2**, 755–767.
- Unsel, M., Marienfeld, J.R., Brandt, P., and Brennicke, A.** (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15**, 57–61.
- van Engelen, F.A., Molthoff, J.W., Conner, A.J., Nap, J.P., Pereira, A., and Stiekema, W.J.** (1995). pBINPLUS: An improved plant transformation vector based on pBIN19. *Transgenic Res.* **4**, 288–290.
- Vermel, M., Guermann, B., Delage, L., Grienberger, J.M., Marechal-Drouard, L., and Gualberto, J.M.** (2002). A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc. Natl. Acad. Sci. USA* **99**, 5866–5871.
- Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D.** (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**, 949–956.
- Wall, M.K., Mitchenall, L.A., and Maxwell, A.** (2004). *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc. Natl. Acad. Sci. USA* **101**, 7821–7826.



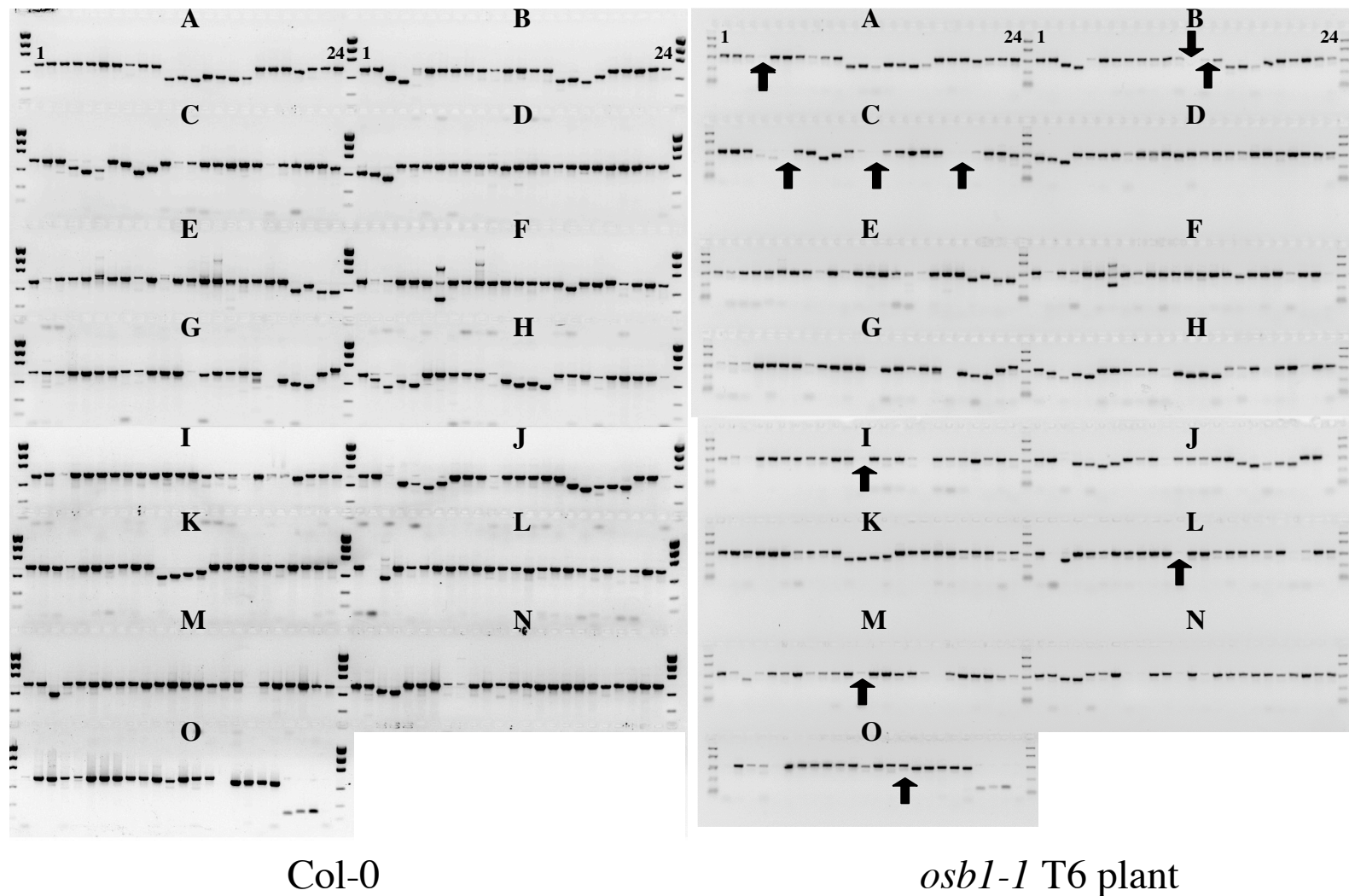
**Supplemental Figure 1. ssDNA-binding activity of *At*-OSB2.**

(A) Analysis by SDS-PAGE and silver staining of the purification of *At*-OSB2 expressed in *N. benthamiana* leaves. The protein was expressed fused to a N-terminal tag composed of the FLAG and Calmodulin Binding Peptides. 1-Total soluble fraction (S20), 2-Protein eluted from Calmodulin-Sepharose, 3-Flow-through from FLAG-antibody column, 4-Washing of FLAG-antibody column, 5-Protein eluted from FLAG-antibody column.

(B) Constant quantities of *At*-OSB2 protein and probes (<sup>32</sup>P-labeled ssDNA, dsDNA or RNA) were incubated with increasing quantities of cold ssDNA competitor and analyzed by EMSA. The nature of the probe (ssDNA, dsDNA or RNA), the presence (+) or absence (-) of *At*-OSB2 protein and the competitor to probe molar ratio (C/P) are indicated. Arrows indicate *At*-OSB2/ssDNA complexes.

(C) Scatchard blot results used to deduce the dissociation constant of *At*-OSB2 for ssDNA ((ACTG)x8 probe). The concentration of active protein used in the experiments was calculated from saturation binding curves.

(D) Affinity of *At*-OSB2 for the different homodeoxyribonucleotides used as competitors in EMSA experiments. Relative binding ratios were determined from phosphorimager quantification results.



**Supplemental Figure 2.** PCR analysis of Col-0 and *osb1* mutant mtDNA. Each PCR amplification was performed with specific primers. Supplemental Table 1 gives the genomic coordinates on the C24 mtDNA genome of the forward primer, as well as the size of each amplification product. The locations of mt genes are indicated in the Table. The products of the amplification experiments were analyzed by agarose gel electrophoresis. The amplifications are presented as 15 groups (A to O) of 24 experiments. Five different *osb1* plants (four *osb1-1*, one *osb1-2*) with clear morphological phenotypes of variegation and distortion were analyzed and the PCR profiles compared to the one of wild type Col-0. Differences found for one *osb1-1* plant, taken as example, are shown by arrows in the Figure and in grey background in Supplemental Table 1.

## VI.3 Annexe à l'analyse des génomes mitochondriaux

**Tableau 15 : Analyse des génomes mitochondriaux complets de 12 écotypes d'*Arabidopsis thaliana***

Ce tableau présente les résultats d'analyse par PCR de 12 écotypes d'*Arabidopsis thaliana*, les écotypes dont les noms sont en gras ont été analysés au moins deux fois. En vert sont indiqués les couples d'amorces dont au moins un se trouve dans une phase codante, en rouge les cas de non amplification de produit de PCR, en gris sont indiqués les cas où un problème technique est à l'origine de la non amplification. Les indications de gènes représentent les phases ouvertes de lecture ou les ARNm (5' et 3' non traduit compris), elles ne correspondent donc pas exactement aux amorces marquées en vert.

### Ecotypes

gènes	coord.	amorces	Jea	Pi-0	Oy-0	Shahdara	Cvi-0	Kondara	LER	Col-0	Ws	Bur-0	Ge-0	Ri-0
	A01	1	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A02	2	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
<b>nad5 ex 4&amp;5</b>	A03	3	non	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A04	4	non	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A05	5	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A06	6	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A07	7	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A08	8	non	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A09	9	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
<b>nad9</b>	A10	10	oui	oui	non	nd	oui	nd	oui	oui	oui	non	oui	non
	A11	11	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A12	12	non	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A13	13	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	non
<b>rpl16</b>	A14	14	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A15	15	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A16	16	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A17	17	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A18	18	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A19	19	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A20	20	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
<b>rps3</b>	A21	21	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A22	22	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A23	23	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	A24	24	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B01	25	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B02	26	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B03	27	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
<b>ccmB</b>	B04	28	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	non	oui
	B05	29	oui	non	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B06	30	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	non	oui
	B07	31	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B08	32	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	non	oui
	B09	33	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B10	34	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
<b>cox2</b>	B11	35	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	B12	36	non	oui	oui	nd	oui	nd	oui	oui	oui	oui	non	oui
	B13	37	oui	non	non	nd	oui	nd	oui	oui	oui	non	oui	non
	B14	38	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	non	oui
	B15	39	oui	non	non	nd	oui	nd	oui	oui	oui	oui	oui	non
	B16	40	oui	oui	oui	nd	oui	nd	oui	oui	oui	non	oui	oui
	B17	41	oui	oui	oui	non	oui	nd	oui	oui	oui	oui	oui	oui
	B18	42	oui	oui	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui
	B19	43	oui	non	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui
	B20	44	oui	oui	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui
	B21	45	oui	oui	oui	non	oui	nd	oui	oui	oui	oui	oui	oui
	B22	46	oui	oui	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui
B23	47	oui	oui	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	
B24	48	oui	non	oui	oui	oui	nd	oui	oui	oui	non	oui	oui	
<b>ccmFc</b>	C01	49	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	C02	50	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	C03	51	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	C04	52	oui	oui	oui	nd	non	nd	oui	oui	oui	oui	oui	oui
	C05	53	oui	oui	oui	nd	oui	nd	oui	oui	oui	oui	oui	oui
	C06	54	non	non	non	nd	non	nd	oui	oui	oui	oui	oui	non
	C07	55	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	C08	56	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	C09	57	oui	oui	oui	nd	oui	oui	non	oui	oui	oui	oui	oui

Ecotypes

gènes	coord.	amorces	Jea	Pi-0	Oy-0	Shahdara	Cvi-0	Kondara	LER	Col-0	Ws	Bur-0	Ge-0	Ri-0
	C10	58	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	C11	59	oui	oui	oui	nd	oui	non	oui	oui	oui	oui	oui	oui
	C12	60	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui
	C13	61	non	non	non	non	oui	non	oui	oui	oui	non	oui	non
	C14	62	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	C15	63	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	C16	64	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	C17	65	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	C18	66	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
<b>rpl5</b>	C19	67	non	oui	oui	oui	oui	non	oui	oui	oui	non	oui	non
	C20	68	non	non	oui	oui	oui	non	oui	oui	oui	oui	oui	non
	C21	69	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	non
	C22	70	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	C23	71	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	C24	72	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>cob</b>	D01	73	oui	oui	oui	nd	oui	non	oui	oui	oui	oui	oui	oui
	D02	74	oui	oui	oui	nd	oui	non	oui	oui	oui	oui	oui	oui
	D03	75	oui	oui	oui	nd	oui	non	oui	oui	oui	oui	oui	oui
	D04	76	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	D05	77	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	D06	78	oui	oui	oui	nd	oui	non	oui	oui	oui	oui	oui	oui
	D07	79	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	D08	80	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	D09	81	oui	oui	oui	nd	oui	oui	oui	oui	oui	oui	oui	oui
	D10	82	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D11	83	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>nad6</b>	D12	84	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D13	85	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	D14	86	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D15	87	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D16	88	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D17	89	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D18	90	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>nad2 ex 1&amp;2</b>	D19	91	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D20	92	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D21	93	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D22	94	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>rpl4</b>	D23	95	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	D24	96	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E01	97	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	E02	98	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	E03	99	oui	oui	oui	non	oui	oui	oui	oui	non	oui	oui	oui
	E04	100	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	E05	101	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	E06	102	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
	E07	103	oui	oui	oui	non	oui	non	oui	oui	oui	oui	oui	oui
	E08	104	oui	oui	oui	non	oui	non	oui	oui	oui	non	oui	oui
	E09	105	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E10	106	oui	oui	oui	oui	oui	non	oui	oui	oui	non	oui	oui
<b>atp6</b>	E11	107	oui	oui	oui	oui	oui	non	oui	oui	non	oui	oui	oui
	E12	108	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E13	109	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E14	110	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E15	111	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E16	112	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E17	113	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E18	114	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E19	115	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E20	116	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E21	117	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>atp8</b>	E22	118	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E23	119	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	E24	120	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	F01	121	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>nad7</b>	F02	122	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	non
	F03	123	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	F04	124	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	F05	125	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	F06	126	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	F07	127	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui

Ecotypes

gènes		coord.	amorces	Jea	Pi-0	Oy-0	Shahdara	Cvi-0	Kondara	LER	Col-0	Ws	Bur-0	Ge-0	Ri-0		
matR	nad5 ex1&2	F08	128	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
		F09	129	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F10	130	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F11	131	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F12	132	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F13	133	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F14	134	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F15	135	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F16	136	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		F17	137	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	F18	138	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	F19	139	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	F20	140	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	F21	141	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	F22	142	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	F23	143	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	
	F24	144	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	G01	145	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	G02	146	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	G03	147	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
G04	148	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G05	149	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G06	150	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G07	151	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G08	152	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui		
G09	153	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui		
G10	154	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G11	155	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G12	156	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G13	157	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non		
G14	158	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G15	159	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G16	160	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G17	161	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G18	162	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non		
G19	163	oui	oui	oui	oui	oui	oui	oui	oui	non	non	oui	oui	oui	oui		
G20	164	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G21	165	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G22	166	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
G23	167	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui		
G24	168	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H01	169	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H02	170	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H03	171	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H04	172	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H05	173	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H06	174	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H07	175	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H08	176	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H09	177	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H10	178	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
H11	179	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
nad4	H12	180	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H13	181	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H14	182	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H15	183	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H16	184	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H17	185	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H18	186	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H19	187	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	
	H20	188	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	
	H21	189	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	H22	190	oui	oui	non	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	
	H23	191	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	
	H24	192	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	
	I01	193	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	I02	194	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	
	I03	195	oui	oui	non	non	non	oui	non	oui	oui	oui	oui	oui	oui	oui	
	I04	196	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	I05	197	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	I06	198	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	I07	199	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
I08	200	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		

Ecotypes

gènes		coord.	amorces	Jea	Pi-0	Oy-0	Shahdara	Cvi-0	Kondara	LER	Col-0	Ws	Bur-0	Ge-0	Ri-0
		109	201	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		110	202	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		111	203	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		112	204	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui
		113	205	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	<b>atp4</b>	114	206	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
<b>nad4L</b>		115	207	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		116	208	oui	non	non	non	oui	non	oui	oui	oui	non	oui	oui
		117	209	oui	non	oui	non	oui	non	oui	oui	oui	oui	oui	non
		118	210	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	<b>nad5 ex5</b>	119	211	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		120	212	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		121	213	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		122	214	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		123	215	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		124	216	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J01	217	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J02	218	oui	oui	oui	oui	oui	oui	oui	oui	non	non	oui	oui
		J03	219	oui	oui	non	oui	oui	oui	oui	non	non	oui	non	non
		J04	220	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	<b>cox3</b>	J05	221	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J06	222	oui	oui	oui	oui	oui	non	non	oui	oui	oui	oui	oui
		J07	223	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J08	224	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J09	225	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J10	226	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J11	227	oui	non	oui	non	oui	non	oui	oui	oui	oui	non	oui
		J12	228	oui	oui	oui	oui	oui	oui	oui	oui	non	non	oui	oui
		J13	229	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	non	non
		J14	230	oui	non	non	non	non	non	oui	oui	oui	oui	non	oui
		J15	231	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J16	232	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non
	<b>ccmFN1</b>	J17	233	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J18	234	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J19	235	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J20	236	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J21	237	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J22	238	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J23	239	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		J24	240	non	non	non	oui	oui	non	oui	oui	oui	oui	non	oui
		K01	241	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K02	242	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K03	243	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K04	244	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K05	245	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K06	246	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non
		K07	247	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K08	248	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K09	249	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K10	250	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K11	251	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	<b>ccmC</b>	K12	252	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K13	253	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non
		K14	254	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K15	255	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K16	256	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui
		K17	257	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K18	258	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K19	259	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K20	260	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K21	261	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K22	262	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K23	263	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		K24	264	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		L01	265	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	non	non
	<b>ccmFN2</b>	L02	266	non	oui	non	oui	oui	oui	oui	non	non	oui	oui	oui
		L03	267	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui
		L04	268	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		L05	269	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		L06	270	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		L07	271	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui

Ecotypes

gènes		coord.	amorces	Je	Pi-0	Oy-0	Shahdara	Cvi-0	Kondara	LER	Col-0	Ws	Bur-0	Ge-0	Ri-0	
<b>nad3</b>	<b>rps12</b>	L08	272	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
		L09	273	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		L10	274	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	L11	275	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L12	276	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L13	277	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	
	L14	278	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	non	
	L15	279	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	
	L16	280	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L17	281	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L18	282	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L19	283	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L20	284	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L21	285	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L22	286	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L23	287	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	L24	288	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	
	M01	289	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	M02	290	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	<b>atp9</b>	M03	291	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui
M04		292	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	
M05		293	oui	non	non	oui	non	non	oui	oui	non	non	oui	oui	oui	
M06		294	oui	non	non	oui	non	non	oui	oui	non	non	oui	oui	oui	
M07		295	oui	non	non	oui	non	non	oui	oui	non	non	oui	oui	oui	
M08		296	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
M09		297	oui	non	non	oui	non	non	oui	oui	non	non	oui	oui	oui	
M10		298	oui	non	non	oui	non	non	oui	oui	non	non	oui	oui	oui	
<b>atp1</b>	M11	299	oui	non	non	oui	non	non	oui	oui	non	non	oui	oui		
	M12	300	oui	non	non	oui	non	non	oui	oui	non	non	non	oui		
	M13	301	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui		
	M14	302	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	M15	303	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui		
	M16	304	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	M17	305	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	M18	306	non	oui	non	non	oui	non	oui	oui	non	non	oui	oui		
	M19	307	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui		
	M20	308	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui		
M21	309	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
M22	310	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
M23	311	oui	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui			
M24	312	non	oui	non	oui	oui	non	oui	oui	oui	oui	oui	oui			
N01	313	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
N02	314	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
N03	315	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
N04	316	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui			
<b>rps7</b>	N05	317	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N06	318	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N07	319	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N08	320	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N09	321	non	non	non	non	oui	non	non	oui	oui	oui	non	non		
N10	322	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
<b>nad1</b>	N11	323	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N12	324	oui	oui	non	non	oui	non	oui	oui	oui	oui	oui	non		
	N13	325	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N14	326	oui	oui	non	oui	oui	non	oui	oui	oui	oui	oui	oui		
	N15	327	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N16	328	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N17	329	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	non		
	N18	330	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui		
	N19	331	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui		
	N20	332	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui		
N21	333	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui			
<b>nad2</b>	N22	334	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N23	335	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	N24	336	oui	non	oui	oui	oui	non	oui	oui	oui	oui	non	oui		
	O01	337	non	non	non	non	oui	oui	oui	oui	oui	non	oui	oui		
	O02	338	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	O03	339	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	O04	340	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		
	O05	341	oui	oui	non	non	oui	non	oui	oui	oui	oui	oui	oui		
	O06	342	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui		



Ecotypes

gènes	coord.	amorces	Jea	Pi-0	Oy-0	Shahdara	Cvi-0	Kondara	LER	Col-0	Ws	Bur-0	Ge-0	Ri-0	
	O07	343	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	O08	344	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	O09	345	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	O10	346	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	O11	347	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	
	O12	348	oui	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	
	O13	349	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	O14	350	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	
	cox1	O15	351	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		O16	352	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		O17	353	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		O18	354	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
		O19	355	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
	O20	356	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	
O21	357	oui	oui	non	oui	oui	oui	oui	oui	oui	oui	non	oui		
O22	358	oui	oui	oui	oui	oui	non	oui	oui	oui	oui	non	oui		
O23	359	oui	non	oui	oui	oui	non	oui	oui	oui	non	oui	oui		