



HAL
open science

Caractérisation de protéines PPR impliquées dans le stress biotique chez *A. thaliana*.

Bastien Malbert

► **To cite this version:**

Bastien Malbert. Caractérisation de protéines PPR impliquées dans le stress biotique chez *A. thaliana*.. Biologie végétale. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACLS522 . tel-01993950

HAL Id: tel-01993950

<https://theses.hal.science/tel-01993950v1>

Submitted on 25 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Caractérisation de protéines PPR impliquées dans le stress biotique chez *A. thaliana*

Thèse de doctorat de l'Université Paris-Saclay
Préparée à l'Université Paris-Sud

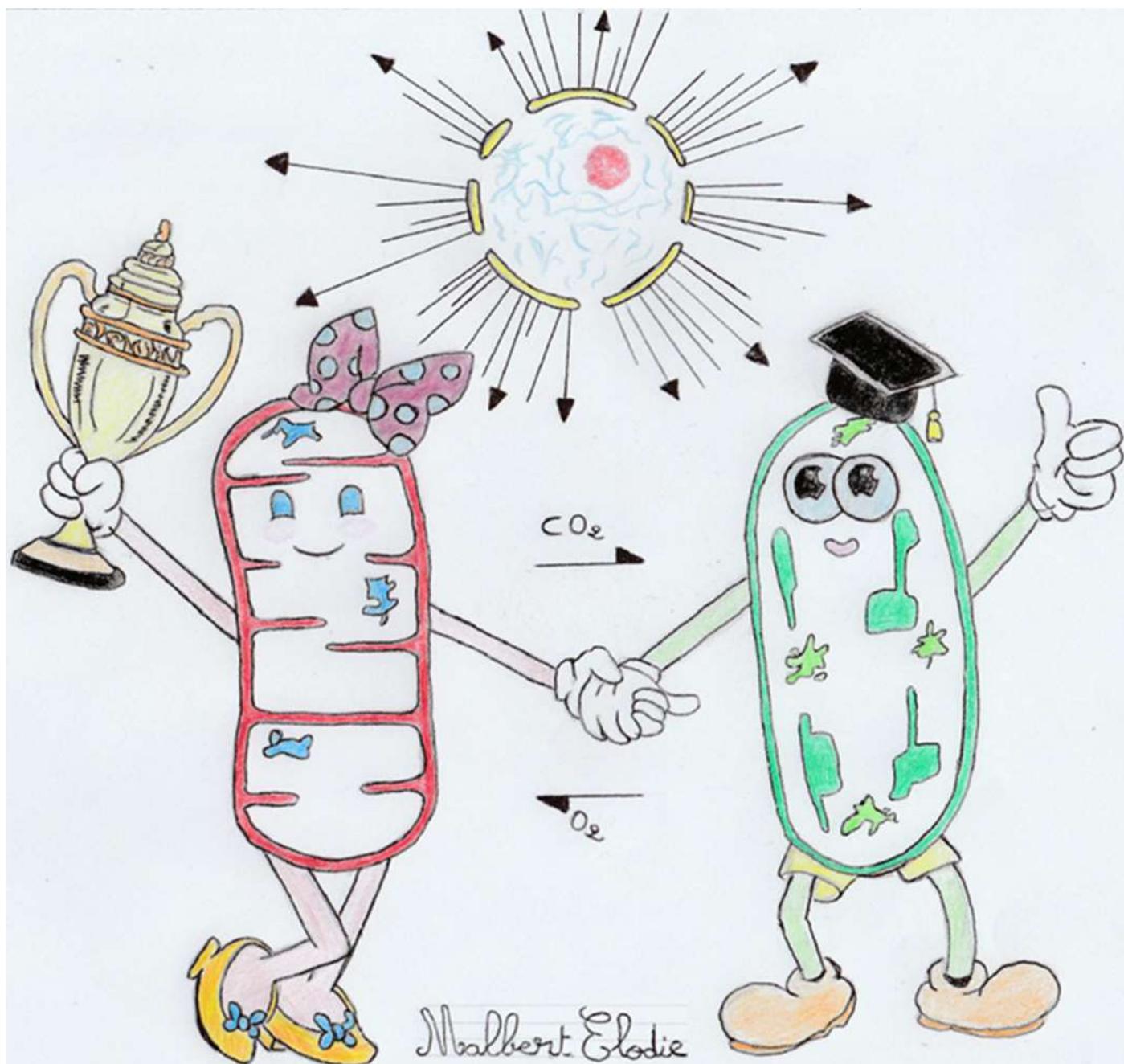
École doctorale n°567 Sciences du Végétal : du gène à l'écosystème
(SDV)

Spécialité : Biologie
Soutenance effectuée à Gif sur Yvette, le 10/12/18, par

Bastien MALBERT

Composition du Jury :

Mme Alia Dellagi Professeur, AgroParisTech (IJPB)	Présidente
Mr Stéphane Genin DR, Université Toulouse 3-CNRS (LIPM)	Rapporteur
Mr Philippe Giegé DR, Université de Strasbourg-CNRS (IBMP)	Rapporteur
Mr Hakim Mireau DR, INRA (IJPB)	Examineur
Mr Yves Choquet DR, Université Pierre et Marie Curie-CNRS (IBPC)	Examineur
Mr Etienne Delannoy CR, INRA (IPS2)	Directeur de thèse
Mme Claire Lurin DR, INRA (IPS2)	Co-Encadrante de thèse



**A mon grand-père Julien qui m'a donné le goût d'apprendre,
Et à l'ensemble de ma famille qui m'a permis de le faire,
Je dédie cette thèse.**

Remerciements

En premier lieu, je souhaite remercier mes encadrants de thèse, Claire Lurin et Etienne Delannoy, pour leur accueil, leur encadrement, leur patience infinie et l'émulation intellectuelle qui m'a permis de mener cette thèse.

Je remercie les membres de l'équipe OGE de l'IPS2, Nathalie Rezé, Livia Merendino, Wojciech Majeran, Dario Monachello, Benoit Castandet, Damien Guillaumot et Kevin Baudry pour leur gentillesse et leur bienveillance pendant ces 3 dernières années.

Je remercie également les membres de la plateforme de transcriptomique pour leur aide. Je remercie particulièrement José pour ces longues discussions pas vraiment scientifiques et plutôt footballistiques.

Je remercie également les stagiaires que j'ai eu la chance de côtoyer ces dernières années, en particulier Félix Barbut, Sébastien Skiada et Dany Vo.

Je remercie les membres de mon comité de thèse, Olivier Vallon, Pascal Ratet et Sébastien Thomine pour leur aide et leurs conseils avisés.

Je remercie par avance l'ensemble du jury de thèse, et particulièrement les deux rapporteurs Stéphane Genin et Philippe Giegé pour le temps consacré et votre expertise.

Je remercie Nemo Peeters pour avoir collaboré avec nous sur l'effecteur de *Ralstonia*. Je remercie également Mathilde Fagard pour la collaboration sur *Erwinia*, pour sa gentillesse et sa disponibilité.

Je tiens également à remercier Séverine Domenichini, Marie-Noelle Brisset, Marie-Christine Soulié, Sylvie Jolivet, Dominique Expert, Marie Dufresne, Julien Lang, Marie-Laure Martin-Magniette, Guillem Rigail, Emmanuelle Issakidis-Bourguet, Anja Krieger-Liszkay, Mathieu Jossier, Jean-Philippe Tamby et l'ensemble de l'équipe serre de l'IPS2 pour leur aide et leur implication aux différentes étapes de la thèse.

Pour finir, je remercie sincèrement l'ensemble de ma famille pour m'avoir soutenu, de la classe préparatoire jusqu'à la fin de cette thèse et mes amis, que je n'ai pas pu voir aussi souvent que je le voulais dernièrement. Enfin, je te remercie Marion, pour avoir partagé mes joies, mes peines, mes réussites et mes échecs pendant toutes ces années.

We choose to go to the moon.

We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard.

Because that goal will serve to organize and measure the best of our energies and skills.

Because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win, and the others, too.

John Fitzgerald Kennedy – 12 septembre 1962 – Université Rice (Houston).
Address at Rice University on the Nation's Space Effort

Abréviations

A : Adénine	MORF : <i>Multiple Organellar RNA editing Factor</i>
aa : Acide aminé	MS : Milieu Murashige et Skoog
AA : Auto-activateur	Mt : Mitochondrie
ABA : Acide abscissique	NaCl : Chlorure de sodium
AD : <i>Activation Domain</i> (domaine d'activation)	NEF : <i>Nucleotide Exchange Factor</i>
ADN : Acide désoxyribonucléique	NEP : <i>Nuclear Encoded RNA Polymerase</i>
ADNc : ADN complémentaire	NPQ : <i>Non-Photochemical Quenching</i>
ADNg : ADN génomique	Nt : Nucléotide
ADN-T : ADN de transfert	ORF : <i>Open Reading Frame</i> (cadre de lecture ouvert)
AGI : Initiative pour le génome d' <i>Arabidopsis</i>	ORRM : <i>Organelle RNA Recognition Motif-containing</i>
ARN : Acide ribonucléique	PAM : <i>Pulse Amplitude Modulation</i>
ARNi : ARN interférent	PAMP : <i>Pathogen Associated Molecular Pattern</i>
ARNm : ARN messager	Pb : Paire de bases
ARNr : ARN ribosomal	PCR : <i>Polymerase Chain Reaction</i> (réaction de polymérisation en chaîne)
ARNt : ARN de transfert	PEP : <i>Plastid-Encoded RNA Polymerase</i>
Art : Artificiel	POPS : Plateforme Transcriptomique de l'IPS2
ATP : Adénosine triphosphate	PPR : <i>Pentatricopeptide Repeat</i>
Bc : <i>Botrytis cinerea</i>	PSI : Photosystème I
BiFC: Complémentation de fluorescence bimoléculaire	PSII : Photosystème II
BLAST : <i>Basic Local Alignment and Search Tool</i>	Pst : <i>Pseudomonas syringae pv. tomato</i>
C : Cytosine	Pt : Plaste
Col0 : <i>Arabidopsis thaliana</i> écotype "Columbia"	PTI : <i>PAMP-Triggered Immunity</i>
DAB : 3-3'-Diaminobenzidine	RIP : <i>RNA editing factor Interacting Protein</i>
DB : <i>Binding Domain</i>	RNAseq : Séquençage des ARN
DE : Différentiellement Exprimé	ROS : <i>Reactive Oxygen Species</i> (composés réactifs de l'oxygène)
dNTP : Désoxyribonucléotide	RRM : <i>RNA Recognition Motifs</i>
DO : Densité optique	Rs : <i>Ralstonia solanacearum</i>
Ea : <i>Erwinia amylovora</i>	RT-qPCR : Rétro-transcription suivie d'une PCR quantitative
ET : Ethylène	SA : Acide salicylique
ETI : <i>Effector-Triggered Immunity</i>	T : Thymine
FDR : <i>False-Discovery Rate</i> (taux de faux positifs)	T3E : <i>Type-III Effector</i> (effecteur de type III)
G : Guanine	T3SS : <i>Type-III Secretion System</i> (système de sécrétion de type III)
GFP : <i>Green Fluorescent Protein</i>	TAIR: <i>The Arabidopsis Information Resource</i>
Glc : Glucose	TMM : <i>Trimmed Mean of M values</i>
HA : Hémagglutinine	TPR : <i>Tetratricopeptide Repeat</i>
Hpa : <i>Hyaloperonospora arabidopsidis</i>	U : Uracile
HR : <i>Hypersensitive Response</i> (réponse hypersensible)	UFC : Unité formant colonie
Hsp : <i>Heat-Shock Protein</i>	UTR : <i>Untranslated region</i>
IPS2 : <i>Institute of Plant Science Paris-Saclay</i>	WT : <i>Wild type</i>
JA : Acide jasmonique	YFP : Yellow Fluorescent Protein
KO : <i>Knock-out</i>	
LB : <i>Lysogeny broth</i>	
MEF : <i>Mitochondrial Editing Factor</i>	
miRNA : Micro ARN	

Sommaire

Remerciements.....	2
Abréviations.....	5
Sommaire.....	6
Avant-propos.....	9
Définitions.....	10
Introduction.....	11
1. Interaction plante-pathogène vue par le pathogène.....	11
1.1. Origines et importance de la phytopathologie.....	11
1.2. Différents types de pathogènes.....	12
1.3. Qu'est-ce qu'un effecteur ?.....	12
2. Interaction plante-pathogène vue par la plante.....	19
2.1. Reconnaissance du pathogène.....	19
2.2. Principaux acteurs de la défense des plantes.....	24
2.3. Rôle des organites dans la défense.....	27
3. Régulation de l'expression génétique des organites pendant un stress biotique.....	28
3.1. Génomique des organites chez <i>Arabidopsis thaliana</i>	28
3.2. Expression génétique chez les organites d' <i>Arabidopsis thaliana</i>	31
3.3. Acteurs de la régulation de l'expression génétique des organites.....	32
3.4. Les protéines PPR.....	35
3.5. Organites et stimuli environnementaux.....	42
4. Projet de thèse.....	45
Chapitre 1 : Proposition d'un pipeline d'analyse de données <i>RNAseq</i> pour le chloroplaste et la mitochondrie.....	47
1. Présentation du pipeline d'analyse de données <i>RNAseq</i>.....	47
1.1. Pipeline appliqué à l'étude du transcriptome du chloroplaste.....	47
1.2. Pipeline appliqué à l'étude du transcriptome de la mitochondrie.....	48
2. Discussion.....	48
2.1. Génération des données.....	48
2.2. Analyse de l'expression des gènes.....	49
2.3. Analyse de couverture.....	52
2.4. Analyse d'épissage.....	53
2.5. Analyse d'édition.....	54

Chapitre 2 : Amélioration du système de prédiction de sites d'édition sous dépendance de PPR E+	55
1. Mise en contexte	55
1.1. Etude de CLB19, DYW2 et NUWA	55
1.2. Analyse transcriptomique des mutants $\Delta nuwa$ et $\Delta dyw2$	56
1.3. Hypothèse et projet de recherche	57
2. Lien fonctionnel entre DYW2 et les PPR E+	58
2.1. Validation du lien fonctionnel entre DYW2 et les PPR E+	58
2.2. Application du critère DYW2 et rangs des sites prédits	63
3. Extension à l'ensemble des PPR E+	68
4. Discussion	70
Chapitre 3 : Caractérisation fonctionnelle de la protéine PPR PGN en condition de stress	74
1. Etat des lieux	74
1.1. Une PPR mitochondriale à l'expression induite par <i>Botrytis cinerea</i>	74
1.2. Un mutant hypersensible au glucose, au NaCl et à l'acide abscissique	75
1.3. L'expression de gènes impliqués dans le fonctionnement de la mitochondrie perturbée chez le mutant Δpgn	76
1.4. Hypothèse de travail	78
2. PGN, une PPR E+ impliquée dans l'édition	78
2.1. Etude <i>in silico</i>	78
2.2. Utilisation de la nouvelle méthode de prédiction pour PPR E+	78
2.3. Analyse du transcriptome	79
2.4. Complémentation fonctionnelle du mutant Δpgn	80
3. PGN, une PPR dont la transcription est induite en condition de stress	81
3.1. En condition de stress biotique	81
3.2. En condition de stress salin	82
4. Discussion	83
Chapitre 4 : Mise en évidence et caractérisation fonctionnelle de protéines bactériennes à motifs PPR	85
1. Mise en évidence de protéines à motifs PPR chez les bactéries	85
1.1. Initiation du projet	85
1.2. Recherche de motifs PPR chez les bactéries	85
1.3. Choix des objets d'étude	90
2. Hypothèse de travail	91
3. Caractérisation de la protéine PPR trouvée chez <i>Ralstonia solanacearum</i>	92
3.1. Présentation du pathogène <i>Ralstonia solanacearum</i> et de sa protéine PPR	92

3.2.	Importance de RipL dans la pathogénicité de <i>Ralstonia solanacearum</i>	93
3.3.	Injection.....	94
3.4.	Localisation subcellulaire.....	94
3.5.	Rôle <i>in planta</i>	95
3.6.	Discussion	103
4.	Caractérisation de la protéine PPR trouvée chez <i>Erwinia amylovora</i>	106
4.1.	Description de l'agent pathogène <i>Erwinia amylovora</i> et de sa protéine PPR.....	106
4.2.	Injection de la protéine codée par EAMY_0820 lors de l'infection	108
4.3.	Importance d'EAMY_0820 dans la pathogénicité d' <i>E. amylovora</i>	109
4.4.	Localisation subcellulaire.....	115
4.5.	Rôle <i>in planta</i>	116
4.6.	Exportation des résultats sur <i>Malus domestica</i> « Golden delicious »	125
4.7.	Discussion	126
	Discussion générale.	129
	Matériels et méthodes	135
1.	Matériel biologique utilisé lors de la thèse	135
1.1.	Espèces végétales utilisées et conditions de cultures	135
1.2.	Bactéries utilisées	135
1.3.	Champignons utilisés	135
1.4.	Souches de levure.....	135
2.	Méthodes	136
2.1.	Préparation de bactéries compétentes	136
2.2.	Génotypage.....	136
2.3.	Clonage et transformation de plantes	139
2.4.	Microscopie.....	143
2.5.	Transcriptomique	143
2.7.	Double Hybride levure	148
2.8.	Analyses bio-informatiques et logiciels utilisés.....	149
	Table des matières.....	151
	Table des figures.....	157
	Table des tableaux.....	161
	Annexes.....	163
	Liste des amorces PCR utilisées pour le clonage.....	189
	Liste des amorces PCR utilisées pour mesurer l'édition.	190
	Liste des amorces utilisées pour le génotypage des lignées d'insertion d'ADN-T.	192
	Références.....	196

Avant-propos

Qui ne s'est jamais réveillé un matin, en se posant une question toute simple, mais ô combien intrigante : quels événements dans ma vie ou celle de mes proches, ont fait que je suis là aujourd'hui, en train de faire ce métier, de mener cette vie ? La plupart des gens parleraient d'influence familiale, de goûts et de couleurs, ou encore de concours plus ou moins réussis, ce que des Pierre Bourdieu et autres sociologues de talent se feraient et feront un plaisir à disséquer. Je me suis souvent posé la même question pour les personnages historiques. Qu'est-ce qui a fait qu'ils se trouvaient en ce lieu, à ce moment précis de l'histoire, et pour quelles conséquences ? Vaste question. Quel rapport avec ce sujet de thèse, me direz-vous. Imaginez une seconde qu'une personne vous réponde que la vie qu'elle mène a, pour partie, été influencée par un agent pathogène de plante. Insensé.

Entre 1845 et 1852, la Grande Famine s'abat sur l'Irlande, résultant d'une relation conflictuelle avec la couronne britannique, de pratiques agricoles inadéquates, mais surtout de l'apparition de *Phytophthora infestans*. Responsable d'une maladie connue sous le nom de mildiou, ce pathogène de la famille des Pythiacées infecte les pommes de terre, principale culture vivrière en Irlande au XIX^{ème} siècle. L'une des conséquences de cette famine fut l'émigration vers l'actuel Royaume-Uni, le Canada, l'Australie ou encore les Etats-Unis. Parmi ces migrants à destination du nouveau continent, se trouvaient les familles Fitzgerald et Kennedy. A la question « qu'est-ce qui m'a amené à être le 35^{ème} président des Etats-Unis ? », le petit John s'est probablement exclamé, un matin dans sa salle de bain –et sans mauvaise foi aucune de ma part- « c'est grâce à *Phytophthora infestans* ! ».

Définitions

La définition choisie du terme « organite » dans le manuscrit est plus restrictive et ne concerne que les organites à génome, les plastes et les mitochondries.

Malgré la grande variété de plastes chez les plantes, le choix a été fait de s'intéresser seulement aux chloroplastes.

Le terme « procaryote » est utilisé dans le manuscrit en opposition aux Eucaryotes. Il se réfère aux domaines des Bactéries et des Archées.

On parle de mutant KO pour un gène donné lorsque, par différentes méthodes, on inactive ou on altère suffisamment son expression pour ne pas avoir de protéine fonctionnelle.

Un pathotest est une évaluation de la sensibilité d'une espèce végétale à un phytopathogène, chacun pouvant être sauvage ou mutant KO pour un gène donné.

La définition de répétition biologique adoptée dans ce manuscrit est une répétition à 24h d'intervalle.

Cette thèse discutant de gènes eucaryotes, procaryotes et endosymbiotiques, une nomenclature arbitraire a été définie afin de différencier clairement les gènes des ARN, protéines ou éventuelles mutations.

Nomenclature :

Eucaryotes :

- Gène : *pgn* ; *aox* ; *sig3*
- ARN : *pgn* ; *aox* ; *sig3*
- Protéine : PGN ; AOX ; SIG3
- Mutant : Δpgn ; Δaox ; $\Delta sig3$

Bactéries et endosymbiontes :

- Gène : *psbT* ; *dspA/E* ; *cox2*
- ARN : *psbT* ; *dspA/E* ; *cox2*
- Protéine : PsbT ; DspA/E ; Cox2
- Mutant : $\Delta psbT$; $\Delta dspA/E$; $\Delta cox2$

Introduction

L'objectif de ce travail de thèse est de caractériser des protéines PPR, impliquées dans l'expression des gènes des organites chez les plantes, en condition de stress biotique. Les stress biotiques correspondent à l'action néfaste d'un organisme pathogène sur la plante. Les protéines PPR que nous étudions peuvent être d'origine endogène, c'est-à-dire que les gènes les codant sont dans la cellule végétale, ou alors d'origine exogène, provenant des pathogènes. Dans ce dernier cas de figure, on parle d'effecteur lorsqu'une protéine de l'organisme pathogène est envoyée dans la cellule végétale afin de faciliter l'infection. L'objectif de cette introduction est, en définissant précisément l'interaction plante-organisme pathogène, le rôle des effecteurs, l'expression des gènes des organites et le rôle connu des protéines PPR, d'avoir les informations nécessaires à la compréhension du travail effectué ces trois dernières années.

1. Interaction plante-pathogène vue par le pathogène

1.1. Origines et importance de la phytopathologie

A l'image des animaux, les plantes tombent malades. Déjà, l'Ancien Testament rapportait des cas de rouille sur le blé (Aggée, 2 :17). Considérée comme un fléau des Dieux, les traitements impliquaient des prières pour Robigus, Dieu des cultures dans la Rome antique, jusqu'aux Rogations, pratiquées par l'Eglise catholique encore très récemment. Bien que cette dernière pratique influença des peintres naturalistes comme Jules Breton (Figure 1), les résultats agronomiques étaient assez peu répétables d'une année sur l'autre. La prédominance de croyances religieuses de l'Antiquité jusqu'à la révolution industrielle (résumée par Orlob, 1971) sur l'origine des maladies des plantes explique en partie la naissance tardive de la phytopathologie telle que nous la connaissons. Son essor commence au XIX^{ème} siècle, indirectement, suite à plusieurs événements. On peut citer en exemples l'utilisation du microscope par A. Van Leeuwenhoek appliquée à la biologie (XVII^{ème} siècle), la réfutation de la théorie de la génération spontanée par, entre autres, Louis Pasteur (1864), et plus cyniquement, l'épidémie de mildiou de la pomme de terre qui toucha l'Europe à cette période. Les découvertes de J. G. Mendel (1866) sur l'hérédité, de J. D. Watson et F. H. Crick sur la structure de la molécule d'ADN (1953) et plus récemment le développement des techniques de séquençage haut débit ouvrent d'incroyables perspectives sur l'étude des interactions plante-pathogène(s). Les paragraphes qui vont suivre détaillent la reconnaissance par la plante d'un pathogène, les systèmes de défense mis en place par la plante, la signalisation associée puis les éventuelles stratégies de contournement du pathogène. Il convient cependant de garder en tête que les mécanismes qui vont être expliqués ci-dessous concernent l'interaction d'une plante avec un et un seul pathogène. Bien que cela soit la meilleure façon de comprendre l'interaction, on reste très éloigné de ce qui peut se passer en champ où les plantes devront gérer une multitude de pathogènes, parfois au même moment, tout en s'adaptant à de potentiels stress abiotiques (température extrêmes, manque d'eau, etc). La façon de donner priorité à la réponse contre l'un ou l'autre pathogène reste pour l'instant inconnue (Bari et Jones, 2009).



Figure 1: Jules Breton, La bénédiction des blés en Artois (1857).

www.musee-orsay.fr

1.2. Différents types de pathogènes

En champ, la plante va être soumise à de nombreux stress biotiques. Les principaux seront les insectes, les herbivores ou encore les nématodes. Les microorganismes, comme par exemple des bactéries, des virus, des oomycètes ou des mycètes, peuvent également représenter un danger pour la plante. Certains pathogènes de plantes peuvent être divisés –théoriquement- en différents groupes suivant leur cycle de reproduction (Moore et al., 2011). Les pathogènes dits biotrophes tirent leur énergie de cellules végétales vivantes, par opposition aux pathogènes nécrotrophes qui vont tuer la cellule végétale lors de l'infection. A mi-chemin entre les deux se trouvent les pathogènes hémibiotrophes qui vont alterner dans leur cycle de la biotrophie et de la nécrotrophie. Ces définitions restent cependant très générales et ne peuvent s'appliquer de façon claire à tous les pathogènes.

1.3. Qu'est-ce qu'un effecteur ?

1.3.1. Définition

La définition d'effecteur choisie est celle proposée par Hogenhout *et al.* (2009). Un effecteur est une protéine ou une petite molécule du pathogène qui va altérer la structure et/ou la fonction de la cellule-hôte. Ces altérations peuvent avoir pour conséquences de faciliter l'infection, de déclencher des mécanismes de défense de la plante, voire les deux (Hogenhout et al., 2009). Cette définition, extrêmement large, permet d'inclure les protéines de pathogènes non impliquées directement dans la pathogénicité et reconnues par la plante, ou encore les différentes toxines dans la catégorie des effecteurs. Cette définition n'inclut pas de notion relative à l'injection ou à la sécrétion. Pour autant, de nombreux effecteurs sont injectés dans la cellule végétale afin de faciliter l'infection, on parle alors de translocation, quand d'autres seront sécrétés dans l'apoplasme.

Après avoir détaillé les principales caractéristiques des effecteurs, je m'intéresserai aux mécanismes permettant leur injection.

1.3.2. Caractéristiques des effecteurs

Dans ce paragraphe, on se focalisera sur le cas particulier des bactéries Gram-négatives et du système de sécrétion de type III (T3SS) pour deux raisons : la première étant qu'il s'agit du mécanisme d'injection d'un pathogène vers une plante le plus étudié, la seconde raison étant que l'essentiel du travail de thèse s'est fait sur

l'étude de bactéries pathogènes Gram-négatives utilisant ce système de sécrétion en particulier. Les effecteurs de type III semblent avoir quelques caractéristiques communes en terme de séquence protéique. Oh et al. (2010) ont résumé les principales caractéristiques, en se basant sur le travail de Petnicki-Ocwieja et al. (2002) sur l'alignement des premiers 40 acides aminés de 28 effecteurs putatifs chez *Pseudomonas syringae* :

- On trouve les acides aminés Ile, Val, Leu, Ala ou Pro en positions 3 ou 4 mais pas les deux, et ils sont généralement précédés par une Pro
- La position 5 est très rarement occupée par une Met, Ile, Leu, Phe, Tyr ou Trp
- On ne trouve pas de Glu ou Asp dans les 12 premières positions
- Les 40 premiers acides aminés sont riches en Ser et Gln, et de manière générale en acides aminés polaires et amphiphiles
- Il y a très peu de Cys entre les positions 5 et 40
- Dans les 40 premiers acides aminés, on ne trouve pas plus de 3 acides aminés Met, Ile, Leu, Val, Phe, Tyr ou Trp consécutifs

Un doute persiste cependant sur le signal de sécrétion et son importance pour la translocation. Jusqu'à présent, il paraissait relativement évident que le signal correspondait aux acides aminés N-terminaux de l'effecteur. Pourtant, Ramamurthi et Schneewind (2005) ont démontré chez *Yersinia enterocolitica* que des mutations silencieuses dans les codons 2, 3, 5 et 7 de l'effecteur YopE empêchaient sa translocation. On ne sait donc toujours pas avec certitude si l'adressage au T3SS est déterminé par une séquence et/ou structure de bases azotées ou d'acides aminés.

1.3.3. Sécrétion et translocation

D'abord supposé comme étant réservés aux bactéries pathogènes, les systèmes de sécrétion d'effecteurs sont également présents chez de nombreux autres pathogènes de plantes. En plus des bactéries Gram-négatives, on peut citer les mycètes, les oomycètes ou encore les nématodes (Hogenhout et al., 2009). L'injection ou la sécrétion consiste en l'envoi d'effecteurs dans différents compartiments cellulaires de la plante infectée (noyau, organites, cytosol, apoplaste, etc.) en utilisant différents systèmes de sécrétion. Par exemple, les oomycètes utilisent un haustorium afin de récupérer les nutriments de la plante comme pour envoyer un certain nombre d'effecteurs dans la plante (Panstruga, 2003; Catanzariti, 2006). De la même façon que dans le paragraphe précédent, je vais me focaliser sur l'étude du T3SS. Selon Puhar et Sansonetti (2014), le T3SS présente une particularité importante: à la différence de la plupart des systèmes de sécrétion qui sécrètent dans le milieu, le T3SS permet la translocation de protéines encodées par la bactérie, à partir de son cytosol vers la membrane plasmique et/ou le cytoplasme de la plante.

Le T3SS est un complexe protéique bactérien, formant ce qui pourrait ressembler à une seringue moléculaire, permettant l'injection d'effecteurs dans la plante infectée. Attention cependant : bien que des bactéries pathogènes utilisent leur T3SS afin d'infecter des plantes et que celui-ci soit crucial pour l'infection, les bactéries commensales en sont aussi pourvues.

Suite à la mise en évidence de fortes similarités de séquences entre les protéines du flagelle et du système de sécrétion de type III, de nombreuses équipes ont appliqué avec succès les méthodes de purification du flagelle bactérien au T3SS. Son architecture a été mise en évidence après isolation, purification et observation en microscopie électronique chez les bactéries zoopathogènes (pour exemple, Kubori, 1998), montrant une base cylindrique surmontée d'un appendice de forme cylindrique (Figure 2). Selon Puhar et Sansonetti (2014), environ 25 protéines sont nécessaires pour constituer le T3SS. On y trouve une ATPase, des structures en forme d'anneaux intégrés dans les membranes internes et externes de la bactérie, un pilus pour les bactéries pathogènes de plantes qui permet le passage à travers l'épaisse paroi des plantes.

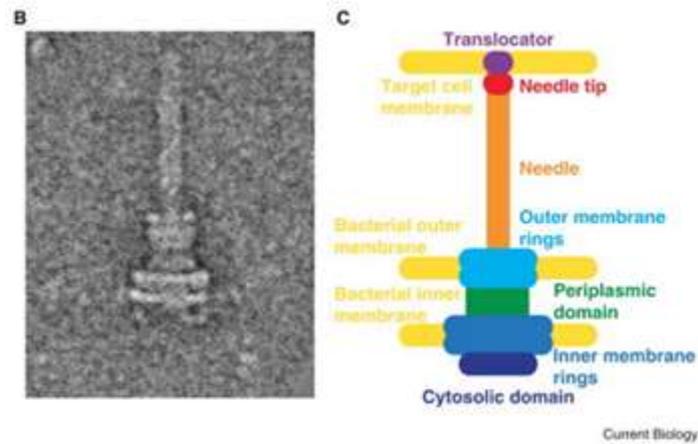


Figure 2: Observation par microscopie électronique après purification (B) et représentation schématique (C) du système de sécrétion de type III de *Salmonella enterica*.

Tiré de Puhar et Sansonetti (2014).

1.3.4. Redondance fonctionnelle des effecteurs

La génétique inverse est l'une des méthodes les plus utilisées afin d'étudier la fonction d'un gène. A partir d'un mutant dans lequel le gène que l'on souhaite étudier n'est plus fonctionnel, et en comparant ce mutant à l'organisme sauvage correspondant, on infère le rôle du gène. Cette stratégie n'a pas permis dans de nombreux cas de mettre en évidence de phénotype marqué pour les mutants simples d'effecteurs de type III, suggérant soit que la protéine n'est pas impliquée dans la pathogénicité, soit une redondance fonctionnelle d'effecteurs (Jamir et al., 2004). Alfano et Collmer (2004) confirment l'analyse en avançant que la plupart des simples mutants d'effecteurs de type III présentent bien des phénotypes en terme de pathogénicité mais trop subtils pour être étudiés. L'étude de certains effecteurs se fait au niveau de la famille entière, à l'aide d'extinctions de gènes multiples, défendant l'hypothèse d'une redondance fonctionnelle récurrente chez les effecteurs de type III (Lahaye et Bonas, 2001). Les rares effecteurs faisant exception sont ceux impliqués dans le T3SS, ou par exemple DspA/E, effecteur majeur d'*Erwinia amylovora* (Launay et al., 2016).

1.3.5. Cibles connues d'effecteurs de type III

Cette redondance fonctionnelle n'empêchera cependant pas la caractérisation de nombreux effecteurs, (Xin et He, 2013) chez la bactérie phytopathogène modèle *Pseudomonas syringae* pv. *tomato* DC3000, dont on peut trouver un résumé dans le schéma en figure 3.

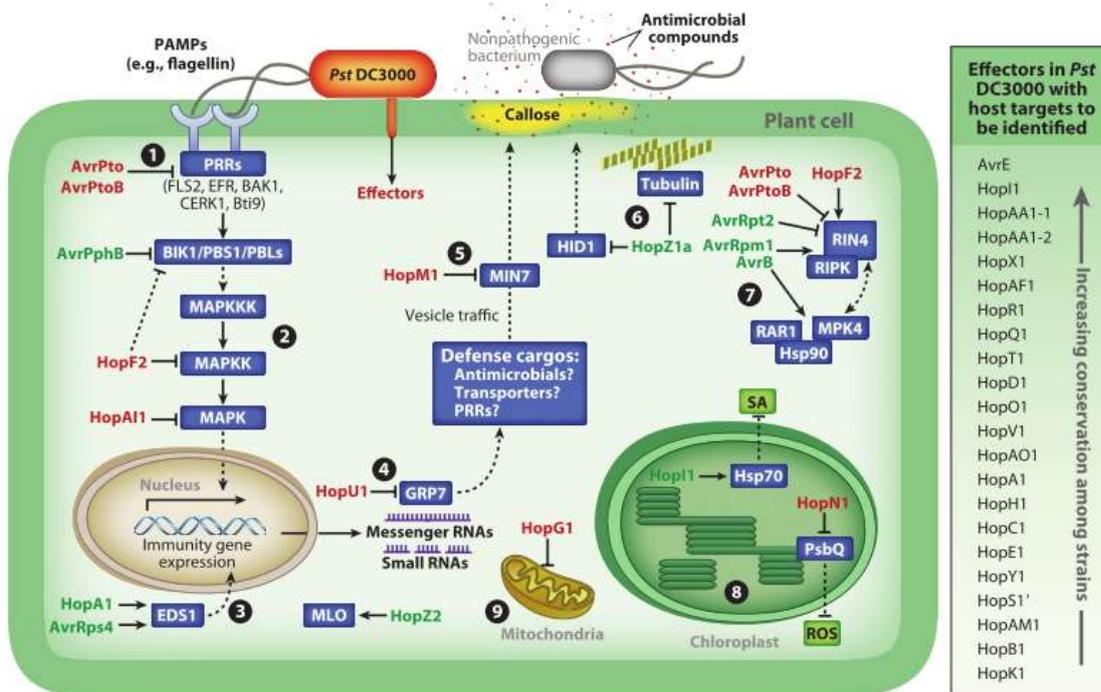


Figure 3: Cibles connues d'effecteurs de type III de *Pseudomonas syringae* DC3000 chez *Arabidopsis thaliana*.

Tiré de Xin et He (2013).

Le paragraphe suivant détaillera 3 exemples d'effecteurs de *Pseudomonas syringae* (*Pst*), HopF2, HopN1, HopG1, choisis pour représenter la multitude de compartiments cellulaires pouvant être affectés par les effecteurs bactériens.

HopF2 cible une protéine d'*Arabidopsis* appelée RIN4 (Robert-Seilaniantz et al., 2006; Wilton et al., 2010). RIN4 est une protéine transmembranaire ciblée par de nombreux effecteurs de *Pst* (AvrB, AvrRpm1 et AvrRpt2) avec pour conséquence une activation des défenses de la plante. Les travaux de Wilton *et al.* montrèrent qu'en plus d'interagir aussi bien *in vitro* qu'*in vivo* avec RIN4, HopF2 semble empêcher le clivage de RIN4 par AvrRpt2, et, avec cela, l'activation des défenses de la plante associée. Les auteurs montrèrent qu'en plus de ne pas déclencher d'ETI (pour *Effector Triggered Immunity*, réaction de défense de la plante suite à la reconnaissance d'un effecteur), HopF2 semble favoriser la croissance de *Pst* chez *Arabidopsis*. Le rôle de cet effecteur n'est pas sans rappeler l'étape 4 du modèle en « zigzag » de Jones et Dangl (2006) (Figure 7) concernant l'acquisition de nouveaux effecteurs afin d'éviter les défenses de la plante, mais rien ne prouve que cet effecteur a été obtenu plus tard dans l'évolution que les 3 autres effecteurs ciblant RIN4.

HopN1 a été étudié pour la première fois en 2004 par López-Solanilla et al. En plus de démontrer l'injection via le T3SS dans *Solanum lycopersicum*, les auteurs remarquèrent que HopN1 permettait de bloquer la mort cellulaire, associée à une réponse hypersensible chez une plante non-hôte comme à la sensibilité chez une plante hôte (López-Solanilla et al., 2004). *In vitro*, les auteurs montrèrent aussi que l'effecteur avait une activité cystéine protéase, et qu'en mutant le site catalytique, ils perdaient le phénotype associé à la mort cellulaire. HopN1 semble donc avoir besoin de son activité protéase dans son rôle d'effecteur. Dans un second temps, Rodríguez-Herva et al. (2012) localisèrent HopN1 dans le chloroplaste en agroinfiltrant des fusions GFP dans *Nicotiana benthamiana* ainsi qu'en transformant transitoirement de l'épiderme d'oignon (*Allium cepa*) par biolistique. Par immunoprécipitation et spectrométrie de masse, une interaction entre HopN1 et PsbQ, une sous unité du photosystème II codée par le noyau fut démontrée. En surexprimant HopN1 dans *Nicotiana tabacum*, le niveau de ROS mesurés pour la plante sauvage avec HopN1 serait 2 à 3 fois inférieur en comparaison avec la plante sauvage et un vecteur vide. Les ROS (pour *Reactive Oxygen Species*, espèces

réactives de l'oxygène hautement réactives) sont des composés dérivés de l'oxygène pouvant endommager les cellules et servir dans la communication en condition de stress. Ce faible niveau de ROS correspond cependant au niveau obtenu lorsque l'on éteint l'expression de *psbQ*. L'extinction du gène codant pour *PsbQ* entraînerait donc une forte diminution dans la production de ROS. La fonction de HopN1 *in planta* lors de l'infection pourrait donc en partie être d'altérer le fonctionnement de cette protéine en la dégradant, limitant ainsi la production de ROS.

HopG1 est également un effecteur de *Pst*, dont la sécrétion *in vitro* fut mise en évidence par Petnicki-Ocwieja et al. (2002). Plus tard, il suscita l'intérêt de Jamir et al. (2004) lors d'un crible d'effecteurs de *Pseudomonas* car il fut montré que HopG1 était capable de bloquer certaines réponses de la plante. Block et al. (2010) continuèrent l'étude de cet effecteur. Ils montrèrent que son expression est induite en condition d'infection, et à l'aide d'une fusion adénylate cyclase, que l'effecteur est injecté dans la plante lors de l'infection. Ils montrèrent aussi que la surexpression de HopG1 dans différentes plantes (*A. thaliana* – *N. Tabacum* – *S. esculentum*) entraînait un phénotype nain et l'infertilité. Cependant, ces phénotypes n'étant pas observés lors de l'infection avec une souche *Pst* sauvage, les auteurs suggèrent qu'ils sont dus à des interactions qui ne peuvent avoir lieu au stade où se déroule l'infection, ou que la cible de HopG1 joue un second rôle, dans le développement de la plante par exemple. En fusionnant l'effecteur à la GFP et après agroinfiltration, Block *et al.* (2010) localisèrent l'effecteur dans la mitochondrie. Fait intéressant, différentes mutations de l'effecteur furent testées, aussi bien en localisation subcellulaire qu'en recherche de phénotype(s) en surexprimant dans *A. thaliana*. Toutes les mutations empêchant l'envoi dans la mitochondrie ne permettaient pas de retrouver le phénotype nain associé à la surexpression de HopG1 pleine longueur. Cela suggère donc qu'au moins une des cibles de HopG1 se situe dans la mitochondrie. La surexpression de l'effecteur dans *N. tabacum* entraîne une importante accumulation de ROS –environ 4 fois plus qu'avec un plasmide vide- et une diminution de la consommation d'oxygène de moitié. Le mécanisme exact de HopG1 dans la mitochondrie n'est pas connu et les auteurs ne tentent pas d'hypothèse sur la ou les cibles putatives.

La mise en évidence d'effecteurs possédant des séquences d'adressage au chloroplaste et à la mitochondrie soulève des questions quant à l'origine de ces séquences chez des bactéries comme *Pseudomonas*. Proviennent-ils d'évènements de transferts horizontaux de gènes ou d'accumulations de mutations ? Le paragraphe suivant détaille deux familles d'effecteurs, connus pour posséder des motifs d'origine eucaryote autres que des séquences d'adressage.

1.3.6. Effecteurs à motif eucaryote

1.3.6.1. F-Box

Les protéines à motif F-box sont des protéines impliquées dans la dégradation des protéines par le protéasome 26S. Formant un complexe (nommé SCF, pour *SKP1* - *CUL1* - *F-box* du nom des protéines qui le composent), les F-box sont connues comme étant le facteur de spécificité pour la reconnaissance du substrat, qui va être ubiquitiné. Cette ubiquitination entraîne la reconnaissance par le protéasome 26S et, de fait, une dégradation ciblée des protéines (Lechner et al., 2006). Le motif F-box est composé de 50 acides aminés et est connu pour jouer un rôle dans les interactions protéine-protéine (Kipreos et Pagano, 2000). On retrouve des protéines de cette famille impliquées dans de nombreux mécanismes dans la plante, de la mise en place des racines latérales au voies de réponses de différentes hormones ou la reconnaissance du pollen par exemple (Lechner et al., 2006). Si pas loin de 700 protéines à motif F-box ont été prédites chez *Arabidopsis thaliana* (Lechner et al., 2006), ou encore 38 chez l'Homme, on en retrouve chez l'immense majorité des Eucaryotes, mais pas chez les Procaryotes (Kipreos et Pagano, 2000). Néanmoins, quelques cas de Procaryotes codant des protéines à motif F-box ont été mis en évidence, de même que des virus. Des gènes codant pour des protéines F-box ont été trouvés chez les bactéries pathogènes de plantes *Ralstonia solanacearum* et *Agrobacterium tumefaciens* mais aucun mécanisme clair n'a été mis en évidence, si ce n'est leur injection *in planta* pour certains d'entre eux (Lechner et al., 2006). L'effecteur F-box le plus étudié est la protéine P0, codée par deux phytovirus, le virus de la jaunisse occidentale de la betterave et le virus de la jaunisse des Cucurbitacées transmise par

pucceron (Pazhouhandeh et al., 2006). Les auteurs ont montré que P0 interagissait avec un homologue de SKP1, une sous-unité du SCF, et que les mutations ponctuelles empêchant cette interaction *in planta* facilitaient la mise en place par la plante de l'extinction de gènes utilisée lors de la défense face à un virus. L'hypothèse des auteurs est que P0 joue dans la plante le rôle d'une protéine à motif F-box pour cibler une protéine impliquée dans la dégradation des petits ARN en réponse à une attaque virale (Pazhouhandeh et al., 2006).

1.3.6.2. TAL

Les effecteurs TAL (pour *Transcription Activator-Like*, souvent nommés effecteurs de la famille *AvrBs3*), sont des protéines découvertes dans le genre *Xanthomonas* (Boch et Bonas, 2010). On trouve également des homologues des TAL chez les bactéries du genre *Ralstonia* (Heuer et al., 2007). Injectés via le système de sécrétion de type III, les protéines TAL sont envoyées dans le noyau de la plante infectée afin de moduler l'expression des gènes de celle-ci. Étudiée dès les années 80, cette famille d'effecteurs bactériens a été décrite à de nombreuses reprises, (Bonas et al., 1989; Boch et Bonas, 2010; Szurek et al., 2001; Boch et al., 2009). On trouve en partie N-terminale de ces protéines une séquence d'adressage au système de sécrétion de type III, et en partie C-terminale un à plusieurs motifs NLS (pour *Nuclear Localization Signal*) d'adressage au noyau, ainsi qu'un domaine AD (pour *Activation Domain*), permettant l'activation de la transcription. Entre ces deux se trouvent des motifs répétés en tandem permettant la fixation à l'ADN, dont le nombre peut varier de 1,5 à 33,5. Ces motifs contiennent en moyenne 34 acides aminés (aa), à l'exception du dernier qui correspond à un demi-motif d'environ 20 aa (Boch et Bonas, 2010).

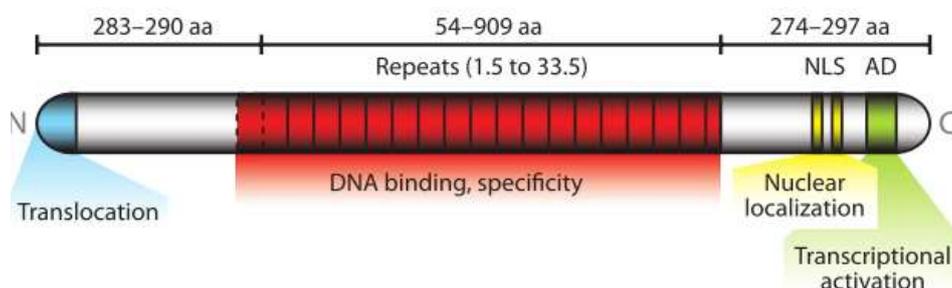


Figure 4: Domaines fonctionnels des effecteurs TAL de *Xanthomonas*.

Issu de Boch et Bonas (2010).

Dès 1992, Herbers et al. réussirent à démontrer que les motifs centraux des effecteurs TAL contrôlaient la spécificité de fixation à l'ADN. En 2009, fut proposé par deux équipes indépendantes un code permettant de prédire la zone fixée par les effecteurs TAL (Boch et al., 2009; Moscou et Bogdanove, 2009). La zone d'ADN fixée par les effecteurs TAL est appelée boîte UPA (pour *UP-regulated by AvrBs3*). Cette boîte UPA peut être localisée à des endroits différents de la boîte TATA, mais si les deux se trouvent à proximité, la gène stérique créée par la fixation d'un effecteur TAL peut modifier la zone d'initiation de la transcription. Bien que certains effecteurs TAL aient été mis en évidence comme régulant positivement la transcription de gènes comme ceux codant pour la protéine Bs3 par exemple, par l'effecteur *AvrBs3* (Romer et al., 2007), il reste très difficile de connaître le ou les gènes de la plante dont la transcription sera manipulée par les effecteurs, tant leur redondance fonctionnelle est importante (Boch et Bonas, 2010).

1.3.6.3. PPR

Les protéines à motifs PPR (pour *Pentatricopeptide Repeat*) sont des protéines présentes chez tous les Eucaryotes. Elles sont connues pour jouer un rôle dans la régulation de l'expression des gènes des organites. Bien qu'elles présentent de nombreuses similitudes, notamment en termes de structure, avec les protéines TPR (pour *Tetratricopeptide Repeat*, présentes chez les Eucaryotes et Procaryotes), les protéines contenant ces motifs n'ont pas du tout la même fonction. Là où les motifs TPR vont permettre l'interaction avec d'autres protéines, les motifs PPR permettent eux de fixer l'ARN. Selon Barkan et Small (2014), les motifs PPR proviendraient d'une divergence des motifs TPR, et non l'inverse, puisque les motifs TPR sont plus répandus dans le vivant. Toujours selon les auteurs, cette divergence TPR-PPR se serait produite très tôt chez les Eucaryotes. Cependant, Barkan et Small (2014) font également remarquer que des bactéries possèdent également des motifs PPR. Étonnamment, ces bactéries –on peut citer *Erwinia amylovora*, *Ralstonia solanacearum* ou encore *Legionella longbeachae* – sont toutes des pathogènes ou symbiotes d'Eucaryotes, suggérant que ces protéines ont plutôt été obtenues par transfert horizontal de gènes. Étant donné le rôle connu des protéines PPR chez les Eucaryotes, la question de leur rôle chez les bactéries demeure sans réponse. On peut avancer l'hypothèse que ces protéines PPR, à l'image des protéines TAL, sont des effecteurs bactériens injectés dans la plante lors de l'infection, pour moduler l'expression des gènes des organites. L'un des arguments forts en faveur de cette hypothèse est la démonstration de l'injection de la protéine PPR trouvée chez *Ralstonia solanacearum*, nommée RipL (Mukaihara et al., 2010). Aucune analyse fonctionnelle de cet effecteur n'a cependant été publiée à ce jour.

1.3.6.4. Hypothèse du transfert horizontal de gènes

Chez les effecteurs TAL, la présence de motifs eucaryotes comme les domaines AD ou NLS laisse perplexe. L'hypothèse proposée par Boch et Bonas (2010) est que les effecteurs TAL auraient évolué à partir d'un évènement de transfert horizontal de gène(s). On retrouve la même hypothèse concernant les protéines PPR (Barkan et Small, 2014; Schallenberg-Rüdinger et al., 2013). Il est habituellement admis que ces évènements de transferts horizontaux de gènes, entre plantes et bactéries par exemple, se font plus souvent lorsque ceux-ci sont associés plus ou moins intimement (Bock, 2010). Dans le cas de *Ralstonia solanacearum*, plusieurs travaux montrent que la bactérie peut être compétente –c'est-à-dire accepter de l'ADN étranger- lors de l'infection. Bertolla et al. (1999) montrèrent que lors de l'infection de *Solanum lycopersicum* par *Ralstonia solanacearum* GMI1000, la bactérie pouvait passer par un stade de compétence et accepter de nouveaux plasmides, venant d'autres bactéries ou présents dans le milieu. Des nuances sont cependant à apporter : Bertolla et al. (1997) n'ont pas réussi à intégrer de cette façon un ADN d'une autre espèce (*E. coli*, *B. cepacia* et *A. tumefaciens* avaient été testés). De nombreux facteurs résumés par Bertolla et Simonet (1999) font que l'insertion d'ADN d'autre espèces reste très limitée. Enfin, bien que proposant de nombreux critères afin de définir ce qu'est un transfert horizontal de gènes, Koonin et al. (2001) rappellent que toutes ces indications restent probabilistes, et que le cas le plus flagrant restera le transfert des gènes des organites endosymbiotiques –chloroplastes et mitochondries- vers le noyau.

Comme nous venons de le voir, les bactéries phytopathogènes possèdent de très nombreux effecteurs, dont certains possèdent des motifs eucaryotes. Une fois au contact de la plante, les bactéries utilisant un T3SS vont pouvoir injecter leurs effecteurs dans la plante afin de favoriser l'infection. Pour autant, la plante ne reste pas sans réaction. Ces réactions se déclenchent-elles seulement lorsque les effecteurs sont déjà dans la cellule végétale ou la plante peut-elle détecter la présence du pathogène avant ? Quelles sont ces réactions de défense de la plante ? Le paragraphe suivant a pour but de répondre à ces questions.

2. Interaction plante-pathogène vue par la plante

2.1. Reconnaissance du pathogène

La sédentarité de la plante l'oblige à faire face à ses agresseurs, sans possibilité de fuir. Elle a donc développé, au fil de l'évolution, de nombreux systèmes de défense, tant physiques que chimiques. Malgré le côté spectaculaire de l'infection de plantes par certains pathogènes, assez peu de végétaux sont infectés dans l'absolu. L'infection serait plutôt le cas particulier et non la norme, comme l'on pourrait le croire. L'interaction entre un pathogène dit compatible et une plante-hôte est appelée interaction homologue, ou interaction compatible. Ici, les défenses de la plante, basales ou induites, ne seront pas suffisantes pour empêcher le pathogène de se multiplier. Le pathogène sera dit virulent. Dans certains cas de figure, l'interaction peut devenir incompatible suite à la présence de gène(s) de résistance, alors même que les défenses basales n'empêchaient pas la multiplication du pathogène *in planta*. S'il s'agit d'une plante non-hôte, on parlera d'interaction hétérologue. A ce moment, les défenses préformées de la plante peuvent être suffisantes pour empêcher ou fortement ralentir la croissance du pathogène (Prell et Day, 2001). Cette information n'est cependant pas nécessairement vraie dans tous les cas. On peut citer par exemple l'interaction non-hôte entre *A. thaliana* et *E. amylovora* où lorsque l'on bloque la synthèse de protéines de la plante, la bactérie se multiplie énormément (Moreau et al., 2012). De manière générale, l'interaction non-hôte reste peu étudiée. Ces définitions sont résumées dans la figure 5.

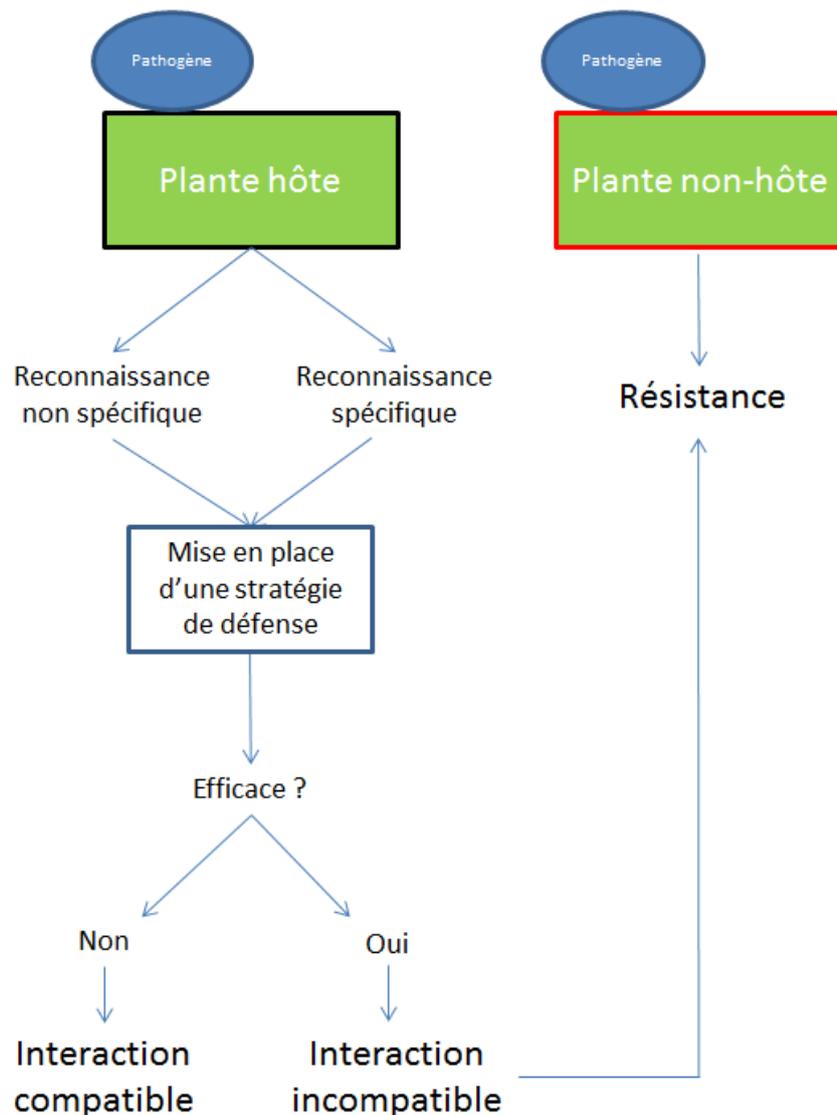


Figure 5: Représentation schématique des différents types d'interactions plante-pathogène.

En plus des défenses préformées de la plante, on peut trouver les défenses basales de la plante. Ces systèmes de défense sont extrêmement complexes, fonctionnent sur plusieurs strates qui peuvent être efficaces contre différents pathogènes (Bari et Jones, 2009). Les paragraphes suivants vont discuter dans les grandes lignes les différentes défenses de la plante en condition de stress biotique, jusqu'à arriver à la description du modèle général actuellement privilégié.

2.1.1. Défenses préformées

Les défenses préformées de la plante peuvent être de deux natures : physiques et chimiques.

Les premières barrières rencontrées par un pathogène seront d'ordre physique : ce sera l'écorce, dans le cas d'un arbre, la cuticule et la paroi pour la plupart des plantes. Ces barrières physiques peuvent être contournées par les pathogènes, par perforation (comme le font les insectes phytophages), ou par passage via une ouverture naturelle (lenticelles, stomates, nectarthodes, hydathodes) ou une blessure (Jones et Dangl, 2006; Bari et Jones, 2009).

Une fois le pathogène dans la plante, il fera face à une seconde barrière, cette fois-ci, de nature chimique. De nombreux métabolites secondaires ont été mis en évidence dans les plantes, sans qu'ils n'aient de rôle évident dans le métabolisme ou la physiologie de la plante. Depuis une cinquantaine d'années, un consensus a émergé, supposant que ces composés étaient impliqués dans la défense des plantes (Becerra, 2015). Certains vont avoir des propriétés antifongiques, antibactériennes voire insecticides. On peut citer entre autres les phytoanticipines (Pedras et Adio, 2008) qui sont synthétisées par la plante de manière constitutive et impliquées dans la défense de la plante contre les pathogènes. Une fois les défenses préformées contournées, le pathogène peut faire face à la défense non spécifique de la plante.

2.1.2. Mise en place de la défense non spécifique

A la différence des mammifères, les plantes ne possèdent pas de cellules de défense mobiles (Jones et Dangl, 2006). Elles doivent donc compter sur le système immunitaire propre à chacune de ses cellules, qui est tout de même capable de reconnaître le non-soi. Les structures reconnues par les plantes sont des motifs associés à des fonctions extrêmement importantes pour le pathogène, donc moins sujettes à mutation(s), et non présentes chez la plante. On parle alors de PAMP (pour *Pathogen-Associated Molecular Pattern*) ou MAMP (pour *Microbe-Associated Molecular Pattern*) dans un contexte plus général. Les PAMP peuvent être de la flagelline, protéine majoritaire du flagelle bactérien, des peptidoglycanes formant la paroi bactérienne, des lipopolysaccharides formant la membrane externe de bactéries Gram négatives ou encore la chitine chez les mycètes (Newman et al., 2013). La reconnaissance des PAMP par la plante se fait au niveau de la surface des cellules végétales, par des protéines PRR (pour *Pattern Recognition Receptor*). Les PRR sont des protéines transmembranaires possédant un domaine de reconnaissance des PAMP (Monaghan et Zipfel, 2012). La reconnaissance basale peut aussi être initiée par la reconnaissance de signaux dérivés de la plante. On parle de DAMP (pour *Damage-Associated Molecular Pattern*), lorsque par exemple des fragments de la paroi cellulaire sont libérés, suite à une attaque de pathogène ou une blessure. Ces DAMP sont alors reconnus par des protéines PRR et leur domaine transmembranaire (Jones et Dangl, 2006).

Cette reconnaissance des PAMP/DAMP par des protéines PRR entraîne de nombreuses modifications dans la cellule végétale : on parle alors de PTI (pour *PAMP-Triggered Immunity*). La PTI entraîne l'activation de voies de signalisation de protéines kinases, comme les *mitogen-associated* (MAPK). Ces dernières induisent une reprogrammation de la transcription dans le noyau afin de répondre à l'attaque. Les facteurs de transcription de la famille WRKY sont particulièrement impliqués à cette étape (Jixin Dong, Chunhong Chen, 2003). Les hormones végétales jouent également un rôle dans la réponse aux pathogènes, en modulant l'expression de gènes de défense par exemple. La cellule peut également produire des composés antimicrobiens. La PTI entraîne également une production accrue de ROS (pour *Reactive Oxygen Species*). Ces espèces dérivées de l'oxygène peuvent être du peroxyde d'hydrogène, de l'oxygène singulet ou encore du radical hydroxyde (Apel et Hirt, 2004). Ils sont générés par plusieurs compartiments cellulaires, dont la paroi, les photosystèmes I et II du chloroplaste, le peroxysome (photorespiration) et la chaîne respiratoire de la mitochondrie. Cette génération de ROS semble se faire en deux étapes, avec une première, de faible amplitude, puis une seconde, beaucoup plus importante qui semblerait être corrélée avec la résistance au pathogène (Torres et al., 2006). Les pathogènes réussissant à infecter les plantes avec succès sont capables de limiter la PTI à l'aide d'effecteurs (Block et Alfano, 2011), ou en modifiant la séquence et/ou la structure de certains PAMP. Pour ce second cas de figure, on peut citer l'exemple de la flagelline 22 de *Ralstonia solanacearum*, dont la séquence n'est pas reconnue par *Arabidopsis thaliana* (Pfund et al., 2004), bien qu'*Arabidopsis* soit capable de reconnaître la flagelline 22 de *Pseudomonas aeruginosa* par exemple (Suarez-Rodriguez et al., 2006). Concernant les effecteurs, même s'ils peuvent limiter ou bloquer la PTI –on parle alors d'ETS pour *Effector Triggered Susceptibility*- ils peuvent également être reconnus par la plante lors de l'infection et induire une seconde vague de défenses.

2.1.3. Mise en place de la défense spécifique

On parle de reconnaissance induite, ou d'immunité spécifique lorsque la plante est capable de reconnaître un effecteur de pathogène, de manière directe ou indirecte (via des produits de dégradation, une interaction avec le complexe effecteur-protéine ciblée, etc.). Cette reconnaissance se fait à l'aide d'une protéine codée par un gène appelé gène R (pour *Resistance gene*). En 1971, Flor proposa un modèle génétique pour expliquer ce type de résistance : le modèle gène-à-gène (Flor, 1971). Ce modèle propose que pour que la plante empêche l'établissement de la maladie, elle doit avoir dans son génome un gène de résistance R correspondant à un gène dit d'avirulence chez le pathogène, c'est-à-dire codant pour un effecteur qui sera reconnu par la protéine codée par le gène R. Selon van der Hoorn et Kamoun (2008), les protéines d'avirulence semblent également contribuer à la virulence du pathogène lors de l'absence du gène de résistance associé, expliquant leur conservation d'un point de vue évolutif chez les pathogènes en possédant. L'interaction entre les deux protéines peut être directe ou indirecte, la protéine R pouvant aussi reconnaître la cible affectée par l'effecteur (Jones et Dangl, 2006). Cette surveillance indirecte de l'intégrité des cibles d'effecteurs est appelée modèle de garde. Proposé par Van der Biezen et Jones en 1998, l'idée fut généralisée par Jones et Dangl en 2006 (Van Der Biezen et Jones, 1998; Jones et Dangl, 2006). A partir du modèle de Flor, de nombreuses interactions physiques entre effecteurs d'avirulence et protéines de résistance furent recherchées. Quelques-unes furent trouvées (Jia et al., 2000), mais de très nombreuses furent impossibles à mettre en évidence. On peut citer l'exemple de RIN4, une protéine transmembranaire d'*Arabidopsis*. Lors de l'infection par *Pst*, RIN4 est ciblée par deux effecteurs, AvrRpm1 et AvrB, qui vont phosphoryler leur cible. Cette modification de la protéine va activer une voie de défense via la protéine de résistance RPM1 et entraîner un mécanisme de défense de la plante, l'ETI (pour *Effector Triggered Immunity*). Un troisième effecteur de *Pst*, AvrRpt2, va cibler RIN4 et cliver la protéine en 3. Dans ce second cas, c'est une seconde protéine de résistance, RPS2, qui va entraîner une réponse de la plante suite au clivage de RIN4 (résumé par Jones et Dangl, 2006).

Au sein d'une même population végétale, il est possible d'avoir un polymorphisme des gènes R, c'est-à-dire des plantes possédant le gène, alors que d'autres non. Dans ce cas de figure et d'un point de vue évolutif, on peut dire que les protéines gardées par les effecteurs sont dans une situation délicate. D'un côté, en l'absence du gène R, la sélection naturelle devrait favoriser une protéine gardée dont l'interaction avec l'effecteur est fragilisée. Tout le contraire d'une plante possédant le gène R associé, où la protéine gardée a tout intérêt à conserver et même à favoriser l'interaction avec l'effecteur afin de favoriser la mise en place de défenses adaptées (van der Hoorn et Kamoun, 2008). Selon les auteurs, afin de débloquer cette situation potentiellement instable évolutivement parlant, les plantes auraient développé des protéines spécialisées dans la reconnaissance des effecteurs : des leurres. On parle alors de modèle du leurre (de l'anglais *Decoy Model*). Obtenues probablement suite à la duplication de gènes codant pour des protéines ciblées, ces protéines leurres ne jouent pas d'autre rôle connu dans la cellule végétale. La théorie de van der Hoorn et Kamoun (2008) est résumée dans la figure 6.

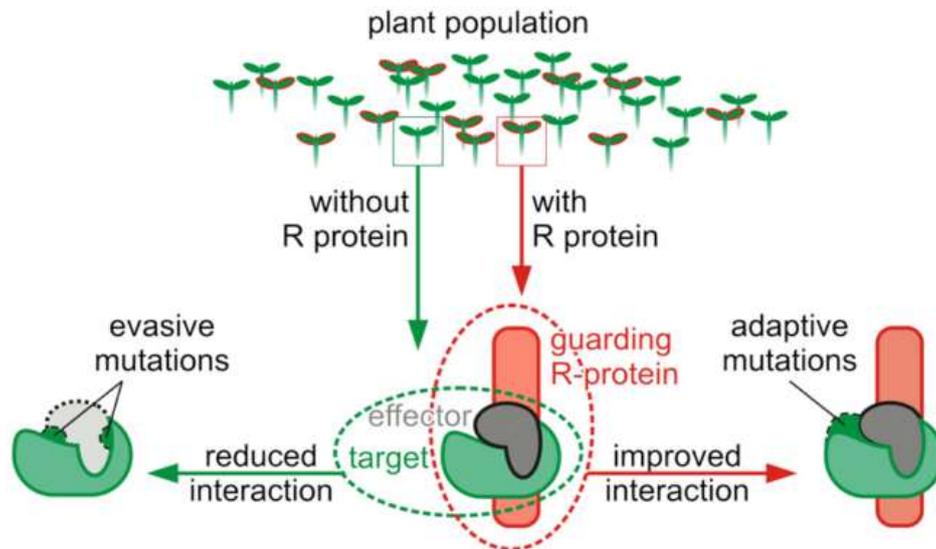


Figure 6: Représentation schématique du modèle proposé par Van der Hoorn et Kamoun (2008).

Tiré de Van der Hoorn et Kamoun (2008).

La reconnaissance d'un ou plusieurs effecteurs par la plante va entraîner l'ETI. La principale différence avec la PTI va être l'intensité de la réponse et sa vitesse (Jones et Dangl, 2006). L'ETI se termine parfois par une réponse hypersensible (ou HR, pour *Hypersensitive Response*) sur le site d'infection. Les auteurs font néanmoins remarquer que la HR n'est pas systématique ni requise pour l'ETI, et que le mécanisme exact bloquant la croissance des pathogènes reste flou, probablement une accumulation des nombreux effets négatifs dus aux défenses mises en place.

On sait cependant que l'activation de la PTI ou de l'ETI va très nettement limiter la croissance du pathogène, en augmentant la résistance de la plante (Bari et Jones, 2009). La chronologie de reconnaissance du pathogène couplé à une réponse rapide et adaptée semblent être la clé entre sensibilité et résistance. En se basant sur les différentes réponses et contre-réponses de la plante et du pathogène lors de l'interaction, Jones et Dangl (2006) proposèrent un modèle général que nous allons détailler dans le paragraphe qui suit (Jones et Dangl, 2006).

2.1.4. Modèle en zigzag

L'interaction plante/pathogène peut s'imaginer d'un point de vue évolutif comme les relations diplomatiques entre les Etats-Unis et l'ex-URSS après la Seconde Guerre Mondiale : une véritable course à l'armement, où les structures ayant obtenu les meilleurs systèmes d'attaque ou de défense seront sélectionnés.

Le système immunitaire des plantes se résume donc en deux parties : la première partie va reconnaître des molécules communes chez les pathogènes et y répondre (PTI) alors que la seconde va reconnaître des facteurs de virulence du pathogène, de façon directe ou indirecte, et y répondre de façon beaucoup plus marquée et rapide (ETI). L'ensemble de ces détections/réactions a été modélisé par Jones et Dangl en 2006, selon un système en 4 étapes appelé modèle en « zigzag » (Figure 7).

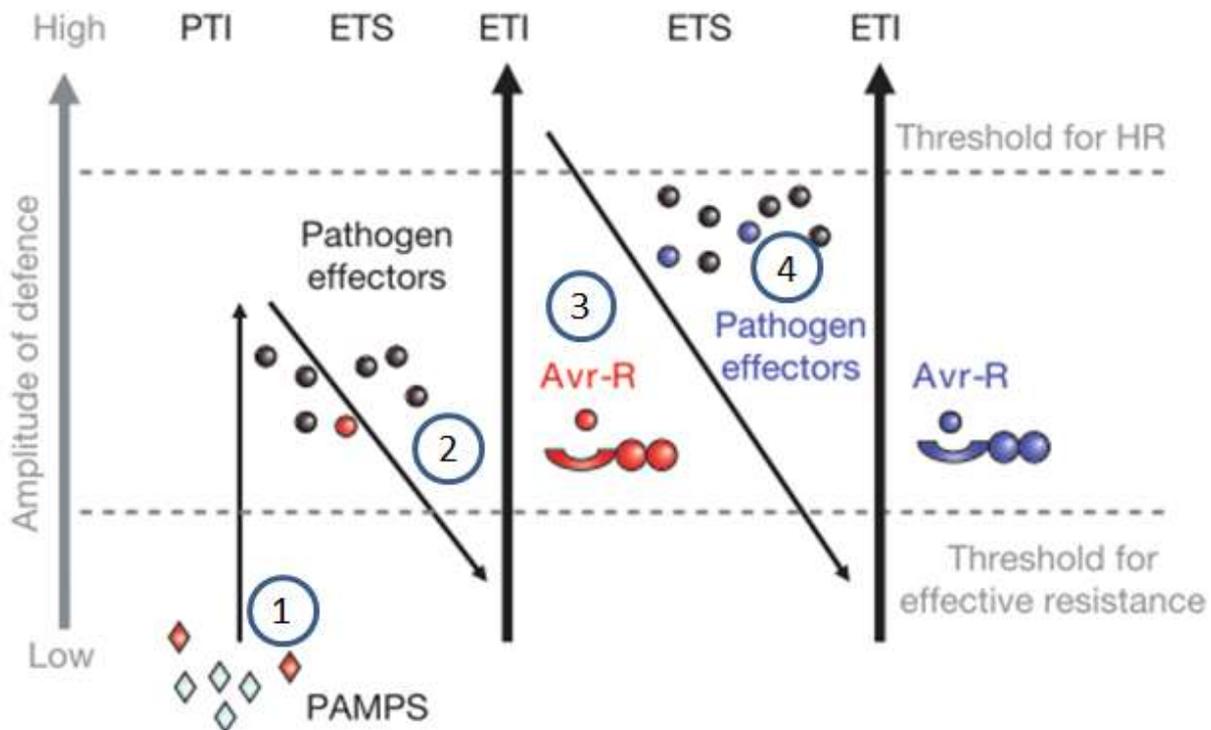


Figure 7 : Représentation du modèle en « zigzag » proposé par Jones et Dangl (2006).

Tiré de Jones et Dangl (2006).

L'étape 1 consiste en la reconnaissance des PAMP/MAMP par les protéines PRR de la plante, déclenchant la PTI. L'étape 2 ne concerne que les pathogènes ayant pu déployer leurs effecteurs avec succès. Ces effecteurs peuvent limiter la PTI et amener à l'ETS. L'étape 3 sera la reconnaissance d'un ou plusieurs effecteurs par des protéines R, entraînant l'ETI. Enfin, l'étape 4 met en évidence l'action de la sélection naturelle chez les pathogènes afin d'éviter le déclenchement de l'ETI, par le gain de nouveaux effecteurs par exemple. Cette stratégie d'évitement peut se faire par mutation(s) dans la séquence des effecteurs ou par l'acquisition de nouveaux effecteurs bloquant l'ETI (Jones et Dangl, 2006).

Un certain nombre de familles de molécules ont été mis en évidence comme jouant un rôle clé dans la reconnaissance du pathogène, la signalisation ou encore la réponse de la plante. Le paragraphe suivant va permettre d'expliquer plus en détail certaines familles de molécules.

2.2. Principaux acteurs de la défense des plantes

2.2.1. Les gènes R

Les gènes R (pour *Resistance genes*) sont, comme nous l'avons vu, les gènes codant pour des protéines de la plante permettant de détecter les effecteurs des pathogènes pour ensuite entraîner l'ETI. La plupart de ces gènes codent pour des protéines faisant partie de la famille des NB-LRR (pour *Nucleotide Binding Leucine Rich Repeat*) et tiennent leur nom de la présence de ces deux domaines (Jones et Dangl, 2006). Les gènes codant pour des NB-LRR sont très nombreux chez les plantes et relativement groupés sur le génome, suggérant qu'ils sont obtenus lors d'évènements de duplication (Marone et al., 2013). On en dénombre 149 par exemple chez *A. thaliana* (Meyers et al., 2003).

2.2.2. Les protéines kinases

Suite à la détection d'un pathogène, il est crucial pour la plante d'envoyer un message au noyau afin d'adapter ses réponses et se défendre. Parmi les voies de signalisation, on trouve les phosphorylations en cascade des MAPK (pour *Mitogen Activated Protein Kinase*). Le modèle minimum de fonctionnement des cascades de phosphorylation est le suivant : une MAPK kinase kinase (ou MAPKKK) va phosphoryler une MAPK kinase (MPAKK) qui elle-même va phosphoryler une MAPK. Cette dernière va ensuite activer une ou plusieurs autres protéines en aval, comme par exemple des facteurs de transcription, tels que les WRKY (Pitzschke et al., 2009). Cette réorganisation de la transcription nucléaire a pour but d'activer les défenses de la plante, avec pour conséquences un renforcement de la paroi ou encore une production de composés antimicrobiens. A noter que chez *Arabidopsis thaliana*, 110 gènes codent pour des composants au moins putatifs des voies de signalisation MAPK, avec environ 20 MAPK, 10 MAPKK et 80 MAPKKK (Pitzschke et al., 2009). La taille de cette famille de protéines suggère l'importance pour la survie de la plante de la transduction du signal d'un stimulus extérieur vers le noyau. Le fait que de nombreux effecteurs de type III de *Pseudomonas* ciblent des MAPK contribue aussi à cette hypothèse (Xin et He, 2013).

2.2.3. Les hormones végétales

Trois hormones végétales sont connues pour jouer un rôle majeur dans la réponse à un stress biotique : l'acide salicylique (SA), l'acide jasmonique (JA) et l'éthylène (ET). De manière générale, SA joue un rôle crucial dans l'activation des défenses de la plante face à des pathogènes biotrophes et hémibiotrophes, alors que JA et ET ont été mis en évidence comme étant associés à des mécanismes de lutte contre les pathogènes nécrotrophes et les insectes herbivores (Bari et Jones, 2009). L'antagonisme existant entre ces deux voies de défense suggère que la réponse de la plante est adaptée au pathogène et à sa stratégie d'infection (Bari et Jones, 2009). En dehors de la réponse à un stress biotique, ces hormones jouent d'autres rôles dans la plante. Par exemple, l'acide jasmonique est impliqué dans la germination (Ranjan et Lewak, 1992) ou l'ouverture des stomates (Popova et al., 1988). L'existence de différentes formes conjuguées fait que l'on parle généralement de jasmonates au sens large. Leur synthèse se fait par la voie des octadécanoïdes, ce sont des dérivés d'acides gras oxygénés. Les étapes initiales de biosynthèse se déroulent dans le chloroplaste, pour se terminer dans le peroxyosome (Creelman et Mullet, 1997). Pour la transduction du signal, on peut noter l'importance de deux protéines, COI1 (*Coronotiv Insensitive 1*) et JAR1 (*Jasmonate Resistant 1*), qui permettraient de conjuguer l'isoleucine à l'acide jasmonique afin de lui donner une forme perceptible par la plante (Thines et al., 2007). JA va induire l'expression de gènes de défense, comme les gènes codant pour des protéines impliquées dans la biosynthèse de métabolites secondaires ou de défensines à l'image de PDF1.2 (Manners et al., 1998).

Concernant l'acide salicylique, il existe deux voies de synthèse : la voie des phénylpropanoïdes, à partir de la phénylalanine, ou à partir du chorismate à travers la voie du shikimate, qui est celle produisant le SA associé au stress biotique (Loake et Grant, 2007). Il est à noter que cette seconde voie se situe pour partie dans le stroma du chloroplaste (Strawn et al., 2007), suggérant des échanges entre le noyau et le chloroplaste en condition de stress biotique. L'une des protéines clés dans la voie de signalisation SA est NPR1 (*Non-expressor of PR genes*). Cette protéine est connue pour interagir avec des facteurs de transcription, entraînant l'expression de gènes sous dépendance de SA et codant pour des protéines dites PR (pour *Pathogenesis Related*) impliquées dans la défense de la plante.

Enfin, l'éthylène est une molécule gazeuse, impliquée dans d'autres mécanismes de la plante, comme la maturation des fruits par exemple (Giovannoni, 2004). Sa biosynthèse est initiée à partir de dérivés de la méthionine (Yang et Hoffman, 1984). Un certain nombre de récepteurs à l'éthylène sont connus chez *Arabidopsis thaliana*. On peut citer par exemple ERF1 (pour *Ethylene Response Factor 1*) qui est un

régulateur positif des voies ET et JA, et de manière générale les membres de la famille ERF (Bari et Jones, 2009).

D'autres hormones sont impliquées dans la signalisation et la réponse à un stress biotique. Sans entrer dans les détails, on peut citer l'auxine, l'acide abscissique, les gibbérellines, les brassinostéroïdes ou les cytokinines. Pour plus d'informations sur les rôles de toutes ces hormones en condition de stress biotique, Bari et Jones (2009) proposent un bon résumé des connaissances sur ce sujet.

2.2.4. Les ROS

Les formes activées de l'oxygène (ou ROS, pour *Reactive Oxygen Species*) sont des dérivés de l'oxygène hautement instables. Cette instabilité provient de la présence d'un électron de valence non apparié (Apel et Hirt, 2004). Les ROS peuvent donc être toxiques pour la cellule en endommageant l'ADN, les lipides ou encore les protéines. Chez les plantes, 4 formes de ROS sont prédominantes : l'oxygène singulet $^1\text{O}_2$, l'anion superoxyde O_2^- , le peroxyde d'hydrogène H_2O_2 et pour terminer le plus réactif, le radical hydroxyle HO \cdot (Waszczak et al., 2018). Étant donné leur caractère instable et leur effet délétère pour la cellule, leur niveau est finement régulé par la plante lorsque les conditions environnementales sont stables. Les ROS peuvent être produits dans à peu près tous les compartiments cellulaires de la plante et, à l'exception de l'apoplaste qui a une faible capacité antioxydante, chaque compartiment est équipé pour détoxifier les ROS (Waszczak et al., 2018). Les principaux compartiments producteurs de ROS dans la cellule sont le chloroplaste, la mitochondrie, le peroxysome mais également la membrane plasmique via la NADPH oxydase (Apel et Hirt, 2004). Doke (1985) est le premier à mettre en évidence le rôle des ROS au service de la plante, démontrant la mise en place par le tubercule de *Solanum tuberosum* d'une très forte augmentation de la production de ROS (en anglais *oxidative burst*) en réponse à une espèce avirulente de *Phytophthora infestans*. Ce "burst" oxydatif a lieu en deux temps face à un pathogène avirulent, avec une seconde production de ROS beaucoup plus marquée, semblant coïncider avec les mises en place de la PTI puis de l'ETI (Levine et al., 1994). Les ROS semblent également jouer le rôle de messagers en condition de stress. En effet, Apel et Hirt (2004) font remarquer que la production de ROS dans les organites entraîne des modifications du transcriptome nucléaire de la plante. Ces modifications transcriptomiques par les ROS passent probablement via des cascades de signalisation, une oxydation directe des composants de voies de signalisation par les ROS ou encore en ciblant directement des facteurs de transcription (Apel et Hirt, 2004). Malgré leur implication dans la réponse hypersensible, Kangasjärvi et al. (2012) ne pensent pas que les ROS fonctionnent comme des agents qui endommageraient directement la cellule, mais plutôt comme des éléments déclencheurs d'un mécanisme de mort cellulaire programmée. La relation entre les ROS, la reprogrammation nucléaire et la réponse hypersensible n'est pas encore entièrement comprise, mais semble être spécifique au pathogène (Kangasjärvi et al., 2012).

Comme nous venons de le voir avec les ROS, les organites peuvent aussi être impliqués en condition de stress biotique. En effet, le chloroplaste est ciblé par des effecteurs de type III comme HopN1 ou encore HopI1 (Xin et He, 2013) de *Pst*, suggérant que le pathogène a un intérêt à cibler des protéines voire, de manière plus générale, des mécanismes s'y déroulant (de Torres Zabala et al., 2015). Il en va de même avec la mitochondrie avec l'effecteur HopG1 de *Pst*. Le chloroplaste est également impliqué dans les voies de synthèse de SA et JA. Pour rappel, le chloroplaste et la mitochondrie sont les « centrales énergétiques » de la cellule végétale, le premier via la photosynthèse et la production de sucre et de dioxygène à partir de dioxyde de carbone, d'eau et de lumière ; le second via la respiration, convertissant l'énergie des nutriments en énergie utilisable directement par la cellule. Les paragraphes suivants vont détailler les différents rôles des organites dans les mécanismes de défense des plantes.

2.3. Rôle des organites dans la défense

2.3.1. Cas du chloroplaste

Le chloroplaste revêt une importance particulière dans la cellule végétale car il est le principal moteur de la croissance via la photosynthèse. Il est également le siège d'étapes cruciales de la biosynthèse de lipides, métabolites secondaires, vitamines ou acides aminés. Nous avons vu que les voies de synthèses de SA et JA, impliquées dans les mises en place de défenses se trouvaient entièrement ou pour partie dans cet organite, et que ces voies étaient modulées par la plante –et parfois par le pathogène- en réponse à une infection. Les organites, mitochondries et chloroplastes, utilisent des molécules informant le noyau de leur état physiologique –par exemple une diminution de l'activité photosynthétique- afin de moduler l'expression des gènes nucléaires et mettre en place une réponse adaptée ; on appelle cette voie de signalisation « le signal rétrograde » (Kleine et Leister, 2016). Parmi ces messagers, les ROS sont les plus connus. Selon Joo et al. (2005), le chloroplaste serait capable de percevoir les ROS apoplastiques et même d'amplifier le signal. Le mécanisme reliant les ROS et la réponse hypersensible demeure pour l'instant inconnu. La lumière revêt une importance toute particulière lors de l'infection par un pathogène, en modifiant la production de ROS, le niveau énergétique et le pouvoir réducteur de la cellule (Kangasjärvi et al., 2012).

2.3.2. Cas de la mitochondrie

Une cellule parenchymateuse infectée va passer de source à puits, suite à la diminution de l'activité photosynthétique. C'est à ce moment particulier que la respiration mitochondriale va jouer un rôle d'autant plus important, afin de fournir à la cellule une partie de ses besoins en énergie, l'autre partie provenant d'une augmentation de l'activité des transporteurs d'hexoses et de la voie des pentoses phosphates (Kangasjärvi et al., 2012). Mais ce rôle ne semble pas être le seul lors d'un stress biotique. A l'image du chloroplaste, et en se basant sur des études de l'apoptose chez les animaux, Amirsadeghi et al. (2007) font l'hypothèse que la mitochondrie joue aussi un rôle dans la perception du pathogène et la réponse hypersensible. Sur des cultures cellulaires d'*Arabidopsis*, Krause et Durner (2004) ont effectué un traitement avec des harpines purifiées de *Pst* (un facteur de virulence) afin d'observer l'effet sur la mitochondrie. Ils montrèrent qu'une partie du « *burst* » oxydatif provenait de la mitochondrie, via la chaîne de transport des électrons, et qu'il était associé à une diminution du niveau cellulaire d'ATP. Ces éléments semblent se passer peu de temps avant la réponse hypersensible (Amirsadeghi et al., 2007). La mitochondrie serait également capable, comme le chloroplaste, de percevoir certains ROS apoplastiques pour ensuite activer, via des cascades de signalisation de MAPK, une réorganisation de l'expression génétique nucléaire (Amirsadeghi et al., 2007). La capacité de détoxification de la mitochondrie semble aussi être modulée en réponse aux pathogènes. Une augmentation de la détoxification des ions superoxyde couplée à une diminution de la détoxification du peroxyde d'hydrogène a été observée en réponse à *Botrytis cinerea* (Kuzniak et Skłodowska, 2004), suggérant une accumulation spécifique de certains ROS. Plusieurs études ont mis en évidence une augmentation de l'expression d'AOX (Alternative Oxydase, qui permet entre autres de réduire l'O₂ en H₂O) suite à l'augmentation du niveau de SA cellulaire (Kuzniak et Skłodowska, 2004). Comme nous l'avons vu précédemment, l'acide salicylique est principalement impliqué dans la réponse aux pathogènes biotrophes et hémibiotrophes et AOX va permettre de limiter pour partie l'accumulation de ROS, qui joueraient le rôle de messagers dans l'activation d'une réponse hypersensible. Est-ce qu'AOX retarde ou limite la réponse hypersensible ou permet-elle face à *Botrytis cinerea* de maintenir un niveau énergétique suffisant dans la cellule afin de mettre en place des défenses ?

L'adaptation des réponses des organites aux différents stress biotiques explique pour partie la complexité des mécanismes observés. Cette complexité augmente d'autant plus lorsqu'on se rend compte que les protéines des organites sont codées pour partie par le noyau, le reste par leur propre génome. Comment s'organise

l'expression des gènes des organites ? Est-elle régulée, et si oui, l'est-elle en condition de stress biotique ? Le paragraphe suivant résume les connaissances à ce jour et propose des éléments de réponse.

3. Régulation de l'expression génétique des organites pendant un stress biotique

Avant de s'intéresser à l'expression des gènes des organites lors de stress biotique, je vais dans un premier temps décrire l'organisation des génomes chloroplastiques et mitochondriaux d'*Arabidopsis*, puis décrire le fonctionnement un peu particulier de leur expression génétique.

3.1. Génomique des organites chez *Arabidopsis thaliana*

L'origine des mitochondries et des chloroplastes a longtemps été discutée par la communauté scientifique. Dès la fin du XIX^{ème} siècle, le botaniste allemand Andreas Schimper supposait que la plante verte pouvait être l'union d'un organisme incolore avec un microbe contenant des pigments verts. Cette hypothèse se heurta à de nombreux écueils. Tout d'abord, il s'avéra impossible de cultiver des chloroplastes et des mitochondries isolés, entrant en opposition avec l'un des postulats de Koch. Un second biais de perception joua contre la théorie de Schimper : les premiers microorganismes étudiés au début du XX^{ème} siècle étaient pour l'immense majorité d'entre eux des pathogènes. La communauté scientifique ne comprenait pas ce que des pathogènes pouvaient offrir à une cellule eucaryote. Dans les années 1970, les travaux –et l'acharnement (Margulis et al. (2012) pour la petite histoire)- de Lynn Margulis relancèrent cette ancienne hypothèse, appelée « théorie endosymbiotique » (Sagan, 1967). Selon son hypothèse, la mitochondrie descendrait d'une bactérie endosymbiotique capable d'effectuer la respiration et le chloroplaste d'une cyanobactérie phagocytée par un protozoaire. Sur la même période, des recherches mirent en évidence que les organites sont isolés du reste de la cellule par plusieurs membranes et qu'ils contiennent de l'ADN (Hagemann, 2010). Bien que cette hypothèse soit aujourd'hui très largement acceptée, certains points sont toujours en discussion comme la date des deux événements endosymbiotiques ou encore l'embranchement monophylétique de la cyanobactérie à l'origine des chloroplastes. Depuis les années 2000 et l'avènement du séquençage, de nombreux génomes d'organites ont été séquencés. Même si ceux-ci présentent des tailles et un nombre de gènes divers et variés, un trait commun semble tous les réunir : la réduction du matériel génétique suite à l'endosymbiose. Cette réduction peut résulter d'une perte ou d'un transfert génétique vers le noyau (Timmis et al., 2004).

Les paragraphes suivants vont détailler l'organisation des génomes chloroplastiques et mitochondriaux. Je fais le choix de me focaliser sur *Arabidopsis thaliana* pour 3 raisons : simplification de mon propos, le fait que ça soit un organisme modèle très répandu et également la principale plante étudiée lors de ma thèse.

Ces deux génomes présentent des caractéristiques communes, à savoir être composés d'une molécule d'ADN double brin, avec plusieurs copies par organites et la présence d'introns. Les deux génomes codent pour une partie de la machinerie transcriptionnelle et traductionnelle des organites, leur conférant un caractère semi-autonome, et la majorité des gènes sont organisés en pseudo-opérons. En comparaison avec les bactéries, on assiste à une simplification relative des génomes des organites, tout du moins pour la taille, accompagnée d'une complexification dans leur fonctionnement. Klaff (1991) montra que la demi-vie d'un ARNm chloroplastique pouvait varier de 6 à 40h, contre 20s à quelques minutes pour les ARNm d'*Escherichia coli* (Klaff, 1991). Les paragraphes suivants vont détailler les spécificités d'organisation des deux génomes des organites d'*Arabidopsis thaliana*.

3.1.1. Génome chloroplastique

Le génome chloroplastique d'*Arabidopsis thaliana* écotype *Columbia* a une taille de 154478 paires de bases et comporte 127 gènes. Bien que l'ADN chloroplastique soit régulièrement représenté comme étant circulaire, il est connu que seule une minorité de l'ADN chloroplastique l'est réellement, la majorité étant linéaire

(Bendich, 2004). Les gènes chloroplastiques codant pour des protéines appartiennent à 3 catégories : les gènes codant pour des protéines impliquées dans l'appareil photosynthétique (Figure 8), transcriptionnel (rpo, ARNt) et traductionnel (ARNr). Seul clpP, gène codant pour une sous-unité de protéase (Shikanai et al., 2001) n'appartient à aucune de ces catégories. Le génome chloroplastique d'*Arabidopsis* comporte également 2 régions inversées répétées, séparées par deux régions, une grande de 84170pb et une plus petite de 17780pb (Sato et al., 1999).

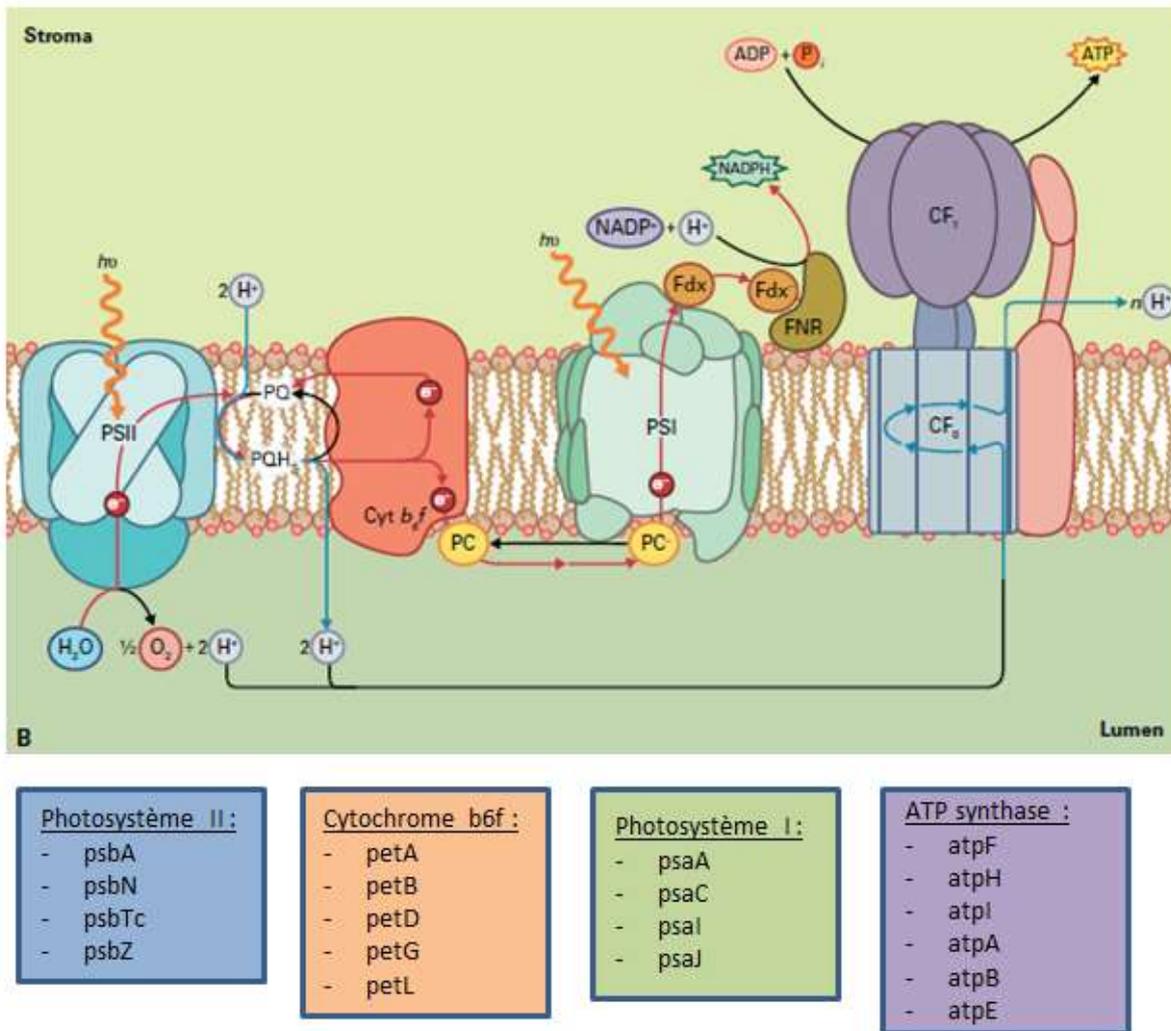


Figure 8 : Représentation schématique de la chaîne de transport des électrons de la membrane thylakoïdienne.

Adapté de Buchanan et al. (2015). Sont représentés sous les complexes protéiques, les gènes chloroplastiques codant pour certaines des sous-unités. PSII : photosystème II ; PQ : plastoquinone ; cyt b6f : cytochrome b6f ; PC : plastocyanine ; PSI : photosystème I ; FDX : Ferrédoxine ; FNR : Ferrédoxine NADP+ Réductase.

3.1.2. Génome mitochondrial

Chez *Arabidopsis thaliana* écotype C24, le génome mitochondrial a une taille de 366750 paires de bases, soit plus de deux fois la taille du génome chloroplastique. Pour autant, ses 60 gènes sont répartis sur 10% à 20% du génome seulement, suggérant la présence de nombreuses zones sans fonction connue. A cela se rajoutent 8% environ d'introns parmi les 60 gènes, 7% de duplications, 4% de retrotransposons d'origine nucléaire et 1% de séquence originaire des plastes (Unsel et al., 1997). La communauté *Arabidopsis* s'est longtemps basé

sur ce génome, jusqu'aux travaux de Davila et al. (2011), puis de Sloan et al. (2018) sur l'écotype Col0. Un certain nombre de SNP (pour *Single Nucleotide Polymorphism*) ont été mis en évidence, mais seulement deux dans des ARNr et 15 dans des introns de gènes codant pour des protéines. Toutes les autres différences étaient intergéniques. Le génome comporte 60 gènes, dont 24 codants pour des complexes/cytochromes de la chaîne respiratoire (Figure 9), 9 protéines ribosomales, 9 ARNr, le reste codant pour des ARNt. Contrairement aux idées reçues, les génomes mitochondriaux sont variables entre les différentes espèces. Chez l'Homme par exemple, le génome mitochondrial fait 16569 paires de bases selon Anderson *et coll.* (1981), soit une taille 20 fois moins importante que celui de la mitochondrie d'*Arabidopsis* (Anderson et al., 1981). Pourtant, le génome mitochondrial humain est tout de même composé de 37 gènes, suggérant une importante complexité dans l'organisation du génome mitochondrial d'*Arabidopsis thaliana*. André et al. (1992) suggèrent que cette complexité serait due à de très nombreux événements de recombinaison.

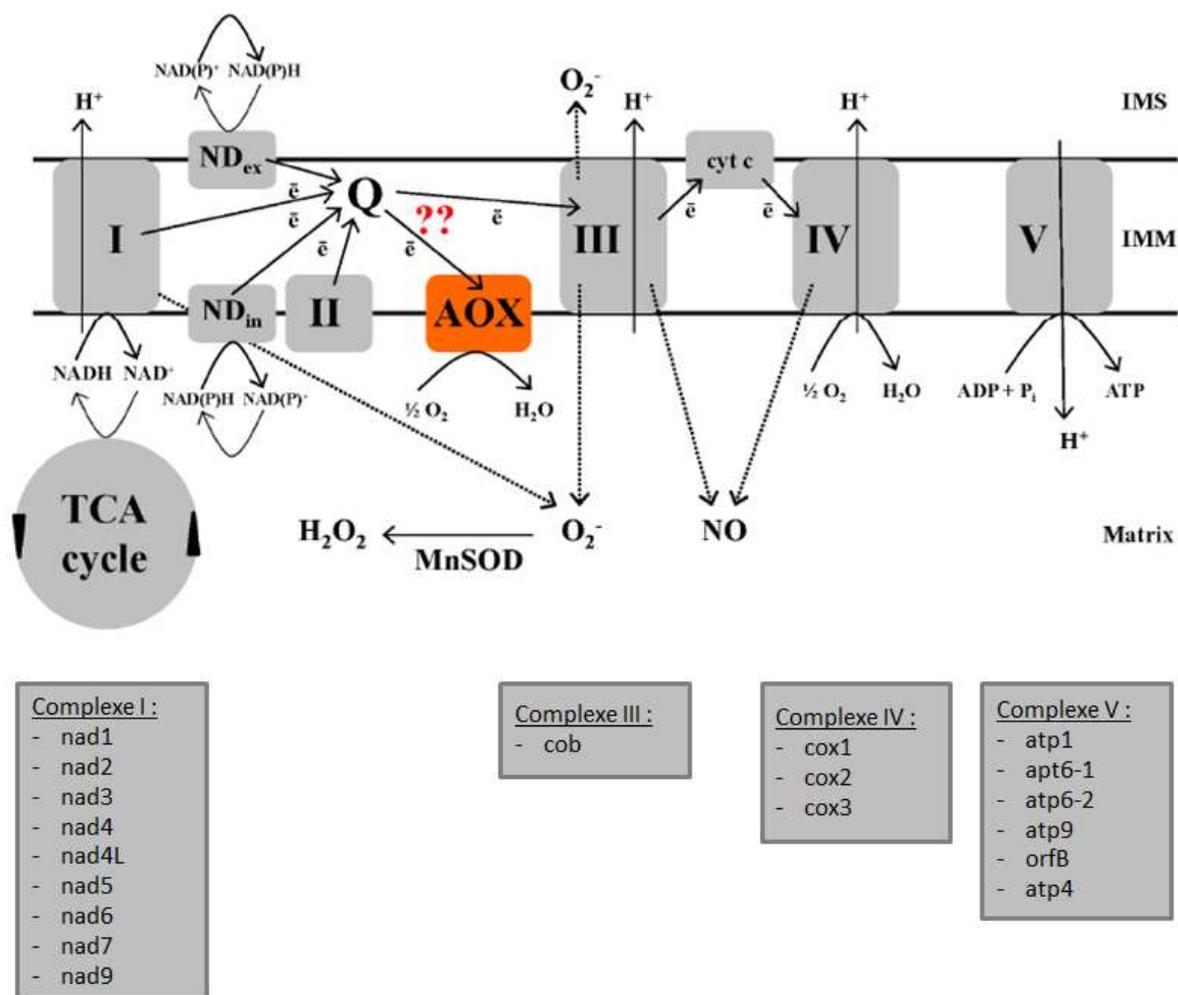


Figure 9 : Représentation schématique de la chaîne de transport des électrons de la mitochondrie.

Adapté de Vanlerberghe (2013). Sont représentés sous les complexes protéiques les gènes mitochondriaux codant pour certaines des sous-unités. IMM : membrane interne de la mitochondrie (*Inner Membrane Mitochondria*) ; Matrix : Matrice ; IMS : espace intra-membranaire (*Inner Membrane Space*) ; Q : ubiquinone ; TCA cycle : cycle de Krebs (*Tricarboxylic Acid cycle*).

A noter que chez les plantes, le transport des électrons peut se faire de façon linéaire, du complexe I ou II au complexe IV, ou de façon alternative via AOX, l'alternative oxydase qui permet l'élimination du surplus

d'électrons. La constitution du gradient électrique de part et d'autre de la membrane interne de la mitochondrie permet de produire de l'ATP via le complexe V ATP synthase.

3.2. Expression génétique chez les organites d'*Arabidopsis thaliana*

3.2.1. Transcription

La transcription dans les deux organites des plantes présente de fortes similitudes, et de grandes disparités avec la transcription nucléaire. On retrouve chez le plaste et la mitochondrie une organisation des gènes la plupart du temps sous forme de pseudo-opérons, c'est-à-dire plusieurs gènes sous la dépendance d'un même promoteur, organisation héritée d'un ancêtre bactérien. Les ARNm sont donc pour la plupart des polycistrons. A la différence de chez leurs ancêtres, les ARN des organites sont soumis à de très nombreuses et très complexes étapes de maturation avant toute traduction, comme de l'épissage ou de l'édition (Small et al., 2013). On trouve aussi des disparités entre le chloroplaste et la mitochondrie, que l'on va détailler.

La mitochondrie ne code pas pour sa propre ARN polymérase (Liere et al., 2011). De ce fait, elle est entièrement dépendante des ARN polymérases codées par le noyau, ou NEP (pour *Nuclear-Encoded RNA polymérase*) que je détaillerai dans le paragraphe suivant (RPOTM et RPOTMP). On peut noter qu'il existe de très nombreux sites d'initiation de la transcription (ou TSS, pour *Transcription Start Sites*) chez la mitochondrie, ce qui laisse supposer que leur nombre permettrait d'assurer la transcription malgré d'éventuels réarrangements génomiques (Weihe et al., 2012). D'autres résultats font même remarquer que les TSS ne se trouvent pas systématiquement en amont d'un gène, suggérant la présence d'ARN non fonctionnels (Kuhn et al., 2009).

De manière similaire à la mitochondrie, on va retrouver dans le chloroplaste les NEP, avec RPOTMP, co-localisée dans les deux organites, et RPOTP, qui va être spécifique aux plastes. Il existe également un autre complexe de transcription, et celui-ci est encodé directement par le chloroplaste. On parle de PEP (pour *Plastid Encoded RNA-polymerase*). Il est intéressant de noter que la quasi-totalité du génome chloroplastique est transcrit en ARN (Zhelyazkova et al., 2012), suggérant que la régulation transcriptionnelle est limitée. De la même façon que dans la mitochondrie, on trouve de nombreux sites d'initiation de la transcription dans le génome chloroplastique. On peut citer en exemple l'opéron *psbB* dont la transcription a été décrite par Zghidi-Abouzid et al. (2011). Outre la présence de sites d'initiation de la transcription devant chaque gène de l'opéron, on trouve également des TSS à l'intérieur même du gène codant pour *psbB*.

3.2.2. Maturation 5' et 3' UTR des ARN

Dans les organites des plantes, la régulation de l'expression des gènes se fait pour partie lors de la transcription et en majorité lors des étapes post-transcriptionnelles dites de maturation de l'ARN (Hammani et Giegé, 2014; Giegé et al., 2000). Dans ce paragraphe, je vais détailler les différentes étapes de maturation post-transcriptionnelles.

Comme vu précédemment, l'origine et l'organisation des génomes de organites induisent la synthèse de transcrits polycistroniques. Lors des étapes de maturation, des enzymes, les ribonucléases –ou RNAses- vont venir cliver l'ARN en 5' et 3' UTR (pour *Untranslated Region*). Il existe deux types de nucléases : les exonucléases, qui viennent digérer les bases azotées situées aux extrémités de l'ARN, et les endonucléases qui elles vont être capables de couper à l'intérieur de l'ARN en reconnaissant certaines séquences. Dégradant l'ARN en 5' et 3', les exonucléases ne sont arrêtées dans leur activité que par deux structures physiques : une structure secondaire de l'ARN d'une part, ou une protéine fixant l'ARN d'autre part, protégeant ainsi certains cadres de lecture (Stern et al., 2010). Cette maturation aura pour conséquence une modulation de la quantité d'ARN avant la traduction.

3.2.3. Edition

Dans les organites des plantes, les ARN peuvent être soumis à une autre étape de maturation appelée édition par conversion. L'édition par conversion consiste à modifier une base de l'ARN par une autre (voir Benne et al. (1986) pour le premier cas étudié). Différents types d'édition existent chez les plantes, mais nous nous intéresserons particulièrement aux phénomènes d'édition des organites. Dans ce cas de figure, l'édition consiste en une conversion d'une cytidine en uridine probablement par désamination. On parle alors d'édition C vers U, ou C→U. Dans de rares cas chez les Angiospermes mais de façon massive chez les Bryophytes et les Ptéridophytes, l'édition U vers C est également possible (Steinhauser et al., 1999) mais je me focaliserai sur le cas le plus commun chez *Arabidopsis thaliana* lors de la thèse. On estime à 67 le nombre de sites d'édition C vers U connus dans le chloroplaste et à 723 chez la mitochondrie d'*Arabidopsis thaliana* ecotype Col0 (Guillaumot et al., 2017). Ces événements d'édition peuvent ouvrir de nouveaux cadres de lecture, modifier des triplets codants et l'acide aminé, ou ne pas avoir de conséquence connue. Etant donné l'importance de l'édition lors de ce travail de thèse, je reviendrai plus en détail sur les mécanismes de fonctionnement et les acteurs impliqués par la suite.

3.2.4. Epissage

L'épissage est un phénomène qui se produit lors de la maturation des ARN. Des excisions suivies de ligations sur l'ARN vont entraîner l'élimination de partie(s) de l'ARN. On appelle intron la ou les séquences d'un gène qui, une fois transcrit, va être éliminée. Pour les ARN codant pour des protéines, la ou les séquences conservées sont appelées exons. Pour la plupart des cas, les introns séparent les exons, le tout étant sur le même ARN. On parle alors d'épissage en *cis*. Suite à de nombreuses recombinaisons des génomes des organites, il arrive que certains exons se retrouvent très éloignés dans le génome, et que leur transcription se fasse sur différents ARN. On parle alors d'épissage en *trans* (Hammani et Giegé, 2014).

3.3. Acteurs de la régulation de l'expression génétique des organites

3.3.1. NEP

Les NEP sont des ARN polymérases encodées dans le noyau impliquées dans la transcription des organites. Deux sont impliquées dans la transcription mitochondriale : RPOTM et RPOTMP (pour *RNA polymerase T7-like*, soulignant l'origine phagique de ces ARN polymérases (Liebers et al., 2017)). Peu de choses sont connues concernant la répartition de la transcription mitochondriale entre RPOTM et RPOTMP. Cependant, la présence d'éventuels facteurs de transcription chez la mitochondrie d'*Arabidopsis* fait encore débat, notamment au sujet de facteurs bien connus chez le chloroplaste et également présent dans la mitochondrie (Beardslee et al., 2002). Ces protéines peuvent être impliquées dans la régulation de l'expression des gènes des organites, comme le supposent Liere et al. (2011). RPOTMP semble être responsable seulement de la transcription de *nad2*, *nad6* et *cox1*, alors que RPOTM transcrit également ces gènes. Les auteurs supposent que cette redondance fonctionnelle a pour but d'assurer une certaine finesse dans la régulation de l'expression de ces gènes. Chez le chloroplaste, on retrouve aussi deux NEP impliquées dans la transcription : RPOTP et RPOTMP. RPOTP semblerait être capable de reconnaître des promoteurs particuliers, puisque sa surexpression montre une augmentation de la transcription pour certains gènes possédant ce promoteur (Liere et al., 2011).

3.3.2. Facteurs sigma

Une spécificité de la transcription des chloroplastes est la présence des facteurs sigma. Au nombre de 6 chez *Arabidopsis thaliana*, ces facteurs de transcription codés dans le noyau sont reconnus par la PEP afin d'initier

la transcription. Il semblerait que ces différents facteurs sigma, à travers l'étude de leurs mutants, aient des rôles assez redondants même si SIG2 et 6 semblent être primordiaux dans les étapes précoces de développement (Liere et al., 2011). Les rôles et spécificités des différents facteurs sigma chez *Arabidopsis* sont résumés dans le tableau I (adapté de Lerbs-Mache, 2011) :

Inspiré de Lerbs-Mache (2010).

Facteur sigma	AGI	Spécificités	Rôles putatifs
SIG1	AT1G64860	Expression induite suite à exposition lumière bleue ou rouge (Onda <i>et al.</i> , 2008). Spécifique pour la transcription de l'opéron <i>psaA</i> (Tozawa <i>et al.</i> , 2007). Interagit en double-hybride avec SIB1, protéine impliquée dans les réponses SA et JA (Morikawa <i>et al.</i> , 2002).	Régulation rapide du PS I via la transcription de l'opéron <i>psaA</i> (Lerbs-Mache, 2010).
SIG2	AT1G08540	Mutant KO déficient en chlorophylle (Privat <i>et al.</i> , 2003). Reconnaissance spécifique des promoteurs de <i>psaJ</i> et <i>trnV-UAC</i> (Privat <i>et al.</i> , 2003).	Régulation de la traduction et de la biosynthèse de la chlorophylle (Lerbs-Mache, 2010).
SIG3	AT3G53920	Reconnait de nombreux promoteurs <i>in vitro</i> mais très peu de conséquences en terme de transcription chez le mutant KO (Lerbs-Mache, 2010). Expression indépendante de la lumière (Privat <i>et al.</i> , 2003). Transcrit spécifiquement <i>psbN</i> et <i>atpH</i> (Zghidi <i>et al.</i> , 2007).	Compléter l'action des autres facteurs sigma (Lerbs-Mache, 2010).
SIG4	AT5G13730	Transcription spécifique de <i>ndhF</i> , et absence de la protéine NdhH, probablement dégradée par absence de complexe avec NdhF (Favory <i>et al.</i> , 2005).	Régulation de la quantité des complexes NDH (Lerbs-Mache, 2010).
SIG5	AT5G24120	Transcription spécifique de <i>psbD</i> , mais peut reconnaître d'autres promoteurs si peu de concurrence (Nagashima <i>et al.</i> , 2004). Expression induite par différents stress abiotiques (Nagashima <i>et al.</i> , 2004).	Protection contre les stress en réparant le centre de réaction du PS II (Lerbs-Mache, 2010).
SIG6	AT2G36990	Mutant KO déficient en chlorophylle (Ishizaki <i>et al.</i> , 2005). Semble impliqué dans la transcription de très nombreux gènes dans les phases initiales de développement (Loschelder <i>et al.</i> , 2006).	Régulation de la transcription lors d'étapes spécifiques du développement de la plante (Loschelder <i>et al.</i> , 2006).

Sans parler de redondance fonctionnelle entre les facteurs sigma, la plupart des promoteurs de gènes et opérons chloroplastiques semblent être reconnus par différents facteurs sigma et finalement assez peu semblent être spécifiques d'une ARN polymérase puis d'un facteur de transcription. On peut rapprocher cette absence de spécificité d'une part à l'absolue nécessité de transcription chloroplastique pour la survie de la plante, et d'autre part à la méthode de reproduction non sexuée des organites. En effet, si l'on suit l'hypothèse du « cliquet » de Muller (1931), la multiplication asexuée risque d'entraîner sur le long terme des mutations altérant la viabilité de l'organisme donc sa disparition. Le manque de spécificité des différents promoteurs peut s'expliquer comme étant une stratégie évolutive permettant d'échapper à cette théorie et d'assurer une transcription des gènes chloroplastiques (Lerbs-Mache, 2011). Mais les mutations et recombinaisons n'ayant pas seulement lieu dans les promoteurs, d'autres mécanismes existent peut-être pour pallier les modifications génomiques, et pas seulement lors de la transcription. Les trois paragraphes suivant vont décrire des familles de protéines connues pour interagir avec l'ARN des organites, participer à sa maturation et de ce fait, jouer un rôle dans la régulation de l'expression des gènes.

3.3.3. RIP/MORF

Les protéines RIP (pour *RNA-editing factor Interacting Protein*) également appelées MORF (pour *Multiple Organellar RNA editing Factor*) constituent une famille de protéines connues pour être impliquées dans l'édition C vers U des ARN des organites. Le tableau ci-dessous résume les différentes RIP/MORF connues à ce jour chez *Arabidopsis thaliana*.

Tableau II : Résumé des différentes protéines RIP connues chez *A. thaliana* et de la proportion de sites

d'édition impactés par leur mutation.

Adapté de Sun et al. (2016). M : Mitochondrie ; P : Plaste ; NA : pas d'information.

Nom		AGI	Localisation subcellulaire (MS/MS)	% de sites d'édition chloroplastiques selon Bentolila <i>et al.</i> (2013)	% de sites d'édition mitochondriaux selon Bentolila <i>et al.</i> (2013)
RIP1	MORF8	AT3G15000	M, P	22%	77%
RIP2	MORF2	AT2G33420	P	100%	NA
RIP3	MORF3	AT3G06790	M	NA	26%
RIP4	MORF4	AT5G44780	NA	NA	NA
RIP5	MORF5	AT1G32580	M	NA	NA
RIP6	MORF6	AT2G35240	NA	NA	NA
RIP7	MORF7	AT1G72530	NA	NA	NA
RIP8	MORF1	AT4G20020	NA	NA	19%
RIP9	MORF9	AT1G11430	P	97%	NA
RIP10	/	AT1G53260	NA	NA	NA

On remarque que sur les 10 protéines RIP/MORF étudiées à ce jour, 5 seulement semblent jouer un rôle dans l'édition, les 5 restantes ayant un effet mineur sur l'édition, une redondance fonctionnelle ou pas d'effet du tout (Sun et al., 2016). Les auteurs remarquèrent également une certaine redondance dans les sites d'édition affectés entre les différentes protéines RIP, à l'exception de RIP1 qui semble contrôler pas moins de 250 sites d'édition mitochondriaux sans redondance avec des protéines de la même famille (Bentolila et al., 2013). Autre fait intéressant, si certaines protéines RIP affectent systématiquement de nombreux sites d'édition et non un nombre très limité, cela sous-tend que ces protéines ne jouent pas un rôle dans la spécificité de reconnaissance des sites, mais plutôt dans un complexe protéique d'édition. L'absence de motifs connus dans les protéines RIP et les interactions observées avec des protéines de familles différentes impliquées dans l'édition appuient cette hypothèse (Sun et al., 2016; Yang et al., 2017).

3.3.4. ORRM

Les protéines ORRM (pour *Organelle RNA Recognition Motif-containing*) sont une famille de protéines codées dans le noyau et envoyées dans le chloroplaste et la mitochondrie (Shi et al., 2017). Ces protéines sont caractérisées par la présence d'un motif RRM (pour *RNA Recognition Motif*) de liaison aux ARN, d'environ 80 acides aminés et très conservé dans la famille. Identifié pour la première fois dans les années 1980, les différents projets de séquençage de génomes montrèrent que ce motif est très répandu dans le vivant, et pas seulement chez les Eucaryotes mais aussi chez les Bactéries (Maris et al., 2005). Les protéines ORRM sont connues pour fixer l'ARN dans les organites des plantes, et jouer un rôle dans l'édition C vers U des ARN. Jusqu'à présent, 6 protéines ORRM ont été décrites chez *Arabidopsis thaliana*. Deux protéines ORRM ont été mises en évidence comme jouant un rôle dans l'édition dans le chloroplaste, ORRM1 (Shi et al., 2017) et ORRM6 (Hackett et al., 2017). En se basant sur les similarités de séquence avec ORRM1, ORRM2, 3, 4 furent mises en évidence (Shi et al., 2015, 2016), tout comme ORRM5 dans un second temps (Shi et al., 2017). Petite spécificité d'ORRM4 : à la différence des mutants affectés dans les autres *orm*, qui ne présentent que quelques défauts d'édition, le mutant $\Delta orrm4$ présente pas moins de 270 altérations d'édition dans la mitochondrie, ce qui en fait un facteur majeur de l'édition dans la mitochondrie (Shi et al., 2016). Bien que les protéines ORRM soient capables de fixer l'ARN, ce résultat suggère que les protéines ORRM peuvent faire partie d'un complexe protéique jouant un rôle dans l'édition, sans jamais être le facteur de spécificité déterminant la base éditée. Dans ce phénomène d'édition, une famille de protéines a été particulièrement bien étudiée, elle aussi connue comme fixant l'ARN, il s'agit des protéines à motifs PPR.

3.3.5. Protéines PPR

Les protéines PPR (pour *pentatricopeptide repeat*) sont des protéines décrites uniquement chez les Eucaryotes et contenant des motifs répétés de 35 acides aminés (Small et Peeters, 2000). Ces motifs permettent à la protéine PPR de lier l'ARN : un motif PPR fixant une base azotée. Codées dans le noyau, leur immense majorité est envoyée dans le chloroplaste et/ou la mitochondrie, avec pour principal rôle la régulation de l'expression des gènes des organites. Décrites pour la première fois par Small et Peeters (2000), ces protéines sont très proches des protéines TPR (pour *tetratricopeptide repeat*), dont le motif permet l'interaction avec d'autres protéines (Blatch et Lässle, 1999). L'une des particularités de cette famille de protéines est leur abondance chez les plantes. Par exemple, Cheng et al. (2016) ont dénombré 496 gènes codant pour des protéines PPR chez *Arabidopsis*, faisant de cette famille l'une des plus abondantes du génome. Malgré de nombreuses études lors de ces 20 dernières années, le rôle exact de la majorité des protéines de cette famille n'est pas encore connu. Bien que de nombreuses protéines nucléaires soient impliquées dans la régulation de l'expression génique des organites, je me suis particulièrement intéressé à la famille des PPR lors de la thèse.

3.4. Les protéines PPR

3.4.1. Caractéristiques

3.4.1.1. Motif PPR et fixation à l'ARN

Par définition, une protéine PPR est une protéine contenant au moins deux motifs PPR (en dehors du cas particulier des DYW-like). On parle alors de protéines à motifs répétés en tandem. Lurin *et al.* mirent en évidence en 2004 que les protéines PPR sont composées de 2 à 26 motifs, avec en moyenne 12 motifs (Lurin et al., 2004). Une analyse plus récente (Cheng et al., 2016) parle de 2 à 29 motifs pour une moyenne légèrement inférieure à 14. Les motifs PPR sont composés de deux hélices- α antiparallèles, et leur répétition forme une structure hélicoïdale (Figure 10). Dans cette structure, à la différence des protéines TPR, le caractère basique et chargé positivement des résidus proches du sillon central favorise l'interaction avec des cibles chargées négativement, comme par exemple les acides nucléiques. Le modèle proposé par Delannoy et al. (2007) suggère qu'un motif PPR permet la reconnaissance d'une base nucléotidique, modèle confirmé depuis par des analyses de cristallographie sur PPR10 entre autres (Yin et al., 2013).

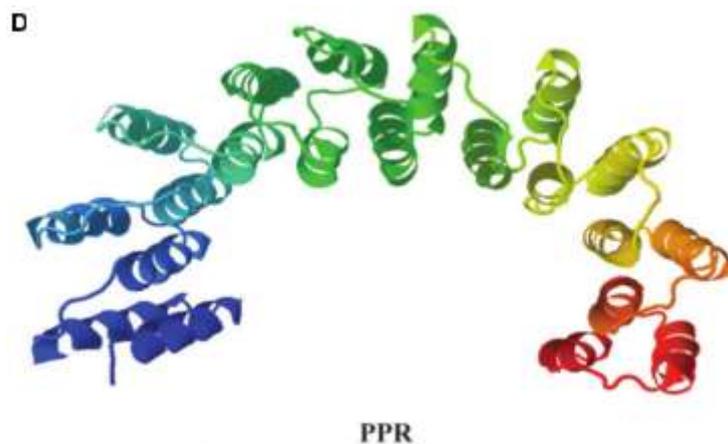


Figure 10 : Structure 3D d'une protéine PPR.

Tiré de Sharma et Pandey (2016).

3.4.1.2. Différentes familles

Selon les motifs qu'elles possèdent, les protéines PPR peuvent appartenir à deux sous-familles : la sous-famille P, pour les protéines PPR « Pures » et la sous-famille PLS pour les protéines PPR qui en plus des motifs « Purs » portent des motifs L pour « Long » et S pour « Short ». Les PPR-P sont uniquement composées de motifs P de 35 acides aminés. Ces motifs sont appelés motifs purs car ils furent les premiers motifs PPR mis en évidence par Small et Peeters (2000). Parmi les 496 protéines PPR d'*Arabidopsis thaliana*, 283 sont des PPR-P (Cheng et al., 2016).

Les PPR-PLS, en plus des motifs purs de 35AA, contiennent des motifs L de 36 aa et des motifs S de 31 aa, avec une succession des motifs qui se présente dans cet ordre dans la plupart des cas. Cheng et al. (2016) dénombrent 213 protéines appartenant à cette sous-famille. Celle-ci est elle-même composée de plusieurs sous-groupes, caractérisés par la présence ou l'absence de motifs particuliers en partie C-terminale de la protéine. Le premier sous-groupe est celui des PPR-PLS, sans motif supplémentaire en partie C-terminale. Cheng et al. (2016) en dénombrent 9 chez *Arabidopsis*. Le second sous-groupe est celui des PPR-PLS possédant un domaine E pour « Elongated » en partie C-terminale (52 protéines chez *Arabidopsis*). Le troisième sous-groupe est celui des E+, avec après les motifs PLS puis E, un domaine E+ (61 protéines chez *Arabidopsis*) et pour finir le sous-groupe des protéines DYW, qui après les différents domaines cités précédemment, contient en partie C-terminale un domaine DYW (91 chez *Arabidopsis*). A la différence des domaines E et E+, fortement dégénérés, le domaine DYW (dont le nom provient des 3 derniers acides aminés du domaine Aspartate-Tyrosine-Tryptophane) est très conservé. L'intérêt de cette classification réside principalement dans le fait que certains sous-groupes sont associés à des rôles connus dans les organites, facilitant ainsi leur étude. Je m'intéresserai à ces rôles un peu plus loin dans l'introduction.

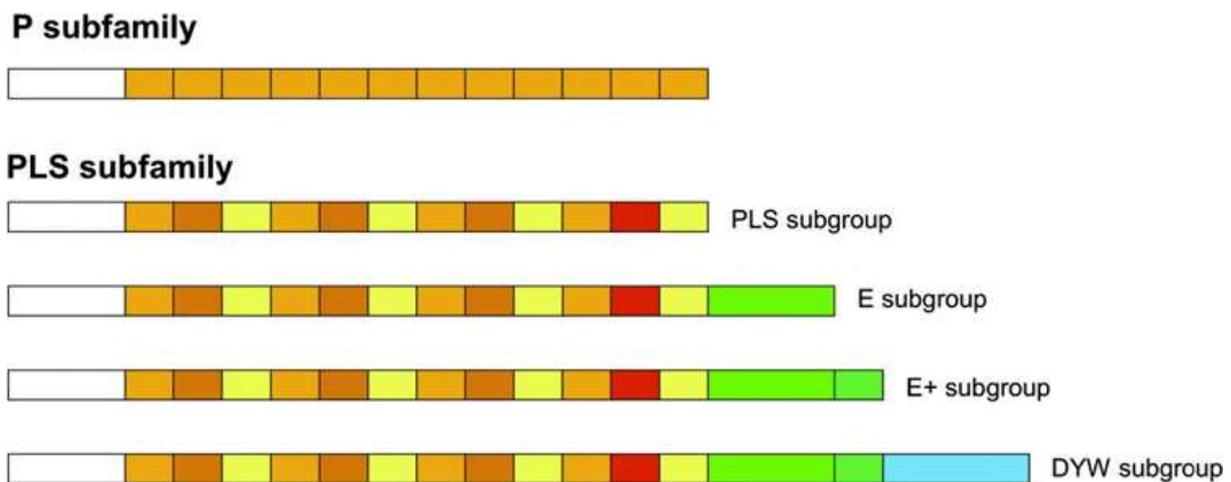


Figure 11 : Représentation schématique des différentes familles de protéines PPR.

Adapté de Lurin et al. (2004).

3.4.1.3. Distribution des protéines PPR dans l'arbre du vivant

On trouve des protéines PPR chez tous les Eucaryotes. Cependant, et en opposition avec les protéines TPR, les protéines PPR sont quasiment absentes chez les bactéries (Figure 12). Il en existe cependant quelques-unes comme nous l'avons vu précédemment, mais aucune n'a été décrite fonctionnellement. Chez les Eucaryotes,

leur abondance dans les génomes est très variable. De 7 gènes codant pour des protéines PPR chez *Homo sapiens* (Lightowers et Chrzanowska-Lightowers, 2008) et 15 chez *Saccharomyces cerevisiae* (Lipinski et al., 2011) à 496 chez *Arabidopsis thaliana*, 703 chez *Zea mays*, 955 chez *Pyrus bretschneideri*, ou même 1544 chez *Malus domestica* mettant en évidence une expansion très supérieure de cette famille de protéines dans le règne des plantes, par rapport au reste des Eucaryotes. Un biais existe cependant dans la recherche des motifs PPR, cette recherche se fait en général par rapport à un consensus du motif trouvé chez les plantes. Comme l'ont mis en évidence Lipinski et al. (2011), le fait d'affiner la recherche de PPR de *S. cerevisiae* à partir de celles déjà connues dans le même organisme et non celles de plantes a permis de passer de 3 à 15 gènes codant pour des protéines PPR dans le génome. Cependant, même si le nombre de PPR dans le vivant en dehors des plantes reste sous-évalué, il est tout de même très peu probable que leur nombre soit supérieur à ces dernières.

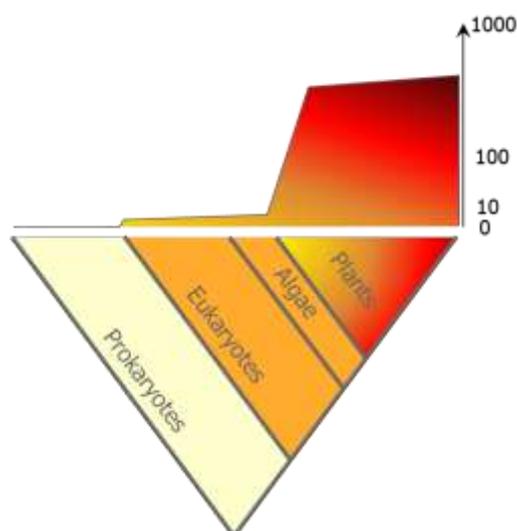


Figure 12 : Proportion de protéines PPR suivant une distribution phylogénétique.

Adapté de Lurin et al. (2004).

3.4.1.4. Code PPR

Initialement, le modèle un motif PPR:une base d'ARN a été proposé suite à l'analyse de deux PPR bien connues : PPR10 (Prikryl et al., 2011) chez *Zea mays*, et MCA1 (Loiselay et al., 2008) chez *Chlamydomonas reinhardtii*. En effet, les auteurs remarquèrent une corrélation entre le nombre de motifs PPR et le nombre de nucléotides fixés. Barkan et Small, (2014) suggèrent qu'il s'agit là des premiers arguments en faveur de l'hypothèse de stœchiométrie 1:1. A l'aide de méthodes bioinformatiques et statistiques, Barkan et al. (2012) ont mis en évidence l'importance des acides aminés 5 et 35 de chaque motif PPR dans la reconnaissance des bases azotées. Il est à noter que pour l'intégralité du manuscrit, la nomenclature de numérotation des acides aminés des protéines PPR est celle proposée par Yin et al. (2013). En position 5 par exemple, on voit une corrélation entre l'asparagine et une pyrimidine, alors que la sérine ou la thréonine sont corrélées avec une purine. A partir d'observations empiriques et de corrélations, un code permettant de prédire la séquence d'ARN fixée par une PPR fut proposé par Barkan et al. (2012). Le code fut ensuite affiné au fur et à mesure que de nouvelles protéines PPR furent décrites (Yagi et al., 2013; Takenaka et al., 2013), et confirmé par les analyses par cristallographie de PPR10 (Yin et al., 2013). Fonctionnant pour bon nombre de protéines PPR, Barkan et Small (2014) préviennent cependant que la réalité est bien plus complexe que l'application d'un simple code. En effet, le caractère dégénéré du code conduit à ce qu'à partir d'un seul doublet d'acides aminés en positions 5 et 35, plusieurs nucléotides puissent être prédits. Ils expliquent également que suivant la sous-

famille de motifs PPR à laquelle on s'intéresse, les prédictions pourraient varier. Enfin, la fixation PPR-ARN doit également s'imaginer dans un contexte dynamique. Les protéines PPR sont probablement en compétition avec d'autres protéines PPR pour la fixation d'un ARN, mais les ARN étant probablement également en compétition pour la fixation à une protéine PPR, l'abondance de chaque partenaire étant également importante, l'ensemble de ces éléments faisant du code un outil prédictif précieux, mais incomplet.

3.4.2. Rôles des protéines PPR dans la maturation des ARN des organites

Les protéines PPR sont impliquées dans de nombreux mécanismes dans la plante, la plupart étant des mécanismes de régulation transcriptionnelle et post-transcriptionnelle. Les paragraphes suivants vont décrire les différents rôles connus à ce jour dans lesquels les protéines PPR sont impliquées. Un exemple sera détaillé pour le chloroplaste et la mitochondrie. Il convient de garder à l'esprit que la fonction principale d'une protéine PPR est la fixation à un ou plusieurs ARN. Suivant la zone et l'intensité de fixation de la molécule d'ARN, les conséquences peuvent varier drastiquement. Les paragraphes suivants vont présenter les conséquences connues des fixations de PPR sur l'ARN, ou en interaction avec d'autres molécules, et chaque cas sera détaillé avec un exemple pour le chloroplaste et un exemple pour la mitochondrie.

3.4.2.1. Transcription

Certaines protéines PPR sont décrites comme fixant l'ADN ou formant des complexes avec des protéines impliquées dans la transcription.

P63 est une protéine PPR mise en évidence par Ikeda et Gray (1999) dans *Zea mays*. Appartenant à la sous-famille des PPR-P, P63 a été localisée dans la mitochondrie. Ikeda et Gray se sont ensuite intéressés aux fractions extraites montrant une activité de fixation à l'ADN mitochondrial. Ils mirent en évidence une fixation de P63 à la région promotrice du gène codant pour Cox2, et améliorant la transcription *in vitro*.

Un autre exemple de protéine PPR impliquée dans la transcription est PTAC2 (Plastidial Transcriptional Active Complex, AT1G74850). Cette PPR-P, localisée dans le chloroplaste, a été décrite par Pfalz, (2006). En plus des motifs PPR, les auteurs mirent en évidence des domaines TPR, connus pour permettre l'interaction protéine-protéine ainsi que des domaines SMR (pour Small MutS Related). Les protéines SMR joueraient un rôle dans la recombinaison et la réparation de l'ADN (Fukui et Kuramitsu, 2011). L'étude du mutant KO $\Delta ptac2$ met en évidence le rôle de PTAC2 avec la Pep (pour *Plastid Encoded RNA Polymerase*, une ARN polymérase encodée dans le chloroplaste, qui va reconnaître des promoteurs différents des NEP décrites précédemment) formant très probablement un complexe protéique permettant la transcription. Son rôle exact n'est malheureusement pas connu avec plus de précision.

Une fois la transcription terminée, l'ARNm est soumis à différentes étapes de maturation. Selon Hammani et Giegé (2014), ces différentes étapes de maturation constituent la majeure partie de la régulation de l'expression des gènes des organites des plantes.

3.4.2.2. Maturation 5' et 3' UTR des ARN

Dans le chloroplaste, l'exemple le plus connu est PPR10, protéine PPR chloroplastique chez *Zea mays*. Pfalz et al. (2009) mirent en évidence la fixation de PPR10 en 5' UTR (de l'anglais *Untranslated Transcribed Region*) de l'ARN codant pour AtpH et en 3' UTR de l'ARN codant pour Psaj. Cette fixation permet la protection de ces deux transcrits contre l'action des exonucléases et permet leur accumulation en vue de la traduction. Dans la mitochondrie, on peut citer l'exemple de MPPR6. Étudiée dans *Zea mays* par Manavski et al. (2012), le mutant KO de cette PPR-P présente un retard de développement, des graines de taille réduite avec une plus faible accumulation d'amidon. La traduction dans la mitochondrie est également sévèrement

affectée dans le mutant. Les auteurs ont mis en évidence un orthologue du gène codant pour MPPR6 chez *A. thaliana* et réussi une complémentation hétérologue sur le maïs. Par coimmunoprécipitation avec de l'ARN, les auteurs mirent en évidence l'interaction entre MPPR6 et le 5' UTR du transcrite codant pour Rps3. Par RT-PCR, une suraccumulation du transcrite fut mise en évidence, mais elle ne semble pas spécifique à rps3 puisque les ARNr 18 et 26S présentent le même profil, cohérent avec les nombreux défauts de traduction dans le mutant. L'accumulation de la protéine Rps3 est beaucoup plus faible dans le mutant et seule une version tronquée non fonctionnelle est présente. Les auteurs supposent que MPPR6 joue un rôle dans la maturation en 5' UTR de *rps3*, s'appuyant sur des données de RT-PCR circulaire. En effet, ils remarquèrent que l'ARN mature *rps3* chez le mutant est plus long que l'ARN mature chez le sauvage. Ils suggèrent donc que MPPR6, en fixant l'ARN, permet la reconnaissance et/ou l'activité d'une endonucléase qui va venir cliver l'ARN. Est-ce que MPPR6 joue le rôle de facteur de reconnaissance séquence-spécifique ? Quelle séquence exacte de l'ARN est reconnue et comment cela entraîne la synthèse d'une protéine clivée non fonctionnelle ? A l'heure actuelle, ces questions demeurent sans réponse. L'intérêt de l'étude de MPPR6 réside aussi dans le fait que les protéines PPR ne se bornent pas à être des clamps moléculaires fixant l'ARN pour bloquer l'action des exonucléases puisqu'elles pourraient faire partie de complexes bien plus compliqués impliqués dans le clivage, jouant probablement le rôle de facteur de spécificité.

3.4.2.3. Epissage

Pour le chloroplaste, on peut citer l'exemple de PPR4. Schmitz-Linneweber et al. (2006) montrèrent que le mutant KO Δ *ppr4* présentait un phénotype macroscopique albinos, et un phénotype moléculaire montrant un défaut de d'épissage sur le transcrite codant pour Rps12. Par co-immunoprécipitation, ils mirent en évidence une interaction entre PPR4 et l'intron 1 de *rps12*, suggérant que cette PPR est impliquée dans l'épissage de cet intron en particulier. Du côté de la mitochondrie, OTP43 (AT1G74900) est une PPR-P décrite par de Longevialle et al. (2007) chez *A. thaliana*. Le mutant KO Δ *otp43* présente de grosses difficultés de germination, les graines de l'homozygote *otp43* sont plus petites et de couleur plus sombre que les graines sauvages ou hétérozygotes et la croissance de la plante est fortement affectée. L'absence de complexe I mitochondrial fut mise en évidence par Blue-Native PAGE, en cohérence avec une accumulation de la protéine AOX, connue chez les plantes pour proposer une voie alternative au complexe I (Vidal et al., 2007). Par RT-PCR, les auteurs trouvèrent un défaut d'épissage au niveau de l'intron 1 du transcrite codant pour *Nad1*. Le défaut fut complétement en transformation stable dans *A. thaliana*, mais également en transitoire après infiltration par une souche d'*Agrobacterium* contenant une copie du gène codant pour OTP43 sur son ADN de transfert, démontrant l'implication d'OTP43 dans l'épissage de *nad1*. De Longevialle et al. (2010) proposent une revue des facteurs nucléaires impliqués dans l'épissage des organites des plantes, dont les protéines PPR.

3.4.2.4. Edition C vers U

Jusqu'à présent, une grande majorité des facteurs impliqués dans l'édition C vers U sont des protéines PPR de la sous-famille PLS. Barkan et Small (2014) résumant très bien le mécanisme : à l'image des PPR-P, les PPR-PLS fixent l'ARN de manière séquence-spécifique, 4 bases en amont du site d'édition. Cette position permet alors une activité cytidine désaminase, qu'aucune expérimentation *in vitro* n'a jamais réussi à reproduire, signifiant d'une part, que l'on ne connaît toujours pas l'enzyme ayant l'activité enzymatique d'édition C vers U, et d'autre part l'existence d'éventuels co-facteurs nécessaires au mécanisme. De fortes présomptions pèsent cependant sur le motif DYW. Salone et al. (2007) remarquèrent que le domaine DYW était composé de résidus ne variant jamais et ressemblant fortement au site actif de cytidines désaminases connues. A l'aide d'une approche phylogénétique, ils montrèrent une corrélation entre la présence de protéines PPR à domaine DYW dans une espèce et la présence d'édition des ARN. Quelques années plus tard, Boussardon et al. (2014) ont démontré que le domaine DYW de la protéine PPR DYW1 (AT1G47580) contenait une signature assez

conservée ressemblant beaucoup aux sites actifs de fixation d'ions zinc que l'on trouve chez des désaminases connues. Après avoir démontré la fixation de zinc par DYW1, à l'aide de mutations ponctuelles, ils montrèrent que certains acides aminés de la signature étaient nécessaires à la fixation du zinc et à l'édition. Un résultat similaire fut obtenu par Wagoner et al. (2015) avec les PPR QED1 (Quintuple Editing Factor 1) et RARE1. Les corrélations n'impliquant pas nécessairement des causalités, il manque toujours en 2018 à la communauté PPR une expérimentation *in vitro* démontrant de façon directe que le domaine DYW joue bien le rôle d'enzyme dans l'édition.

Malgré ces faisceaux d'indices pointant vers le domaine DYW, Barkan et Small (2014) font remarquer que de nombreuses publications mettent en évidence l'obligation d'avoir un domaine E pour que l'édition fonctionne, alors que le domaine DYW semble être facultatif. Le caractère non obligatoire du domaine DYW pourrait provenir du recrutement d'autre(s) protéine(s) PPR possédant un domaine DYW fonctionnel, complétant ainsi des PPR-PLS dépourvues de DYW. C'est l'hypothèse avancée par Boussardon et al. (2012) à travers la caractérisation de DYW1, une PPR qui présente la particularité de ne pas avoir l'enchaînement de domaines PPR (P – L – S)ⁿ – E – E+ – DYW caractéristique des PPR-DYW. DYW1 possède un signal d'envoi au chloroplaste en partie N-terminale, un domaine qualifié par les auteurs de PPR-like suivi d'un domaine DYW. On peut qualifier cette protéine de DYW-like. Les auteurs ont démontré que le mutant KO $\Delta crr4$ (AT2G45350) présentait exactement le même phénotype moléculaire que le mutant KO $\Delta dyw1$, à savoir un défaut d'édition sur le site *ndhD-1*. L'histoire devient encore plus intéressante quand ils montrèrent une interaction *in planta* de DYW1 et CRR4 et que les 2 mutants peuvent être complétés par une fusion traductionnelle entre CRR4 et DYW1. Le modèle proposé par les auteurs est donc que CRR4 et DYW1 font partie du même complexe d'édition, et que cette interaction permet de « reformer » une PPR-PLS-DYW complète et fonctionnelle. Ce papier ouvre la voie à une nouvelle façon de voir l'édition pour les PPR E et E+, car, comme ils le rappellent, 50% des PPR-PLS n'ont pas de domaine DYW. Une seconde protéine DYW-like a été caractérisée par Guillaumot et al. (2017), il s'agit de DYW2 (AT2G15690). Le mutant KO $\Delta dyw2$ est embryon léthal, mais à l'aide d'un promoteur embryon spécifique, les auteurs réussirent à exprimer DYW2 chez l'hétérozygote afin d'obtenir des graines homozygotes viables pour le mutant $\Delta dyw2$. A l'aide du séquençage des ARN (*RNAseq*), 39 défauts d'édition furent trouvés dans le chloroplaste (sur 67 sites connus) et 353 défauts dans la mitochondrie (sur 723 sites connus) chez le mutant $\Delta dyw2$. Parmi les sites affectés chez $\Delta dyw2$, les auteurs ont recherché si des PPR d'édition étaient déjà connues pour agir sur ces sites. Parmi la quarantaine de sites connus comme dépendant de PPR-PLS-DYW, aucun n'est affecté chez le mutant $\Delta dyw2$, à l'inverse des sites dépendants de PPR-PLS E+ connus dont 100% sont affectés. Guillaumot et al. (2017) suggèrent une fois encore que des PPR contenant seulement un domaine DYW, comme les DYW-like, permettraient à des PPR-PLS –donc sans domaine DYW– de faire de l'édition, ces dernières jouant probablement le rôle de facteur de spécificité en fixant l'ARN.

Ces phénomènes d'édition peuvent avoir des conséquences différentes pour la plante, remplaçant un acide aminé par un autre (édition du transcrit *ndhD* par OTP85, changeant une proline en leucine, (Hammani et al., 2009)), modifiant la structure et/ou la stabilité de l'ARN (exemple de sites d'édition de *rps10* facilitant la mise en place de structures secondaires dans les introns, nécessaires à l'épissage, (Castandet et al., 2010)), créant des codons d'initiation de la traduction (Kotera et al., 2005), ou ne pas avoir de conséquence (exemple d'OTP81 éditant dans l'intron de *rps12* décrit par Hammani et al. (2009)).

3.4.2.5. Traduction

La maturation 5' et 3' UTR des ARN décrite précédemment a pour conséquence une accumulation de transcrits, influençant le rendement de la traduction dans les organites. Dans ce paragraphe, nous allons plutôt nous intéresser au rôle des protéines PPR dans la biosynthèse des sous-unités ribosomales et dans la reconnaissance des sites d'initiation de la traduction.

Chez *Arabidopsis thaliana*, PPR336 (AT1G61870) est une PPR-P mitochondriale décrite par Uyttewaal et al. (2008). Outre sa petite taille, l'une des caractéristiques de cette PPR est son fort niveau d'expression, en comparaison avec la plupart des autres protéines PPR qui sont de manière générale assez peu exprimées (Lurin et al., 2004). Le double mutant KO $\Delta ppr336/\Delta ppr336\text{-like}$ (un proche homologue de *ppr336* chez *A. thaliana*) ne présente pas de phénotype macroscopique mais les auteurs mirent en évidence l'association entre PPR336 et les polysomes, suggérant l'implication de cette PPR au moins dans les étapes préliminaires de la traduction. Chez *Zea mays*, CRP1 (AT5G42310 pour son homologue chez *Arabidopsis thaliana*) est une PPR pure chloroplastique caractérisée par Fisk et al. (1999), soit 1 an avant la mise en évidence de la famille des protéines PPR par Small et Peeters (2000). Les auteurs mirent en évidence des homologues de CRP1 chez les mycètes et impliqués dans la traduction chez la mitochondrie. Ils s'intéressèrent donc à la traduction chloroplastique dans un mutant $\Delta crp1$ et remarquèrent que CRP1 est nécessaire à la traduction de *petA* et *petD*, mais aussi à la maturation de l'ARNm de *PetD*. Les auteurs considèrent cependant que ce dernier rôle serait plutôt une conséquence d'un défaut de traduction, en proposant l'hypothèse que CRP1 serait nécessaire à la fixation du ribosome sur l'ARNm, et que cette fixation modifierait la structure de l'ARNm, permettant ainsi aux enzymes nécessaires à la maturation de jouer leur rôle (Fisk et al., 1999).

Après avoir présenté une PPR impliquée dans chaque étape de l'expression génétique de la transcription à la traduction incluse, dans les deux organites (Figure 13), je souhaiterais présenter un dernier exemple, représentant parfaitement les différentes conséquences que la fixation d'une protéine de cette famille sur l'ARN peut avoir sur la plante. PPR5 (AT4G39620) est une protéine PPR chloroplastique nécessaire à l'embryogénèse chez *Arabidopsis thaliana* (Beick et al., 2008). Les mutants $\Delta ppr5_1$ et $\Delta ppr5_2$ chez le maïs ne sont pas embryon-léthaux mais les jeunes pousses sont trop affectées pour survivre. Les feuilles présentent un phénotype albinos négativement corrélé au niveau de transcrit *ppr5*. Les auteurs montrèrent que la transcription chloroplastique était largement affectée chez ces mutants, d'où le phénotype albinos. De plus, ils remarquèrent l'absence totale de transcrit *trnG*, épissé ou non, alors que la transcription de ce gène n'est pas affectée. Par co-immunoprécipitation, les auteurs démontrèrent que PPR5 fixe l'intron de *trnG*, le protégeant ainsi d'un clivage entraînant ensuite la non-accumulation du transcrit par dégradation. L'exemple de cette PPR est, à mes yeux, très représentatif de la diversité des conséquences que peut avoir la fixation de l'ARN d'un organite par une protéine PPR. Le défaut de traduction était évident sur ce mutant, le fait que la PPR fixe dans un intron suggérait également un problème d'épissage et la dégradation des transcrits *trnG* suggérait un rôle dans la maturation 5' ou 3' UTR de l'ARN. Il est extrêmement compliqué de déterminer les défauts primaires des défauts secondaires chez les mutants PPR, d'où ma volonté d'insister via cet exemple sur le rôle des protéines PPR : fixer l'ARN, avec différentes conséquences.

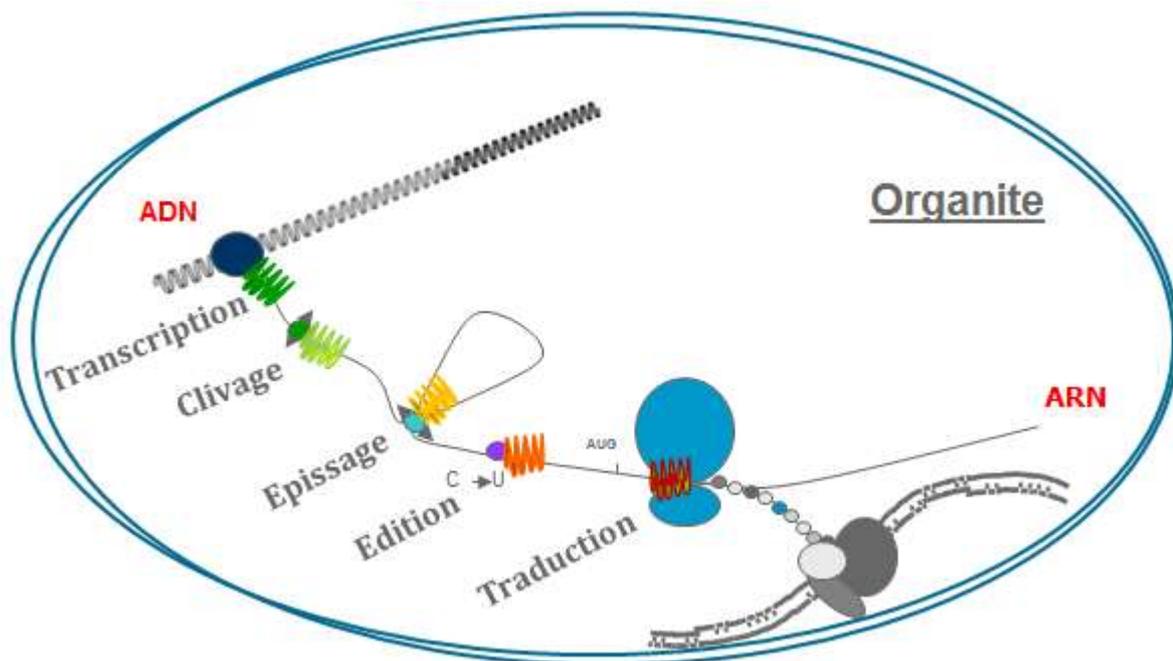


Figure 13 : Représentation schématique de l'organisation de l'expression des gènes des organites de la transcription à la traduction.

Adapté de Andrés et al. (2007).

Comme cela vient d'être détaillé, les protéines PPR sont impliquées dans de nombreux événements de maturation des ARN des organites. Ces événements participent à la régulation de l'expression des gènes des organites. En effet, les variations de quantité, stabilité voire de séquence des ARN vont avoir des conséquences sur les protéines, puis sur l'activité du chloroplaste et de la mitochondrie. Etant donné leur importance, il est primordial pour la plante de gérer l'expression des gènes des organites lors de stress (Taylor et al., 2009). Le paragraphe suivant détaille les connaissances à ce jour sur les variations d'expression des gènes des organites en réponse à des stimuli environnementaux, en particulier les stress biotiques.

3.5. Organites et stimuli environnementaux

3.5.1. Modifications de l'expression génétique des organites en réponse à un stress biotique

De manière générale, très peu de données sur le transcriptome du chloroplaste et la mitochondrie existent, la plupart se focalisant sur le noyau. Cette limitation est principalement technique. En effet, les ARNm nucléaires sont pourvus d'une queue poly-adénylée (ou queue poly(A)) en extrémité 3', ce qui permet en biologie moléculaire de les extraire très facilement à l'aide de fragments d'ADN poly(T) (ou oligo dT) et récupérer seulement les messagers nucléaires. Les ARN chloroplastiques et mitochondriaux n'ayant pas de queue poly(A), ils ne sont pas observés dans la plupart des expérimentations. Les paragraphes qui suivent présentent les quelques données connues à ce jour.

Dès 1993, Mullet faisait état de nombreuses modifications dans la transcription du chloroplaste de l'orge après exposition à la lumière (Mullet, 1993). Ce genre de données concernant les stress biotiques n'arrivera que bien plus tard dans les années 2000 (Meyer et al., 2001; Bilgin et al., 2010; de Torres Zabala et al., 2015), et seulement pour les gènes nucléaires codant des protéines adressées au chloroplaste. Lors de l'infection, la transcription des gènes codant pour des protéines envoyées dans le chloroplaste va être très largement remodelée. Cette modulation semble être variable suivant le type de pathogène reconnu par la plante. Lors de

l'interaction incompatible *Nicotiana tabacum-Phytophthora nicotianae*, la régulation négative de l'appareil photosynthétique de la plante sera très rapide (de l'ordre de l'heure), suivie par une activation totale des défenses et une réponse hypersensible (Scharte et al., 2005). Par opposition, l'interaction compatible entre *Phaseolus vulgaris* et *Colletotrichum lindemuthianum* entraîne des nécroses et une diminution de l'activité photosynthétique seulement dans les étapes tardives de l'infection (Meyer et al., 2001). Il semblerait que cette reprogrammation permette une mise en place plus rapide et efficace des défenses de la plante, à savoir l'expression de gènes de défense ou la synthèse de composés antimicrobiens, via les ROS comme messagers moléculaires (Kangasjärvi et al., 2012). Selon Dietz et Pfannschmidt (2011), l'état redox du chloroplaste pourrait même être à l'origine de ces modulations de l'expression des gènes de l'organite.

Du côté de la mitochondrie, aucune réorganisation majeure de la transcription n'a été mise en évidence jusqu'à présent en condition de stress. Law et al. (2012) virent un pic d'expression des gènes nucléaires encodant des protéines envoyées dans la mitochondrie lors de la germination, d'abord codant pour le métabolisme des ARN et l'import, puis pour les fonctions bioénergétiques de l'organite. Une fois ce pic passé, la transcription mitochondriale demeure très stable, malgré les stress. Plusieurs hypothèses peuvent expliquer cette stabilité. Il est possible qu'aucune forme de régulation de l'expression des gènes de la mitochondrie ne se fasse au niveau transcriptionnel et que les mécanismes soient pour partie au niveau traductionnel ou post-traductionnel. Une autre possibilité est qu'aucune variation de l'expression des gènes de la mitochondrie n'est tolérée par la plante, et que la transcription est très finement régulée. Récemment, Zubo et al. (2014) ont démontré que des inhibiteurs de certains complexes de la chaîne de transport des électrons ou d'AOX modifiaient la transcription et la quantité d'ARN de la mitochondrie, et ce de façon différente suivant le complexe touché. Cette étude ouvre la porte à une éventuelle régulation redox de l'expression des gènes de la mitochondrie, à l'image du chloroplaste. Cependant, toutes ces données, obtenues à l'aide d'inhibiteurs ou de mutants, ne mettent pas en évidence des régulations de l'expression des gènes de la mitochondrie mais plutôt des mécanismes de compensation suite à un défaut. Enfin, une dernière possibilité est une éventuelle régulation post-transcriptionnelle, via des protéines interagissant avec l'ARN et impliquées dans sa maturation. Cette hypothèse étant celle qui m'intéresse lors de ma thèse, en particulier concernant la famille des protéines PPR, le paragraphe suivant est dédié aux différentes protéines PPR impliquées plus ou moins directement à une réponse à un stress biotique.

3.5.2. Protéines PPR impliquées dans la réponse aux stress biotiques

Les protéines PPR étant codées par le noyau, on peut profiter des nombreux jeux de données de la communauté scientifique pour observer certaines d'entre elles en condition de stress biotique. Les paragraphes suivants détaillent les différentes protéines PPR étudiées dans la littérature en condition de stress biotique.

3.5.2.1. PPRL

PPRL est une protéine PPR mitochondriale dont le rôle dans la défense des plantes fut décrit par Katiyar-Agarwal et al. (2006). A la recherche de nat-siRNA (*natural antisens small interfering RNA*) endogènes de la plante induits suite à l'infection par une bactérie pathogène, les auteurs ont mis en évidence nat-siRNA *atgb2*, dont la séquence est complémentaire de la partie 3' de *PPRL* (AT4G35850). Il a été démontré que l'action de nat-siRNA *atgb2* entraîne un *silencing* de *PPRL* suite à une infection par *Pseudomonas syringae* DC3000, mais seulement pour les souches contenant le gène d'avirulence *avrRpt2*. Les auteurs ont également remarqué que la surexpression de *PPRL in planta* entraîne un retard dans la mise en place d'une réponse hypersensible après infection par *Pst (avrRpt2)*. Ils suggèrent que *PPRL* agit comme un régulateur négatif de la voie de résistance *RPS2* de la plante. En dehors de sa localisation subcellulaire mitochondriale, très peu d'éléments sur le mode d'action de cette PPR sont connus. *PPRL* est une PPR pure, donc probablement impliquée dans des phénomènes de maturation des ARN autres que l'édition. Elle a été également vue comme interagissant en

double hybride levure avec AT1G26460, une seconde PPR pure (Dreze et al., 2011), prédite pour être envoyée dans la mitochondrie. Selon les auteurs, son mécanisme d'action pourrait donc également impliquer une interaction ARN-protéine, voire protéine-protéine avec une seconde PPR, avec pour conséquence(s) une régulation du *burst* oxydatif et/ou de la réponse hypersensible.

3.5.2.2. PPR1/PPR2

De façon similaire, Park et al. (2014) se sont intéressés aux cibles de miRNA400 (microRNA 400) chez *Arabidopsis thaliana*. Cet ARN non codant a été mis en évidence par Fahlgren et al. (2007) pour être plus faiblement exprimé par *Arabidopsis* après infection par *Pst*. En surexprimant miRNA400, les auteurs ont remarqué que la plante était plus sensible à *Pst* et *Bc*, avec des niveaux de H₂O₂ très nettement supérieurs dans la lignée de surexpression en comparaison du sauvage après infection par *Pseudomonas*. Par analyse bioinformatique puis validation par RT-qPCR, Park *et coll.* démontrèrent que deux PPR étaient ciblées et leur ARNm clivé par miRNA400 : PPR1 (AT1G06580) et PPR2 (AT1G62720). PPR1 est une PPR pure (Cheng et al., 2016) et prédite pour être localisée dans la mitochondrie. PPR2 est également une PPR pure localisée dans la mitochondrie (Yang et al., 2011). Le mutant KO Δ *ppr2* montre une sensibilité à *Pst* plus marquée que la lignée sauvage, mais sans atteindre le niveau de sensibilité du surexprimeur de miRNA400 selon les auteurs. Lors de l'infection par *Pseudomonas* ou par *Botrytis*, l'expression de miRNA400 semble être régulée négativement dès 12 heures après infection, avec pour probable conséquence une régulation positive de l'expression des gènes codant pour PPR1 et PPR2. PPR2 semble donc être impliquée dans un mécanisme de défense face à *Pst* et *Bc*. Malheureusement, les auteurs n'ont pas présenté de résultat sur un mutant KO Δ *ppr1*. Etant donné que le mutant KO Δ *ppr2* ne montre pas une sensibilité aux pathogènes testés aussi importante que la lignée de surexpression miRNA400, on peut supposer que ce miRNA a d'autres cibles dans la plante impliquées dans la défense, une cible non encore mise en évidence, ou déjà connue comme PPR1.

3.5.2.3. SLO2

Décrite en 2014 par Zhu *et coll.* pour son lien dans les stress biotiques et abiotiques, SLO2 est une PPR E+ mitochondriale impliquée dans l'édition chez *Arabidopsis thaliana* (Zhu et al., 2014). L'étude du mutant KO Δ *slo2* a permis de mettre en évidence des défauts de croissance et une sensibilité marquée à l'absence de sucrose par rapport à une lignée sauvage, typique d'un mutant de complexe I (Zhu et al., 2012a). Les complexes I, III et IV sont très affectés dans le mutant, ce qui est cohérent avec les nombreux défauts d'édition mitochondriaux trouvés (sur les transcrits *mttb*, *nad4l*, *nad7* et *nad1*). Le mutant présente une hypersensibilité à l'ABA et une insensibilité à l'éthylène, une sensibilité accrue au sel et au stress osmotique durant la germination. Outre les stress abiotiques, le mutant Δ *slo2* montre également une sensibilité plus marquée à *Bc* et des niveaux d'H₂O₂ également supérieurs comparé à une lignée sauvage. Cependant, à la différence de PPR1, PPR1 ou encore PPR2, le niveau d'expression du gène codant pour SLO2 ne semble pas varier en réponse à un stress biotique (*Pst* en l'occurrence) ou à différents traitements hormonaux (MeJA, SA, GA et IAA).

3.5.2.4. PPR40

PPR40 (AT3G16890) est une protéine PPR composée de 13 motifs purs (Cheng et al., 2016) et décrite par Zsigmond et al. (2008). L'étude de son mutant KO montra une croissance ralentie en condition standard et une germination retardée avec des effets plus marqués encore en présence d'ABA en comparaison d'une lignée sauvage d'*Arabidopsis*. Le mutant est également plus sensible au stress oxydatif et a un niveau basal de peroxyde d'hydrogène plus élevé. Les auteurs ont localisé la protéine dans la mitochondrie de deux façons différentes, d'une part par microscopie à épifluorescence et d'autre part par Western Blot après un Blue-Native PAGE sur des mitochondries purifiées. L'avantage de cette seconde méthode est qu'elle a également

permis de montrer l'association de PPR40 avec le complexe III de la mitochondrie, qui semble très affecté mais pas totalement absent chez le mutant KO $\Delta ppr40$. Les auteurs ont donc logiquement regardé l'édition, l'épissage et l'accumulation de transcrit de Cob, seule protéine du complexe III encodée par la mitochondrie, mais sans trouver le moindre défaut. On peut imaginer un rôle de PPR40 dans la stabilisation de l'ARNm cob en vue de la traduction mais aucun Western Blot n'a été fait pour voir le niveau d'accumulation de la protéine Cob dans le mutant $\Delta ppr40$. Dans un second temps, Laluk et al. (2011) ont mis en évidence une sensibilité accrue du mutant $\Delta ppr40$ à *Bc*, mais le niveau de transcrits de PPR40 ne semble pas varier suite à différents traitements et stress (analyse Genevestigator, puis confirmé par RT-qPCR par Zsigmond *et al.*).

3.5.2.5. PGN

PGN (PPR for Germination on NaCl) est une PPR d'*Arabidopsis* (Laluk et al., 2011). Suite à l'analyse transcriptomique du mutant *bik1* (Botrytis-Induced Kinase 1, une kinase impliquée dans la résistance à *Bc* et la sensibilité à *Pst* selon Veronese (2006)), les auteurs remarquèrent une augmentation de l'expression du gène codant pour PGN, suggérant que PGN serait une potentielle cible de BIK1. Des analyses de RT-qPCR ont également montré une augmentation de l'expression du gène codant pour PGN chez le sauvage en réponse à une infection par *Bc*. L'étude du mutant KO Δpgn montre une sensibilité accrue à *Alternaria brassicicola* et *Bc* mais pas à *Pst*, suggérant une implication de PGN dans un mécanisme de défense spécifique contre les pathogènes nécrotrophes. Le mutant présente des difficultés de germination sur milieu complété avec du NaCl, du glucose et de l'ABA, et son niveau basal d'ABA est plus élevé en condition standard qu'une lignée sauvage. Une suraccumulation de peroxyde d'hydrogène chez le mutant par rapport au sauvage est également vue après un stress NaCl. Une fusion GFP a permis aux auteurs de localiser la protéine dans la mitochondrie. De plus, un pathotest *Bc* réalisé en obscurité continue sur le mutant met en évidence une sensibilité encore plus marquée. Alors qu'en condition lumière la détoxification des ROS se fait par le chloroplaste et la mitochondrie, à l'obscurité la plante compte principalement sur la mitochondrie. L'hypothèse des auteurs est que PGN est impliqué dans la régulation redox de la mitochondrie. Aucun mécanisme moléculaire n'est proposé par Laluk *et coll.* mais si l'on tient compte de l'appartenance de PGN à la sous-famille des PPR E+, il est fort probable que cette PPR soit impliquée dans l'édition de transcrits mitochondriaux.

4. Projet de thèse

Après avoir décrit les cas généraux des interactions plante-pathogène, l'importance des organites dans la mise en place des défenses de la plante et des exemples de modifications de la transcription en condition de stress, nous supposons qu'il existe une forme de régulation de l'expression des gènes des organites lors de l'attaque d'un pathogène. Par régulation, j'entends qu'il existe un élément régulé par un élément régulateur en condition de stress. L'hypothèse générale de la thèse est qu'il existe des mécanismes de régulation de l'expression génétique des organites, en lien avec les protéines PPR, en réponse au stress biotique. Ces protéines PPR peuvent être des protéines endogènes de la plante, c'est-à-dire encodées par le noyau et dont l'expression sera induite ou réprimée lors de l'interaction. Ces protéines PPR peuvent également être exogènes, c'est-à-dire être encodées par le pathogène puis injectées dans la cellule végétale, à l'image d'un effecteur.

Andéol Falcon de Longevialle, ancien doctorant dans l'équipe OGE (Organelle Gene Expression - IPS2), a remarqué que des motifs PPR avaient été mis en évidence dans les banques de données de protéines, chez des bactéries. Il a donc initié la caractérisation de deux protéines PPR de bactéries pathogènes de plantes, l'une chez *Erwinia amylovora*, la seconde chez *Ralstonia solanacearum*. La poursuite de la caractérisation de ces deux protéines représente l'essentiel du travail effectué lors de ma thèse. En parallèle, je me suis intéressé à une PPR E+ endogène d'*Arabidopsis thaliana* que j'ai décrite précédemment, PGN. Afin de caractériser ces 3 PPR, nous avons eu besoin d'outils spécifiques à nos objets d'étude. Ces outils sont un pipeline d'analyse de

données de séquençage d'ARN et, plus spécifiquement pour PGN, un perfectionnement de la méthode de prédiction des sites d'édition dépendants de PPR E+. Nous allons donc dans un premier temps nous intéresser aux méthodes développées lors de la thèse puis ensuite voir leurs applications dans deux cas concrets, ceux d'une PPR endogène et de deux PPR exogènes impliquées dans les interactions plante-pathogène.

Chapitre 1 : Proposition d'un pipeline d'analyse de données *RNAseq* pour le chloroplaste et la mitochondrie.

1. Présentation du pipeline d'analyse de données *RNAseq*

L'étude de l'implication des protéines PPR dans les événements de maturation post-transcriptionnelle nécessite des outils spécifiques afin de comprendre au mieux quel(s) ARN sont fixés par la PPR et les conséquences de cette fixation. Pour cela, nous utilisons la technologie de séquençage des ARN. Cette méthode nous permet d'accéder à une grande partie du transcriptome de la plante à un moment donné. A partir de ces données, nous pouvons identifier les gènes différentiellement exprimés, les variations d'édition, de couverture ou d'épissage, entre une lignée mutante ou de surexpression et une lignée sauvage. L'article suivant détaille les outils moléculaires, bioinformatiques et statistiques utilisés dans l'analyse du transcriptome chloroplastique d'*Arabidopsis thaliana*. Avant de rentrer dans de plus amples détails, quelques définitions sont posées.

Définitions :

RNAseq : Séquençage des ARN à haut débit après fragmentation et rétro-transcription.

Reads : Fragments de séquences obtenus après séquençage. On suppose que leur quantité est proportionnelle à l'abondance de l'ARN leur correspondant.

Alignement : Alignement des *reads* sur le génome de référence de l'espèce. Permet d'obtenir la position d'origine du transcrit.

Trimming : Contrôle qualité des *reads*. Etape d'élimination des *reads* trop courts ou de mauvaise qualité.

DE : Pour *Différentiellement Exprimé*. Se dit d'un gène lorsque son expression est statistiquement différente entre une condition contrôle et une condition testée.

Log2FoldChange : ou Log-ratio est le logarithme binaire du rapport de l'expression d'un gène entre deux conditions.

p-value : Valeur mesurant la cohérence entre l'hypothèse nulle et les données. Probabilité d'observer nos données si l'hypothèse nulle est vraie.

1.1. Pipeline appliqué à l'étude du transcriptome du chloroplaste



Bioinformatic Analysis of Chloroplast Gene Expression and RNA Posttranscriptional Maturation Using RNA Sequencing

Bastien Malbert, Guillem Rigail, Veronique Brunaud, Claire Lurin, and Etienne Delannoy

Abstract

Sequencing of total RNA enables the study of the whole plant transcriptome resulting from the simultaneous expression of the three genomes of plant cells (located in the nucleus, mitochondrion and chloroplast). While commonly used for the quantification of the nuclear gene expression, this method remains complex and challenging when applied to organellar genomes and/or when used to quantify posttranscriptional RNA maturation. Here we propose a complete bioinformatic and statistical pipeline to fully characterize the differences in the chloroplast transcriptome between two conditions. Experimental design as well as bioinformatics and statistical analyses are described in order to quantify both gene expression and RNA posttranscriptional maturation, i.e., RNA splicing, editing, and processing, and identify statistically significant differences.

Key words Chloroplast transcriptome, Organellar RNAseq, Differentially expressed, Splicing, Editing, Processing

1 Introduction

Reminiscent of the cyanobacterial ancestral endosymbiont genome expression, chloroplastic transcription is mainly polycistronic [1]. After transcription, in addition to cleavages, RNA molecules can be submitted to several maturation steps such as splicing, editing or modifications of their 5' and 3' ends [2]. All of these events, from transcription to translation, are highly regulated and linked. Until about 10 years ago, the best method to simultaneously quantify expression, splicing and cleavages of a chloroplastic gene was *northern* blotting. Transcript editing was separately accessed using cDNA sequencing or a more quantitative method such as Poison Primer Extension. With the development of Next Generation Sequencing (NGS) technologies allowing high-throughput RNA sequencing, it is now possible to quantify all RNA molecules

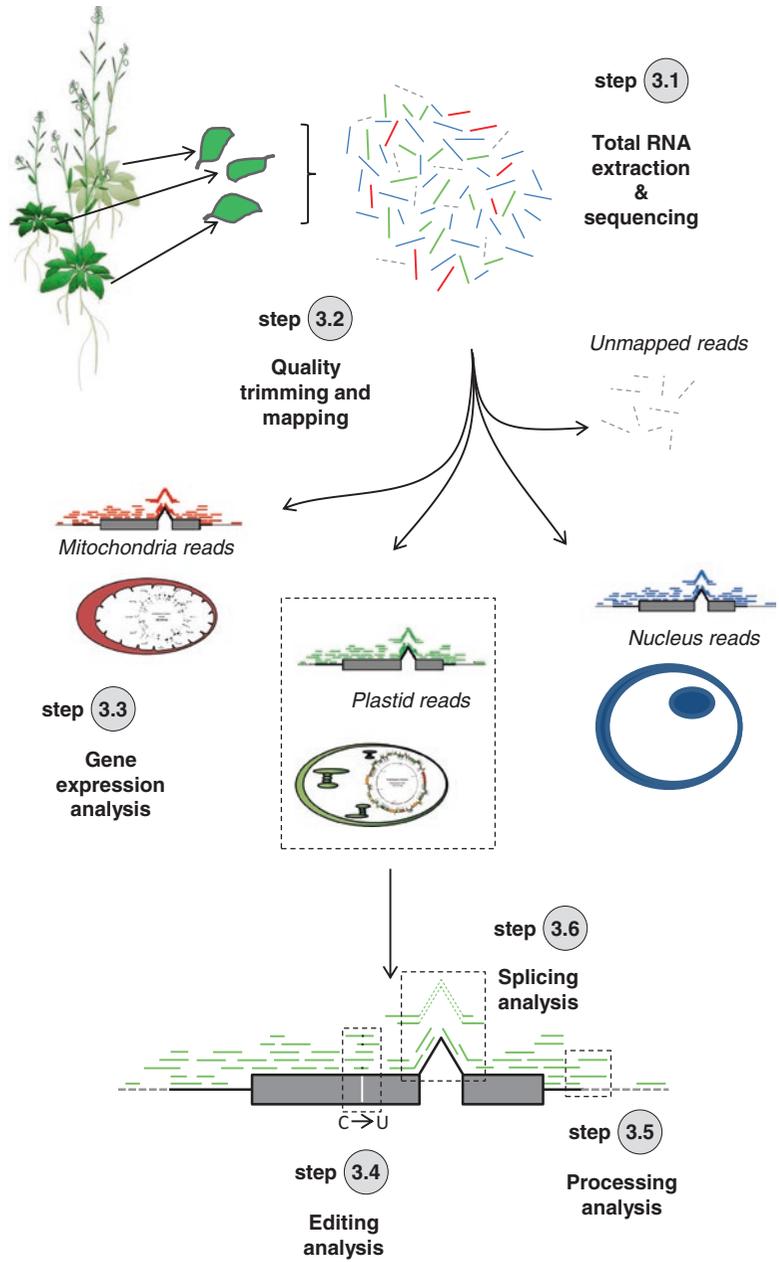


Fig. 1 Workflow of the plastid gene expression analysis pipeline

corresponding to all genes and maturation events in a single sequencing library [3]. Here, we describe a complete bioinformatical and statistical pipeline to quantify each maturation steady state of plastid genome expression in any plant for which the plastid genome is available.

This pipeline (Fig. 1) which is derived from [4] starts after the generation of the sequencing datasets which can be achieved using commercially available kits or through sequencing platforms or

companies. The quality of reads is first assessed in order to remove low quality reads and the remaining reads are then mapped on the genome. From this mapping, four procedures allow for the analysis of gene expression, RNA processing, RNA editing, and RNA splicing.

2 Materials

1. Computer running on Linux with at least 50Go of RAM, 1To of hard drive space and multiple cores.
2. Trimmomatic (<http://www.usadellab.org/cms/index.php?page=trimmomatic>).
3. STAR (<https://github.com/alexdobin/STAR>).
4. Samtools (<http://samtools.sourceforge.net/>).
5. Bedtools (<http://bedtools.readthedocs.io/en/latest/>).
6. featureCounts (<http://bioinf.wehi.edu.au/featureCounts/>).
7. pysam (<http://pysam.readthedocs.io/en/latest/api.html/>).
8. R (<https://www.r-project.org/>) with the libraries edgeR and data.table.
9. The Chloroseq suite (<https://github.com/BenoitCastandet/chloroseq>) [5].

3 Methods

3.1 Experimental Design and Generation of the Sequencing Datasets

1. To be statistically relevant, the experimental design should include the proper controls and at least three biological replicates corresponding to three independent experiments.
2. The RNA extraction can be performed with any protocol yielding intact total RNA, free of genomic DNA and pure enough for the subsequent sequencing library construction. For *Arabidopsis thaliana* and several other plant species, NucleoZOL (Macherey-Nagel) can be used, followed by a treatment with the DNase Max kit (Qiagen).
3. The plastid mRNAs are not polyadenylated, so the protocols used to construct the sequencing libraries must not select polyA RNAs but allow the analysis of all RNAs. However, given their very high abundance, ribosomal RNAs should be depleted. The TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina) can be used successfully on leaf samples.
4. Finally, the sequencing can be performed. Single-end 75–100 bases long sequencing is sufficient for this analysis. The sequencing depth will vary depending on the type of tissue but mostly on the focus of the study. When focusing on plastid

gene expression in leaves, 5 million reads per sample is sufficient given the abundance of plastid RNAs in this tissue. On the other hand, when focusing on nuclear gene expression, over 30 million reads should be estimated per sample. To avoid any possible sequencing bias, all samples should be run on the same sequencing lane. In the described protocol, samples are sequenced on a NextSeq500 (Illumina) to generate at least 30 million single-end 75 base-long reads per sample.

5. At the end of the sequencing run, base calling, demultiplexing and conversion of the sequence files to the fastq format are performed to generate one fastq file per sample.
6. All steps from the sequencing library construction to the generation of the fastq files can be done by most sequencing platforms or companies.

3.2 Quality Trimming and Mapping of the Reads

Once the fastq files are generated, the bioinformatics analysis starts by cleaning the reads to keep only high confidence bases and reads. These reads are then mapped onto the genome. These steps require computers with large amounts of RAM.

3.2.1 Quality Trimming Using Trimmomatic (v0.32)

1. In the folder containing the gunzip-compressed RNA-seq fastq file (named “input.fastq.gz” in this example) and the Illumina adapter file “TruSeq3-PE.fa” (provided with the trimmomatic software), run the command:

```
<script>
java -jar /usr/local/bio/Trimmomatic/trim-
momatic-0.32.jar SE -phred33 input.fastq.
gz output.fastq.gz ILLUMINACLIP:TruSeq3-
SE.fa:2:30:10 LEADING:15 TRAILING:15
SLIDINGWINDOW:4:20 MINLEN:30
</script>
```

2. This command will (*see Note 1*):
 - Remove adapters (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10).
 - Remove leading low quality or N bases (below quality 15) (LEADING:15).
 - Remove trailing low quality or N bases (below quality 15) (TRAILING:15).
 - Scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 20 (SLIDINGWINDOW:4:20).
 - Drop shorter than 30 bases (MINLEN:30).
 - Generate the “output.fastq.gz” file.

3.2.2 Mapping of the Reads

1. To access most aspects of plastid gene expression, a mapping of the RNA-seq reads on the genome is required. Various tools are available, such as STAR. If possible, the mapping will be done against the entire genome because the sequenced RNA is produced by the entire genome. It also avoids the erroneous mapping of some nuclear reads on the plastid genome. The fasta file(s) of the genome sequence and the GTF file of the annotation can be retrieved from <ftp://ftp.ensemblgenomes.org/pub/release-36/plants/>.
2. STAR (as many other mappers) requires the creation of an index based on the genome sequence and annotation before proceeding to the actual mapping.
3. Run the command:

```
<script>
STAR --runMode genomeGenerate --genomeDir.
--sjdbOverhang 74 --genomeFastaFiles ge-
nome_sequence.fa --sjdbGTFfile genome_anno-
tation.gtf
</script>
```
4. This command creates the index files in the current folder using the genome_sequence.fa and genome_annotation.gtf files. The --sjdbOverhang option should be set to read length-1. Then start the actual mapping:

```
<script>
STAR --readFilesCommand gunzip -c --out-
FilterMultimapScoreRange 0 --alignIntron-
Max 1 --outSAMtype BAM SortedByCoordinate
--alignEndsType EndToEnd --outFileNamePre-
fix Input --genomeDir path/to/mapping/index/
folder --readFilesIn Input.fastq.gz
</script>
```
5. The --outSAMtype BAM SortedByCoordinate option creates a bam file with the alignments sorted by coordinates which is required for some subsequent analyzes. The --alignIntronMax 1 option prevents STAR from creating de novo splicing sites. The --alignEndsType EndToEnd option ensures a stringent mapping (which requires good quality reads). The --outFilterMultimapScoreRange 0 option allows outputting only the best alignments for each read (*see Notes 2 and 3*).
6. The alignment file is then indexed:

```
<script>
samtools index Input_Aligned.sortedByCoord.
out.bam
</script>
```
7. All steps of Subheading 3.2 should be done for all the samples of interest.

3.3 Gene Expression Analysis Including-Mitochondrial and Nuclear Genes

3.3.1 Counting

1. Run the command:


```
<script>
featureCounts -M -s 2 -t exon -g gene_id -a
genome_annotation.gtf -o Input_featurecounts.
txt Input_Aligned.sortedByCoord.out.bam
</script>
```
2. This command will count alignments overlapping exons (parameter `-t exon`) and summarize the counts per gene (parameter `-g gene_id`).
3. The `-M` option allows counting reads mapping at multiple sites (*see Note 4*). The `-s 2` option allows `featureCounts` to take into account the fact that the reads generated by this protocol are in the reverse orientation.

3.3.2 Statistical Analysis

1. This statistical analysis pipeline is designed for an experiment comparing 2 conditions (mutant/wild-type or treated/control) with 3 biological replicates following the recommendations of [6] (*see Note 5*).
2. The count files are named `mut1_featurecounts.txt`, `mut2_featurecounts.txt`, `mut3_featurecounts.txt`, `WT1_featurecounts.txt`, `WT2_featurecounts.txt` and `WT3_featurecounts.txt`.

3. Using R:

```
<script>
rm(list=ls()); gc();
## read all files and merge them #####
#####
###
library(edgeR)
files <- list.files(pattern="featurecounts.txt")
samp.order <- c(grep("WT", files),
grep("mut", files))
files <- files[samp.order]
samp.names <- gsub("_.*", "", files)
for (i in 1:length(files))
{
  dat.table <- read.table(files[i],h=F,skip=2,
colClasses=c("character", rep("NULL", 5),
"integer"))
  names(dat.table) <- c("ID", samp.names[i])
  if(i == 1) data <- dat.table else
  data <- merge(data, dat.table, sort=FALSE,
all=TRUE, by="ID");
  rm(dat.table)
}
## removing non expressed and poorly expressed
genes #####
i.tmp <- rowSums(data[, -1]) > 0
```

```
countTable <- data[i.tmp, -1]
rownames(countTable) <- data[i.tmp, 1]
countPerMillions <- 1e6 * t(t(countTable) /
colSums(countTable))
ridx <- rowSums(countPerMillions > 0.001) >= 3
countTable <- countTable[ridx, ]
cat("Number of analyzed genes after edgeR
filtering: ", nrow(countTable), "\n")
## Creating the design matrix #####
#####
#####
group <- rep(c("WT", "mut"), each=3)
batch <- rep(paste("repbio", 1:3, sep="_"), 2)
batch <- as.factor(batch)
group <- as.factor(group)
group <- relevel(group, ref="WT")
design <- model.matrix(~batch+group)
cds <- DGEList(countTable, group =
paste(batch, group))
## normalization and differential analysis
#####
#####
cds <- calcNormFactors(cds)
cds <- estimateDisp(cds, design)
fit <- glmFit(cds, design)
lrt <- glmLRT(fit, coef=4)
## generation of the output file #####
#####
#####
# mut_WT_GLM_EdgeR_CompleteListe.txt
# with columns: gene ID, normalized WT
count, normalized mut count,
# and log2FoldChange, adjusted p-value (FDR)
N <- cds$samples$lib.size
f <- cds$samples$norm.factors
TMM <- N*f / mean(N*f)
norm <- scale(cds$count, center=FALSE,
scale=TMM)
baseMeanA <- rowMeans(scale(cds$counts[,
which(group == "WT") ], center = FALSE,
scale = TMM[which(group == "WT") ]))
baseMeanB <- rowMeans(scale(cds$counts[,
which(group == "mut")], center = FALSE,
scale = TMM[which(group == "mut")]))
final <- data.frame(id = rownames(norm),
WT = baseMeanA,
mut = baseMeanB,
```

```

log2FoldChange = log2(baseMeanB) -
log2(baseMeanA),
PValue = lrt$table$PValue,
FDR = p.adjust(lrt$table$PValue, "BH" )
write.table(final, "mut_WT_GLM_EdgeR_
CompleteList.txt", row.names=FALSE, col.
names=TRUE, sep="\t", quote=FALSE)
## saving the normalization factors for
later use #####
#####
GE_SF<-cds$samples$norm.factors
save(GE_SF, file="GeneExpression_SizeFactors.
RData")
</script>

```

3.4 Editing Analysis

1. This pipeline which is based on the same experimental design as previously described doesn't require the previous knowledge of the editing sites. It will test all potential editing sites.
2. It starts with the generation of a pileup file for each sample. "Name_of_plastid_chromosome" should be replaced by "Pt" when using the araport11 annotation for *Arabidopsis thaliana*.

3. Run the command:

```

<script>
pysamstats -d -D 100000000 -f genome_sequence.
fa -c Name_of_plastid_chromosome --type vari-
ation_strand Input_primary.bam > Input_pysam.
txt
</script>

```

4. The previous command line should be repeated for each sample to create the mut1_pysam.txt, mut2_pysam.txt, mut3_pysam.txt, WT1_pysam.txt, WT2_pysam.txt, and WT3_pysam.txt files.
5. The following R script will extract all the potential editing sites, i.e., Cs or Gs, with at least six edited reads over the six samples; with an average editing rate >0.5% for WT or mutant and then perform the statistical analysis. For the normalization of the editing counts, there are two options: when analyzing gene expression on the whole genome, the size factors should be calculated from this analysis and saved in the GeneExpression_SizeFactors.RData file.
6. Else, size factors can be calculated using the editing counts. These two options are described in the following script but only the first one is implemented.

7. Using R:

```

<script>
rm(list=ls()); gc();

```

```

## colums of interest #####
#####
#####
# pos, ref, T_rev, C_rev, A_fw, G_fw
# to take into account the fact that the
reads generated
# by this protocol are in the reverse ori-
entation.
colClasses <- rep("NULL", 63)
colClasses[c(2, 3, 48, 42, 35, 53)] <- NA
colOrder <- c(1, 2, 5, 4, 3, 6)
files <- list.files(pattern="_pysam.txt")
samp.order <- c(grep("WT", files),
grep("mut",files))
files <- files[samp.order]
samp.names <- gsub("_.*", "", files)
for (i in 1:length(files))
{
  dat.table <- read.table(files[i], h=T,
colClasses=colClasses)[, colOrder]
  colnames(dat.table)[3:6] <-
paste(colnames(dat.table)[3:6], samp.
names[i], sep="_")
  posG <- grep("G", dat.table[, 2])
  posC <- grep("C", dat.table[, 2])
  dat.G <- dat.table[posG, c(1, 2, 5, 6)]
  dat.C <- dat.table[posC, 1:4]
  colnames(dat.G) <- colnames(dat.C) <-
c("pos","ref",
paste0("Edited_", samp.names[i]),
paste0("Unedited_", samp.names[i]))
  # edited= A_fw, unedited=G_fw
  # edited= T_rev, unedited=C_rev
  dat.All <- merge(dat.C, dat.G, all=TRUE)
#merge(dataC,dataG,all=TRUE)
  if(i == 1) data <- dat.All else {
    data <- merge(data, dat.All[,-2], sort=FALSE,
all=TRUE, by="pos")
  }
  rm(dat.table)
}
data$pos <- paste("C", data$pos, sep="")
## selection of the potential editing sites
#####
#####
data <- data[rowSums(is.na(data)) == 0, ]
colEdited <- grep("Edited", colnames(data))
edited <- which(rowSums(data[, colEdited]) > 5)

```

```

data <- data[edited, ]
## Assumes corresponding non edited column just
after the edited column #####
colEdited.WT <- grep("Edited_WT",
colnames(data))
colEdited.mut <- grep("Edited_mut",
colnames(data))
rate.WT <- rate.mut <- numeric(nrow(data))
for(i in colEdited.WT) rate.WT <- rate.
WT + (data[, i]/(data[, i]+ data[, i+1]))/
length(colEdited.WT )
for(i in colEdited.mut) rate.mut <- rate.
mut + (data[, i]/(data[, i]+ data[, i+1]))/
length(colEdited.mut)
## keep sufficiently edited position and
format #####
#####
ok=which(rate.WT >=0.005 | rate.mut>=0.005)
data <- data[ok, ]
rownames(data) <- data[, 1]
data <- data[, -c(1,2)]
## statistical analysis #####
#####
#####
require(edgeR)
nSample <- colnames(data)
genotype <- gsub(".*_|[1-9]", "", nSample)
editing <- substr(nSample, 1, 1)
batch <- substr(nSample, nchar(nSample),
nchar(nSample))
genotype <- as.factor(genotype)
editing <- as.factor(editing)
batch <- as.factor (batch)
genotype <- relevel(genotype, ref="WT")
editing <- relevel(editing, ref="U")
design <- model.matrix(
~batch+genotype*editing)
cds <- DGEList( data, group = paste(batch,
genotype, editing))
## Normalizing using whole genome gene ex-
pression #####
#####
load("GeneExpression_SizeFactors.RData")
i_edited <- grep("Edited", colnames(data))
cds$samples$norm.factors[c(i_edited, i_ed-
ited+1)] <- rep(GE_SF, 1)

```

```

## Normalizing using sum Edited and
Unedited #####
#####
#i_edited <- grep("Edited", colnames(data))
#dataSF <- data[, i_edited] + data[, i_edited+1]
#genotypeSF <- genotype[i_edited]
#batchSF <- batch[i_edited]
#designSF <- model.matrix(
~batchSF+genotypeSF)
#cdsSF <- DGEList(dataSF, group =
paste(batchSF,genotypeSF))
#cdsSF <- calcNormFactors(cdsSF)
#cds$samples$norm.factors[c(i_edited, i_edit-
ed+1)] <- rep(cdsSF$samples$norm.factors, 1)
## Differential analysis #####
#####
#####
cds <- estimateDisp(cds, design)
fit <- glmFit(cds, design)
lrt <- glmLRT(fit, coef=6)
final <- data.frame(data,
PValue=lrt$table$PValue)
final <- data.frame(final, Bonferroni=p.adjust(
final$PValue,"bonferroni"))
## output file #####
#####
#####
write.table(final, file="Edit.txt", col.
names=TRUE, row.names=TRUE,sep="\
t",quote=FALSE)
## saving the normalization factors for
later use #####
#####
#ED_SF<- cdsSF$samples$norm.factors
#save(ED_SF,file="Editing_SizeFactors.
RData")
</script>

```

3.5 Coverage Analysis (Processing)

1. A study of the read coverage can highlight processing changes such as modifications of transcript extremities which may not be detected by the gene based differential analysis described earlier. However, instead of considering all the bases of the sequenced reads, we will only consider the starts of the reads for two reasons. First, as the reads generated by this protocol are almost all of the same length (75 bases), detecting the start of a read means that the next 74 bases are covered as well. Second, even though the statistical analysis of the coverage is not yet stabilized (and not provided in this protocol), most

statistical models will assume that counts at every base pair are independent.

2. From a biological perspective transcription at a particular base pair is likely to be dependent of transcription at the previous base pair, but counting a read at all the base pairs it covers further increase the dependence between counts at neighboring bases. It also smooths the profile reducing the perceived variance. Overall this smoothing and increased “auto-correlation” makes the statistical analysis more complex and makes it harder to distinguish true signal from artifacts.
3. The pipeline starts with extraction of the read start for each read mapping on the plastid genome.
4. Run the command:


```
<script>
bedtools bamtobed -i Input_primary.bam|awk
'{print $1 "\t" $2 "\t" $6}'|grep "Name_of_
plastid_chromosome" >coverage.bed
</script>
```
5. Use R to summarize the coverage per base. To keep the bases with no coverage, it is necessary to know the length of the plastid genome. 154478 is the length of the *Arabidopsis thaliana* plastome.
6. Run the command:


```
<script>
rm(list=ls()); gc();
library(data.table)
genome_size <- 154478
data <- fread("coverage.bed", sep="\t",
showProgress = FALSE)
colnames(data) <- c("Pt", "pos", "strand")
i_plus <- which(data$strand == "-")
i_minus <- which(data$strand == "+")
coverage <- data.frame(
  pos = 0:genome_size,
  Minus = tabulate(data$pos[i_minus]+1,
nbins=genome_size+1),
  Plus = tabulate(data$pos[ i_plus]+1,
nbins=genome_size+1)
)
rm(data);gc();
write.table(coverage, file="coverage.
txt",sep="\t")
</script>
```

3.6 Splicing Analysis

1. This analysis uses the analysis 2 of the ChloroSeq suite. This suite provides the annotations and script files required for the analysis of the *Arabidopsis thaliana* chloroplast splicing sites.

When studying a different plastome, the equivalent of the TAIR10_ChrC_exon.gff3, TAIR10_ChrC_introns.gff3 and TAIR10_ChrC_splice_sites_sort.gff3 files must be provided.

2. As the reads generated by this protocol are in the reverse orientation, the annotation files must be modified to reverse the strand of the annotations by inverting “+” and “-”. In addition, as the araport11 annotation file is used, the plastome name is “Pt” instead of “ChrC” in the ChloroSeq annotation files. These modifications (strand and plastome name) create the araport11_Pt_splice_sites_sort.gff3 araport11_Pt_exon.gff3 and araport11_Pt_introns.gff3 files.
3. The prep_bam.sh and get_splicing_efficiency.sh files of the ChloroSeq suite should be in the same folder as the annotation files and the mapping file.
4. The following command generates the splicing_efficiency.txt file with the columns: Chromosome, intron start, intron end, spliced count, unspliced count, splicing efficiency.


```
<script>
chloroseq.pl -a 2 -b Input_primary.bam -e
araport11_Pt_exon.gff3 -i araport11_Pt_in-
trons.gff3 -n Pt -s araport11_Pt_splice_
sites_sort.gff3
</script>
```
5. When experimental design is composed of six samples (three replicates of mutant and three replicates of wild-type), the following R script can be used.
6. Files should be named WT1_splicing.txt to WT3_splicing.txt and mut1_splicing.txt to mut3_splicing.txt. As the number of splicing event is low, the calculation of the size factors for normalization by edgeR is not adapted. So either the size factors should be estimated during the whole genome gene expression analysis and saved in the GeneExpression_SizeFactors.RData file (recommended and implemented in the following script) or the size factors should be estimated during the editing analysis.
7. Run the command:

```
<script>
rm(list=ls()); gc();
## load data #####
#####
files <- list.files(pattern="_splicing.txt")
samp.order <- c(grep("WT", files), grep("mut",
files))
files <- files[samp.order]
samp.names <- gsub("_.*", "", files)
for (i in 1:length(files))
```

```

{
  raw <- read.table(files[i], header=FALSE,
colClasses=c("NULL", NA, NA, NA, NA,
"NULL"))
  colnames(raw) <- c("int_start", "int_end",
paste0("Spliced_", samp.names[i]),
paste0("Unspliced_", samp.names[i]))
  if(i == 1) data <- raw else {
    data <- merge(data, raw[, -2], sort=FALSE,
all=TRUE, by="int_start")
  }
}
data <- data[, -2]
row.names(data) <- data[, 1]
data <- data[,-1]
## statistical analysis and experimental
design ##
require(edgeR)
nSample <- colnames(data)
genotype <- gsub(".*_|[1-9]", "", nSample)
splicing <- substr(nSample, 1, 1)
batch <- substr(nSample, nchar(nSample),
nchar(nSample))
genotype <- as.factor(genotype)
splicing <- as.factor(splicing)
batch <- as.factor(batch)
genotype <- relevel(genotype, ref="WT")
splicing <- relevel(splicing, ref="U")
design <- model.matrix( ~batch +
genotype*splicing)
cds <- DGEList( data, group = paste(batch,
genotype, splicing))
## Normalizing using whole genome gene ex-
pression #####
#####
load("GeneExpression_SizeFactors.RData")
i_spliced <- grep("Spliced", colnames(data))
cds$samples$norm.factors[c(i_spliced, i_
spliced+1)] <- rep(GE_SF, 1)
## Normalizing using editing analysis #####
#####
load("Editing_SizeFactors.RData")
#i_spliced <- grep("Spliced",
colnames(data))
#cds$samples$norm.factors[c(i_spliced, i_
spliced+1)] <- rep(ED_SF, 1)
## Differential analysis ##

```

```
cds <- estimateGLMCommonDisp(cds,design)
cds <- estimateGLMTagwiseDisp(cds,design)
fit <- glmFit(cds,design)
lrt <- glmLRT(fit, coef=6)
final <- data.frame(raw,
PValue=lrt$table$PValue)
final <- data.frame(final, Bonferroni=p.adjust(
final$PValue,"bonferroni"))
## output file ##
write.table(final, file="splicing.txt",
col.names=TRUE, row.names=FALSE, sep="\t",
quote=FALSE)
</script>
```

4 Notes

1. The remaining rRNA reads are not removed because the depletion is usually very efficient and because tools such as sort-meRNA tend to remove some genuine nonribosomal plastid reads in addition to ribosomal reads.
2. It is possible to parallelize the mapping using the parameter `--runThreadN` to run it on multiple cores.
3. There is a counting option included in STAR but it is not adapted for the study of plastid gene expression because it discards reads mapping at multiple loci which includes all reads transcribed from the inverted repeats.
4. Although not recommended in general, it is often required for the analysis of plastid gene expression because of the quadripartite structure of the plastid genome of numerous plant species.
5. This statistical pipeline is optimized for the simultaneous analysis of a large number of genes (nuclear and organelle genes). Users should be cautious when using it for the analysis of plastid genes only.

Acknowledgments

This work has benefited from a French State grant (LabEx Saclay Plant Sciences-SPS, ANR-10-LABX-0040-SPS), managed by the French National Research Agency under an “Investments for the Future” program (ANR-11-IDEX-0003-02).

References

1. Barkan A (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol* 155:1520–1532
2. del Campo EM, Sabater B, Martín M (2002) Post-transcriptional control of chloroplast gene expression. *J Biol Chem* 277:36457–36464. <https://doi.org/10.1074/jbc.M204500200>
3. Han Y, Guo S, Muegge K et al (2015) Advanced applications of RNA sequencing and challenges. *Bioinform Biol Insights* 9:29. <https://doi.org/10.4137/BBI.S28991>
4. Guillaumot D, Lopez-Obando M, Baudry K et al (2017) Two interacting PPR proteins are major *Arabidopsis* editing factors in plastid and mitochondria. *Proc Natl Acad Sci U S A* 114:8877–8882. <https://doi.org/10.1073/pnas.1705780114>
5. Michel JSM, Hotto AM, Strickler SR, Stern DB, Castandet B (2018) A guide to the chloroplast transcriptome analysis using RNA-Seq. *Meth Mole Bio* 1829:00–00. <https://github.com/BenoitCastandet/chloroseq>
6. Rigaille G, Balzergue S, Brunaud V et al (2016) Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief Bioinform* 19(1):65–76. <https://doi.org/10.1093/bib/bbw092>

1.2. Pipeline appliqué à l'étude du transcriptome de la mitochondrie

Il est à noter que le même script a été utilisé dans l'analyse du transcriptome mitochondrial, en modifiant certaines parties :

Pour l'édition :

- l'annotation « Pt » a été remplacée par « Mt »
- le fichier genome_sequence.fa doit également comprendre la séquence du génome mitochondrial d'*Arabidopsis*

Pour la couverture :

- la taille du génome doit être modifiée. Le génome mitochondrial de Col0 fait 366750pb

Pour l'épissage :

L'analyse d'épissage nécessitant la suite ChloroSeq, il est nécessaire d'utiliser une autre méthode pour la mitochondrie (Castandet et al., 2016a). Nous avons téléchargé les fichiers TAIR10_ChrM_exon.gff3, TAIR10_ChrM_introns.gff3 et TAIR10_ChrM_splice_sites_sort.gff3 et utilisé le script suivant :

```
<script>
intersectBed -abam WT1_Aligned.sortedByCoord.out.bam -b TAIR10_ChrM_splice_sites_sort.gff3 -s | intersectBed -abam stdin -b TAIR10_ChrM_introns.gff3 | intersectBed -abam stdin -b TAIR10_ChrM_exon.gff3 > splice_junctions.out

samtools view -h splice_junctions.out | awk '$6 ~ /N/ || $1 ~ /^@/' | samtools view -bS - | coverageBed -abam stdin -b TAIR10_ChrM_introns.gff3 | sort -k 4,4n | cut -f 4,5,10 > spliced_coverage.txt

samtools view -h splice_junctions.out | awk '$6 !~ /N/ || $1 ~ /^@/' | samtools view -bS - | coverageBed -abam stdin -b TAIR10_ChrM_introns.gff3 | sort -k 4,4n | cut -f 4,10 > unspliced_coverage.txt

join spliced_coverage.txt unspliced_coverage.txt | awk '{print $1 "\t" $2 "\t" $3 "\t" $4}' > WT1_mito_splicing_efficiency.txt
</script>
```

L'analyse statistique est ensuite la même que pour le chloroplaste.

2. Discussion

De nombreux choix, paramètres ou méthodes statistiques ne sont pas détaillés dans la publication. Les paragraphes suivants ont pour but de donner plus de détails sur les tests et paramètres utilisés, tout en mettant en évidence leurs éventuelles limites.

2.1. Génération des données

La sélection des ARN polyA+ n'est pas une option pour l'analyse du transcriptome des organites car leurs ARN ne sont pas polyadénylés. Cependant, cette sélection est également une méthode efficace pour éliminer les ARN ribosomiaux qui représentent sinon 95 à 99% des *reads*. C'est pourquoi les banques sont construites

en effectuant une déplétion des ARN ribosomiaux. Au lieu de sélectionner les ARN d'intérêts, les ARN indésirables sont éliminés. Cela permet de disposer de tous les ARN, qu'ils soient matures ou non. L'intérêt pour nous est important car nous souhaitons justement observer les différentes étapes de la maturation des ARN.

Disposant d'un génome bien connu, nous avons choisi de travailler en séquençage *Single End* car le surcoût d'un séquençage *Paired End* n'est pas compensé par la légère augmentation de longueur et de qualité des séquences. Malgré la bonne qualité du séquençage, certaines séquences de faible qualité peuvent perturber l'étude d'évènements rares et/ou peu couverts. Afin de limiter ce problème on effectue un *trimming* après séquençage. Cela consiste en un contrôle qualité des *reads* sur la base du score de fiabilité de chaque base. Les *reads* sont ensuite placés sur le génome et pas seulement le transcriptome lors de l'alignement par le logiciel STAR (Dobin et al., 2013). L'alignement sur génome présente l'intérêt de permettre l'observation des évènements intergénomiques et introniques. C'est d'autant plus intéressant lorsque l'on travaille sur le chloroplaste et la mitochondrie puisque, chez *Arabidopsis thaliana*, près de la moitié des *reads* alignés sur le génome des organites l'est en région intergénomique et que la quasi-totalité du génome est transcrit. Cet alignement est fait sur un génome annoté, c'est-à-dire qu'il contient toutes les informations connues à ce jour (position des exons, des introns, évènements d'édition, etc.). Cela permet de faciliter le mapping tout en limitant très largement les artefacts d'épissage associés à des erreurs de mapping du logiciel. La dernière étape est celle du comptage du nombre d'alignements sur une entité génétique (gène, intron...). Ce nombre ne correspond pas forcément au nombre de *reads* car certains d'entre eux peuvent s'aligner à plusieurs endroits dans le génome (« *multihit* »). Contrairement à d'autres équipes, nous avons choisi de conserver ces *multihits* car sinon les régions inversées répétées du plaste ne seraient pas analysées de même qu'une partie du génome mitochondrial en raison de la copie présente dans le chromosome 2 chez l'écotype Col0.

2.2. Analyse de l'expression des gènes

La première chose à faire pour commencer une analyse différentielle d'expression des gènes est, à la suite de l'alignement des *reads* sur le génome annoté, de réaliser un comptage de ces *reads* pour chaque gène. Bien que cela semble simple, il existe de très nombreux paramètres de comptages, et de nombreux logiciels le réalisant. De nombreuses questions se posent. Par exemple, comment attribuer un *read* à un gène lorsque celui-ci est en bordure de la séquence codante ? Quel poids donner à un *read* lorsque celui-ci est aligné à plusieurs positions sur le génome en même temps ? Nous avons utilisé le logiciel *featureCounts* dans notre analyse (Liao et al., 2014), en donnant la même valeur, c'est-à-dire 1, à chaque alignement indépendamment du nombre de fois où le *read* correspondant a été aligné. La raison de ce choix est que le pipeline d'analyses statistiques utilisé par la suite n'accepte en entrée que des nombres entiers (pour des raisons mathématiques et non informatiques). Il serait intéressant de comparer si l'utilisation d'un poids proportionnel au nombre d'alignements d'un même *read* sur le génome permet une analyse plus fine malgré un arrondi vers l'entier le plus proche afin de permettre l'analyse statistique.

Une fois les tables de comptage obtenues, l'analyse statistique peut commencer. L'objectif de l'analyse de l'expression des gènes est de trouver des gènes DE entre le contrôle et, dans notre cas de figure, un mutant. Il existe 32735 ORF d'*Arabidopsis* répertoriés par l'AGI (*Arabidopsis Genome Initiative*), qui sont comparés entre deux conditions, avec 3 réplicats biologiques par condition. La comparaison étant réalisée gène par gène, il n'est pas nécessaire de réaliser une correction sur la taille du gène. On se retrouve donc dans des conditions de tests multiples trop lourds pour être réalisés par ANOVA ou test de Student. De plus, en se basant classiquement sur un risque acceptable d'erreur de type I (ou faux-positif) à 5%, à partir de 20 tests on obtient 64% de chances de trouver un résultat significatif par chance. Il est donc nécessaire de trouver une méthode statistique adaptée à la dimension des approches transcriptomiques tout en limitant de façon nette les erreurs de type I.

On pose donc l'hypothèse nulle H_0 : « la différence d'expression entre les deux traitements est nulle » et l'hypothèse alternative H_1 : « la différence d'expression entre les deux traitements est non nulle »

La suite de l'analyse consiste ensuite à filtrer les gènes peu ou pas exprimés et on calcule un facteur de normalisation afin de tenir compte des différentes tailles de banques de séquençage. Historiquement, ce facteur de normalisation (différent pour chaque échantillon) était calculé de façon très simple en divisant le nombre de *reads* pour un gène donné dans un échantillon donné par le nombre de *reads* total de cet échantillon. Le problème que posait cette méthode de calcul des facteurs de normalisation était qu'elle partait de l'hypothèse que la variance était la même pour tous les gènes, ce qui est faux. Imaginons que l'échantillon 2 a deux fois plus de *reads* que le 1. En appliquant cette méthode, on divise par deux les *reads* dans l'échantillon 2. Mais si cette différence est due à quelques gènes très exprimés dans l'échantillon 2 et non dans le 1 ? Alors la quantification de l'expression de tous les gènes dans l'échantillon 2 en pâtirait. La méthode utilisée dans notre publication est celle du TMM pour *Trimmed Mean of M-values* (Robinson et Oshlack, 2010). Les hypothèses sur lesquelles se base ce modèle sont que la majorité des gènes ne sont pas différentiellement exprimés et que le total de *reads* est extrêmement dépendant de quelques gènes très exprimés. Ce modèle élimine donc du calcul du facteur de normalisation (et non de l'analyse différentielle) les gènes ayant une expression et un Log2FoldChange trop extrêmes par rapport à la moyenne de tous les gènes.

Afin d'appliquer des tests statistiques connus, l'étape suivante est de choisir la distribution ressemblant le plus à des données de *RNAseq*. La distribution utilisée pour des données de *RNAseq* est une loi binomiale négative. Elle permet de prendre en compte la surdispersion due aux réplicats biologiques en passant par un modèle linéaire généralisé : on parle alors de régression binomiale négative de type 2. Pour expliquer cela de façon plus simple, prenons l'exemple de Y , qui suit une loi binomiale négative de paramètres μ et \emptyset :

$$Y \sim NB(\mu, \emptyset)$$

$$\text{Avec } E(Y) = \mu \text{ et } V(Y) = \mu(1 + \emptyset\mu)$$

Comme on peut le voir, l'espérance de Y ne permet pas d'observer tous les effets, comme l'effet génotype ou réplicat. L'idée est de décomposer la moyenne en passant par un modèle linéaire généralisé via une fonction de lien, appelons la λ :

$$E(\lambda(Y)) = \text{constante} + \text{effet génotype} + \text{effet réplicat}$$

L'intérêt majeur du modèle linéaire généralisé dans notre cas de figure est de nous permettre d'évaluer directement l'effet génotype suite à la décomposition du signal en plusieurs facteurs, tout en tenant compte de l'effet réplicat (Rigaill et al., 2016).

EdgeR permet donc de calculer les différents facteurs de normalisation et d'effectuer l'analyse différentielle. On obtient pour chaque gène une *p-value* brute, c'est-à-dire la probabilité d'observer nos données si l'hypothèse nulle H_0 est vraie. On parle ici de *p-value* brute, par opposition à la *p-value* ajustée correspondant à celle obtenue après correction des tests multiples. L'étude de la distribution de ces *p-values* brutes permet de s'assurer de la qualité de l'analyse statistique (Figure 14).

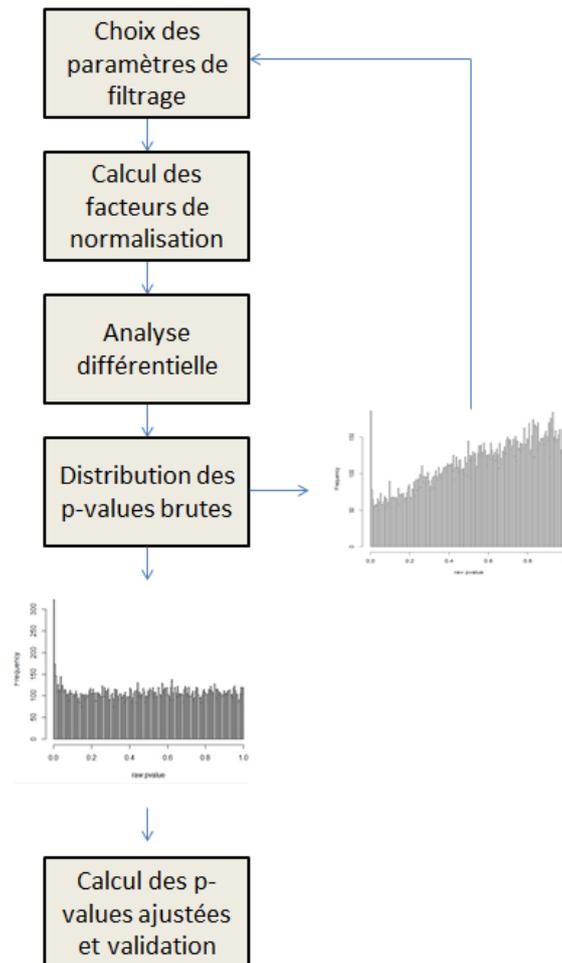


Figure 14 : Représentation schématique des principales étapes de l'analyse différentielle.

Le résultat attendu est une distribution uniforme de la fréquence des p -values, avec idéalement un enrichissement à 0, suggérant la présence hypothétique de gènes différentiellement exprimés. Si ce n'est pas le graphique observé, cela signifie qu'il y a probablement un problème dans l'analyse et que le modèle choisi ne correspond pas à nos données. Il est possible de modifier les paramètres de filtrage afin de rapprocher nos données du modèle utilisé, voire de réfléchir à un éventuel effet qui aurait été oublié. Si la distribution des p -values est homogène hors proximité de 0, on peut alors calculer les p -values ajustées afin de déterminer les gènes différentiellement exprimés et tenir compte des tests multiples, comme nous l'avons vu précédemment.

Le principe des corrections de p -values est simple : après la définition d'un critère de sélection et d'un objectif, les valeurs de nos p -values sont modifiées pour répondre à notre critère de sélection. Différentes corrections permettent de calculer les p -values ajustées. La première est celle de Bonferroni. L'objectif derrière cette correction est de n'avoir aucun faux positif. Bien que très fiable, cette méthode est très stringente et risque de nous faire passer à côté d'un certain nombre de gènes DE. La seconde méthode est celle de Benjamini-Hochberg. L'objectif est de ne pas avoir plus de $x\%$ de faux positifs parmi les gènes proposés comme DE, avec x habituellement fixé à 5% , et appelé FDR (pour *False Discovery Rate*). Ici, le risque est tout autre : on passera à côté de beaucoup moins de gènes DE qu'après une correction de Bonferroni, mais $x\%$ de nos gènes DE risquent d'être des faux positifs. Cette seconde méthode de correction est celle choisie dans notre analyse différentielle d'expression des gènes.

2.3. Analyse de couverture

L'idée derrière l'analyse de couverture est de regarder la répartition des *reads* tout au long du génome indépendamment de l'annotation, pour mettre en évidence des défauts pas nécessairement observables avec l'analyse d'expression des gènes. Cette notion est d'autant plus pertinente lorsque l'on s'intéresse à la transcription des organites, dont la quasi-totalité du génome est transcrite. L'analyse de couverture permet donc de voir les variations de couverture entre deux conditions pour un gène donné (Figure 15) mais également dans les zones non codantes.



Figure 15 : Représentation schématique d'un cas hypothétique et extrême démontrant l'importance d'une analyse de couverture en complément d'une analyse d'expression différentielle.

La publication propose un pipeline pour représenter ces données, mais pas de traitement statistique permettant de mettre en évidence des variations d'accumulation/de couverture entre deux conditions. Bien que nécessitant encore quelques ajustements, une méthode a été développée à l'IPS2 par Guillem Rigai pour mettre en évidence ces variations. La méthode s'appelle l'analyse par segmentation. L'idée est de compter les débuts de chaque *read* le long du génome ce qui permet d'obtenir un signal. Il est inutile de conserver l'intégralité du *read* car dans leur majorité, ils font environ 75pb, ce qui implique qu'il y aura du signal sur les bases suivantes et induire de l'autocorrélation, c'est-à-dire une période pendant laquelle le signal sera déterministe, lié au signal sur les bases précédentes, risquant au final de brouiller les données. L'analyse de segmentation détermine ensuite les changements abrupts dans la moyenne de ce signal, appelés ruptures. Pour cela, le signal est approximé comme une fonction constante par morceaux, et un bruit. Ainsi, ce signal le long du génome est découpé en différents morceaux de façon à ce que chaque morceau ait une moyenne homogène. C'est à cette étape que réside l'essentiel de la difficulté de cette méthode. En effet, de très nombreux découpages sont possibles et le bruit peut s'avérer très problématique de même que la prise en compte des répétitions biologiques. Mais les développements de cette méthode dépassent de très loin le cadre de ma thèse. Pour notre étude, j'ai appliqué l'analyse de segmentation sur chacun des échantillons, indépendamment. A partir de ces données, un modèle commun regroupant toutes les ruptures trouvées sur les différents échantillons est construit : à partir du moment où l'on trouve une rupture dans un échantillon, à la base près, celle-ci est représentée dans le modèle commun. L'idée derrière cette analyse est de refaire une annotation, en recherchant des zones de signal homogène, constant, et ce quel que soit l'échantillon. Sur la base de ce modèle commun, de cette « annotation », on réalise les comptages par segment sur chacun des échantillons et l'analyse statistique est la même que pour l'analyse de l'expression des gènes. En comparant le mutant à une condition contrôle, l'objectif est de trouver des segments dont l'intensité diffère selon la

condition observée, pouvant indiquer des variations dans le début/la fin de la transcription, des zones non exprimées, etc.

Nous avons utilisé cette même analyse par segmentation pour la recherche d'empreintes PPR en analysant les petits ARN (Annexe S1). En se fixant, les protéines PPR vont potentiellement protéger le site de fixation de la dégradation par des exonucléases. En comparant deux conditions, sauvage et mutant/surexprimeur, l'idée est de trouver l'ARN fixé par la PPR et la zone de fixation à quelques bases près. Le protocole est très similaire à celui décrit plus haut mais avec quelques modifications. Comme les petits ARN sont de taille variable, la segmentation est faite séparément sur les débuts et les fins de *reads*. Ensuite, l'apparition/disparition d'une empreinte doit se traduire par un segment de débuts de *reads* suivi d'un segment de fins de *reads* augmentant/baissant ensemble dans les mêmes proportions.

2.4. Analyse d'épissage

L'idée est de comparer les efficacités d'épissage de différents sites, entre un mutant et un contrôle. L'efficacité d'épissage est calculée en comparant sur un même site les ARN épissés et non épissés. A partir des données de *RNAseq* dont l'alignement a été réalisé sur le génome, les *reads* exoniques passant à travers l'intron, représentant les ARN épissés, et les *reads* à cheval entre l'intron et l'exon, représentant les ARN non épissés sont extraits (Figure 16).

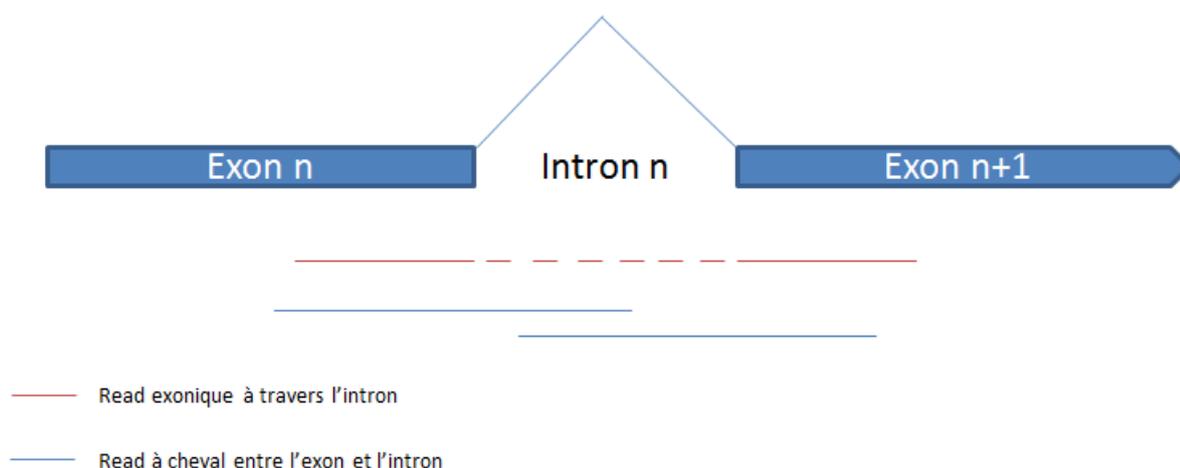


Figure 16 : Représentation schématique des différents types de *reads* à proximité d'un évènement d'épissage.

En pratique, on calcule cette valeur en divisant la quantité de *reads* exoniques à travers un intron par la somme de ces *reads* et de la moitié des *reads* à cheval entre l'exon et l'intron.

$$\text{Efficacité d'épissage} = \frac{\text{reads exoniques à travers l'intron}}{\text{reads exoniques à travers l'intron} + (\text{reads à cheval entre l'exon et l'intron})/2}$$

Etant donné que la taille des *reads* est autour de 75pb et qu'il n'existe pas d'intron de moins de 75pb dans les 37 évènements connus dans le chloroplaste et la mitochondrie d'*A. thaliana* Col0, cela signifie que l'on va trouver des *reads* à cheval entre l'intron et l'exon du côté de l'exon n comme du côté de l'exon n+1. Afin de ne pas les prendre en compte deux fois, cette valeur est divisée par deux.

L'analyse statistique réalisée par la suite est, comme pour l'analyse d'édition, basée sur l'analyse statistique de l'expression des gènes. Deux différences majeures sont cependant à mettre en évidence. La première sera la décomposition de l'espérance :

$$E(\lambda(Y)) = \text{constante} + \text{effet réplicat} + \text{effet génotype} + \text{effet épissage} + \text{interaction}(\text{génotype} * \text{épissage})$$

Dans ce cas de figure, on s'intéresse à l'effet épissage en fonction du génotype, c'est-à-dire leur interaction. Cela signifie que l'action simultanée des deux variables n'est pas additive. Afin de mieux comprendre la notion d'interaction, on peut prendre l'exemple d'un chien laissé seul dans la cuisine, puis d'un chat laissé lui aussi seul dans la cuisine. Il est assez peu probable que la présence simultanée du chien et du chat dans la cuisine soit équivalente à la somme de leurs présences individuelles. Il y a donc interaction entre les effets chien et chat. Dans notre analyse, nous avons donc dû modifier notre modèle en conséquence pour l'analyse d'épissage par rapport à l'analyse d'expression.

La seconde grande différence avec l'analyse d'expression est le nombre d'évènements d'épissage étudiés. Comme la valeur est très faible dans les organites et que l'on ne travaille qu'avec 3 échantillons par condition, il est délicat de modéliser la variance à partir de si peu d'évènements. On peut alors se poser la question de l'intérêt d'une méthode statistique aussi lourde quand des tests comme l'ANOVA pourraient suffire à notre analyse. L'intérêt d'utiliser un modèle linéaire généralisé est que, même si notre modélisation de la variance peut en souffrir, il est toujours plus intéressant de conserver les informations disponibles sur tous les évènements d'épissage plutôt que de ne regarder que la variance par site d'épissage comme le ferait une ANOVA. L'analyse étant moins fine que pour l'expression des gènes, le choix a été fait d'effectuer une correction de tests multiples de Bonferroni afin d'être plus stringent.

2.5. Analyse d'édition

L'analyse d'édition va rechercher des variations d'édition entre les différentes conditions. L'avantage de notre méthode est qu'elle ne requiert pas de fichier d'annotation répertoriant les différents sites d'édition. Elle permet donc de trouver les défauts *de novo*, ou non encore répertoriés. La sélection des sites potentiels d'édition est basée sur la présence d'au moins 6 *reads* édités sur les 6 échantillons (3 réplicats biologiques par condition) avec un taux d'édition strictement supérieur à 0.5% chez le contrôle ou le mutant. Ces seuils, arbitraires, ont été obtenus de façon empirique en observant les graphiques de *p-values* brutes, à l'image de ce que nous avons fait pour l'analyse d'expression. L'analyse statistique qui s'en suit est identique à celle effectuée pour l'épissage, avec les mêmes avantages et les mêmes limites, à un niveau moindre cependant car il y a plus d'évènements d'édition (67 connus dans le chloroplaste et 723 chez la mitochondrie d'*Arabidopsis thaliana* écotype Col0 selon Guillaumot et al. (2017)) que d'épissage.

Le développement de ces outils par le travail conjoint de la plateforme de transcriptomique, de l'équipe *Genomic Network* et de l'équipe *Organellar Gene Expression* a très grandement contribué à mon travail de thèse, notamment au chapitre suivant qui propose une méthode affinée de prédictions de site(s) de fixation sur l'ARN pour les PPR du sous-groupe E+.

Chapitre 2 : Amélioration du système de prédiction de sites d'édition sous dépendance de PPR

E+

1. Mise en contexte

1.1. Etude de CLB19, DYW2 et NUWA

CLB19 (pour *Chloroplast Biogenesis 19*) est une PPR E+ décrite comme nécessaire à l'édition des transcrits *rpoA* et *clpP* dans le chloroplaste (Chateigner-Boutin et al., 2008). Afin de mieux comprendre le fonctionnement et la composition de l'éditosome d'*A. thaliana*, Guillaumot et al. (2017) ont utilisé une protéine CLB19 fusionnée à deux étiquettes, séparées par un site de clivage (Figure 17).

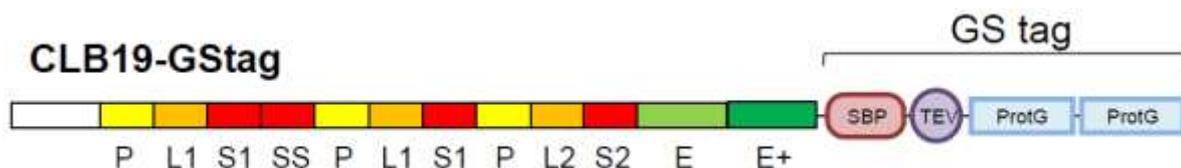


Figure 17 : Représentation schématique de la protéine de fusion CLB19-GS tag.

Tiré de Guillaumot et al. (2017).

Les auteurs ont ensuite utilisé la méthode TAP (pour *Tandem Affinity Purification*) afin de récupérer les protéines interagissant *in vivo* avec CLB19. Le principal intérêt de cette méthode est de faire une première immunoprécipitation, un clivage pour éliminer la première étiquette, puis une seconde immunoprécipitation, assurant ainsi un certain niveau de pureté des extraits obtenus. Deux PPR furent mises en évidence dans leurs travaux : NUWA (AT3G49240) et DYW2 (AT2G15690). NUWA est une PPR pure, alors que DYW2 fait partie des PPR-DYW (Figure 18).

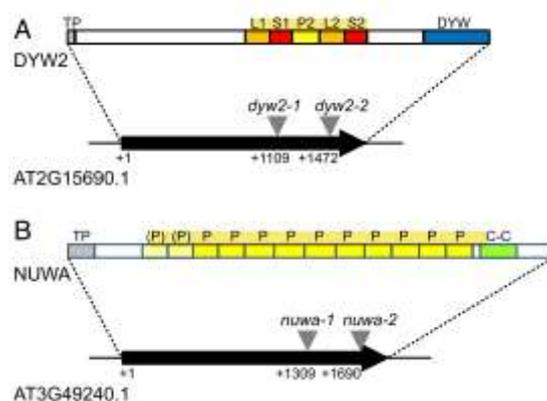


Figure 18 : Représentation schématique des PPR interagissant avec CLB19, par la méthode TAP.

Tiré de Guillaumot et al. (2017).

L'interaction NUWA-DYW2 fut également confirmée en double hybride chez la levure. En revanche, dans ce système, les interactions avec CLB19 n'ont pas pu être confirmées et il faut noter qu'un crible de la banque de 12000 protéines d'*Arabidopsis* n'a permis d'identifier aucun interagissant de CLB19. Concernant les localisations subcellulaires, si CLB19 est une PPR chloroplastique, NUWA et DYW2 colocalisent dans le chloroplaste et la mitochondrie.

1.2. Analyse transcriptomique des mutants $\Delta nuwa$ et $\Delta dyw2$

Deux lignées d'insertion d'ADN-T pour NUWA et deux pour DYW2 furent étudiées. Les deux protéines semblent être primordiales pour la plante dans les étapes embryonnaires puisque aucun homozygote $\Delta nuwa$ et $\Delta dyw2$ ne put être obtenu. Pour contourner l'embryoléthalité, les auteurs complémentèrent les lignées hétérozygotes $\Delta nuwa$ et $\Delta dyw2$ avec NUWA et DYW2 sous contrôle d'un promoteur ABI3. Ce promoteur est sensé ne fonctionner que lors de l'embryogénèse, et Guillaumot et al. confirmèrent l'extinction du gène au stade adulte par RT-qPCR. Ils réalisèrent ensuite un séquençage des ARN nucléaires, chloroplastiques et mitochondriaux après déplétion ribosomale des deux mutants. Avec les méthodes d'analyse décrites dans le chapitre 1, ils observèrent de nombreuses choses. 392 sites d'édition –chloroplastiques et mitochondriaux– furent trouvés différentiellement édités chez $\Delta dyw2$ par rapport au Col0, contre 223 chez $\Delta nuwa$ avec 166 en commun aux deux (Figure 19). Il est nécessaire de remarquer qu'un site différentiellement édité ne signifie pas nécessairement que son taux d'édition chez le mutant diminue. Fait extrêmement intéressant : 17 sites dépendants de PPR E+ sur 35 analysés sont affectés chez $\Delta nuwa$, et tous le sont chez $\Delta dyw2$.

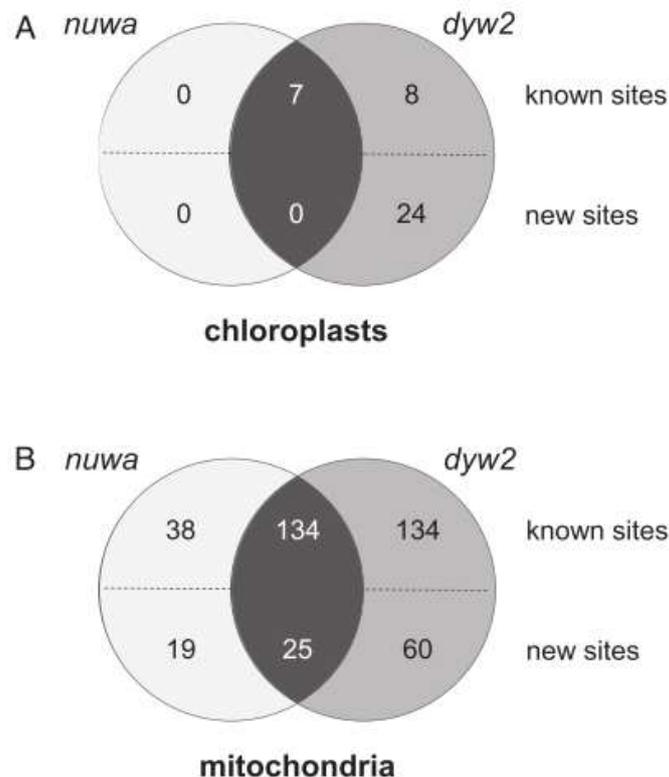


Figure 19 : Résumé des défauts d'édition trouvés chez les mutants d'insertion $\Delta nuwa$ et $\Delta dyw2$.

Tiré de Guillaumot et al. (2017).

On considère un site d'édition comme étant dépendant de DYW2 et NUWA lorsque le défaut d'édition observé chez le mutant pour ce site correspond à une diminution de 10% ou plus de l'édition observée chez la plante sauvage. Par exemple, pour un site donné édité à 90% chez Col0, un niveau d'édition inférieur ou égal à 81% chez le mutant $\Delta dyw2$ signifie que ce site est dépendant de DYW2. Pour rappel, 67 sites d'édition sont connus à l'heure actuelle chez le chloroplaste et 723 chez la mitochondrie (Guillaumot et al., 2017).

A la suite de cette publication, un défaut dans le pipeline d'analyse bio-informatique lié à l'alignement des *reads* sur la mitochondrie a été mis en évidence. Certains *reads* s'alignant aussi bien sur le génome nucléaire

que mitochondrial n'étaient pris en compte que pour le génome nucléaire, créant artificiellement des zones non couvertes sur la mitochondrie. Ce biais n'invalide pas les résultats obtenus par Guillaumot et al. (2017) mais sa correction permet maintenant d'avoir des informations sur le niveau d'édition de tous les sites connus jusqu'à présent.

Suite à cette nouvelle analyse, nous disposons de nouvelles données :

- 272 sites sont différenciellement édités chez *Anuwa*, dont 215 ont une baisse d'édition inférieure ou égale à 10% de l'édition chez le sauvage. Ces derniers sont dits dépendants de NUWA
- 531 sites sont différenciellement édités chez *Δdyw2*, dont 291 ont une baisse d'édition inférieure ou égale à 10% de l'édition chez le sauvage. Ces derniers sont dits dépendants de DYW2

La nouvelle analyse couvre l'ensemble des sites ayant une PPR associée. Parmi les sites dépendants de DYW2, on retrouve les 38 sites publiés de PPR E+, et seuls 3 sites d'édition associés à des PPR non E+ connues font figure d'exception car aussi dépendants de DYW2 :

- M234052, dépendant de PPME, une PPR pure (Leu et al., 2016)
- M17869, dépendant de MEF19, une PPR E (Takenaka et al., 2010)
- M233897, dépendant de MEF57, une PPR DYW (Andrés-Colás et al., 2017)

Parmi les sites dépendants de NUWA, on retrouve 17 des 38 sites E+ publiés, mais aussi les sites dépendants de MEF20 (PPR E), MEF31 (PPR E), MEF57 (PPR DYW) et 3 sites de MEF32 (PPR E). Enfin, parmi les sites dépendants à la fois de DYW2 et NUWA, on trouve 17 sites E+ et le site de MEF57.

1.3. Hypothèse et projet de recherche

Pour rappel, le motif DYW est largement suspecté de jouer le rôle d'enzyme ayant l'activité d'édition par la communauté scientifique. Cependant, comme le rappellent Barkan and Small (2014), des PPR sans domaine DYW, à l'image des PPR E+, sont capables d'éditer l'ARN. L'hypothèse avancée par Guillaumot et al. est que NUWA et DYW2, en s'associant à une PPR E+ jouant le rôle de facteur de spécificité c'est-à-dire de reconnaissance de la zone d'ARN qui sera fixée, font partie de l'éditosome des sites E+ d'*Arabidopsis* et apporteraient l'activité cytidine déaminase portée par le domaine DYW (Figure 20).

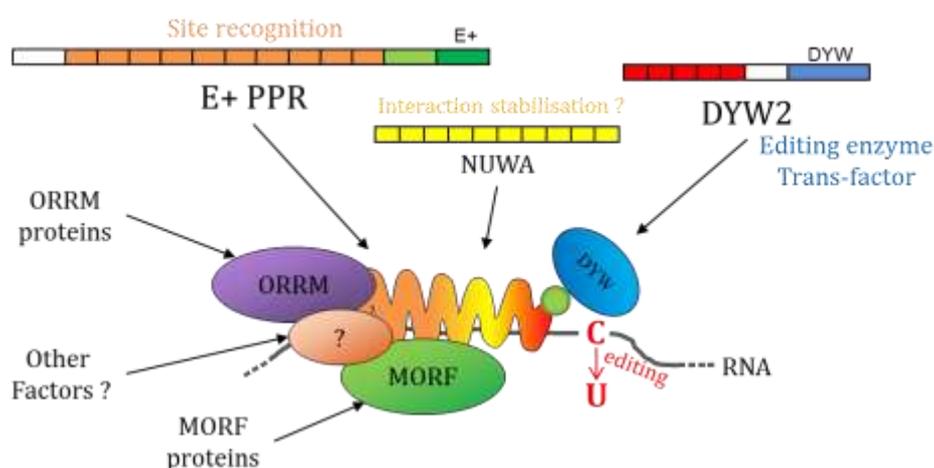


Figure 20 : Modèle de l'éditosome des PPR E+ chez *Arabidopsis thaliana* proposé par Guillaumot et al.

Illustration de D. Guillaumot et K. Baudry.

L'association fonctionnelle entre un site d'édition et une protéine PPR est classiquement identifiée par génétique inverse à l'aide éventuellement du code PPR qui permet de cibler la recherche de défauts d'édition. Le code PPR permet de donner à chaque motif un score de fixation pour chaque base azotée. En réalisant une moyenne de ces scores, on peut alors classer les sites d'édition connus et ne tester que ceux avec les meilleures moyennes. Bien que non parfait car incomplet, le code, obtenu empiriquement, est le meilleur moyen de déterminer les sites putatifs d'édition d'une PPR, si l'on veut les cibler par RT-PCR. En effet, à la différence d'une approche par séquençage des ARN totaux, l'approche par RT-PCR ne peut pas être exhaustive dans la mitochondrie. Dans ce cas de figure, la sélection des sites putatifs est nécessaire.

L'hypothèse avancée par Guillaumot et al. est que tous les sites non affectés chez $\Delta dyw2$ et $\Delta nuwa$ ne sont pas des sites sous dépendance d'une PPR E+. Par opposition, tous les sites affectés chez ces mutants sont potentiellement sous dépendance, au moins partielle, d'une PPR E+. Cependant, avec la nouvelle analyse, seulement 46% (17/38) des sites E+ dépendent de NUWA avec en plus 26% (6/23) de faux positifs. A l'inverse, 100% (38/38) des sites E+ sont dépendants de DYW2 avec 7% (3/41) de faux-positifs. L'intersection ou l'union de ces 2 jeux de données ne font pas mieux.

A partir de ces informations, nous formulons deux hypothèses :

- La première hypothèse est que tous les sites dépendants de PPR E+ sont aussi dépendants de DYW2.
- La seconde hypothèse est que si la première hypothèse est vraie, alors la prise en compte de cette information permet d'améliorer grandement le rang -c'est-à-dire la position au classement des moyennes d'un site donné- d'un site effectivement dépendant d'une PPR E+ donnée.

2. Lien fonctionnel entre DYW2 et les PPR E+

2.1. Validation du lien fonctionnel entre DYW2 et les PPR E+

La première hypothèse formulée précédemment est que toutes les PPR E+ sont dépendantes de DYW2. Elle est basée sur 38 sites d'édition de PPR E+ connues et décrites dans la littérature, à savoir CLB19 (Chateigner-Boutin et al., 2008), CWM1 (Hu et al., 2016), SLO2 (Zhu et al., 2012a), MEF21 (Takenaka et al., 2012), SLO1 (Sung et al., 2010), COD1 (Dahan et al., 2014), AHG11 (Murayama et al., 2012), MEF13 (Glass et al., 2015), MEF12 (Härtel et al., 2013), OTP72 (Chateigner-Boutin et al., 2013), AEF1 (Yap et al., 2015), MEF25 (Arenas-M et al., 2013), GRS1 (Xie et al., 2016), SLG1 (Yuan et Liu, 2012), CRR21 (Okuda et al., 2014), et OTP80 (Okuda et al., 2014).

Pour renforcer cette hypothèse, nous pouvons ajouter d'autres sites d'édition à notre analyse. Parmi eux, il y a deux sites d'édition non publiés dépendant respectivement de GRS1 et CLB19. Concernant la PPR GRS1, le nouveau site gracieusement communiqué par Xiongbo Peng est M219388 (*mttB_377*). Pour CLB19, un troisième site, C43350, situé dans l'intron 2 d'*ycf3*, nous a gracieusement été communiqué par Ian Small. A ces sites non publiés, nous avons aussi ajouté deux mutants d'insertion d'ADN-T dans les gènes codant pour les PPR E+, AT1G56570 (Salk_141937) et AT5G08305 (Salk_058773).

AT1G56570 (Salk_141937) n'a pas été complètement choisie au hasard puisque l'allèle sauvage du gène ici muté code pour la protéine PGN. Décrite dans la littérature (Laluk et al., 2011) et présentée en introduction, cette protéine fait l'objet d'une description plus détaillée dans le chapitre suivant. Par opposition, AT5G08305 (Salk_058773) a été choisie au hasard parmi les protéines PPR E+ dont la ou les cibles n'étaient pas connues. Les paragraphes suivants vont détailler les résultats obtenus pour les éditosomes de ces deux mutants de PPR E+.

2.1.1. Analyse *RNAseq* du transcriptome d'AT1G56570 (Salk_141937)

Deux défauts d'édition ont été mis en évidence par le séquençage des ARN totaux et notre pipeline d'analyse pour Salk_141937 (Tableau III).

Tableau III : Sites différenciellement édités chez le mutant Salk_141937 (AT1G56570) en comparaison avec Col0

Position	Annotation	Edition Col0	Edition Salk_141937	<i>P-value</i> ajustée (Bonferroni)
M8348	<i>cox2_742</i>	100%	0%	2.6614E-203
M165765	<i>nad6_leader_73</i>	84%	0%	3.8687E-112

Les deux défauts d'édition sont dans la mitochondrie. Aucun défaut d'édition n'a été trouvé dans le chloroplaste. En les comparant aux données de *Δdyw2*, on obtient le tableau IV.

Tableau IV : Sites différenciellement édités chez le mutant SK_141937 (AT1G56570) en comparaison avec Col0 avec ajout du critère DYW2.

Les deux valeurs d'édition pour Col0 correspondent respectivement aux données obtenues par Guillaumot et al. et notre expérimentation.

Position	Moyenne	Annotation	E+	Edition Col0	Edition <i>Δdyw2</i>	Edition Salk_141937
M165765	0.435384615	<i>nad6_leader_73</i>	1	81-84%	3%	0%
M8348	0.430769231	<i>cox2_742</i>	1	100%	11%	0%

La première chose à remarquer est que les deux défauts d'édition trouvés chez Salk_141937 correspondent à des sites affectés chez le mutant *Δdyw2*, corroborant notre hypothèse de départ. De plus, si l'édition est totalement abolie pour *cox2_742* et *nad6_leader_73* chez le mutant Salk_141937, elle ne l'est pas entièrement chez le mutant *Δdyw2*, suggérant deux choses : soit le domaine DYW n'est pas entièrement responsable de l'activité d'édition, et PGN garde une activité d'édition en fonctionnant seule mais cette hypothèse va à l'encontre d'une bonne partie de la littérature (Wagoner et al., 2015; Boussardon et al., 2014) ; soit ces sites sont édités par d'autres protéines, qui ont cependant besoin de la présence de PGN. Cette hypothèse suggère la présence d'une ou plusieurs autres protéines possédant un domaine DYW et participant à l'édition de sites E+, avec une efficacité moindre en comparaison de DYW2.

2.1.2. Analyse *RNAseq* du transcriptome d'AT5G08305 (Salk_058773)

11 défauts d'édition ont été mis en évidence pour ce mutant à travers le séquençage des ARN totaux et l'analyse via notre pipeline (Tableau V).

Tableau V : Sites différenciellement édités chez le mutant Salk_058773 (AT5G08305) en comparaison avec Col0

Position	Annotation	Edition Col0	Edition Salk_058773	<i>p-value</i> ajustée (Bonferroni)
M17869	<i>ccmB_566</i>	91%	7%	1.13051E-34
M17884	<i>ccmB_551</i>	96%	1%	8.82297E-44
M189896	/	4%	0%	1.37766E-06
M215126	<i>nad4_437</i>	98%	49%	7.25897E-65
M219378	<i>mtb_387</i>	43%	6%	2.74914E-22

M23217	rps3_1470	72%	0%	1.12388E-57
M308476	ccmC_179	96%	1%	7.91157E-86
M362007	/	90%	2%	8.89E-176
M362343	atp4_138	92%	1%	6.93E-207
M362349	atp4_144	8%	27%	1.8961E-15
M49473	/	1%	0%	1.26516E-06

Tous les défauts d'édition trouvés sont dans la mitochondrie. Aucun défaut chloroplastique n'a été mis en évidence. Les données de Colcombet et al. (2013) proposent une prédiction de localisation subcellulaire mitochondriale, mais aucune fluorescence n'avait été détectée pour la fusion de la protéine pleine longueur ou du peptide d'adressage avec la RFP. Sur les 11 défauts trouvés, 5 sont totaux ou presque (M17884, M23217, M308476, M362007 et M362343), 3 correspondent à des défauts assez marqués (M17869, M215126, M219378), 2 à des défauts statistiquement significatifs mais probablement non biologiquement pertinents (M189896 et M49473) et enfin 1 défaut correspondant à une augmentation de l'édition (M362349). Bien qu'étonnant ce n'est cependant pas une nouveauté de trouver des augmentations d'édition chez certains mutants (Shi et al., 2017; Guillaumot et al., 2017).

En les comparant aux données de $\Delta dyw2$, on obtient le tableau VI.

Tableau VI : Sites différenciellement édités chez le mutant SK_058773 (AT5G08305) en comparaison avec Col0 avec le critère DYW2.

La valeur d'édition pour Col0 correspond à la moyenne des données obtenues par Guillaumot et al. et notre expérimentation. Le symbole « / » correspond à des données pour lesquelles la couverture était insuffisante lors du *RNAseq* $\Delta dyw2$.

Position	Moyenne	Annotation	E+	Edition Col0	Edition $\Delta dyw2$	Edition Salk_058773	PPR associée	Sous-famille
M215126	0.479	nad4_437	1	98.5%	71%	49%	-	-
M362343	0.464	atp4_138	1	88.5%	4%	1%	-	-
M17884	0.434	ccmB_551	1	94%	4%	1%	-	-
M23217	0.426	rps3_1470	1	72.5%	0%	0%	-	-
M308476	0.394	ccmC_179	1	96%	3%	1%	-	-
M17869	0.392	ccmB_566	1	92.5%	13%	7%	MEF19	E
M219378	0.369	mttb_387	1	35%	1%	6%	-	-
M362007	0.353	-	1	88,5%	3%	2%	-	-
M362349	0.21	atp4_144	0	9.5%	46%	27%	-	-
M189896	/	-	/	3.5%	/	0%	-	-
M49473	/	-	/	0.6%	/	0%	-	-

Ces données sont extrêmement riches en enseignements. Tout d'abord, sur les 11 défauts présentés ici, 8 sont dépendants de DYW2, 2 n'ont pas une couverture suffisante sur $\Delta dyw2$ pour pouvoir conclure et un, M362349, est indépendant de DYW2. Ce défaut correspond à une hausse de l'édition. Cependant, si on regarde de près, ce site d'édition possède dans son site de fixation, M362343, qui lui présente un défaut d'édition majeur. Bien que ce site ne corresponde pas à notre définition de « dépendant de DYW2 », son édition monte de 11 à 46% chez $\Delta dyw2$. Toutes ces données me permettent de conclure que l'augmentation d'édition en M362349 est probablement un défaut secondaire, non associé directement à l'action de la PPR

E+ codée par AT5G08305 mais à la conséquence du défaut d'édition de M362343. On peut supposer que la PPR éditant M362349 a une meilleure affinité de fixation pour l'ARN non édité en M362343 que pour l'édité.

Un autre site d'édition affecté présente un intérêt particulier : M17869. Chez notre mutant Salk_058773, l'édition de ce site diminue de 84 points. Ce site est déjà connu, et décrit comme étant dépendant de MEF19, une PPR E. Ce site a été présenté en introduction de ce chapitre comme étant l'une des 3 exceptions trouvées dans les données de DYW2, car il s'agit d'un site sous dépendance d'une PPR non E+. Il est intéressant de remarquer que deux sites d'édition sont présents dans la zone de fixation à proximité : M17881 et M17884. L'édition de M17884 est abolie dans notre mutant. Il est possible que ce défaut d'édition modifie la fixation de MEF19, donc l'édition de M17869. Dans ce cas de figure, on peut supposer que MEF19 a une affinité plus grande pour la base éditée que la base non éditée. Afin de mieux comprendre ces données, nous avons contacté Mizuki Takenaka, impliqué dans la caractérisation de MEF19 (Takenaka et al., 2010). Il s'avère que son équipe travaille sur la caractérisation de la même protéine, AT5G08305, mais avec une lignée T-DNA différente (Salk_110420). La PPR encodée par AT5G08305 a été nommée MEF37. Ils ont trouvé les mêmes défauts d'édition, à l'exception de nad4_437 qui n'a pas été analysé. Ils ont également effectué une complémentation fonctionnelle avec MEF37, et vérifié sur 3 sites (atp4_138, ccmB_551 et rps3_1470) que l'on retrouvait des niveaux d'édition comparables à Col0. Enfin, et de façon extrêmement élégante, ils ont réalisé une seconde complémentation fonctionnelle de $\Delta mef37$ avec une version mutée de MEF19, où le motif supposé fixer la base éditée a une plus grande affinité pour la cytosine. Les résultats de cette complémentation originale ne sont pour le moment pas encore connus.

Concernant le site M215126, si son édition chute à 49% chez le mutant Salk_058773, elle baisse seulement à 71% chez $\Delta dyw2$. On a donc un écart d'un peu plus de 20 points entre le mutant $\Delta dyw2$ et notre mutant. On peut imaginer une redondance fonctionnelle entre deux PPR dont notre E+ pour ce même site. Cependant, si c'était le cas, l'absence de la PPR E+ devrait libérer la zone de fixation sur l'ARN avec pour conséquence un niveau d'édition situé entre celui de $\Delta dyw2$ et Col0. Or, ce n'est pas ce que nous observons. L'hypothèse la plus simple pour expliquer notre observation est un petit peu plus complexe. Nos données suggèrent que la PPR E+ encodée par AT5G08305 a une activité d'édition seule (peu probable en se basant sur la littérature) ou avec une autre protéine *DYW-like* différente de DYW2. Cela expliquerait pourquoi le mutant $\Delta dyw2$ est moins affecté que le mutant Salk_058773 pour l'édition de ce site. Cependant, en l'absence de la PPR E+ que nous caractérisons, le site est toujours édité à hauteur de 49%. Il existe donc une troisième PPR impliquée dans l'édition de ce site. A partir de nos données, rien ne nous permet de savoir si cette troisième PPR est dépendante ou non de DYW2. Un double mutant Salk_058773/ $\Delta dyw2$ nous permettrait d'avoir une réponse mais demeure techniquement délicat à réaliser suite à l'embryoléthalité de $\Delta dyw2$.

2.1.3. Résumé des résultats obtenus et approche probabiliste

A partir des résultats connus dans la littérature et des nouvelles informations obtenues dans le cadre de la thèse, on peut compléter le tableau VII.

Tableau VII : Résumé des sites connus et trouvés comme dépendants de PPR E+ et dépendance à DYW2.

PPR E+	Nombre de sites connus	Nombre de sites connus dépendants de DYW2
CLB19	3	3
CWM1	3	3
SLO2	6	6
MEF21	1	1
SLO1	2	2
COD1	3	3
AHG11	1	1
MEF13	8	8
MEF12	1	1
OTP72	1	1
AEF1	2	2
MEF25	1	1
GRS1	5	5
SLG1	1	1
CRR21	1	1
OTP80	1	1
PGN	2	2
MEF37	8	8

Comme expliqué lors de l'introduction, tous les sites publiés dépendants de PPR E+ sont aussi dépendants de DYW2. A travers notre analyse, nous avons rajoutés 2 sites communiqués par des collaborateurs, et 10 sites trouvés via l'analyse par *RNAseq* de deux lignées de mutants E+. Ces 12 sites sont également dépendants de DYW2. Finalement, sur 50 sites connus comme étant dépendants des PPR E+, tous sont dépendants de DYW2. Toutes ces données mises bout –à-bout ne prouvent pas de façon absolue que tous les sites E+ sont dépendants de DYW2 mais permettent de formuler une hypothèse très forte.

Je propose de représenter cette expérimentation comme un problème d'urne :

Une urne contient 795 boules, 531 sont blanches et 291 noires (Tableau VIII). On effectue 18 tirages successifs avec remise, d'une ou plusieurs boules simultanément, soit 50 tirages au total. Les boules blanches représentent les sites indépendants de DYW2, les boules noires les sites dépendants, et chaque tirage correspond au(x) site(s) de chaque PPR E+ testée. Pour CLB19 par exemple, on tire simultanément 3 boules, puis on les remet et on tire simultanément trois boules pour CWM1, etc. On suppose que chaque boule est indiscernable au toucher et que chaque boule a la même probabilité d'être tirée.

Tableau VIII : Répartition des sites d'édition observés par Guillaumot et al. puis ajustés par une nouvelle analyse et répartition par dépendance.

Sites non dépendants de DYW2	531
Sites dépendants de DYW2	291
Sites connus dont la couverture est insuffisante chez $\Delta dyw2$	0
Total des sites testés	795

On calcule la probabilité de ne tirer que des boules noires :

$$p(50N) = \prod_1^p \frac{\binom{291}{np}}{\binom{795}{np}}$$

Avec p , protéine PPR à np sites d'édition.

$$p(50N) = \frac{\binom{291}{3}}{\binom{795}{3}} \times \frac{\binom{291}{3}}{\binom{795}{3}} \times \frac{\binom{291}{6}}{\binom{795}{6}} \times \frac{\binom{291}{2}}{\binom{795}{2}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{3}}{\binom{795}{3}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{8}}{\binom{795}{8}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{1}}{\binom{795}{1}}$$

$$\times \frac{\binom{291}{2}}{\binom{795}{2}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{5}}{\binom{795}{5}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{1}}{\binom{795}{1}} \times \frac{\binom{291}{2}}{\binom{795}{2}} \times \frac{\binom{291}{8}}{\binom{795}{8}}$$

$$p(50N) = 1,22304.10^{-22}$$

Si l'on revient à nos sites d'édition, en supposant que les PPR E+ aient la même probabilité d'édition pour chaque site, la probabilité de ne tirer que des sites dépendants de DYW2 pour les PPR E+ est légèrement supérieure à 1.10^{-22} .

Couplé aux très nombreuses données présentées jusqu'ici, nous sommes à présent largement convaincus qu'absolument tous les sites dépendants de PPR E+ sont également dépendants de DYW2. Cependant, tous les sites dépendants de DYW2 ne sont pas nécessairement des sites à E+, étant donné, comme nous avons pu le voir, les possibles défauts secondaires, dont il est difficile d'évaluer la proportion.

Après avoir mené cette démonstration, nous nous intéressons à notre seconde hypothèse, à savoir que la prise en compte de cette information permet d'améliorer grandement le rang d'un site effectivement dépendant d'une PPR E+ donnée par application du code.

2.2. Application du critère DYW2 et rangs des sites prédits

2.2.1. Construction du système de prédiction

Le système de prédiction de fixation des protéines PPR sur l'ARN fonctionne a priori pour toutes les sous-familles de PPR. Cependant, les mutants de PPR d'édition présentent l'intérêt d'avoir un phénotype moléculaire évident. Notre matrice est une matrice poids-position qui, en fonction des données de Takenaka et al. (2013), va donner une probabilité de fixation de motifs PPR sur chacune des bases azotées.

Concernant la liste des sites d'édition pris en compte dans la prédiction, nous nous sommes basés sur :

- les sites différentiellement édités chez *Anuwa* et *Δdyw2* et avec une édition supérieure ou égale à 5%
- les sites avec une édition strictement supérieure à 5% et avec une couverture moyenne strictement supérieure à 20 *reads* par échantillon chez *Anuwa* et *Δdyw2*
- les sites de l'étude de *Δorrm5* et *Δmef8* (Diaz et al., 2017; Shi et al., 2017) avec une édition supérieure ou égale à 5%

Cependant, les sites d'édition situés au-delà de la position 128215 du chloroplaste ne sont pas considérés car étant dans la région inversé répétée, c'est comme si on les prenait en compte deux fois.

En appliquant ces critères, on arrive à la somme de 795 sites d'édition, 100 dans le chloroplaste et 695 dans la mitochondrie. Parfois, il se trouve qu'un autre site d'édition peut se trouver dans la zone de fixation de la PPR (zone -30 à -4 des sites d'édition). Nous avons tenu compte de cela et ajouté les différentes possibilités de zone de fixation pour chaque site. On arrive à 1979 sites de fixation potentiels.

Afin de comprendre notre système de prédiction, je propose d'étudier l'exemple de la PPR E+ CLB19.

2.2.2. Exemple de CLB19

CLB19 est l'une des 4 PPR E+ chloroplastiques connues à ce jour. Elle est impliquée dans l'édition de *clpP* et *rpoA* (Chateigner-Boutin et al., 2008) ainsi que de l'intron 2 de *ycf3* (I. Small, communication personnelle). Selon les données de Cheng et al. (2016), CLB19 a les 13 motifs suivants : 35-P1-L1-S1-SS-P1-L1-S1-P2-L2-S2-E1-E2-E+ (Figure 21).

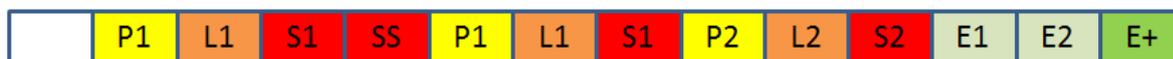


Figure 21 : Représentation schématique de la PPR E+ CLB19.

En appliquant les données de notre matrice détaillée ci-dessus aux motifs des sous-familles P, L et S impliqués dans l'interaction avec l'ARN, on obtient ainsi les scores de fixation de chaque base azotée pour chaque motif de CLB19 (Tableau IX).

Tableau IX : Probabilités de fixation d'une base azotée pour chaque motif de CLB19.

Base ARN	P1	L1	S1	SS	P1	L1	S1	P2	L2	S2
A	0.73	0.08	0.64	0.15	0.04	0.08	0	0.09	0.7	0.25
C	0.05	0.31	0.06	0.23	0	0.31	0.78	0.25	0.16	0.25
G	0.22	0.16	0.15	0.07	0.93	0.16	0.14	0.14	0.14	0.25
U	0	0.45	0.16	0.55	0.03	0.45	0.09	0.52	0	0.25

Dans un second temps, on croise ces informations avec les séquences en amont des sites d'édition chloroplastiques et mitochondriaux connus. Comme expliqué plus haut, certaines zones de fixation contiennent des sites édités. Plusieurs versions de la zone de fixation peuvent donc coexister *in vivo*, et elles

sont prises en compte dans notre liste de sites, sous la forme d'un tiret et d'un chiffre après la coordonnée génomique du site (Tableau X).

Tableau X : Premiers sites prédits pour être fixés par CLB19 en appliquant le code PPR.

Sites d'édition	P1	L1	S1	SS	P1	L1	S1	P2	L2	S2	P1	L1	S1	SS	P1	L1	S1	P2	L2	S2	moyenne
C43350	A	U	A	U	G	C	C	U	A	A	0.73	0.45	0.64	0.55	0.93	0.31	0.78	0.52	0.7	0.25	0.586
C69942	A	G	A	A	G	C	C	C	A	A	0.73	0.16	0.64	0.15	0.93	0.31	0.78	0.25	0.7	0.25	0.49
C37161	A	U	U	C	G	U	C	G	A	U	0.73	0.45	0.16	0.23	0.93	0.45	0.78	0.14	0.7	0.25	0.482
M209879-1	C	C	A	U	G	C	C	G	A	A	0.05	0.31	0.64	0.55	0.93	0.31	0.78	0.14	0.7	0.25	0.466
M209879-2	C	C	A	U	G	C	C	G	A	A	0.05	0.31	0.64	0.55	0.93	0.31	0.78	0.14	0.7	0.25	0.466
C78691	A	C	A	C	G	U	G	C	A	A	0.73	0.31	0.64	0.23	0.93	0.45	0.14	0.25	0.7	0.25	0.463
M308873-1	A	U	G	U	U	U	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.45	0.78	0.52	0.7	0.25	0.461
M308873-2	A	U	G	U	U	U	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.45	0.78	0.52	0.7	0.25	0.461
M185273-1	U	U	A	C	G	U	C	A	A	G	0	0.45	0.64	0.23	0.93	0.45	0.78	0.09	0.7	0.25	0.452
M308873-3	A	U	G	U	U	C	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.31	0.78	0.52	0.7	0.25	0.447
M308873-4	A	U	G	U	U	C	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.31	0.78	0.52	0.7	0.25	0.447
M185273-2	U	C	A	C	G	U	C	A	A	G	0	0.31	0.64	0.23	0.93	0.45	0.78	0.09	0.7	0.25	0.438

Les sites commençant par M correspondent aux sites mitochondriaux et les sites commençant par C aux sites chloroplastiques. On obtient ainsi un classement, basé sur la moyenne des scores de fixation de chaque domaine, des sites potentiellement sous dépendance de la protéine PPR qui nous intéresse, en l'occurrence CLB19, suivant la moyenne des probabilités de fixation d'un motif à la base correspondante en amont de chaque site. Ce classement permet ensuite de calculer les rangs de chaque site, avec par exemple ici C43350 qui arrive au rang 1, C69942 au rang 2, etc. Cependant, les sites possédant des sites d'édition dans la zone putative de fixation de la PPR peuvent être représentés de très nombreuses fois, entraînant une augmentation artificielle des rangs des sites qui les suivent. Pour le calcul des rangs, nous avons fait le choix de ne conserver que la version possédant la meilleure moyenne des scores si plusieurs versions de la zone de fixation existent.

Enfin, en se basant sur notre hypothèse, on peut rajouter un niveau d'information supplémentaire, à savoir si le site d'édition est potentiellement dépendant d'une PPR E+. (Tableau XI).

Tableau XI : Premiers sites prédits pour être fixés par CLB19 en appliquant le code PPR et le critère E+

Sites d'édition	P1	L1	S1	SS	P1	L1	S1	P2	L2	S2	P1	L1	S1	SS	P1	L1	S1	P2	L2	S2	moyenne	E+
C43350	A	U	A	U	G	C	C	U	A	A	0.73	0.45	0.64	0.55	0.93	0.31	0.78	0.52	0.7	0.25	0.586	1
C69942	A	G	A	A	G	C	C	C	A	A	0.73	0.16	0.64	0.15	0.93	0.31	0.78	0.25	0.7	0.25	0.49	1
C37161	A	U	U	C	G	U	C	G	A	U	0.73	0.45	0.16	0.23	0.93	0.45	0.78	0.14	0.7	0.25	0.482	0
M209879-1	C	C	A	U	G	C	C	G	A	A	0.05	0.31	0.64	0.55	0.93	0.31	0.78	0.14	0.7	0.25	0.466	1
M209879-2	C	C	A	U	G	C	C	G	A	A	0.05	0.31	0.64	0.55	0.93	0.31	0.78	0.14	0.7	0.25	0.466	1
C78691	A	C	A	C	G	U	G	C	A	A	0.73	0.31	0.64	0.23	0.93	0.45	0.14	0.25	0.7	0.25	0.463	1
M308873-1	A	U	G	U	U	U	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.45	0.78	0.52	0.7	0.25	0.461	0
M308873-2	A	U	G	U	U	U	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.45	0.78	0.52	0.7	0.25	0.461	0

M185273-1	U	U	A	C	G	U	C	A	A	G	0	0.45	0.64	0.23	0.93	0.45	0.78	0.09	0.7	0.25	0.452	1
M308873-3	A	U	G	U	U	C	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.31	0.78	0.52	0.7	0.25	0.447	0
M308873-4	A	U	G	U	U	C	C	U	A	U	0.73	0.45	0.15	0.55	0.03	0.31	0.78	0.52	0.7	0.25	0.447	0
M185273-2	U	C	A	C	G	U	C	A	A	G	0	0.31	0.64	0.23	0.93	0.45	0.78	0.09	0.7	0.25	0.438	1

2.2.3. Analyse des rangs avec et sans le critère E+

A l'image de ce que nous avons montré pour CLB19, le tableau XII résume les rangs moyens des PPR E+ publiées et étudiées lors de la thèse.

Tableau XII : Comparaison des rangs moyens des sites d'édition connus de PPR E+ avec et sans le critère E+.

PPR E+	Nombre de sites connus	Rang moyen des cibles connues		
		obtenu en appliquant le code PPR	obtenu en rajoutant le critère E+	Rangs gagnés en moyenne
CLB19	3	3	2.67	0.33
CWM1	3	7.33	2.33	5
SLO2	6	271.17	80.67	190.5
MEF21	1	1	1	0
SLO1	2	6.5	3.5	3
COD1	3	7.33	6.33	1
AHG11	1	3	3	0
MEF13	8	28.63	12.38	16.25
MEF12	1	25	12	13
OTP72	1	1	1	0
AEF1	2	3.5	2	1.5
MEF25	1	199	61	138
GRS1	5	15.6	10.4	5.2
SLG1	1	1	1	0
CRR21	1	1	1	0
OTP80	1	1	1	0
PGN	2	2	1.5	0.5
MEF37	8	25.63	11.63	14

Sur un certain nombre de PPR E+, l'application du critère permet de gagner un nombre considérable de rangs en moyenne. Sans surprise, cela correspond aux PPR pour lesquelles le code ne fonctionne que très mal (MEF25, MEF13, MEF12 et surtout SLO2). Pour les PPR où le code fonctionne très bien (OTP80, CRR21, SLG1, OTP72, AEF1, AHG11, MEF21 et CLB19), le critère n'apporte pas ou peu d'intérêt, les sites dépendants sortant déjà dans les premiers rangs de la prédiction.

L'application du critère E+ permet d'éliminer près de 70% des sites d'édition potentiels dans l'analyse (644 sites E+ potentiels contre 1979 sites au total) ce qui permet mécaniquement d'améliorer les rangs. Cependant, notre prédiction reste très dépendante de la performance du code PPR qui semble assez variable d'une PPR à une autre mais également d'un site à un autre comme l'illustrent les résultats du mutant Salk_058773 (MEF37). Le tableau XIII détaille les 10 premiers sites prédits et les sites dépendants de la protéine encodée par MEF37.

Tableau XIII : Classement des premiers sites prédits pour MEF37 en se basant sur le code PPR et le critère E+.

Les deux valeurs d'édition pour Col0 correspondent respectivement aux données obtenues par Guillaumot et al. et notre expérimentation. En rouge sont représentés les sites affectés chez Salk_058773. Le rang corrigé ne tient compte que des sites dépendants de DYW2.

Position	Moyenne	Annotation	E+	Edition Col0	Edition <i>Adyw2</i> si dépendant	Edition Salk_058773	Rang	Rang corrigé
M215126	0.479	nad4_437	1	99-98%	71%	49%	1	1
M362343	0.464	atp4_138	1	85-92%	4%	1%	2	2
M18297	0.45	ccmB_138	0	4%	-	non détecté	3	-
M17884	0.434	ccmB_551	1	94%	4%	1%	4	3
M235424	0.432	nad5_242	1	97%	3%	97%	5	4
M209909	0.432	nad4_1101	0	88%	-	90%	6	-
M235858	0.427	nad5_676	0	99-100%	-	99%	7	-
M23217	0.426	rps3_1470	1	75-70%	0%	0%	8	5
M362600	0.424	atp4_395	1	98-99%	31%	99%	9	6
M18263	0.419	ccmB_172	0	73-88%	-	88%	10	-
...
M308476	0.394	ccmC_179	1	96%	3%	1%	28	13
...
M17869	0.392	ccmB_566	1	94-91%	13%	7%	33	16
...
M219378	0.369	mttb_387	1	27-43%	1%	6%	46	22
...
M362007	0.353	/	1	87-90%	3%	2%	83	31
...
...
M362349	0.21	atp4_144	0	11-8%	-	27%	659	-

Un certain nombre de sites arrive assez loin dans le classement, même s'il est amélioré avec le critère DYW2. Cela pose la question du bon fonctionnement du code PPR pour certains sites qui auraient été très difficiles à trouver avec une approche par RT-PCR, même avec le critère DYW2. Le code peut-être aussi questionné en ce qui concerne le site M235424, dépendant de DYW2 mais non affecté chez le mutant Salk_058773. Il est étonnant de le voir sortir en très bonne position dans le code, entouré de sites affectés, mais ne pas l'être alors qu'il est dépendant de DYW2. L'hypothèse la plus plausible est la concurrence avec une autre PPR E+. A partir de nos données de prédictions, il est très facile de construire une matrice inverse pour déterminer non pas quels sont les meilleurs sites prédits pour une PPR E+ mais, à l'inverse, quelle est la meilleure PPR E+ prédite pour un site donné. En effectuant l'analyse, on voit que la PPR E+ la mieux prédite pour éditer M235424 est AT3G04750 (avec une moyenne de 0.45), suivie de près par MEF37, la PPR E+ dont nous caractérisons l'éditotype (0.43), AT5G56310 (0.42) et AT4G04370 (0.37). Ces 4 PPR sont assez détachées du reste des PPR E+ puisque la cinquième au classement a une moyenne de 0.34. Seule la lignée d'insertion

ADN-T dans AT4G04370 a été testée pour le moment mais cela a permis d'identifier un défaut total d'édition à ce site chez ce mutant.

Concernant le site M17869, nous avons discuté plus tôt ce résultat puisqu'il s'agit d'une des 3 exceptions de PPR non E+ dépendantes de DYW2, puisque ce site est connu pour être dépendant de MEF19, une PPR E. Nous avons fait l'hypothèse qu'il pouvait s'agir d'un défaut secondaire, dû à un autre défaut sur M17884. L'analyse du rang semble confirmer cette hypothèse puisqu'après application du code et prise en compte du critère DYW2, M17884 arrive en 3^{ème} position contre une 16^{ème} position pour M17869. Une autre possibilité est d'appliquer le code PPR à MEF19.

En appliquant le code PPR à MEF19, le site M17869 n'arrive qu'en 41^{ème} position. Cependant, c'est avec la version éditée en M17884 que le meilleur score de prédiction est obtenu (0.35 pour la version éditée contre 0.33). L'édition du site M17881 ne change rien à la prédiction. L'hypothèse la plus probable reste donc que le site M17869 est sous dépendance directe de MEF19, et que le défaut observé dans le mutant Salk_058773 ($\Delta mef37$) est un défaut secondaire dû au défaut initial sur le site M17884. Cette hypothèse explique pourquoi ce site mal classé est différentiellement édité chez notre mutant et montre l'importance de tenir compte des sites édités dans les zones de fixation des PPR pour effectuer nos prédictions. Cependant, il est difficile d'expliquer pourquoi ce site est édité par MEF19 car il est également assez loin dans le classement. En appliquant un code inverse, c'est-à-dire en éliminant de la prédiction les sites dépendants de DYW2, et en considérant suite à notre hypothèse que M17869 n'est pas dépendant de DYW2, ce site remonte à la 21^{ème} place des sites prédits pour MEF19, ce qui reste un moins bon classement que pour notre PPR encodée par AT5G08305. Etant donné qu'aucun autre défaut n'est connu pour MEF19, cette hypothèse reste la plus plausible. La complémentation fonctionnelle de $\Delta mef37$ avec une version mutée de MEF19, où le motif supposé fixer la base éditée a une plus grande affinité pour la cytosine nous apporterait de plus amples informations.

Après avoir renforcé le consensus DYW2-PPR E+, après avoir montré que la prise en compte de ce critère facilite la prédiction de sites potentiellement affectés chez un mutant, nous avons tenté de l'appliquer à l'ensemble des PPR E+ connues, en commandant des mutants d'insertion ADN-T disponibles pour celles-ci.

3. Extension à l'ensemble des PPR E+

Afin de tester notre hypothèse sur un jeu de données plus large, des lignées de mutants d'insertion ADN-T correspondant aux PPR E+ listées (Cheng et al., 2016) ont été commandées. Sur les 57 protéines E+ identifiées dans cette publication, 19 ont déjà été caractérisées et nous disposons de 34 lignées mutantes sur les 38 restantes. Après obtention d'individus homozygotes, l'idée est de tester par RT-PCR les 5 premiers sites proposés par le code PPR couplé au critère DYW2. Seuls les sites d'édition pour un nombre très limité de mutants ont été testés, soit par manque de temps, soit parce que nous n'avons pas réussi à obtenir de mutants homozygotes. L'annexe S2 rassemble les lignées pour lesquelles nous n'avons pas obtenu d'homozygotes. Pour les 3 lignées chez lesquelles nous n'avons pas pu identifier de plante homozygote parmi 24 plantes issues d'une plante hétérozygote et ceci sur deux générations, il est statistiquement improbable que l'homozygote soit viable. Des constructions avec des promoteurs induits lors de l'embryogénèse, à l'image du travail de Guillaumot et al. pourraient être une solution pour contourner ce problème.

Le tableau XIV résume les données obtenues sur les lignées pour lesquelles nous avons pu avoir un ou plusieurs individus homozygotes pour le gène codant pour la PPR E+ concernée.

Tableau XIV : Résumé des données obtenues sur les lignées d'insertion d'ADN-T sur des gènes codants pour des PPR E+, pour lesquelles nous disposons d'individus homozygotes.

Les lignées soulignées correspondent à celles analysées en *RNAseq*.

Lignées avec HO disponible		Rang de la prédiction après correction E+				
Lignée mutante utilisée	AGI	1	2	3	4	5
Salk_064675	AT1G13410	<i>ccmC_467</i>	<i>rps4_77</i>	<i>ccmB_576</i>	<i>nad4_977</i>	<i>nad5_553</i>
Salk_027874	AT1G32415	<i>ccmC_458</i>	M30601	M254963	<i>mtb_144</i>	<i>ccmC_421</i>
<u>Salk_141937</u>	<u>AT1G56570</u>	<i>nad6_leader_73</i>	<i>cox2_742</i>	M307990	<i>nad6_169</i>	<i>nad1_265</i>
Salk_204996C	AT1G69350	<i>nad1_580</i>	<i>mtb_705</i>	<i>rpl5_317</i>	<i>nad4_1405</i>	<i>nad2_1309</i>
Salk_066720	AT1G74400	M237091	M219657	<i>nad2_1309</i>	M20228	M331859
Salk_071391	AT2G21090	<i>nad7_213</i>	<i>nad4_1405</i>	<i>rps3_1352</i>	M231342	<i>ccmFn1_143</i>
Salk_202435C	AT2G37320	<i>nad7_38</i>	<i>nad2_394</i>	M20228	M331859	M346754
Salk_085377	AT2G40720	<i>nad4_977</i>	<i>ccmC_650</i>	M241468	M270995	M332097
Salk_036259	AT3G04750	<i>nad7_209</i>	<i>nad5_242</i>	<i>nad4_158</i>	<i>ccmFc_175</i>	<i>nad4_1405</i>
SAIL_205G08	AT3G29230	<i>ccmB_71</i>	<i>cox2_278</i>	<i>nad3_211</i>	<i>nad7_213</i>	M243096
Salk_024585	AT3G53360	<i>nad1_580</i>	<i>ccmFc_1327</i>	<i>cob_924</i>	<i>nad1_308</i>	<i>mtb_144</i>
Salk_025427	AT4G04370	<i>cox3_314</i>	<i>mtb_387</i>	M191273	<i>nad1_580</i>	<i>nad1_308</i>
Salk_088120	AT4G19191	<i>rpl5_47</i>	<i>matR_1771</i>	<i>cox3_257</i>	M219343	<i>ccmC_473</i>
<u>Salk_058773</u>	<u>AT5G08305</u>	<i>nad4_437</i>	<i>atp4_138</i>	<i>ccmB_551</i>	<i>nad5_242</i>	<i>rps3_1470</i>
Salk_104072	AT5G27110	<i>ccmB_566</i>	<i>nad3_250</i>	M234082	<i>ccmFc_1280</i>	<i>ccmFc_175</i>
SAIL_15_E07C1	AT5G42450	<i>mtb_665</i>	<i>ccmFc_146</i>	<i>nad7_38</i>	M161631	<i>atp4_395</i>
Salk_201330	AT5G56310	<i>nad9_190</i>	<i>nad2_1309</i>	<i>ccmB_566</i>	M219657	<i>rpl5_47</i>

Défaut proportionnel à celui observé chez <i>Δdyw2</i>	Défaut non proportionnel à celui observé chez <i>Δdyw2</i>	Pas de défaut observé	Non testé ou problème technique
--	--	-----------------------	---------------------------------

L'analyse est réalisée par RT-PCR et lecture du chromatogramme après séquençage par la méthode Sanger. L'édition est donc estimée nettement moins précisément qu'avec une approche par *RNAseq*. Pour le moment, les résultats obtenus ne concernent qu'une seule répétition biologique et un seul réplicat technique. Les sites pour lesquels aucun défaut n'est observé montrent cependant que nos résultats sont cohérents avec les niveaux d'édition connus par *RNAseq*. Aucun défaut macroscopique évident n'a été mis en évidence chez les différents mutants homozygotes obtenus.

Suite au changement de génome Col0, à la nouvelle analyse des données des mutants *Δdyw2* et *Δnuwa* et à la prise en compte des sites d'édition dans les zones de fixation, les prédictions des 5 meilleurs sites ont été amenées à changer rapidement lors de la fin de la thèse. De fait, de nombreux sites n'ont pas encore été testés comme on peut le voir sur le tableau XVI, mais des sites se situant hors des 5 premiers prédits ont été testés, et ont même montré des défauts pour certains. Les défauts mis ainsi en évidence concernent AT1G13410 (défaut sur *ccmFc_1337*, classé 32^{ème}, diminution de l'édition de 30% environ, cohérent avec la baisse chez *Δdyw2*) ;

AT1G32415 (défaut sur *mttB_406*, classé 15^{ème}, diminution de l'édition de 20-30%, cohérent avec la baisse chez $\Delta dyw2$) et AT2G21090 (défaut sur *matR_461*, classé 8^{ème}, avec une diminution de l'édition de 90%, cohérent avec la baisse chez $\Delta dyw2$). Au sujet du site *nad5_242* (M235424) non dépendant d'AT5G08305, nous avons testé Salk_025247 (AT4G04370), qui arrive en 4^{ème} position sur la matrice inverse. Il s'avère que l'édition tombe à 0% chez ce mutant, expliquant pourquoi elle n'est pas affectée chez $\Delta mef37$ alors qu'il est bien classé et dépendant de DYW2. Nous n'avons pas encore regardé si ce site était affecté dans les 2 autres lignées bien classées dans la matrice inverse par manque de temps, même si cela s'avère assez peu probable qu'il le soit.

Concernant AT4G19191, l'édition du site *matR_1771* est entre 0 et 10% à la lecture du chromatogramme chez sa lignée d'insertion ADN-T, alors qu'elle est à 49% chez le mutant $\Delta dyw2$ et à 80% chez Col0. On se retrouve donc dans un cas similaire à celui de *nad4_437* chez MEF37, mais ce résultat reste à confirmer.

4. Discussion

Pour résumer, l'ensemble des défauts trouvés chez des PPR E+ correspondent à des sites dépendants de DYW2, renforçant largement notre hypothèse de départ selon laquelle tous les sites à E+ dépendent de DYW2. Mais si tous les sites E+ dépendent de DYW2, cela ne signifie pas pour autant que la réciproque est vraie, à savoir que tous les sites dépendants de DYW2 sont des sites de PPR E+. 3 contre-exemples avaient été présentés en introduction :

- M17869, dépendant de MEF19, une PPR E (Takenaka et al., 2010)
- M233897, dépendant de MEF57, une PPR DYW (Andrés-Colás et al., 2017)
- M234052, dépendant de PPME, une PPR pure (Leu et al., 2016)

Par chance, nous avons étudié la PPR E+ responsable du défaut d'édition de M17869 et, comme discuté précédemment, nous avons conclu que cela correspondait à un défaut secondaire.

Concernant MEF57, il faut remarquer qu'une interaction physique a été observée entre DYW2 et MEF57 *in planta* (Andrés-Colás et al., 2017). En effectuant une immunoprécipitation avec la PPR E+ SLO2, les auteurs ont ainsi identifié parmi d'autres protéines DYW2 et MEF57. Même si le rôle exact de MEF57 dans l'éditosome des PPR E+ n'est pas clair, il semblerait en tout cas que cette PPR DYW en fasse partie. Il n'est donc pas particulièrement étonnant de la retrouver dans les cas particuliers ici.

Pour M234052 (*nad1_898*), nous n'avons pour le moment aucune donnée permettant d'expliquer pourquoi ce site est dépendant de DYW2. Il est à noter qu'un second site dépendant de PPME, M234091 (*nad1_937*), n'est lui pas dépendant de DYW2 (Figure 22).

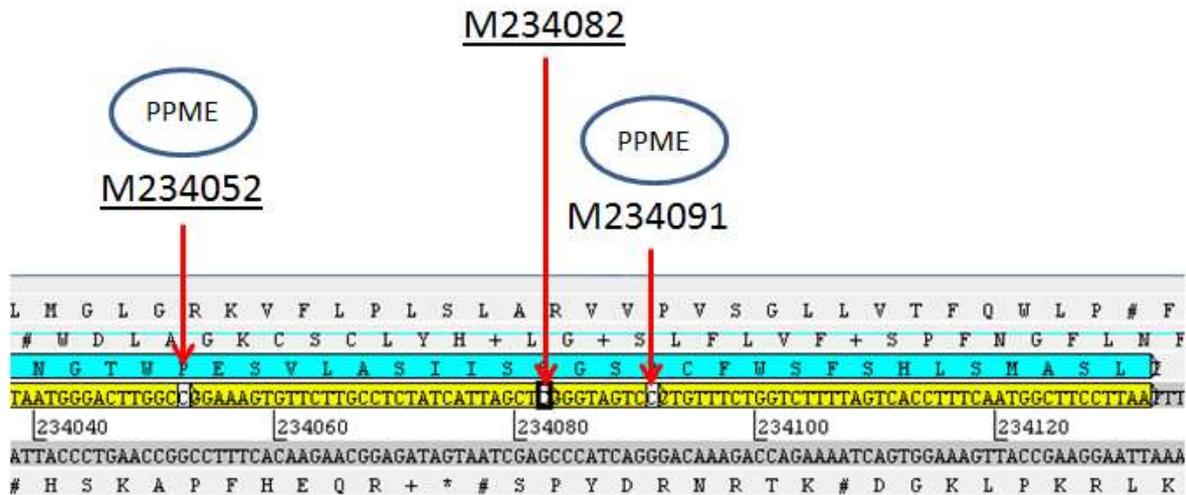


Figure 22 : Partie 3' de la séquence codant pour Nad1 avec ses sites d'édition.

La PPR connue pour être impliquée dans l'édition d'un site est représentée dans un cercle bleu. Les sites soulignés sont dépendants de DYW2.

M234082, un site dépendant de DYW2, n'est pas la zone de fixation pour l'édition de M234052 donc il est très peu probable que M234052 soit un défaut secondaire de M234082. Si on fait l'hypothèse qu'en l'absence de DYW2, la PPR E+ éditant M234082 ne libère pas le site et empêche l'édition de M234052, alors le site M234091 devrait être affecté chez le mutant $\Delta dyw2$, ce qui n'est pas le cas. Une autre possibilité serait que PPME fonctionne comme NUWA, une PPR pure impliquée potentiellement dans la stabilisation d'un complexe d'édition. Cela s'avère également improbable puisque Leu et al. (2016) ont démontré que PPME se fixait bien en amont des deux sites qu'elle édite. Je n'ai donc pas d'explication à l'heure actuelle concernant le lien entre PPME et DYW2.

Une fois le lien entre DYW2 et les PPR E+ confirmé, nous nous sommes intéressés à l'intérêt d'utiliser le critère E+ dans notre système de prédiction. L'intérêt est évident puisqu'il nous permet d'éliminer près de 70% des sites d'édition connus lorsque l'on caractérise une PPR E+, et 30% des sites lorsque l'on s'intéresse à une PPR d'édition non E+. Pourtant, lorsque l'on s'intéresse à certaines PPR, SLO2 ou MEF37 par exemple, on se rend compte que même si notre critère permet de gagner un nombre considérable de rangs, il ne compense pas les limites du code PPR. Une autre information intéressante est que le rang ne semble pas être positivement corrélé à l'importance du défaut d'édition. On le voit très clairement avec l'exemple de MEF37 puisque le premier site à être affecté ne l'est que de moitié. Sans un bon code, le critère E+ ne limite que le choix des possibles. Le code PPR peut se délimiter en deux parties : la construction de la matrice d'une part, donnant les probabilités de fixation de chaque motif sur une base azotée ; et le système de prédiction, c'est-à-dire l'utilisation de ces données pour classer les sites d'édition.

Concernant la matrice utilisée, nous utilisons celle de Takenaka et al. (2013), basée sur 42 PPR de différentes espèces. En juillet 2018, nous disposons de 79 PPR d'édition caractérisées chez différentes espèces (*A. thaliana*, *Z. mays*, *O. sativa*, *P. patens*). Afin d'améliorer la matrice, la première chose à faire serait de la mettre à jour avec les données dont nous disposons. Ensuite, toutes les données observées par Takenaka et al. ne sont pas nécessairement prises en compte. Les auteurs ont utilisé un test statistique pour voir si pour un binôme donné d'acides aminés 5 et 35 d'un motif PPR, on avait un enrichissement pour au moins un nucléotide. Ce système élimine de fait la prise en compte de couples d'acides aminés rares dans la matrice. C'est le problème dans lequel se trouve SLO2 puisque sur ses 15 motifs, 5 n'ont pas d'information. Il est

difficile de garder un maximum d'information dans la matrice tout en tenant compte de la fréquence des observations des couples acides aminés-nucléotides, sans donner trop de poids dans un sens comme dans un autre. Si le système proposé par Takenaka et al. n'est pas parfait et en particulier pour SLO2, je ne vois pour le moment pas d'autre alternative simple. A terme, on peut imaginer via une approche *machine learning* (apprentissage automatique) une amélioration de la matrice utilisée (Libbrecht et Noble, 2015). En retirant chaque information une à une, on teste son apport réel à la matrice en la testant sur des données connues. Il serait donc possible de classer chaque information par rapport à son influence par rapport aux données connues et établir un seuil à partir duquel la non-prise en compte de l'information serait trop coûteuse en terme de prédiction.

Concernant le système de prédiction, nous utilisons un système détaillé à travers l'exemple de CLB19 et qui est différent que celui de Takenaka et al. De nombreux sites d'édition ont été ajoutés à la prédiction, et nous effectuons une sélection sur la moyenne des scores obtenus. Nous avons également ajouté l'information sur la présence de sites d'édition dans la zone de fixation ou non de la PPR afin d'en tenir compte. Notre système présente toutefois des limites. Comme nous avons pu le voir, le changement d'une seule base dans la séquence de fixation peut abolir l'édition du site en aval. Il s'agit de l'exemple du site de MEF19 suite au défaut du mutant $\Delta mef37$. L'utilisation de la moyenne ne permet pas de prendre en compte ce genre d'évènements. En d'autres termes, la moyenne comme outil mathématique n'est peut-être pas adaptée puisque les bases dont le score est très faible et pouvant être moléculairement réductrices peuvent n'avoir qu'une influence très limitée sur la moyenne. Des données récentes suggèrent également que certains motifs jouent un rôle moins important dans la fixation que d'autres. Dans le cas des PPR PLS, le travail de Kindgren et al. (2015) montre que des mutations des acides aminés impliqués dans la reconnaissance des nucléotides des motifs L n'a pas de conséquence sur l'affinité de fixation chez CLB19. Encore plus surprenant, les travaux récents de Miranda et al. (2017) s'intéressent à la fixation de la PPR la plus étudiée à ce jour, PPR10. Dans la caractérisation de cette PPR pure, les auteurs démontrèrent par exemple qu'en position -1 par rapport au début de la zone de fixation de PPR10, la base G était contre-sélectionnée. L'hypothèse avancée par les auteurs est que la présence d'un G en -1 modifie la structure tridimensionnelle de façon à rendre inaccessible à la PPR des bases démontrées comme étant primordiales à la fixation. En réalisant des substitutions d'une base sur l'ARN, les auteurs ont également démontré que l'affinité de fixation de PPR10 à l'ARN était très variable suivant la base mutée. Si certaines bases revêtent une importance primordiale dans la fixation, d'autres peuvent être mutées sans conséquence majeure sur l'efficacité de fixation. Ces données sont extrêmement intéressantes mais n'existent jusqu'à présent que pour PPR10. Une étude similaire sur d'autres PPR permettrait peut-être d'établir un système de prédiction nettement plus complet, en tenant compte de ces informations. Si la moyenne ne permet pas de tenir compte pleinement de ces informations, on peut par exemple calculer la probabilité de fixation d'une PPR en faisant le produit des probabilités de fixation pour chaque motif, tout en remplaçant les zéros par une valeur strictement positive mais suffisamment faible pour sanctionner la trop faible probabilité de fixation. Il serait intéressant pour la communauté PPR de s'intéresser aux nombreux travaux effectués notamment sur les prédictions de fixation des facteurs de transcription qui, en plus d'utiliser des matrices poids-position, intègrent des données supplémentaires comme des motifs particuliers (Chekmenev et al., 2005). Actuellement, nous ne tenons compte que de l'ARN et de la protéine PPR que l'on étudie pour tenter de déterminer quel site elle va éditer. On peut imaginer un système nettement plus complexe qui tiendrait compte des compétitions entre protéines PPR pour une zone de fixation, et même d'autres protéines fixant l'ARN dans les organites à condition que l'on ait des informations sur leur fixation. Rien n'indique non plus que les PPR d'édition comme MEF37 ne se fixent pas ailleurs. En effet, l'intérêt des PPR d'édition est que l'on peut savoir assez facilement où elles se fixent en observant des défauts d'édition. Il est possible que des PPR d'édition se fixent dans des zones ne portant pas de cytosine en +4, dans ce cas la fixation n'aura pas de conséquence en terme d'édition mais pourrait entrer en compétition avec la fixation d'une autre cible par la PPR ou d'une autre protéine par l'ARN. On pourrait également envisager de tenir compte de la structure de

l'ARN, si celle-ci est connue, pouvant modifier l'affinité de fixation ou encore la possibilité que certaines parties d'ARN soient rapidement dégradées par les nucléases et ne représentent pas une concurrence à la fixation pour les PPR dans la durée. La prise en compte de toute cette complexité biologique est encore très limitée voire inexistante dans les prédictions de fixation des protéines PPR mais devra probablement être prise en compte pour obtenir un code efficace.

Si notre approche classique est de rechercher pour une PPR donnée les sites qu'elle pourrait éditer, nous avons aussi utilisé une matrice inverse, c'est-à-dire que pour un site donné, on recherche les PPR ayant le meilleur score de fixation. Une information à apporter cependant : lorsque l'on recherche pour une PPR donnée quel site d'édition correspond le mieux au code PPR, nous comparons une PPR à n motifs, donc potentiellement fixant n nucléotides à x séquences. On se retrouve à calculer une moyenne de n individus. Ce n'est pas ce que l'on fait dans une matrice inverse puisque pour un site donné, nous allons comparer les moyennes des 57 PPR E+ qui portent 7 à 25 motifs PPR. Il est très délicat de comparer la moyenne des scores d'une PPR à 7 motifs avec une PPR à 25 motifs, et cette information n'est pas prise en compte dans nos matrices inverses.

Pour résumer, l'ensemble de la discussion semble tourner autour des nombreux défauts et approximations du code PPR encore existants avec pour conséquence possible pour le lecteur de se détourner des approches classiques par RT-PCR avec le code pour se focaliser sur le *RNAseq*. Des papiers récents confirment la nécessité de passer à une approche plus globale. On peut par exemple citer la publication de Rojas et al. (2018), démontrant que même pour les PPR les plus étudiées au monde –ici en l'occurrence il s'agit de PPR10, PGR3 et PPR5- de nouvelles zones de fixation sont encore découvertes aujourd'hui alors que celles-ci ont fait l'objet de très nombreuses recherches. Pour autant, il serait extrêmement dommageable pour la communauté de se passer du code PPR car à partir du moment où nous disposerons d'un système de prédiction fiable et mettant en évidence tous les sites associés à une PPR donnée, cela signifiera que nous aurons parfaitement compris le fonctionnement moléculaire de la fixation d'une protéine PPR à un ou plusieurs ARN. Enfin, puisque la plupart des défauts partiels sont souvent vus à tort comme des défauts secondaires et rarement publiés, leur prise en compte dans le futur permettrait très probablement de renforcer la matrice du code PPR.

Parmi les PPR E+ présentes chez *Arabidopsis thaliana*, l'une d'entre elles a particulièrement attiré notre attention. PGN (pour *Pentatricopeptide repeat protein for Germination on NaCl*) est une protéine PPR décrite par Laluk et al. (2011) comme ayant un rôle dans la résistance aux stress abiotiques et aux champignons nécrotrophes. Le chapitre suivant va détailler l'état des connaissances sur cette PPR et les expérimentations que nous avons effectué.

Chapitre 3 : Caractérisation fonctionnelle de la protéine PPR PGN en condition de stress.

1. Etat des lieux

1.1. Une PPR mitochondriale à l'expression induite par *Botrytis cinerea*

Dans de précédents travaux (Dhawan et al., 2009), et dans le but de mieux comprendre les mécanismes de défense de la plante face aux nécrotrophes, Laluk et coll. comparèrent le transcriptome du mutant *Abik1* à des plantes sauvages. BIK1 (pour *Botrytis-Induced Kinase 1*) est une protéine kinase dont la transcription est induite en présence de *Botrytis* (Veronese et al., 2006). A travers l'étude de son mutant, les auteurs montrèrent qu'en réponse à une souche avirulente de *Pseudomonas syringae* (AvrRpm1), les réponses du mutant restaient inchangées par rapport à la plante sauvage, suggérant que BIK1 serait impliquée dans les défenses basales et non spécifiques de la plante. En plus d'une sensibilité marquée à *Alternaria brassicicola* et *Botrytis cinerea*, deux pathogènes nécrotrophes, *Abik1* est plus résistant à *Pst*. Lors de l'étude du transcriptome de *Abik1*, Laluk et al. (2011) se sont intéressés à PGN (AT1G56570) comme étant une cible putative de BIK1 car l'expression de *pgn* augmentait significativement chez le mutant *Abik1* par rapport au sauvage, tout comme chez les plantes sauvages en réponse à *Bc* (Figure 23).

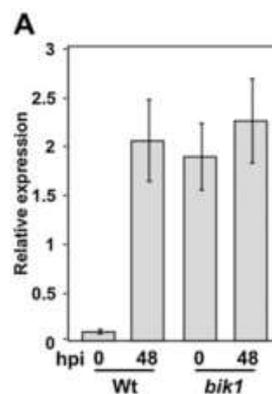


Figure 23 : Expression du gène *pgn* chez le sauvage (Wt) et le mutant *Abik1* (*bik1*) à 0 et 48h après infection par *Bc*.

Adapté de Laluk et al. (2011).

Les auteurs montrèrent par la suite que le mutant *Apgn* présentait une sensibilité plus marquée à *Bc* que le sauvage (Figure 24). Il est à noter que les temps d'infections (5 et 7 jours suivant la méthode choisie) sont assez longs par rapport à ce que nous avons l'habitude de faire (48 à 72h).

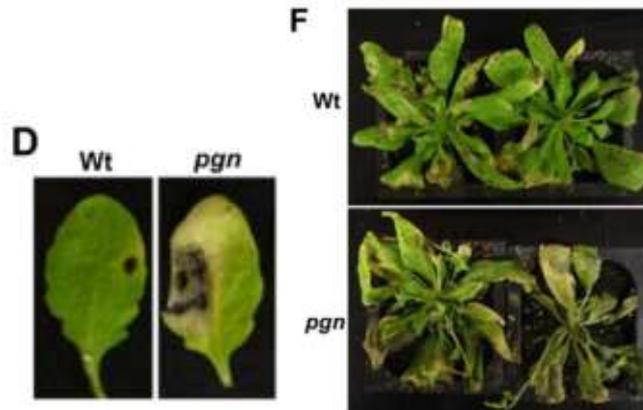


Figure 24 : Symptômes observés après inoculation avec *Bc* par la méthode goutte, après 5 jours (D) et pulvérisation après 7 jours (E), chez le sauvage (Wt) et le mutant Δpgn (pgn).

Adapté de Laluk et al. (2011).

A l'aide d'une fusion GFP, les auteurs ont également démontré par agroinfiltration dans *N. benthamiana* et observation au microscope à épifluorescence une localisation mitochondriale de la protéine PGN. Le contrôle utilisé était la mCHERRY, un marqueur mitochondrial qui co-localisait avec la GFP dans leur expérience (Figure 25).

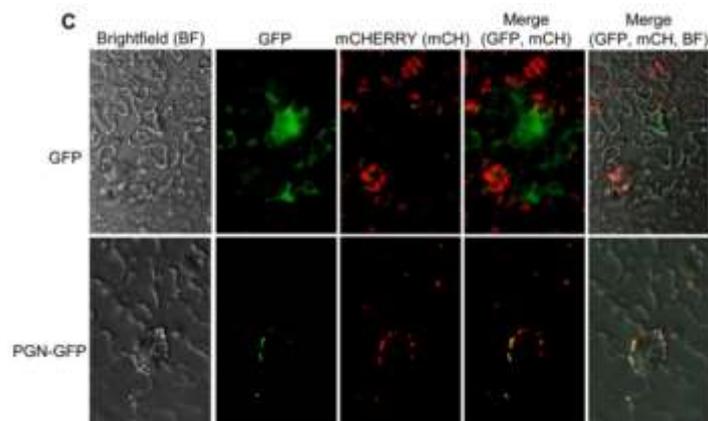


Figure 25 : Observations au microscope à épifluorescence d'épiderme de *N. benthamiana* exprimant transitoirement PGN-GFP.

Adapté de Laluk et al. (2011).

1.2. Un mutant hypersensible au glucose, au NaCl et à l'acide abscissique

Le mutant Δpgn est également associé à une sensibilité accrue au NaCl, au glucose et à l'ABA (Figure 26).

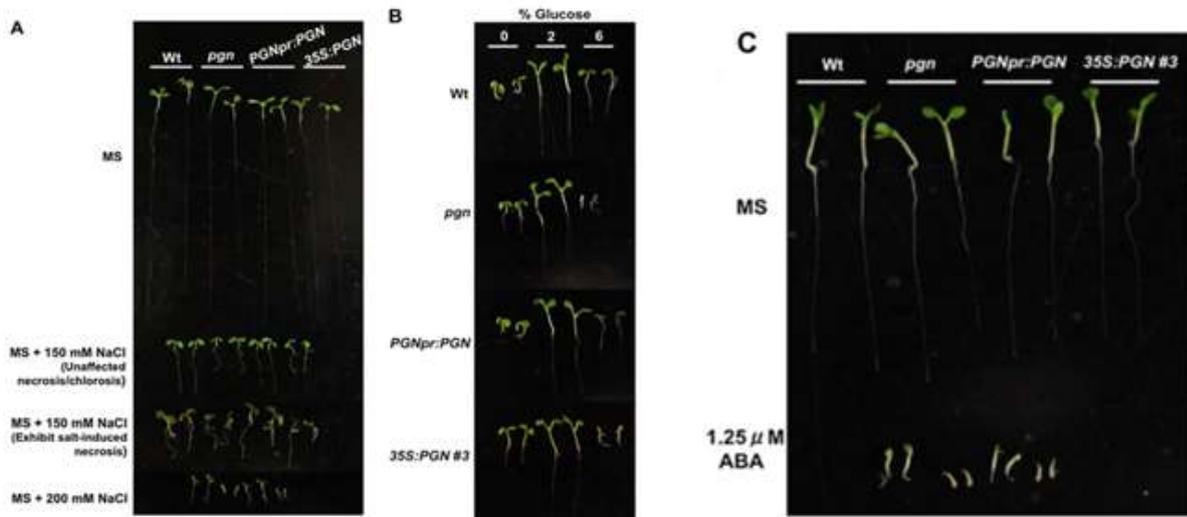


Figure 26 : Hypersensibilité du mutant *Δpgn* au NaCl (A), au glucose (B) et à l'ABA (C) après 10 jours.

Adapté de Laluk et al. (2011).

Les auteurs ont effectué deux types de complémentation, l'une avec le promoteur natif de *pgn* et la seconde à l'aide d'un promoteur fort 35S. Fait étonnant, si le promoteur natif permet une complémentation du mutant en présentant une sensibilité proche de celle du sauvage, le surexprimeur avec le promoteur 35S présente quant à lui le même phénotype que le mutant KO dans les différents stress. Il est cependant à noter qu'une seconde différence existe, autre que le promoteur. En effet, selon le matériel et méthode des auteurs, le clonage a été effectué dans un plasmide de destination pCAMBIA 99-1 avec une étiquette 3xHA. Bien que les amorces choisies montrent que le clonage est en stop, donc sans l'étiquette 3xHA, les auteurs indiquent que les mêmes amorces ont servi au clonage pCAMBIA 99-1 avec une étiquette GFP, ce qui suggère un cadre de lecture ouvert. Les auteurs n'ayant pas répondu à nos nombreuses sollicitations, nous faisons l'hypothèse qu'il n'existe non pas une différence entre les deux lignées de complémentation mais deux, à savoir le promoteur, et l'étiquette en position C-terminale.

1.3. L'expression de gènes impliqués dans le fonctionnement de la mitochondrie perturbée chez le mutant *Δpgn*

Après avoir observé une localisation subcellulaire mitochondriale et avec les nombreuses PPR décrites comme étant impliquées dans les différentes étapes de maturation des ARN des organites, Laluk et al. (2011) se sont intéressés à l'expression de différents gènes de la mitochondrie ou nucléaires dont la protéine est envoyée dans la mitochondrie (Figure 27).

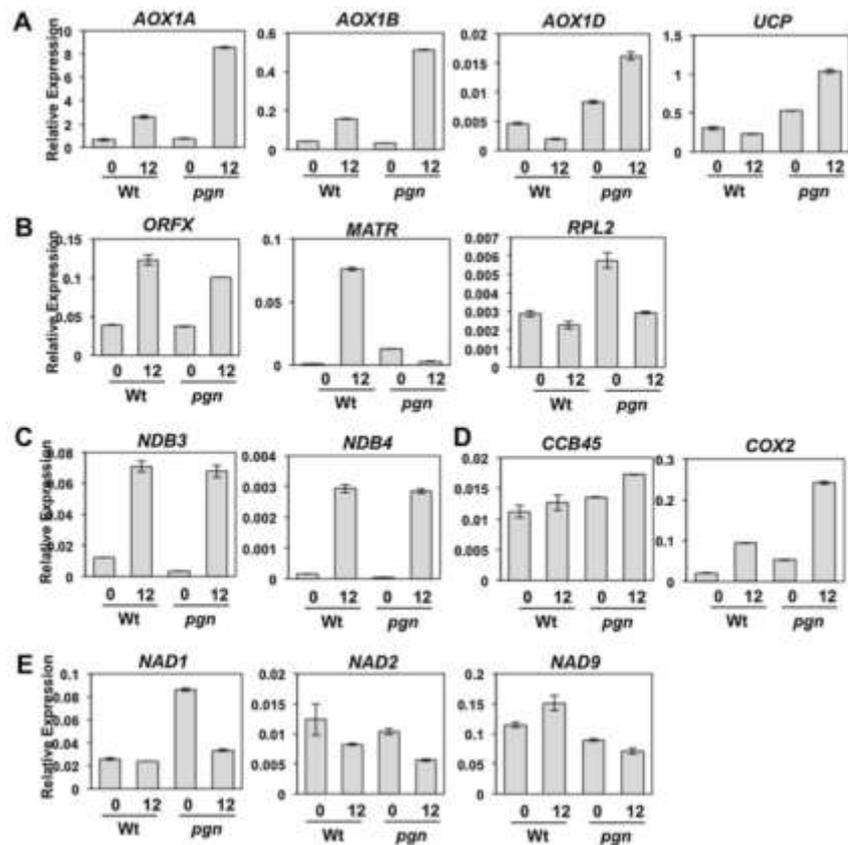


Figure 27 : Profils d'expression de gènes associés à la mitochondrie à 0 et 12h après traitement NaCl chez le sauvage (Wt) et le mutant Δpgn (Δpgn). Gènes associé au transport d'électrons (A), à la transcription (B), à l'oxydation/réduction (C), à la régulation du cytochrome c (D) et aux sous-unités de la NAD(P)H déshydrogénase (E).

Tiré de Laluk et al. (2011).

Malgré les nombreuses variations d'expression observées, les auteurs ne parvinrent pas à mettre en évidence de mécanisme d'action clair pour PGN. Par coloration au DAB (3-3'-Diaminobenzidine), les auteurs ont aussi observé une suraccumulation de peroxyde d'hydrogène chez le mutant Δpgn en réponse à 200mM de NaCl, en comparaison avec une plante sauvage (Figure 28).

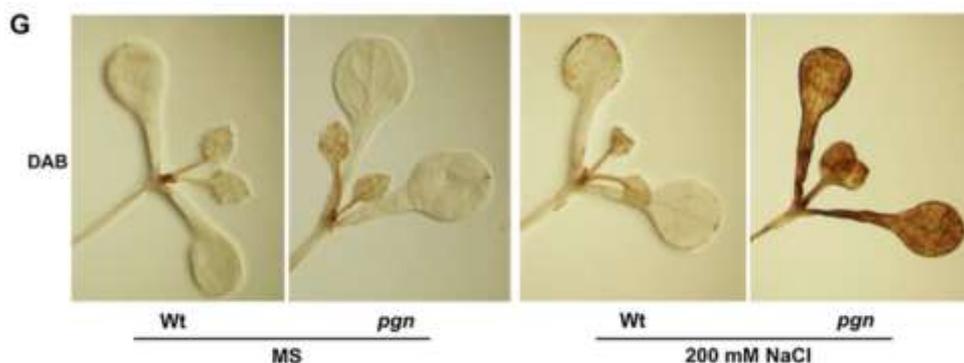


Figure 28 : Observation de l'accumulation de peroxyde d'hydrogène par coloration DAB chez le sauvage (Wt) et le mutant Δpgn (Δpgn) avec ou sans traitement au NaCl.

Tiré de Laluk et al. (2011).

1.4. Hypothèse de travail

PGN est donc une PPR mitochondriale, appartenant au sous-groupe des PPR E+ connu pour être impliqué dans l'édition. A partir de toutes ces données, provenant pour l'immense majorité du travail de Laluk et al. (2011), l'hypothèse de travail qui nous a semblé la plus évidente est que PGN est une PPR d'édition dans la mitochondrie d'*Arabidopsis thaliana*. L'intérêt principal de l'étude de PGN est l'induction de son expression en réponse à *Bc*. Les PPR étant connues pour être impliquées dans la maturation des ARN des organites, et cette maturation étant la principale voie de régulation de l'expression des gènes des organites (Hammani et Giegé, 2014), cela mettrait en évidence une régulation de l'expression de gènes mitochondriaux en réponse à *Botrytis*. A partir des données obtenues dans le chapitre précédent, nous avons recherché d'éventuels défauts d'édition dans le mutant SALK_141937 utilisé par Laluk et al. (2011). Les auteurs de la publication n'ayant jamais répondu à nos sollicitations, nous avons commencé les clonages et la sélection du mutant de zéro.

2. PGN, une PPR E+ impliquée dans l'édition

2.1. Etude *in silico*

A l'aide de l'outil TPRpred (Karpenahalli et al., 2007), nous avons dans un premier temps vérifié si PGN était bien une protéine PPR. Elle est prédite à 100% pour être une PPR, et l'outil détecte exactement 10 motifs, avec un paramètre d'e-value à 10^{-2} (Figure 29).

Per-protein P-value for being PPR : **0.0E+00**
Probability for being PPR : **100.00%**

Repeat	Begin	Alignment	End	P-value
PPR	77	-VAWTAMI TGYASSNYNARAWECFHEMVKQGTSPNE--	111	2.1e-13
PPR	112	-FTLSVLKSCRNKVLAYGALVHGVVVKLGMEGSL--	146	4.6e-05
PPR	179	-VTWTLITGFTHLGDGIGGLKMYKQMLLENAEVTP--	213	1.2e-10
PPR	214	-YCITIAVRASASIDSVTTGKQIHASVTKRGFQSNL--	248	7.3e-07
PPR	249	-PVMNSILDLYCRGYSLEAKHYFHEMEDKDLITWN--	283	4.8e-13
PPR	314	-YTFTSLVAACANIAALNCGQLHGRIFRRGFNKNV--	348	1.2e-08
PPR	381	-VSWTSMHIGYSGHYGAEAVELFDKHWSSGIRPDR--	415	0.0e+00
PPR	416	-IVFMAVLSACRHAGLVEKGLKYFNVHSEYGINPD--	450	2.0e-10
PPR	452	-DIYNCVVDLLGRAGKIGEAYELVERMPFKPDESTW--	486	4.9e-12
PPR	519	-GTYVHLSYIYAAEGKWVDFARVRKMMRMGNKKEA--	553	2.7e-04

Figure 29 : Résultat affiché pour PGN (AT1G56570) par l'outil TPRpred.

En s'appuyant sur les données de Cheng et al. (2016), PGN est une PPR portant 16 motifs (44AA-SS-P1-L1-S1-P1-L1-S1-P1-L1-S1-P2-L2-S2-E1-E2-E+) appartenant au sous-groupe des E+. Ce sous-groupe de PPR est connu pour son implication dans les mécanismes d'édition C vers U des organites.

2.2. Utilisation de la nouvelle méthode de prédiction pour PPR E+

Comme décrit dans le chapitre précédent, à partir de l'amélioration du code nous avons déterminé les 5 sites pour lesquels la fixation de PGN était la plus probable et nous les avons testés par RT-PCR. Nous étions tombés sur deux défauts d'édition, l'un sur le 5' UTR (pour *Untranslated region*) de *nad6*, le second dans la séquence codante de *cox2* (Tableau XV).

Tableau XV : Sites différentiellement édités chez le mutant Salk_141937 (AT1G56570, Δpgn) en comparaison avec Col0 avec ajout du critère DYW2.

Position	Annotation	E+	Edition Col0	Edition <i>Adyw2</i>	Edition Salk_141937	PPR associée	Sous-famille
M165765	<i>nad6_leader_73</i>	1	81-84%	3%	0%	-	-
M8348	<i>cox2_742</i>	1	100%	11%	0%	-	-

Ces deux gènes mitochondriaux codent pour des protéines impliquées dans la chaîne de transport des électrons de la mitochondrie (mETC). Nad6 (ATMG00270) est impliquée dans le complexe I NADH déshydrogénase, alors que Cox2 (ATMG00160) fait partie du complexe IV cytochrome c oxydase.

L'édition de *cox2* en position 742 entraîne un changement d'acide aminé de l'arginine (R) vers le tryptophane (W). Un alignement multiple de séquences protéiques de Cox2 codées par différents génomes mitochondriaux montre que c'est le W qui est conservé chez les non-embryophytes, qui sont dépourvus d'édition (Annexe S3). On peut donc penser que l'édition de ce site chez *A. thaliana* permet de restaurer un acide aminé conservé comme c'est souvent le cas pour les sites d'édition. De plus, l'alignement des séquences protéiques de Cox2 codées par les génomes mitochondriaux des Embryophytes (Annexe S4) montre que 45% des séquences protéiques représentées ont déjà un W, suggérant que l'édition de la position 742 du transcrite de *cox2* n'est pas très conservée. L'impact de l'édition dans le 5'UTR de *nad6* est plus difficile à estimer car il est en dehors de la partie codante. Cependant, elle peut avoir des conséquences sur la structure de l'ARN, donc sur son accumulation et sa traduction. Des alignements multiples de séquences génomiques de *nad6* issues de différentes espèces végétales sur BlastN, en partant de la séquence d'*A. thaliana*, montrent que cette région est assez conservée alors qu'à la position d'édition, le C est parfois remplacé par un T (Annexe S5). Certaines espèces, à l'image de *Nicotiana glauca*, possèdent deux copies de *nad6*, l'une avec une version génomique avec C dans le 5' UTR, l'autre avec un T. Cela suggère que ce nucléotide peut jouer un rôle. Cependant nous ne disposons pas d'assez de données pour estimer la conservation de l'édition de ce site.

2.3. Analyse du transcriptome

L'analyse *RNAseq* a confirmé que ces deux défauts d'édition étaient bien les seuls, suggérant que les défauts associés à l'absence d'une protéine PGN fonctionnelle étaient associés à l'un des deux défauts d'édition, ou à une interaction entre les deux défauts.

L'analyse d'expression différentielle met également en évidence 80 gènes différentiellement exprimés, tous nucléaires (Annexe S6). Une analyse d'enrichissement GO ne montre aucun enrichissement pour les fonctions biologiques, localisations subcellulaires, domaines protéiques ou encore processus biologique.

Cela signifie que les défauts d'édition des sites *nad6* et *cox2* n'entraînent pas de défaut d'accumulation du transcrite. Nous avons vérifié que la couverture était suffisante pour pouvoir maintenir cette affirmation (Tableau XVI).

Tableau XVI: Analyse d'expression différentielle de *nad6* et *cox2*.

Annotation	AGI	Comptages normalisés pour Col0	Comptages normalisés pour Δpgn	log2FoldChange	BH
<i>nad6</i>	ATMG00270	14463.52953	14073.0182	-0.039487884	0.998914049
<i>cox2</i>	ATMG00160	12806.54084	13350.18523	0.059978916	0.998914049

A l'aide d'une analyse de couverture, nous avons également vérifié qu'il n'existait pas de défaut d'accumulation ou de maturation de ces deux transcrits.

2.4. Complémentation fonctionnelle du mutant *Δpgn*

Afin de valider l'implication de PGN dans l'édition des sites *nad6_leader_73* et *cox2_742*, nous avons réalisé deux complémentations fonctionnelles dans le fond génétique *Δpgn* : l'une avec une version stop de la protéine PGN, la seconde avec un cadre de lecture ouvert. Dans les deux cas, le clonage a été effectué à l'aide de vecteur Gateway pGWB14, contenant un promoteur fort 35S et une étiquette 3xHA. L'idée en plus de compléter était de tester le résultat étonnant mais très intéressant et méritant un approfondissement de Laluk et al. (2011) suggérant qu'un niveau particulier d'expression de *pgn* était nécessaire à une bonne complémentation. Un clonage sous dépendance d'un promoteur 35S de PGN avec une étiquette GFP a également été effectué mais suite à un problème de transformation, nous n'avons pas pu obtenir les plantes transformées dans le temps imparti de la thèse. Encore une fois, l'idée derrière cette construction était de tester l'influence d'une étiquette en partie C-terminale dans la complémentation d'un mutant d'édition.

Les résultats obtenus sont présentés dans la figure 30 :

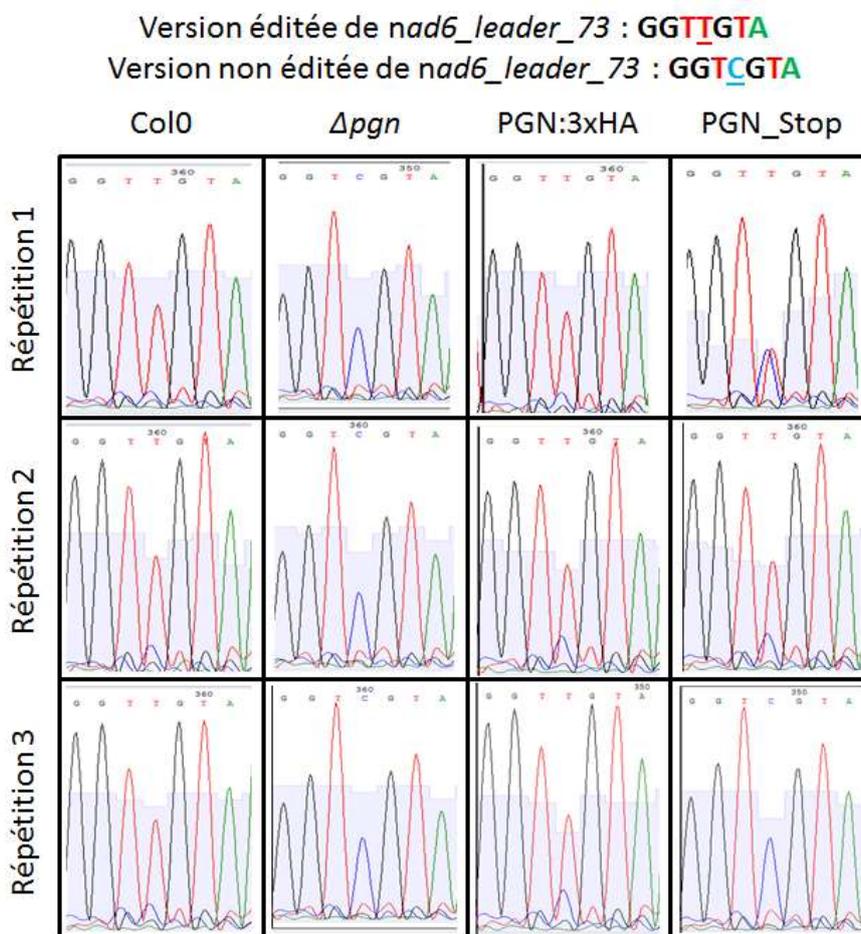


Figure 30 : Chromatogrammes du site *nad6_leader_73* chez *A. thaliana* Col0, *Δpgn*, et les lignées *Δpgn* complémentées avec PGN avec étiquette 3xHA et sans étiquette.

Les répétitions sont des répétitions biologiques. Mesures faites sur feuilles de plantes de 4 semaines en jours longs. Plantes de génération T2 issues d'un même évènement de transformation.

Nos résultats montrent que la complémentation fonctionne, démontrant que PGN est bien impliqué dans l'édition de *nad6*, et très probablement de *cox2* même si cela n'a pas été regardé ici. Autre résultat intéressant : l'utilisation d'un promoteur 35S permet la complémentation, ce qui montre qu'il n'est pas nécessaire d'utiliser le promoteur natif et d'avoir un niveau de transcription finement contrôlé. La présence de l'étiquette 3xHA n'affecte pas non plus l'édition. Toutefois, dans un cas sur 3, la version stop de PGN ne semble pas compléter ; dans un autre cas la complémentation est très partielle mais elle fonctionne dans le dernier cas.

Afin de comprendre ce résultat, nous avons testé les différents niveaux d'expression de *pgn*. Deux jeux d'amorces ont été utilisés pour *pgn* en RT-qPCR : celui proposé par Laluk et al., où celles-ci sont situées de part et d'autre de l'unique intron (82pb) ; le second que nous avons désigné de part et d'autre du T-DNA (Annexe S7). La première constatation est que pour les deux jeux d'amorce, nous avons une amplification même dans le fond génétique Δpgn . Si son niveau est toujours plus bas que tous les autres, les courbes de fusion de la RT-qPCR suggèrent pourtant qu'il n'y a qu'un seul et même produit. Nous n'avons aucune explication à ce phénomène. Il conviendra donc de considérer l'expression de *pgn* avec précaution.

On remarque que pour les 3 répétitions, l'expression de PGN avec l'étiquette 3xHA est très forte. Ça n'est pas le cas de la version stop de PGN, mais même si l'expression est moins forte, elle reste supérieure à celle du Col0, donc cela n'explique pas l'absence de complémentation. Pour la suite de l'analyse, les amorces utilisées sont celles situées de part et d'autre du T-DNA.

3. PGN, une PPR dont la transcription est induite en condition de stress

Les données issues de Laluk et al. (2011) montrent que la transcription de *pgn* est induite 48h après infection de *Botrytis cinerea* (Figure 23). Ces données montrent aussi que l'expression du gène codant pour Cox2 augmente en condition de stress salin (Figure 27). A partir des informations sur l'édition obtenues lors de la thèse et des données publiées sur PGN, une hypothèse très attrayante serait de supposer que l'expression du gène codant pour PGN est induite en condition de stress, avec pour conséquence(s) une augmentation de l'édition de *nad6* et/ou *cox2* ou une augmentation de la quantité de transcrits avec un niveau d'édition stable, suggérant une augmentation de l'activité d'édition. Afin de se rapprocher au maximum des conditions de Laluk et al. (2011), nous avons reproduit leurs résultats en condition de stress salin (Annexe S8) et pathotest *Botrytis cinerea* (Annexe S9).

3.1. En condition de stress biotique

Nous avons donc infecté avec le pathogène des lignées Col0 et Δpgn avec la méthode du spray, c'est-à-dire en pulvérisant des spores de *Bc* sur les plantes. L'intérêt de cette méthode, en comparaison avec la méthode plus classique de la goutte consistant à déposer une goutte contenant les spores du pathogène sur une feuille et mesurer la taille de la lésion, est qu'elle ne pose pas la question de la zone de prélèvement puisque toute partie de la plante pulvérisée peut être prélevée. Cependant, cette méthode présente un inconvénient majeur, à savoir sa reproductibilité, car même si l'on pulvérise toujours la même quantité de la même hauteur, diluée de la même façon, il est bien plus difficile de maîtriser l'infection en comparaison avec la méthode de la goutte. Des prélèvements ont été réalisés à 0 et 48h après infection sur feuille, à l'aide d'un emporte-disque et nous avons effectué des analyses par RT-qPCR de l'expression des gènes codant pour PGN, Cox2 et Nad6 (Figure 31) mais également des RT-PCR pour mesurer le niveau d'édition du site *nad6_leader_73* chez Col0.

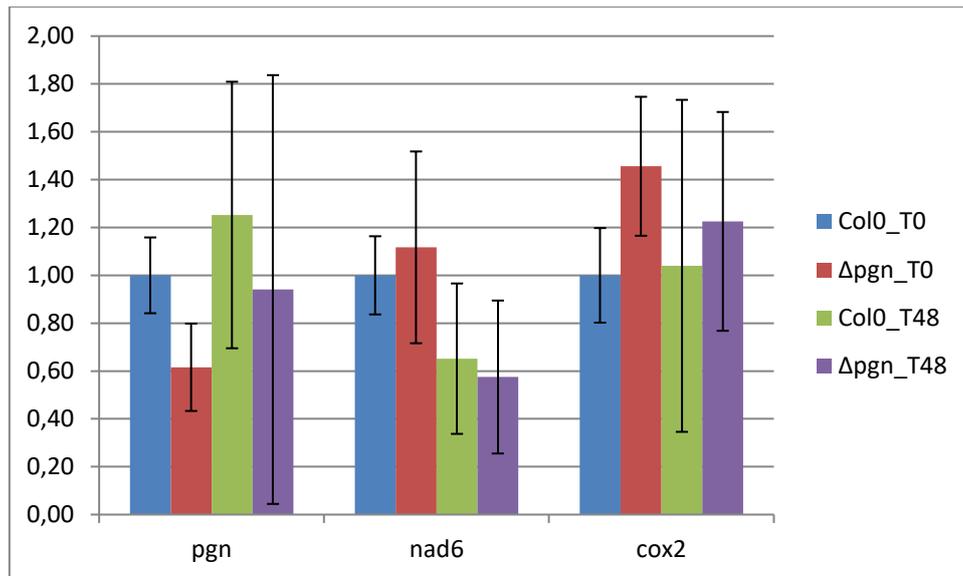


Figure 31 : Représentations graphiques des données de RT-qPCR effectuées sur Col0 et Δpgn avant infection (T0) et à 48h après infection (T48) par *Botrytis cinerea*.

Chaque condition génotype de plante/infection/répétition est composée de deux plantes par répétition biologique et de trois répétitions biologiques, leur variabilité étant représentée par les écart-types. Chaque analyse en qPCR a été réalisée en triplicats techniques avec analyse de courbe de fusion. Les données sont corrigées par la méthode du $\Delta\Delta C_t$ sur l'actine de la plante et comparé à la moyenne pour Col0 non traité.

Nous n'avons trouvé aucune variation d'édition de *nad6_leader_73* chez le Col0 dans toutes les conditions et répétitions biologiques. Par RT-qPCR, nous avons vérifié l'infection en validant la présence du pathogène à l'aide d'amorces sur l'actine de *Botrytis*, mais les niveaux d'infections sont assez variables d'un échantillon à l'autre. Nous ne voyons pas d'induction de l'expression de *cox2* ou de *nad6* après infection, quel que soit le fond génétique. Nous ne retrouvons pas non plus l'induction de *pgn* décrite par Laluk et al. (2011). Dans les différentes répétitions biologiques, nous n'avons pas vu de corrélation positive entre le niveau d'infection et les niveaux d'expression des différents gènes. Nous testons tout de même la même hypothèse en condition de stress abiotique, en stress salin.

3.2. En condition de stress salin

De la même façon, nous avons mesuré l'édition de *nad6_leader_73* et les niveaux de transcrits *pgn*, *nad6* et *cox2* chez Col0 et Δpgn en condition de stress ou non. Dans la condition sel, nous avons séparé les plantes présentant un phénotype chlorotique de celles qui n'en avaient pas (Annexe S10). Les résultats de RT-qPCR sont représentés sur la figure 32.

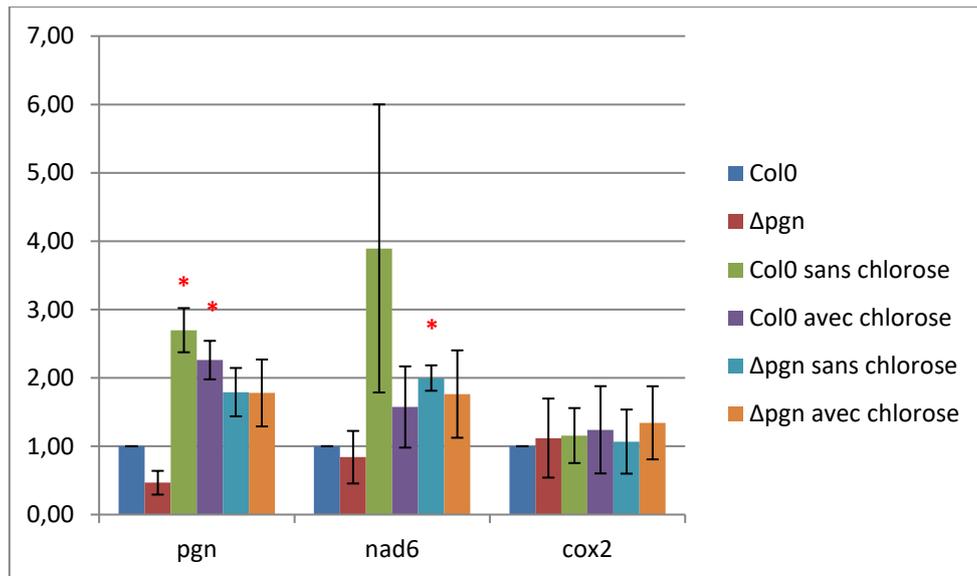


Figure 32 : Représentations graphiques des données de RT-qPCR effectuées sur plantules de 10 jours Col0 et Δ pgn, sur 1/2 MS 1% sucrose avec et sans 150mM NaCl

Chaque condition génotype de plante/stress est composée d'un pool de plantules par répétition biologique et de trois répétitions biologiques, leur variabilité étant représentée par les écart-types. Chaque analyse en qPCR a été réalisée en triplicats techniques avec analyse de courbe de fusion. Les données sont corrigées par la méthode du $\Delta\Delta$ Ct sur l'actine de la plante et comparé à la moyenne pour Col0 non traité. Le test statistique utilisé est un test de Student par rapport à Col0 sans stress, une étoile indiquant que la *p-value* est inférieure à 0,05 après correction de Bonferroni.

Tout d'abord, les données d'édition sur Col0 dans toutes les conditions de stress et répétitions ne montrent aucun différence par rapport à la condition contrôle sans stress salin. L'expression du gène codant pour Nad6 est induite en condition de stress salin sans chlorose chez Δ pgn. La tendance est la même chez Col0 sans chlorose, mais la variabilité marquée sur l'une des répétitions biologiques empêche de conclure dans le sens d'une différence. La tendance est aussi à une induction de pgn, statistiquement significative chez Col0 en condition de stress salin. Ces données suggèrent une corrélation positive entre l'expression de pgn et celle de nad6 dans ces conditions.

4. Discussion

Pour résumer, nous avons démontré que PGN est impliqué dans l'édition de *nad6* et très probablement de *cox2*, même si cela reste à confirmer sur les lignées de complémentation. Concernant l'hypothèse de la régulation fine de PGN nécessaire à jouer son rôle, en particulier en condition de stress, elle ne tient pas puisque la complémentation fonctionne sous dépendance d'un promoteur 35S et que le promoteur natif n'a pas été utile. Nous n'avons observé aucune variation d'édition en condition de stress, avec *Botrytis* comme en stress salin, chez Col0. Même si nous avons une tendance suggérant une corrélation positive entre l'expression des gènes pgn et nad6, suggérant que le premier est induit pour répondre à une demande croissante d'édition, nous n'avons rien vu de tel en condition de stress biotique. Nous n'avons pas non plus retrouvé l'induction de pgn vue par Laluk et al. alors que nous retrouvons bien le phénotype de sensibilité à *Bc* associé au mutant. De manière générale, si une régulation de l'expression de pgn existe, elle ne semble servir qu'à maintenir des niveaux d'édition élevés pour *cox2* et *nad6*, sans que cela passe par des niveaux d'édition intermédiaires et une régulation fine. Concernant les données de RT-qPCR, ce que l'on obtient avec *Botrytis* illustre les

difficultés en phytopathologie d'utiliser des approches de transcriptomiques. Afin d'améliorer cette expérimentation, il est nécessaire de stabiliser nos conditions d'infection afin d'obtenir des niveaux d'infection plus proches les uns des autres.

D'un point de vue mécanistique, on est donc face à une protéine PPR éditant deux sous-unités de complexes mitochondriaux, les complexes I (pour *nad6*) et IV (pour *cox2*). L'absence d'édition pour les deux sites n'entraîne pas de défaut d'accumulation ou de couverture pour les deux transcrits. Des analyses par Western Blot permettraient de tester si l'accumulation de ces deux protéines est similaire à celle observée chez des plantes de type sauvage. L'idéal serait de faire des gels en deux dimensions en condition non dénaturante (2D-Blue-Native PAGE) pour observer les différents complexes de la chaîne de transport des électrons (Eubel et al., 2003). En cas de défaut d'accumulation d'un ou plusieurs complexes, cela permettrait de savoir quel défaut d'édition est responsable du phénotype. Il est également possible que ce phénotype provienne de l'interaction des deux défauts. Afin de formuler des hypothèses, je m'appuie sur l'excellent travail de Colas Des Francs-Small et Small (2014), résumant les mutants de maturation des ARN mitochondriaux, avec leurs phénotypes associés chez *A. thaliana* et d'autres espèces végétales.

Dans le cas de défauts d'édition hors des séquences codantes, il est possible d'imaginer que cela modifie l'affinité de fixation d'une protéine de liaison à l'ARN nécessaire à la maturation. RPF5 est une PPR pure associée à la maturation de *nad6*, *atp9* et de l'ARNr 26S (Hauler et al., 2013). Le mutant $\Delta rfp5$, à l'image Δpgn , présente un taux de germination inférieur à celui de Col0. Impossible cependant d'identifier le (ou les) défaut(s) responsable(s) du phénotype chez le mutant $\Delta rfp5$. En plus de démontrer qu'en l'absence de la PPR il n'y avait pas d'accumulation du transcrit *nad6*, les auteurs ont proposé par application du code PPR une zone putative de fixation, qui se trouve en 5' UTR de *nad6*. Elle est cependant assez loin du site d'édition dépendant de PGN, il s'avère donc assez improbable que le défaut d'édition affecte la fixation de RPF5. De plus, l'accumulation de *nad6* n'est pas affectée chez Δpgn . Trois hypothèses restent alors possibles : le défaut d'édition crée une modification de la structure de l'ARN ou augmente/diminue l'affinité d'une protéine de fixation à l'ARN, les deux pouvant entraîner des défauts de traduction. La dernière hypothèse est que le défaut d'édition est sans effet.

Du côté de *cox2*, le défaut d'édition entraîne un changement d'acide aminé, passant d'un acide aminé apolaire (tryptophane) à un acide aminé chargé positivement (arginine). De la même façon que pour Nad6, il serait intéressant d'étudier l'accumulation de la protéine Cox2. Ces deux acides aminés sont conservés parmi de nombreuses espèces végétales, qu'elles fassent de l'édition ou non. Il est connu depuis plusieurs décennies que l'édition C vers U contribue pour partie à un maintien de l'information génétique chez les organites des plantes (Gualberto et al., 1989). De la même façon que pour Nad6, il est primordial de s'intéresser à la protéine Cox2 afin de déterminer si ce défaut d'édition et cette variation d'acide aminé modifie la stabilité de la protéine, mais aussi du complexe IV de la mitochondrie.

De manière générale, un bon nombre de mutants d'édition dans la mitochondrie ont été caractérisés moléculairement, sans toutefois montrer de phénotype macroscopique en condition standard. On peut citer par exemple MEF25, MEF10, REME1, MEF1, MEF22, MEF18 ou encore MEF9 (Colas Des Francs-Small et Small, 2014). A l'image de PGN, un certain nombre de défauts d'édition mitochondriaux ne semblent pas entraîner de phénotype en conditions standards de croissance, mais semblent perturber fortement les réponses de la plante en condition de stress.

Après s'être intéressé à une protéine PPR endogène d'*A. thaliana* concernant les aspects de régulation de l'expression des gènes des organites en condition de stress biotique, nous allons regarder dans le chapitre qui vient quel(s) rôle(s) peuvent jouer dans la plante deux PPR exogènes. Ces deux PPR sont codées par des bactéries pathogènes de plante, la première est *Ralstonia solanacearum* et la seconde *Erwinia amylovora*.

Chapitre 4 : Mise en évidence et caractérisation fonctionnelle de protéines bactériennes à motifs PPR

1. Mise en évidence de protéines à motifs PPR chez les bactéries

1.1. Initiation du projet

Comme nous avons pu le voir dans l'introduction, on trouve des protéines à motif PPR chez de très nombreux Eucaryotes, mais aucune n'a jamais été étudiée chez les bactéries (Barkan et Small, 2014). La première publication présentant plusieurs protéines PPR chez les bactéries est celle de Schallenberg-Rüdinger et al. (2013). Plusieurs gènes codant pour des protéines PPR ont été trouvés dans les génomes séquencés de plusieurs bactéries : *Rhodobacter sphaeroides*, une bactérie photosynthétique ; *Ralstonia solanacearum*, *Erwinia pyrifoliae* et *Erwinia amylovora*, des bactéries pathogènes de plante et *Simkania negevensis*, une bactérie pathogène d'animaux. Schallenberg-Rüdinger et al. font également remarquer que la majorité de ces bactéries vivent à l'état naturel en contact proche avec des Eucaryotes, suggérant également l'hypothèse d'un transfert horizontal de gènes pour expliquer l'obtention de gènes codant pour des protéines PPR. Néanmoins, le caractère hautement dégénéré des motifs PPR nous a amené à une seconde hypothèse : les PPR sont peut-être présentes chez les Bactéries, avec un motif tout aussi dégénéré mais de manière différente que celui des Eucaryotes. Le motif PPR consensus étant construit sur les PPR de plantes, il est possible que nous soyons en train de passer à côté de nombreuses protéines PPR chez les bactéries, du fait de leur éventuelle évolution en parallèle des Eucaryotes. La stratégie généralement utilisée pour détecter de nouveaux motifs PPR est celle du Modèle Caché de Markov (ou HMM, pour *Hidden Markov Model*). Cette méthode a été utilisée pour détecter les protéines PPR chez les plantes (Lurin et al., 2004), affiner cette détection dans un second temps (Cheng et al., 2016) voire même détecter des motifs PPR dans des espèces non végétales (Lipinski et al., 2011). En nous basant sur les protéines PPR vues par Schallenberg-Rüdinger et al. (2013) et en utilisant pour partie la méthodologie développée par Lipinski et al. (2011) chez la levure, nous allons tester notre hypothèse et rechercher les protéines à motifs PPR chez les bactéries.

1.2. Recherche de motifs PPR chez les bactéries

1.2.1. Méthodologie

Bien que des motifs PPR existent chez les Archées (du genre *Methanosarcina* par exemple), nous ne nous intéressons qu'aux bactéries. La méthodologie utilisée est pour partie inspirée de Lipinski et al. (2011) : à partir des données de Schallenberg-Rüdinger et al. (2013), les protéines PPR de *Ralstonia solanacearum*, *Rhodobacter sphaeroides*, *Simkania negevensis*, *Erwinia pyrifoliae* et *Erwinia amylovora* sont récupérées. L'outil TPRpred pour la détection des motifs PPR (Karpenahalli et al., 2007) est ensuite utilisé. Cet outil donne également une valeur de confiance en pourcentage concernant la protéine testée, pour les PPR et TPR. A cause de la forte proportion de protéines TPR chez les bactéries, et afin de ne pas ajouter du bruit à notre consensus bactérien, un seuil arbitraire a été choisi pour cette valeur de confiance : si la différence entre la valeur de confiance PPR et TPR est strictement supérieure à 80 points, alors la protéine sera considérée comme étant une protéine PPR. Le seuil de 80 points a été choisi en se basant sur les différences de confiance PPR – TPR des 5 protéines PPR bactériennes proposées par Schallenberg-Rüdinger et al. (2013), la protéine ayant la plus faible différence de confiance étant à 83 points. Enfin, je pars du principe qu'une protéine PPR doit avoir au moins deux motifs PPR. Toutes les protéines ayant un seul motif ne sont pas prises en compte. Les motifs sont extraits, alignés à l'aide du logiciel MUSCLE pour générer un consensus qui sera ensuite utilisé pour une recherche avec HMMER3 sur la base de données UniProtKB. Les nouveaux motifs sont ensuite ajoutés en se basant sur les mêmes règles. Les cycles continuent jusqu'à ne plus trouver de nouvelles

protéines PPR bactériennes. Concernant la recherche de protéines PPR à l'aide du consensus PPR, la recherche a été réalisée à l'aide de l'alignement des motifs PPR purs d'*Arabidopsis thaliana* publié par Cheng et al. (2016).

En résumé :

- Etude des protéines PPR bactériennes listées par Schallenberg-Rüdinger et al. (2013)
- Vérification sur TPRpred pour les seuils de confiance PPR et TPR
- Extraction des motifs pour les protéines dont le seuil de confiance (PPR-TPR)>80points
- Alignement multiple des motifs bactériens avec MUSCLE
- HMMER3 contre UniProtKB
- Tester toutes les protéines avec une p-value <0.01 sur TPRpred pour les seuils de confiance PPR et TPR
- Ajouter les motifs et recommencer un cycle
- Continuer les cycles jusqu'à ne plus trouver de nouvelles protéines PPR bactériennes

1.2.2. Résultats

En se basant sur le consensus d'*A. thaliana*, on trouve 21 protéines PPR bactériennes, résumées dans le tableau XVII. Les protéines existant dans de nombreuses souches de la même espèce (*Ralstonia solanacearum* par exemple) ne sont représentées qu'une seule fois. Étonnamment, la PPR de *Rhodobacter sphaeroides* n'est pas retrouvée avec ces paramètres alors que toutes les autres proposées par Schallenberg-Rüdinger et al. (2013) le sont.

Tableau XVII : Espèces bactériennes contenant des gènes codants pour des protéines à motifs PPR obtenues par la matrice HMM d'*A. thaliana* et leur phylogénie.

Phylum	Classe	Ordre	Famille	Genre	Espèce
<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Actinomycetales</i>	<i>Actinomycetaceae</i>	<i>Actinomyces</i>	<i>Actinomyces sp.</i>
		<i>Frankiales</i>	<i>Frankiaceae</i>	<i>Frankia</i>	<i>Frankia sp.</i>
<i>Chlamydiae</i>	<i>Chlamydiae</i>	<i>Chlamydiales</i>	<i>Simkaniaceae</i>	<i>Simkania</i>	<i>Simkania negevensis</i>
<i>Chlorobacteria</i>	<i>Dehalococcoidia</i>	<i>Dehalococcoidales</i>	non classé	<i>Dehalococcoidia</i>	<i>Dehalococcoidia bacterium</i>
<i>Chloroflexi</i>	non classé	non classé	non classé	non classé	<i>Chloroflexi bacterium</i>
<i>Cyanobacteria</i>	<i>Cyanophyceae</i>	<i>Synechococcales</i>	<i>Synechococcaceae</i>	<i>Synechococcus</i>	<i>Synechococcus sp.</i>
<i>Gemmatimonadetes</i>	non classé	non classé	non classé	non classé	<i>Gemmatimonadetes bacterium</i>
	non classé	non classé	non classé	non classé	<i>Candidatus Cloacimonetes bacterium</i>
<i>Planctomycetes</i>	<i>Planctomycetia</i>	<i>Planctomycetales</i>	<i>Planctomycetaceae</i>	non classé	<i>Planctomycetaceae bacterium</i>
<i>Proteobacteria</i>	<i>Actinobacteria</i>	<i>Coriobacteriales</i>	non classé	<i>Actinobacteria</i>	<i>Actinobacteria bacterium</i>
	<i>Alphaproteobacteria</i>	non classé	non classé	non classé	<i>Alphaproteobacteria bacterium</i>
	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>	<i>Burkholderiaceae</i>	<i>Ralstonia</i>	<i>Blood disease bacterium</i>
					<i>Ralstonia solanacearum</i>
					<i>Ralstonia syzygii</i>
			<i>Comamonadaceae</i>	<i>Variovorax</i>	<i>Variovorax sp.</i>
	<i>Epsilonproteobacteria</i>	<i>Campylobacteriales</i>	<i>Helicobacteraceae</i>	<i>Helicobacter</i>	<i>Helicobacter pylori</i>
<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Erwiniaceae</i>	<i>Erwinia</i>	<i>Erwinia amylovora</i>	
				<i>Erwinia pyrifoliae</i>	
				<i>Erwinia sp.</i>	
<i>Synergistetes</i>	<i>Synergistia</i>	<i>Synergistales</i>	non classé	non classé	<i>Synergistales bacterium</i>
<i>Thermodesulfobacteria</i>	<i>Thermodesulfobacteria</i>	<i>Thermodesulfobacteriales</i>	<i>Thermodesulfobacteriaceae</i>	<i>Thermodesulfator</i>	<i>Thermodesulfator autotrophicus</i>

En se basant sur les protéines bactériennes à motifs PPR listées par Schallenberg-Rüdinger et al. (2013), on trouve 15 protéines PPR bactériennes, listées dans le tableau XVIII, dont celles utilisées en entrée.

Tableau XVIII : Espèces bactériennes contenant des gènes codants pour des protéines à motifs PPR obtenues par la matrice HMM bactérienne et leur phylogénie.

Phylum	Classe	Ordre	Famille	Genre	Espèce
<i>Bacteroidetes</i>	<i>Sphingobacteria</i>	<i>Sphingobacteriales</i>	<i>Haliscomenobacteraceae</i>	<i>Phaeodactylibacter</i>	<i>Phaeodactylibacter xiamenensis</i>
<i>Chlamydiae</i>	<i>Chlamydiae</i>	<i>Chlamydiales</i>	<i>Simkaniaceae</i>	<i>Simkania</i>	<i>Simkania negevensis</i>
<i>Chlorobacteria</i>	<i>Dehalococcoidia</i>	<i>Dehalococcoidales</i>	<i>non classé</i>	<i>Dehalococcoidia</i>	<i>Dehalococcoidia bacterium</i>
<i>non classé</i>	<i>non classé</i>	<i>non classé</i>	<i>non classé</i>	<i>non classé</i>	<i>Candidatus Cloacimonetes bacterium</i>
<i>Proteobacteria</i>	<i>Actinobacteria</i>	<i>Bacteroidetes</i>	<i>non classé</i>	<i>Bacteroidetes</i>	<i>Bacteroidetes bacterium</i>
		<i>Coriobacteriales</i>	<i>non classé</i>	<i>Actinobacteria</i>	<i>Actinobacteria bacterium</i>
	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	<i>Rhodobacteraceae</i>	<i>Rhodobacter</i>	<i>Rhodobacter sphaeroides</i>
	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>	<i>Burkholderiaceae</i>	<i>Ralstonia</i>	<i>Blood disease bacterium</i>
					<i>Ralstonia solanacearum</i>
			<i>Ralstonia syzygii</i>		
	<i>Comamonadaceae</i>	<i>Variovorax</i>	<i>Variovorax sp.</i>		
	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Erwiniaceae</i>	<i>Erwinia</i>	<i>Erwinia amylovora</i>
					<i>Erwinia pyrifoliae</i>
					<i>Erwinia sp.</i>
<i>Xanthomonadales</i>	<i>non classé</i>	<i>Xanthomonadales</i>	<i>Xanthomonadales bacterium</i>		

Le recoupement des protéines PPR bactériennes obtenues en fonction de la méthode est résumé par la figure 33.

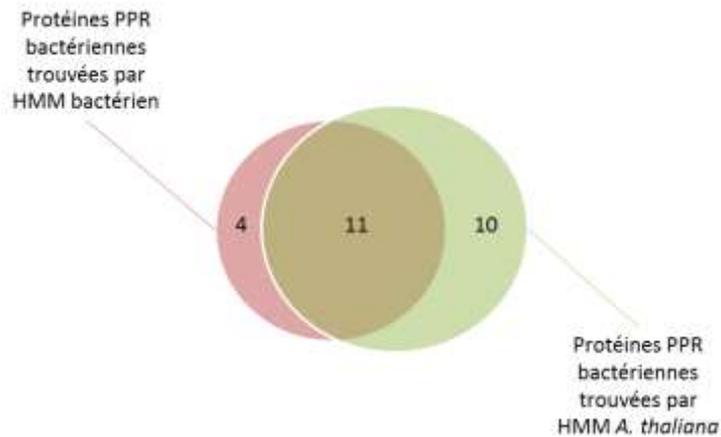


Figure 33 : Diagramme de Venn représentant les protéines PPR bactériennes selon leur méthode de mise en évidence.

Les deux cribles ne trouvant pas les mêmes gènes, les deux consensus ont alors été comparés (Figure 34).

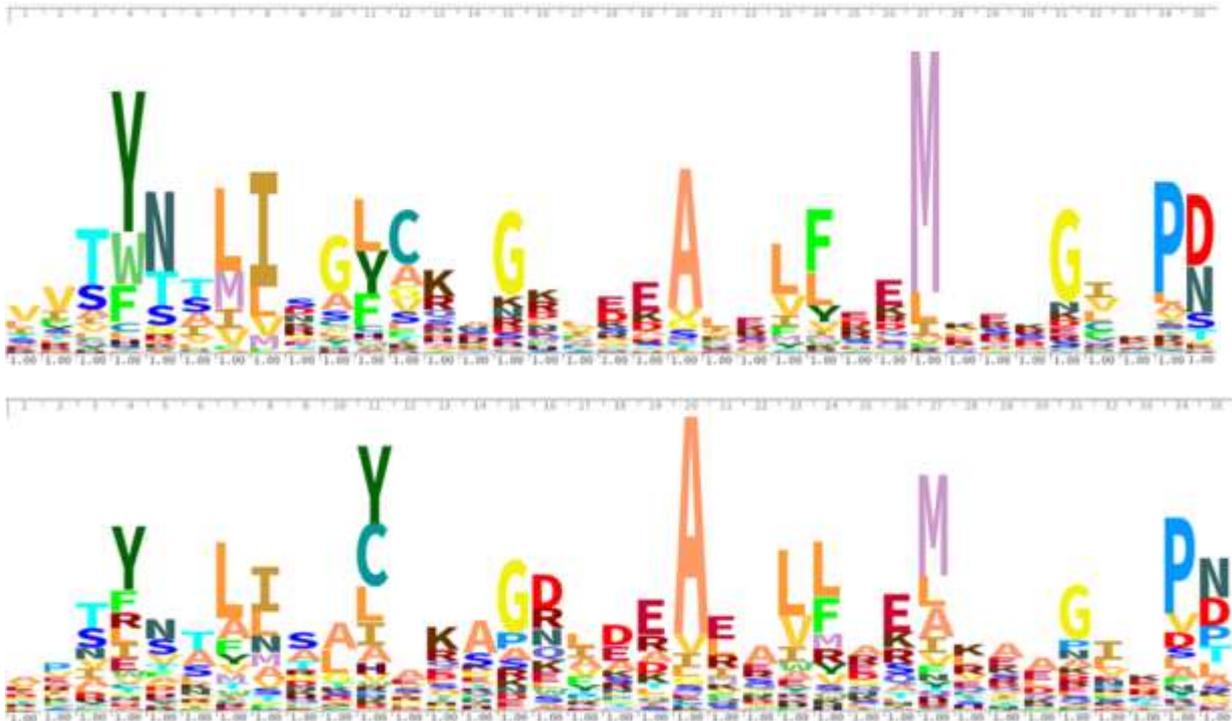


Figure 34 : Consensus du motif pur chez *A. thaliana* basé sur les données de Cheng et al. (2016) (en haut) et sur les motifs trouvés chez les protéines PPR bactériennes quelle que soit leur méthode de mise en évidence (en bas).

On retrouve un consensus proche de celui d'*Arabidopsis*, avec un certain nombre d'acides aminés conservés présents dans les deux consensus (les aa 2, 3, 4, 7, 8, 20 ou 27 par exemple). Fait intéressant, les aa 5 et 35, déterminants de l'ARN à fixer, semblent être plus dégénérés dans le consensus bactérien que dans celui d'*Arabidopsis*. Mais ces aa ne semblent pas être les seuls à être plus dégénérés dans le consensus bactérien que dans le consensus *Arabidopsis*. Par dégénéré ici, outre le fait que de nombreux acides aminés peuvent être présents à chaque position, j'entends que pour le consensus bactérien, il n'y a souvent pas d'aa se démarquant de façon nette des autres, comme on peut souvent le voir dans le consensus d'*Arabidopsis* (en positions 1, 10 ou 12 par exemple).

Parmi les protéines trouvées, la majorité d'entre elles étaient annotées comme étant des protéines PPR, mais pas seulement. Certaines étaient aussi annotées comme étant des protéines TPR, quand d'autres pouvaient être annotées comme TPR sur une base de données et PPR sur une autre. La figure suivante résume les proportions obtenues en fonction des méthodes utilisées.

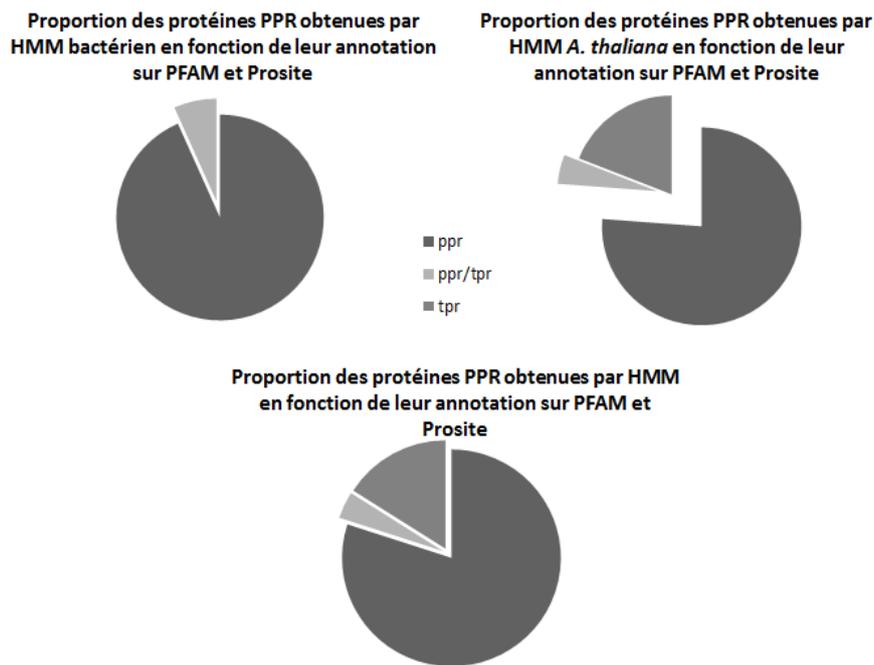


Figure 35 : Proportion de protéines PPR bactériennes trouvées selon leur annotation et en fonction de leur méthode d'obtention.

Première constatation, l'immense majorité des PPR bactériennes mises en évidence sont bien annotées comme étant des protéines PPR. Ensuite, aucune PPR n'est annotée TPR parmi les protéines mises en évidence par le crible HMM bactérien. Cela ne peut pas s'expliquer par la stringence de nos critères de sélection car avec les mêmes critères, le HMM basé sur *A. thaliana* en trouve. Le consensus plante, moins dégénéré, semble donc permettre de récupérer des protéines PPR annotées à tort TPR. De manière générale, l'annotation automatique TPR/PPR est très délicate du fait de la très forte similarité entre ces motifs. En résumé, en faisant la somme des protéines PPR mises en évidence par les deux méthodes, 25 protéines PPR ont été trouvées chez des bactéries.

En observant la phylogénie des différentes bactéries trouvées, on se rend compte qu'elles appartiennent à de nombreux phylum : 10 différents sur les 25. Aucune information ne laisse supposer une apparition des protéines PPR chez les Bactéries de façon unique et tôt dans l'évolution. Tout au plus, on peut supposer des apparitions ci et là chez certaines familles qui semblent très bien représentées, comme les *Erwiniaceae* ou encore les *Burkholderiaceae*. Parmi les 25 bactéries à protéines PPR, on trouve de nombreux pathogènes de plantes, quelques parasites se développant chez les personnes immunodéprimées ou encore une bactérie symbiotique. De manière générale, toutes ces bactéries sont en interaction proche avec un ou plusieurs hôtes Eucaryotes.

1.2.3. Discussion

Le crible met en évidence 25 bactéries possédant au moins une protéine à motifs PPR. L'hypothèse d'une éventuelle cohorte de protéines PPR bactériennes ayant un caractère dégénéré différent de celui des plantes ne s'avère donc pas être vérifiée. Il s'est avéré néanmoins intéressant d'utiliser un HMM bactérien puisqu'il a permis de mettre en évidence 4 protéines qui n'ont pas été vues en utilisant le HMM basé sur *A. thaliana*. Le caractère plus dégénéré du consensus bactérien, provenant probablement d'une plus faible quantité de motifs

le composant peut expliquer pourquoi on trouve moins de protéines qu'avec le consensus d'*Arabidopsis*. En s'intéressant aux annotations existantes sur PFAM et Prosite pour les protéines PPR trouvées, on se rend compte qu'une majorité d'entre elles étaient déjà annotées comme étant des protéines PPR, mais que l'on arrive à enrichir notre liste avec des protéines annotées TPR ou sur lesquelles un doute persistait.

Le seuil arbitraire de confiance PPR – TPR fixé à 80 points est extrêmement stringent. Il permet d'éviter que le consensus bactérien ne dérive trop vers les TPR. En appliquant ce seuil, on élimine le risque de faux positifs, mais on augmente très probablement de manière importante le nombre de faux négatifs. A titre de comparaison, passer d'un seuil de 80 points à 50 points permet de passer de 25 à 36 protéines PPR bactériennes dont 28 avec le HMM bactérien seul. Il est donc très probable que ce seuil nous fasse passer à côté de nombreuses PPR bactériennes, mais il nous assure que les PPR que nous avons sorties en sont bien. L'utilisation du HMM basé sur *A. thaliana* nous permet aussi de retrouver, à l'exception de celle de *Rhodobacter sphaeroides*, les protéines PPR bactériennes proposées par Schallenberg-Rüdinger et al. (2013). Le choix de travailler seulement sur le consensus d'*Arabidopsis* est également discutable. Ce choix a été fait par soucis de simplicité mais un consensus des motifs purs Eucaryotes aurait été probablement plus pertinent. On peut aussi se questionner sur la pertinence du choix du consensus de PPR pures. Même si les premières PPR bactériennes mises en évidence sont des pures, rien ne permet de supposer de l'absence de PPR d'autres sous-familles chez les Bactéries. Il serait donc intéressant de refaire les mêmes recherches avec des consensus des différentes sous-familles. Une dernière limite de notre analyse est l'utilisation de TPRpred. TPRpred est basée sur les motifs PPR connus, principalement chez les plantes. Sa matrice n'est pas adaptée aux motifs purs seulement mais à la plupart des sous-familles de motifs PPR. De plus, si l'on s'en sert d'outil de validation, il est difficile de démontrer qu'il existe de nombreuses protéines PPR bactériennes ayant dérivé du consensus plante si l'on se sert d'un outil construit principalement sur le consensus plante pour valider les protéines que l'on sort. Il est fort probable que l'utilisation de TPRpred explique la proximité des consensus et la surreprésentation des protéines TPR chez les bactéries fait qu'elles représentent un risque de biais assez conséquent dans la recherche de protéines PPR. Il serait cependant possible de modifier la matrice de TPRpred en remplaçant le consensus PPR plante par le consensus PPR bactérien afin de voir si cela permet de trouver plus de protéines PPR bactériennes.

Les protéines PPR n'étant pas aussi répandues chez les Bactéries que chez les Plantes, deux hypothèses peuvent expliquer leur présence : la première serait une évolution de certaines protéines TPR bactériennes, pour donner des protéines PPR, et la seconde un transfert horizontal de gènes, probablement à partir d'Eucaryotes. Le fait que la plupart des bactéries mises en évidence avec des protéines PPR soient à un moment de leur cycle de vie en interaction proche avec un Eucaryote et la diversité phylogénétique des bactéries tend à faire pencher la balance plutôt vers la seconde hypothèse. Afin d'y apporter plus de poids, il serait intéressant de regarder le biais d'usage du code et la proportion de GC dans les gènes codant ces protéines, et de la comparer à la moyenne de tout ou partie du génome de la bactérie suivant la localisation du gène (comparer au chromosome bactérien si le gène se trouve sur celui-ci, ou à un éventuel (méga)plasmide si le gène s'y trouve). Pour les bactéries en interaction directe avec des Eucaryotes, une éventuelle recherche d'orthologues pourrait être intéressante.

1.3. Choix des objets d'étude

Parmi les différentes bactéries mises en évidence, on trouve plusieurs bactéries pathogènes d'animaux du genre *Simkania* ou *Helicobacter* par exemple. Plus intéressant pour nous, plusieurs bactéries sont connues pour être des pathogènes de plantes, du genre *Erwinia* et *Ralstonia*. *Erwinia amylovora* est une bactérie Gram négative responsable du feu bactérien, une maladie touchant certaines *Rosaceae* et causant d'importants dégâts aux cultures de pommiers et poiriers (Piqué et al., 2015). *Ralstonia solanacearum* est une bactérie gram négative responsable d'une pourriture bactérienne. Ses hôtes comprennent de nombreuses *Solanaceae*, tomate

et pomme de terre (*Solanum lycopersicum* et *tuberosum*) pour les plantes d'intérêt agronomique, mais également *Arabidopsis thaliana*, facilitant son étude. Outre le fait que ce sont deux agents pathogènes de plantes, ces bactéries ont également en commun l'utilisation d'un T3SS, comme celui décrit dans l'introduction, leur servant à injecter des effecteurs dans la cellule végétale afin de faciliter l'injection. Pour leur intérêt agronomique, les outils et connaissances existant pour leur étude et les outils dont nous disposons pour étudier leurs hôtes, nous avons fait le choix de nous focaliser sur ces deux bactéries. Toutes nos expérimentations ont été réalisées sur la souche GMI1000 de *Ralstonia solanacearum* et la souche CFBP1430 d'*Erwinia amylovora*.

2. Hypothèse de travail

Nous avons confirmé *in silico* la présence des protéines à motif pentatricopeptide chez différentes bactéries, en particulier *Erwinia amylovora* et *Ralstonia solanacearum* pour celles qui nous intéressent, probablement obtenues par transfert horizontal de gènes. Ces deux bactéries utilisent un système de sécrétion permettant la translocation de protéines bactériennes dans la cellule végétale pour faciliter l'infection. Ensuite, les protéines possédant les motifs PPR sont connues chez les Eucaryotes pour jouer un rôle dans la maturation post-transcriptionnelle des organites à génome. A partir de toutes ces informations, notre hypothèse de travail est que ces protéines PPR jouent un rôle dans la pathogénicité de la bactérie, probablement via sa translocation dans la cellule végétale lors de l'infection, pour aller moduler l'expression génétique des organites. L'hypothèse sous-jacente est que, étant donné que ces agents pathogènes infectent de nombreux hôtes et si ces protéines ciblent les organites, leur cible est potentiellement conservée parmi les plantes hôtes. La figure 36 résume nos hypothèses de travail et les différentes étapes d'analyse qui seront nécessaires pour les tester : validation de l'injection (1), localisation subcellulaire (2) et protéines de la plante interagissant avec les protéines bactériennes (3), rôle potentiel dans les organites sur les différentes étapes de maturation post-transcriptionnelle (4) et conséquences sur la physiologie de la plante et sa résistance aux maladies (5).

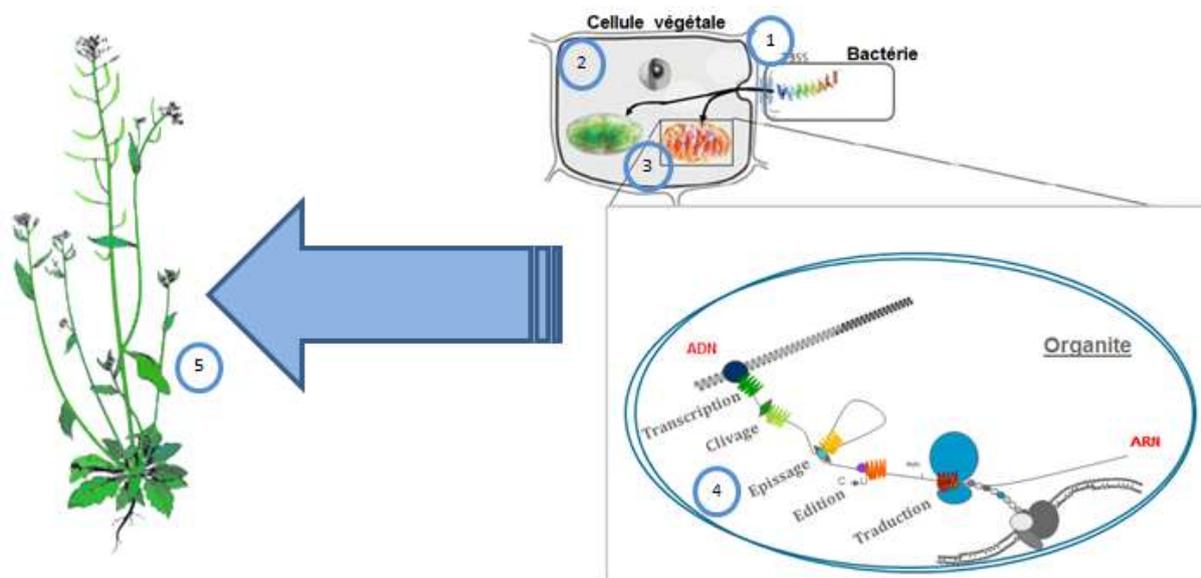


Figure 36 : Représentation schématique des différentes étapes de caractérisation du rôle des PPR bactériennes de *Ralstonia solanacearum* et *Erwinia amylovora*.

Dans un premier temps, je vais présenter les résultats obtenus concernant la protéine PPR de *Ralstonia solanacearum*, puis je présenterai les données obtenues sur la PPR d'*Erwinia amylovora*.

3. Caractérisation de la protéine PPR trouvée chez *Ralstonia solanacearum*

3.1. Présentation du pathogène *Ralstonia solanacearum* et de sa protéine PPR

Ralstonia solanacearum est une bactérie du sol capable d'infecter plus de 200 plantes. L'infection commence au niveau de la racine, continue dans le xylème et infecte la plante via son système vasculaire, jusqu'à empêcher le transport d'eau (Mukaihara et al., 2010). Le transport de l'eau est bloqué par des EPS (pour *Extracellular Polysaccharides*), entraînant le flétrissement des plantes. La bactérie possédant une phase biotrophe et une phase nécrotrophe dans son cycle d'infection est considérée comme hémibiotrophe (Collmer et al., 2009). La souche sur laquelle nous travaillons, GMI1000, possède un chromosome bactérien de 3716kb assorti d'un mégaplasme de 2094kb (Salanoubat et al., 2002). Le gène codant pour la protéine PPR, nommé RipL, est situé sur le mégaplasme à 235kb entre deux gènes codant pour des effecteurs, RipH2 et RipH3 (Peeters et al., 2013). Le gène fait 4122pb et est composé à 69% de G+C, proche des 66.86% de moyenne pour le mégaplasme (Salanoubat et al., 2002). Selon Occhialini et al. (2005), RipL (nommé RSp0193 dans le papier) possède en partie N-terminale un signal d'export vers le système de sécrétion de type III, ainsi qu'une séquence proche des *hrp-box* dans sa région promotrice, suggérant une activation de sa transcription en condition d'infection.

RipL est présente dans sa forme non tronquée dans 42 des 77 souches de *Ralstonia solanacearum* listées par le site Ralsto T3E (Peeters et al., 2013). Il est intéressant de noter que parmi ces 42 souches se trouve au minimum un représentant par phylotype (groupes phylogénétiques reflétant une origine géographique) de *Ralstonia*.

Selon les données de notre collaborateur Nemo Peeters au LIPM à Toulouse, l'annotation de RipL proposée dans les bases de données telles qu'UniprotKB ne proposerait pas le bon codon d'initiation. Il est assez délicat de déterminer le bon codon d'initiation de la traduction chez les effecteurs de Type III, d'autant plus chez *R. solanacearum* qui peut présenter des codons d'initiation alternatifs comme GTG ou TTG. En se basant sur différents homologues de RipL dans différentes souches de *R. solanacearum* et en observant les séquences conservées, un nouveau codon d'initiation a été défini, à 17 aa de celui proposé dans les bases de données, les premiers aa très conservés étant MWRSN. L'ensemble du travail de thèse a été effectué en se basant sur ce codon d'initiation.

RipL possède 5 motifs PPR purs en partie N-terminale selon la méthodologie de Cheng et al. (2016). Aucun autre motif connu n'a pu être mis en évidence sur sa partie C-terminale. Les protéines PPR étant connues pour fixer l'ARN, on peut tenter d'appliquer dessus le code PPR tel que présenté dans le chapitre 2. Il marche plutôt bien pour cette protéine PPR puisque sur les 5 motifs, 4 ont une prédiction de fixation supérieure à 70% (Tableau XIX). Il demeure cependant très difficile de chercher une séquence fixée par une protéine PPR sur la base de 4 nucléotides.

Tableau XIX: Application du code PPR à RipL.

P : motif PPR pur. AA5 et 35 : acides aminés numéros 5 et 35 de chaque motif PPR.

	P	P	P	P	P
AA5	Lys	Ser	Ser	Ser	Ser
AA35	Asn	Asn	Asn	Asn	Asp
A	0.25	0.7	0.7	0.7	0.06
U	0.25	0.13	0.13	0.13	0.1
C	0.25	0.14	0.14	0.14	0
G	0.25	0.04	0.04	0.04	0.85
Séquence fixée putative	N	A	A	A	G

3.2. Importance de RipL dans la pathogénicité de *Ralstonia solanacearum*

Afin d'étudier l'importance de RipL lors de l'infection de *Ralstonia*, différents pathotests ont été réalisés en collaboration avec Nemo Peeters au LIPM (Laboratoire des Interactions Plantes-Microorganismes). Des plantes *Arabidopsis thaliana* Col0 et *Solanum lycopersicum* Marmande ont été infectées avec *Ralstonia solanacearum* GMI1000 sauvage ou mutée pour RipL ($\Delta ripL$). L'expérimentation menée en triplicats biologiques sur *Arabidopsis* n'a pas montré de différence de virulence dans nos conditions et sur cet hôte (Figure 37).

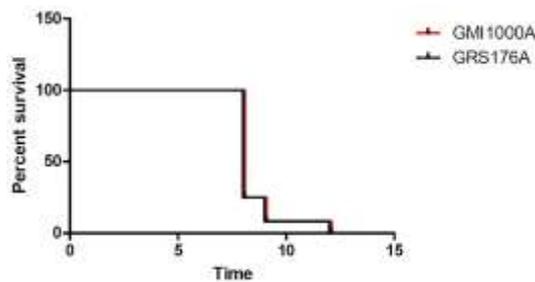


Figure 37 : Pourcentage de survie en fonction du temps d'*Arabidopsis thaliana* infecté par *Ralstonia solanacearum* sauvage (GMI1000A, en rouge) ou $\Delta ripL$ (GRS176A, en noir). Réplicat biologique représentatif.

Il en est de même pour les expérimentations menées sur *Solanum lycopersicum* (Figure 38).

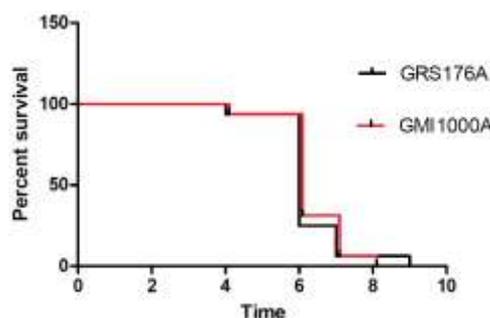


Figure 38 : Pourcentage de survie en fonction du temps de *Solanum lycopersicum* infecté par *Ralstonia solanacearum* sauvage (GMI1000A, en rouge) ou $\Delta ripL$ (GRS176A, en noir). Réplicat biologique représentatif.

Comme nous avons pu le discuter en introduction, il est très courant qu'un simple mutant KO ne présente pas de phénotype en pathotest. Cela ne signifie pas pour autant qu'il ne joue pas un rôle dans la pathogénicité mais alors contre un spectre d'hôtes différent de celui testé. Il est également possible qu'une redondance fonctionnelle existe avec un ou plusieurs autres effecteurs.

3.3. Injection

Malgré de nombreuses données suggérant l'injection de RipL (*hrp-box* dans le promoteur, peptide d'adressage au T3SS en partie N-terminale de la protéine), celle-ci n'a jamais été démontrée pour la souche GMI1000. Elle a cependant été démontrée pour la souche RS1000 par Mukaihara et al. (2010). Etant donné que je n'ai pas réussi à trouver de séquence pour RipL chez RS1000 dans les bases de données, je n'ai pu effectuer d'alignement de séquences et comparer leur identité. Cependant, il est relativement admis dans la littérature que les deux souches ont des répertoires d'effecteurs de Type III très proches (Peeters et al., 2013) ce qui nous permet de supposer que RipL est très probablement transloquée dans la cellule végétale lors de l'infection par GMI1000.

3.4. Localisation subcellulaire

RipL est prédit par Predotar (v. 1.04) pour être envoyé ailleurs dans la cellule (0.98) que dans le chloroplaste (0), la mitochondrie (0.02) ou le réticulum endoplasmique (0). Les prédictions de TargetP voient RipL ailleurs (0.477) plutôt que dans le chloroplaste (0.215), la mitochondrie (0.274) ou vers la sécrétion (0.003).

A partir de cette étape et afin de limiter l'extinction de gène, faciliter les clonages et son expression dans la plante, nous avons synthétisé le gène codant pour RipL en modifiant le biais d'usage de codons afin qu'il corresponde à celui d'*A. thaliana*. Ce gène artificiel est appelé ArtRipL dans la suite du manuscrit.

En réalisant des agroinfiltrations de *N. benthamiana* avec différents construits, nous n'avons pas réussi à avoir de signal. Nous avons également effectué des enrichissements chloroplastiques, mitochondriaux et nucléaires de lignées d'*Arabidopsis thaliana* surexprimant ArtRipL avec une étiquette 3xHA en partie C-terminale, sans pouvoir avoir de signal en *Western Blot* sur extraits totaux et enrichis. A ce jour, la localisation subcellulaire de RipL n'est toujours pas connue. Le tableau XX liste toutes les stratégies tentées sans succès à ce jour.

Tableau XX : Liste des différentes approches testées pour la localisation subcellulaire de RipL.

Méthode	Matériel végétal	Construction			Commentaires
		Promoteur	Insert	Etiquette	
Agroinfiltration	<i>Nicotiana benthamiana</i> sauvage	p35S	ArtRipL ₁₋₁₃₇₃	dsRed	Avec et sans inhibiteur de <i>silencing</i> ; observations à 3, 4 et 5 jours ; microscopie à épifluorescence et confocale ; sur feuille entière ou protoplaste
			ArtRipL ₁₋₁₀₀	dsRed	
			GFP		
	<i>Arabidopsis thaliana</i> sauvage		ArtRipL ₁₋₁₃₇₃	dsRed	Sans inhibiteur de <i>silencing</i> , observation à 5 jours sur feuille entière au microscope confocal
	ArtRipL ₁₋₁₀₀				
Enrichissement sur gradient de Percoll et Western Blot	<i>Arabidopsis thaliana</i> surexpression	pNOS	ArtRipL ₁₋₁₃₇₃	3xHA	Sur extraits totaux ; enrichissement chloroplastes ; mitochondries et noyaux ; thylakoïdes cassés et cytosol

Il est cependant à noter que les pré-séquences fusionnées à la GFP sur feuille entière à 3 jours ont montré un signal dans le noyau et les travées cytosoliques, assez classique pour de la GFP diffusant librement. Un Western Blot contre la GFP avait été effectué afin de déterminer si la taille attendue correspondait à une GFP

seule ou fusionnée à notre pré-séquence mais des problèmes techniques nous ont empêché de conclure et l'expérimentation n'a pas été renouvelée par manque de temps.

3.5. Rôle *in planta*

Afin de déterminer le rôle dans la plante de RipL, nous avons testé différentes approches chez *Arabidopsis thaliana*, qui présente l'avantage d'être une plante hôte de *Ralstonia* sur laquelle existent de très nombreux outils. Afin de savoir si RipL détournait des protéines de plante de leur fonction, nous avons effectué un crible en double-hybride levure contre la banque de protéines d'*Arabidopsis*. Afin de savoir si RipL jouait le rôle d'une protéine PPR dans la plante, nous avons surexprimé ArtRipL de façon stable dans *Arabidopsis* puis nous avons analysé son transcriptome à la recherche de modifications en lien avec l'action potentielle d'une protéine PPR. Enfin, et en lien avec l'hypothèse sous-jacente d'une cible conservée de RipL parmi les différentes espèces végétales, et donc potentiellement importante dans la défense des plantes, nous avons effectués des pathotests sur ces lignées de surexpression.

3.5.1. Double Hybride levure

Le double-hybride levure (ou Y2H pour *Yeast Two-Hybrid*) a été réalisé en suivant la méthodologie détaillée par Guillaumot et al. (2017). De façon simplifiée, on fusionne les protéines que l'on souhaite tester en partie C-terminale avec un domaine de fixation à l'ADN (DB pour *Domain Binding*) ou d'activation de la transcription (AD pour *Activation Domain*). L'interaction entre deux protéines réassemblera un facteur de transcription nécessaire à l'expression d'un gène codant pour un marqueur de sélection. Le criblage s'effectue contre une banque de protéines d'*A. thaliana*. La fusion DB avec ArtRipL est auto-activatrice (c'est-à-dire que sans interaction avec un domaine AD, la croissance était possible) et n'a pu passer dans le crible contre la banque AD. Nous avons donc découpé la protéine en 3 parties égales d'environ 450 acides aminés, puis en 2 deux parts égales d'environ 680 acides aminés, que nous avons fusionnées chacune à un domaine DB. Il s'est avéré que la plupart des versions tronquées étaient aussi auto-activatrices, à l'exception des parties centrale et C-terminale lorsque la protéine était coupée en 3 (Figure 39). Ces deux constructions ont donc été testées en crible contre la banque AD. A cause des nombreux auto-activateurs, nous avons également testé tous nos construits en fusion AD contre la banque DB.

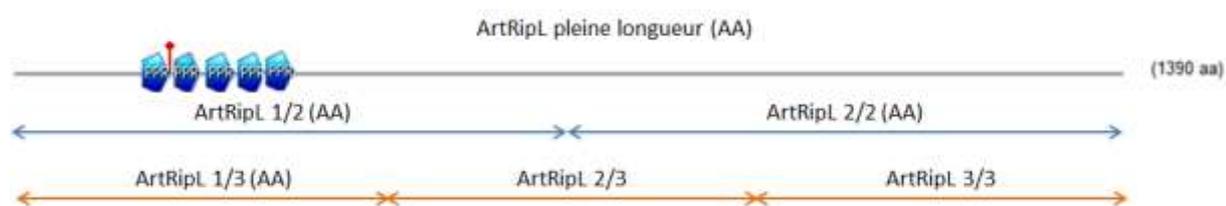


Figure 39 : Représentation schématique des différents construits utilisés en double-hybride levure.
AA : Auto-activateur.

Les différents résultats sont présentés dans le tableau XXI ci-dessous.

Tableau XXI : Liste des interactants obtenus après cribles en double hybride levure en fonction de la banque et du construit testé.

MS/MS : Spectrométrie de Masse en Tandem.

Banque criblée	Construit testé	AGI interactants	Localisation subcellulaire	Description TAIR
Construits AD testés contre la banque DB	ArtRipL 1/2	AT3G19120	noyau	PIF / Ping-Pong family of plant transposase
		AT1G66500	noyau	Pre-mRNA cleavage complex II
	ArtRipL 2/2	∅	∅	∅
		ArtRipL 1/3	∅	∅
	ArtRipL pleine longueur	AT3G19120	noyau	PIF / Ping-Pong family of plant transposase
		AT1G49850	noyau	RING/U-box superfamily protein
		AT2G42260	noyau	Encodes a novel plant-specific protein of unknown function
		AT5G65683	noyau	Zinc finger (C3HC4-type RING finger) family protein
		AT1G75990	cytoplasme, appareil de golgi et membrane plasmique	PAM domain (PCI/PINT associated module) protein
		AT5G42970	cytoplasme	Encodes subunit 4 of COP9 signalosome complex
AT1G72670		cytosquelette	IQ-domain 8	
Construits DB testés contre la banque AD	ArtRipL 2/3	AT5G51440	chloroplaste et mitochondrie	HSP20-like chaperones superfamily protein
	ArtRipL 3/3	AT1G04260	appareil de golgi et membrane plasmique	Encodes protein that interacts with CaMV movement protein
		AT4G25680	appareil de golgi	PPPDE putative thiol peptidase family protein
		AT1G14340	appareil de golgi	RNA-binding (RRM/RBD/RNP motifs) family protein
		AT5G24170	membrane plasmique	Got1/Sft2-like vesicle transport protein family
		AT2G38360	cytoplasme, appareil de golgi, vacuole et membrane plasmique	Prenylated RAB acceptor 1.B4
		AT3G50940	mitochondrie	P-loop containing nucleoside triphosphate hydrolases superfamily protein
		AT4G00270	noyau	GeBP is a non canonical, nuclear localized leucine zipper transcription factor that binds G-elements
		AT1G55190	appareil de golgi et vacuole	PRA1 (Prenylated rab acceptor) family protein
		AT3G13720	membrane plasmique	PRA1 (Prenylated rab acceptor) family protein
		AT4G25660	Réticulum endoplasmique et membrane plasmique	PPPDE putative thiol peptidase family protein
		AT5G16550	Cytoplasme, réticulum endoplasmique et membrane plasmique	voltage-dependent L-type calcium channel subunit
		AT5G59210	noyau	Myosin heavy chain-like protein

Prédit par SUBAcon
Localisé par microscopie à fluorescence
Localisé par MS/MS

De nombreuses protéines d'*Arabidopsis* interagissent avec RipL dans la levure, en particulier avec la partie C-terminale, ne possédant pas de motifs PPR. Chez *Arabidopsis*, ces protéines semblent avoir des localisations subcellulaires assez variées.

Une analyse stringente sur Thalemine avec analyses statistiques et corrections de tests multiples par la méthode de Holm-Bonferroni ne révèle pas d'enrichissement particulier de localisation subcellulaire, fonction moléculaire ou encore processus biologique.

Après cette analyse, il m'a paru intéressant de savoir si ces protéines étaient connues pour interagir avec de nombreuses autres protéines, d'*Arabidopsis* afin de tenir compte du fait qu'elles aient tendance à interagir avec beaucoup de protéines, ou alors qu'elles soient plutôt spécifiques. Le tableau XXII liste le nombre d'interactants trouvés pour nos interactants chez *A. thaliana*. Afin de mieux comprendre ce tableau, il faut savoir que les protéines de la banque sont réparties en deux groupes : le premier groupe correspond à celles qui ont été testées dans un crible géant contre elles-mêmes, donnant lieu à la publication de Dreze et al. (2010). Nous disposons donc de nombreuses informations pour ces protéines. Le second groupe de protéines n'a pas donné lieu à des cribles contre lui-même, mais ces protéines font partie de la banque en interne. Chaque protéine testée, à l'image de RipL, va donc l'être contre des protéines du groupe 1 ou du groupe 2. Il est à noter que les protéines du groupe 1 n'ont pas été criblées contre les protéines du groupe 2. Pour la banque AD, le groupe 1 est composé de 9209 protéines fusionnées au domaine AD, et le groupe 2 du groupe 1

avec 3406 protéines fusionnées au domaine AD en plus. Pour la banque DB, le groupe 1 est composé de 7921 protéines fusionnées au domaine DB, et le groupe 2 est composé du groupe 1 auquel on ajoute 3026 protéines fusionnées au domaine DB. Cette banque est plus petite car les auto-activateurs ne peuvent en faire partie.

Tableau XXII : Nombre d'interactants trouvés chez *A. thaliana* par Dreze et al. (2010) et la plateforme InterATOME pour les interactants d'ArtRipL.

« - » signifie que la protéine fait partie du groupe 2 et n'a pas été testée en crible contre son groupe. Elle a cependant fait partie des cribles InterATOME.

AGI interactants de RipL	Nombre d'interactants trouvés chez <i>A. thaliana</i>	Testé par Dreze et al. (2010) ?
AT1G04260	0	oui
AT1G14340	17	oui
AT1G49850	16	oui
AT1G55190	18	oui
AT1G66500	1	non
AT1G72670	-	non
AT1G75990	3	oui
AT2G38360	6	oui
AT2G42260	10	oui
AT3G13720	35	oui
AT3G19120	7	oui
AT3G50940	-	non
AT4G00270	-	non
AT4G25660	45	oui
AT4G25680	17	oui
AT5G16550	0	oui
AT5G24170	-	non
AT5G42970	0	oui
AT5G51440	2	oui
AT5G59210	17	oui
AT5G65683	28	oui

Si une ou deux protéines possèdent un grand nombre d'interactants, ce n'est pas le cas de l'immense majorité d'entre elles pour les interactions observées. Cela suggère donc que ces interactions sont pour la plupart spécifiques à RipL.

Cependant, ces interactions ne semblent pas être seulement spécifiques à RipL, mais à de nombreux effecteurs (Figure 40).

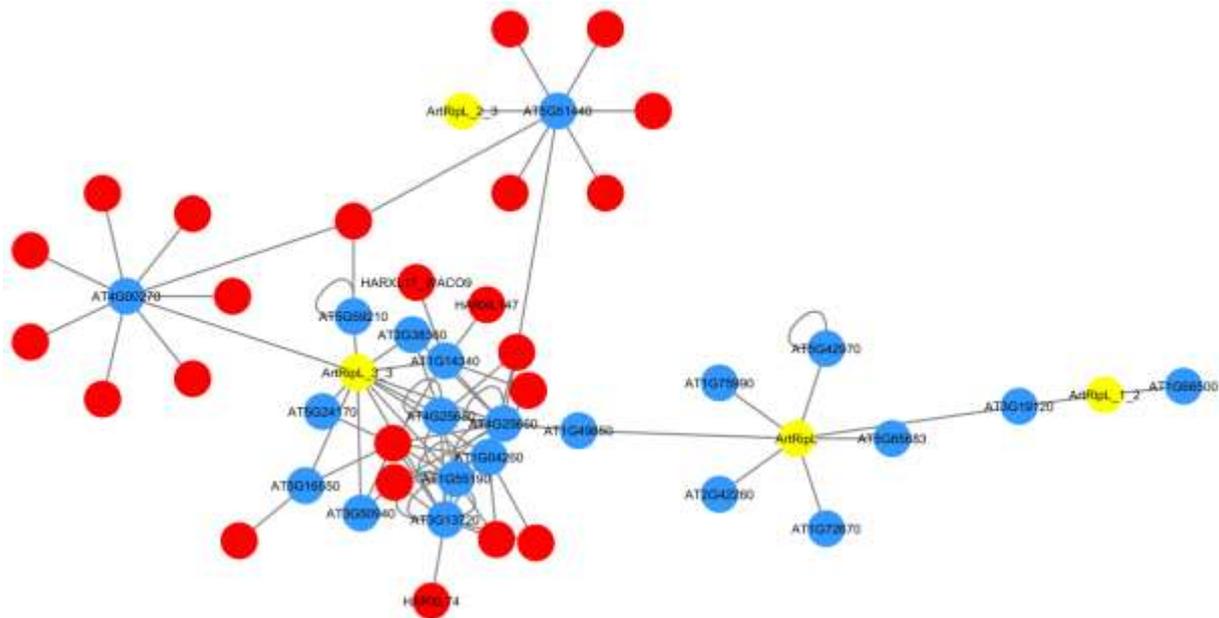


Figure 40 : Interactome d'ArtRipL chez *Arabidopsis thaliana*.

Les traits gris représentent une interaction directe vue en double hybride levure. Les nœuds bleus sont des protéines d'*A. thaliana*, les nœuds jaunes sont des fragments ou la protéine entière de RipL et les nœuds rouges des effecteurs dont l'interaction a été mise en évidence par Mukhtar et al. (2011) ou la plateforme InterATOME.

On remarque ici que parmi les protéines interagissant avec ArtRipL, un grand nombre interagit aussi avec d'autres effecteurs. Ce résultat, extrêmement intéressant, suggère que ces protéines peuvent être des cibles communes de plusieurs effecteurs. Les paragraphes suivants détaillent les candidats les plus ciblés par les effecteurs.

La protéine encodée par AT4G00270 est ciblée par 8 effecteurs dont RipL. Assez peu étudiée, cette protéine nommée GeBP (pour *GLABROUS1 enhancer-binding protein*) serait un facteur de transcription avec un domaine glissière à leucine (Chevalier et al., 2008). Ses cibles ne sont pas encore connues mais elle semblerait être impliquée dans la régulation des cytokinines et partagerait une redondance fonctionnelle avec d'autres protéines d'*A. thaliana*.

La protéine encodée par AT5G51440 est ciblée par 8 effecteurs dont RipL. Nommée Hsp23.5, cette protéine est une sHsp (pour *small Heat-Shock Protein*) localisée dans la mitochondrie. L'expression du gène a été montrée comme étant induite 1h après traitement avec l'harpine, une protéine d'*E. amylovora* (Krause et Durner, 2004). Les auteurs suggèrent que les sHsp jouent un rôle dans la protection de la plante en condition de stress car elles interagiraient avec le cytochrome c et empêcheraient sa fuite dans le cytosol. Leur hypothèse est principalement basée sur les connaissances du déclenchement de l'apoptose chez les animaux (Hoerberichts et Woltering, 2003). Ces informations en font une cible de première importance pour *R. solanacearum*, qui effectue une phase nécrotrophe lors de son cycle d'infection.

En plus d'être ciblées par plusieurs effecteurs, certaines protéines sont prédites ou décrites pour porter des motifs similaires ou appartenir à la même famille protéique. Les paragraphes suivants vont détailler deux exemples trouvés dans l'interactome de RipL.

La protéine encodée par AT1G04260 est ciblée par 7 effecteurs dont RipL. Présente dans le groupe 1 de protéines, elle a été testée par Dreze et al. (2010) contre environ 2700 protéines d'*A. thaliana* sans mettre en évidence d'interaction. Nos résultats couplés à ceux de la plateforme InterATOME et les données publiées suggèrent vraiment un comportement particulier dans l'interaction entre cette protéine et les effecteurs. La première description de cette protéine fut proposée par Huang et al. (2001). Les auteurs mirent en évidence une interaction en double hybride levure entre la protéine encodée par AT1G04260, qu'ils nommèrent MPI7 (pour *Movement-Protein Interacting 7*) avec la protéine de mouvement MP du virus de la mosaïque du chou-fleur. Le gène codant pour MPI7 est exprimé dans tous les organes de la plante, et ce de manière uniforme. La fusion de la protéine avec la GFP a montré que la protéine fusionnée localisait dans le cytoplasme avec des enrichissements faisant penser à une localisation dans l'appareil de Golgi (Huang et al., 2001). Si Huang et al. (2001) avaient déjà remarqué une forte identité de séquence entre MPI7 et des protéines de mammifères, Alvim Kamei et al. (2008) expliquent que MPI7 appartient à la famille des protéines à motif PRA1, pour *Prenylated Rab Acceptor 1*. Il en existe 19 chez *A. thaliana* et leur rôle supposé serait d'être impliqués dans la régulation du transport de vésicules entre les différents compartiments cellulaires (Alvim Kamei et al., 2008). Trois autres protéines appartenant à la même famille ont été mises en évidence comme interagissant avec RipL. Ces protéines, encodées par AT3G13720, AT1G55190 et AT2G38360, interagissent respectivement avec 5, 4 et 1 effecteurs, dont RipL.

Les protéines encodées par AT4G25660 et AT4G25680 interagissent respectivement avec 4 et 2 effecteurs dont RipL. Annotées comme des protéines PPPDE (*Putative Thiol Peptidase Family Protein*), aucune caractérisation fonctionnelle n'a été proposée chez *Arabidopsis*. Chez la souris, la protéine PPPDE1 est une protéine de l'appareil de Golgi impliquée dans la carcinogénèse (He et al., 2013).

Enfin, d'autres protéines semblent être des cibles de choix, compte tenu de leur rôle connu ou putatif dans la plante.

La protéine encodée par AT1G14340 interagit avec RipL, mais également avec deux autres effecteurs. Bien qu'étant annotée comme une protéine de fixation à l'ARN, ce qui demeure très intéressant en interaction avec la protéine PPR et effecteur RipL, aucune caractérisation fonctionnelle de la protéine n'existe encore.

La protéine encodée par AT2G42260 interagit également avec RipL. Elle a été décrite comme étant nécessaire à la mitose, mais pas seulement. La caractérisation de son mutant KO a montré une plus forte tolérance aux ultraviolets B comparé à une plante sauvage (Hase et al., 2006). Il est cependant difficile de conclure à partir de ces données sur les éventuelles conséquences de la fixation de cette protéine par un effecteur lors de l'infection.

La protéine encodée par AT3G19120, nommée AtRAD54, interagit aussi avec RipL. Cette protéine a été décrite chez *A. thaliana* comme étant impliquée dans la réparation de l'ADN (Klutstein et al., 2008). Sa surexpression limite les dommages de l'ADN par les rayons γ , en faisant une cible d'intérêt lors de l'infection.

Après l'étude détaillée de l'interactome de RipL dans la plante hôte *A. thaliana*, nous nous intéressons mécanismes moléculaires de RipL dans la plante. La présence de motifs PPR nous pousse à nous intéresser à sa transcription, d'abord dans les organites à la recherche d'un effet éventuellement direct, puis dans le noyau.

3.5.2. Analyses transcriptomiques

Afin de mettre en évidence un éventuel défaut lié au rôle d'une PPR dans le chloroplaste et la mitochondrie d'*Arabidopsis* en présence de RipL, nous avons surexprimé celle-ci dans la plante avec un promoteur pNOS et une étiquette 3xHA en partie C-terminale. Comme nous avons pu le voir en introduction, les protéines PPR peuvent être impliquées dans l'édition, le clivage, la transcription, la traduction ou encore l'épissage. A partir

de données de séquençage sur ARN totaux et à l'aide du script d'analyse présenté dans le chapitre 1, nous pouvons avoir un accès –parfois partiel selon la couverture de séquençage- aux données d'accumulation de transcrits, de couverture, d'édition et d'épissage. Cependant, il ne faut pas oublier que le rôle premier d'une protéine PPR est de fixer l'ARN. Afin de déterminer le ou les ARN fixés par RipL, nous avons effectué un séquençage des petits ARN à la recherche d'empreintes PPR, c'est-à-dire de fragments d'ARN protégés par les protéines PPR de la dégradation par des nucléases. Nous avons travaillé sur des plantes d'*Arabidopsis* de 4 semaines en jours courts, et les prélèvements ont été effectués sur feuille. Les paragraphes suivants présentent les résultats obtenus après analyse des ARN totaux et des petits ARN, sur les génomes des organites dans un premier temps, puis sur le génome nucléaire.

3.5.2.1. Analyse du transcriptome des organites

Suite au séquençage des ARN totaux, aucun défaut n'a été mis en évidence concernant l'édition dans le chloroplaste et la mitochondrie de la lignée de surexpression. Il en est de même pour les défauts de couverture. Aucun gène chloroplastique ou mitochondrial n'a non plus été vu comme différentiellement exprimé dans nos conditions. Pas de défaut d'épissage non plus pour le chloroplaste. Concernant la mitochondrie, une légère augmentation de l'épissage de l'intron 2 de *nad7* a été observée (Tableau XXIII).

Tableau XXIII: Epissage de l'intron 2 de *nad7*.

% : pourcentage d'épissage pour un réplikat biologique (Rep) donné.

	Col0			ArtRipL			Bonferroni
	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3	
nad7 intron2	1%	4%	2%	4%	5%	6%	0.036312205

Bien que statistiquement significative, la différence entre la lignée de surexpression et Col0 ne nous a pas parue biologiquement pertinente suite à la lecture du pourcentage d'épissage.

Un séquençage des petits ARN a également été effectué sur la lignée de surexpression et sur Col0. En fixant l'ARN, les protéines PPR peuvent protéger celui-ci de la dégradation par les exonucléases. L'objectif de l'observation des petits ARN est de potentiellement mettre en évidence une ou des empreintes PPR, présentes dans la lignée de surexpression et non dans Col0 afin de déterminer éventuellement la zone de fixation de RipL. Cependant, si toutes les empreintes PPR sont présentes dans les petits ARN, tous les petits ARN, produits de dégradation, ne sont des empreintes PPR. La difficulté, dans l'analyse de ces données, est donc de faire le tri entre les produits de dégradation de l'ARN en lien avec les protéines PPR, et ceux qui ne le sont pas. Après application de la méthodologie de segmentation présentée précédemment, on obtient les données suivantes (Annexes S11 et S12).

L'analyse par segmentation des petits ARN de RipL met en évidence de nombreux segments à la hausse. Ces segments correspondent parfois à des séquences intergéniques, mais pour l'immense majorité à des séquences codant pour des ARNt chez la mitochondrie. On trouve également de nombreux segments différentiellement représentés en 5' et 3' UTR de gènes. Toutes ces variations sont imputables à RipL, puisqu'il s'agit de la seule différence avec notre condition contrôle. Cependant, il demeure difficile de déterminer si ces variations sont provoquées par une action directe de fixation de RipL sur l'ARN, ou par une conséquence secondaire de l'action de RipL entraînant des modifications dans l'expression génique chloroplastique et mitochondriale.

Pour déterminer cela, nous pouvons nous appuyer sur les données de *RNAseq* sur ARN totaux et chercher des défauts en lien avec une éventuelle accumulation de petits ARN à proximité. Cependant les ARNt ne font pas partie de l'analyse d'ARN totaux puisqu'ils sont trop petits pour entrer dans la composition des banques de séquençage. Il est donc difficile de conclure directement à partir de nos données de segmentation, mais cela donne néanmoins des pistes tangibles à suivre concernant une éventuelle zone de fixation et/ou une conséquence moléculaire de l'action de RipL. Une approche par Northern Blot ou par RT-qPCR de l'accumulation des ARNt et des ARNr mis en évidence par l'analyse serait pertinente.

Pour les segments trouvés correspondant à autre chose que des ARNt, rien dans nos données sur ARN totaux ne suggère de défaut, en particulier à proximité des gènes pour lesquels nous avons vu des variations de couverture précédemment. Cela ne signifie pas pour autant que la PPR ne s'y fixe pas, et il est plausible que des défauts en termes de traduction existent. L'idéal serait, dans un premier temps de regarder les séquences des différents segments et de les comparer avec les prédictions de fixation de RipL, puis dans un second temps, de faire du séquençage d'ARN en cours de traduction (Riboseq) pour éventuellement détecter des défauts de traduction.

3.5.2.2. Analyse du transcriptome nucléaire

Concernant l'analyse d'expression différentielle, elle met en évidence 179 gènes différentiellement exprimés, dont ArtRipL après correction de p-values par la méthode de Benjamini-Hochberg (Annexe S13). En effectuant une analyse sur Thalemine, après tests statistiques et corrections des tests multiples, on retrouve des enrichissements entre autres pour les réponses aux stimuli (3^E-14), aux composés dérivés de l'oxygène ($3,4^E-13$), aux stress (4^E-11), les réponses de défense de la plante (1^E-9), aux stimuli biotiques ($1,6^E-9$), à la chitine ($2,6^E-6$), aux bactéries (2^E-4), aux fungi (3^E-4), au stress oxydatif ($5,6^E-4$) ou encore les réponses immunitaires ($1,6^E-5$). Aucun enrichissement concernant la localisation subcellulaire n'a été observé.

L'analyse MapMan du métabolisme ne montre pas de défauts majeurs. Après avoir effectué une correction de p-values avec la méthode de Benjamini-Yekutieli, le test de la somme des rangs de Wilcoxon suggère des différences par rapport à la condition contrôle au niveau des centres de réaction à la lumière de la chaîne photosynthétique ($<1^E-20$), du cycle de Calvin ($4,52^E-3$), du métabolisme secondaire des flavonoïdes (0,0208) ou encore de la photorespiration ($3,31^E-3$). En effectuant une analyse MapMan de la même façon, mais ciblée sur le chloroplaste, on trouve après correction des variations liées au photosystème II ($1,02^E-8$) et au photosystème I ($1,64^E-4$). Rien ne ressort après correction de l'analyse ciblée sur la mitochondrie. Concernant les stress, on retrouve après analyse un effet sur la signalisation ($3,88^E-5$) mais aussi les stress abiotiques ($2,02^E-2$).

Suite à ces informations, en particulier sur la photosynthèse, nous avons effectué des dosages de chlorophylles, mais ils n'ont pas mis en évidence de différence significative entre nos lignées de surexpression et notre condition contrôle (Annexe S14). Le rendement du photosystème II a également été estimé par PAM fluorescence (Annexe S15), ne mettant aucune différence significative en évidence. Il est cependant difficile de conclure à partir des données de PAM fluorescence puisque suite à un problème technique, une seule des deux répétitions biologiques s'est avérée utilisable et la seconde n'a pu être refaite par manque de temps.

3.5.2.3. Limites liées à la surexpression d'un gène bactérien *in planta*

Comme expliqué précédemment, afin de limiter l'extinction de l'expression de RipL et en particulier l'expression transcriptionnelle, nous avons fait synthétiser un gène avec un biais d'usage de codon proche de celui des plantes. Cependant, bien que l'expression de l'ARNm demeure quantifiable par RT-qPCR, la détection sur extrait total de plante d'ArtRipL avec une étiquette 3xHA s'est avérée non concluante par

Western Blot anti-HA. Afin de vérifier si ArtRipL était la cible d'ARN interférents, nous avons extrait de nos données de séquençage des petits ARN les *reads* anti-sens s'alignant sur ArtRipL, de 21 et de 24 nucléotides.

En faisant la moyenne des trois réplicats biologiques, 1106 *reads* de 21nt s'alignent en anti-sens d'ArtRipL, et 218 de 24nt. A titre de comparaison, nous nous sommes intéressés à AT3G18110, une PPR pure d'*A. thaliana* de taille et de niveau d'expression proches de ceux d'ArtRipL. Sur les mêmes échantillons, on ne trouve aucun petit ARN anti-sens de 21 et 24 nt associé à ce gène. Il existe donc une forte extinction post-transcriptionnelle pour ArtRipL dans la plante.

3.5.3. Pathotests

En lien avec notre hypothèse suggérant que la cible de RipL pouvait être très conservée chez les plantes, nous avons testé deux événements de transformation indépendants de notre lignée de surexpression dans *Arabidopsis thaliana* en pathotest face à différents pathogènes.

Nous n'avons pas vu dans nos conditions de sensibilité accrue de notre lignée de surexpression par rapport à du Col0 à *Ralstonia solanacearum*, *Hyaloperonospora arabidopsidis* et *Pseudomonas syringae* sauvages (Figures 41 à 43).

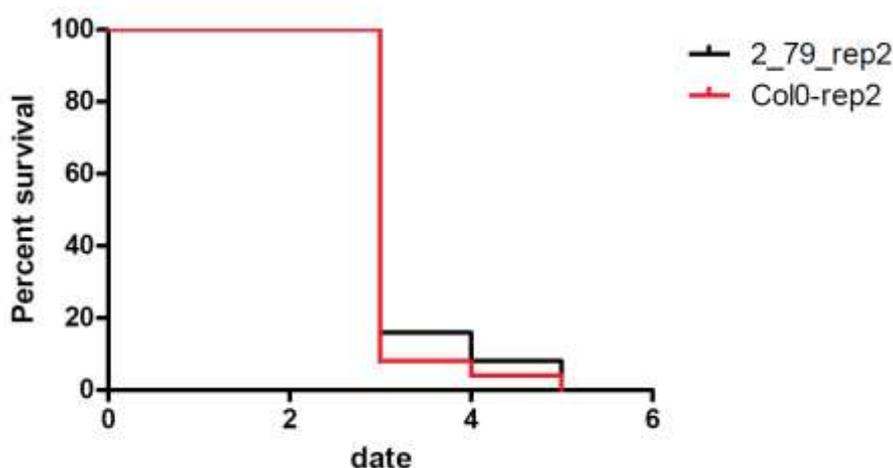


Figure 41 : Pourcentage de survie en fonction du temps d'*Arabidopsis thaliana* Col0 (en rouge) ou surexprimant ArtRipL, infecté par *Ralstonia solanacearum* sauvage.

Réplicat biologique représentatif parmi les 3 effectués.

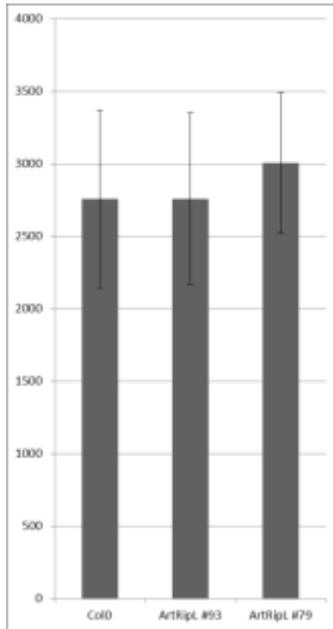


Figure 42 : Comptage de spores de *Hyaloperonospora arabidopsidis* par mg de matière fraîche végétale 7 jours après infection chez *A. thaliana*.
 Réplicat biologique représentatif parmi les 3 effectués.

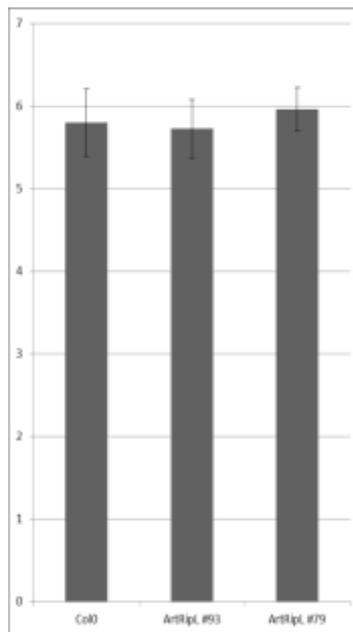


Figure 43 : Comptage de colonies de *Pseudomonas syringae* par mg de matière fraîche végétale, représentés en échelle Log10, 48h après infection.
 Réplicat biologique représentatif parmi les 3 effectués.

3.6. Discussion

Les données, connues dans la littérature ou obtenues lors de la thèse concernant RipL sont résumées sur la figure 44.

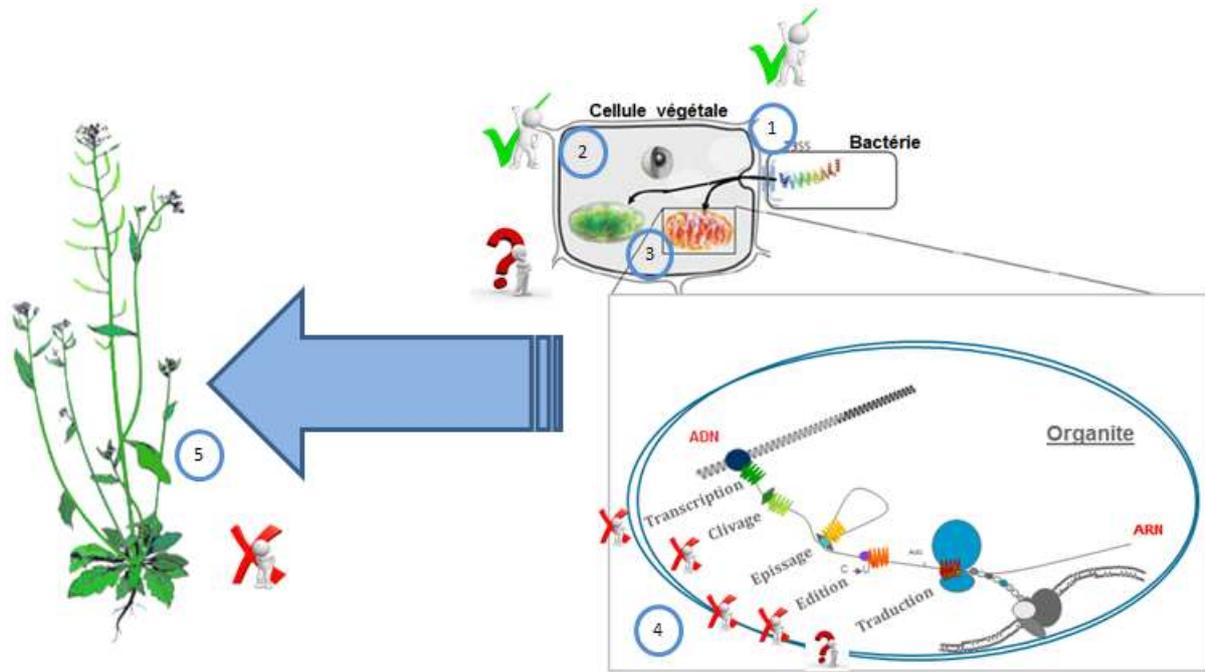


Figure 44 : Résumé des résultats obtenus lors de la thèse sur la caractérisation de RipL.

Les croix vertes font références aux expérimentations ayant mis en évidence un mécanisme concernant RipL, les points d'interrogation aux expérimentations non effectuées ou non concluante et les croix rouges à celles n'ayant pas mis en évidence de défaut ou de mécanisme. Validation de l'injection (1), localisation subcellulaire (2), protéines de la plante interagissant avec les protéines bactériennes (3), rôle potentiel dans les organites sur les différentes étapes de maturation post-transcriptionnelle (4) et conséquences sur la physiologie de la plante et sa résistance aux maladies (5).

Pour commencer, le mutant bactérien *ΔripL* ne présente pas de différence en condition d'infection avec la souche sauvage, chez *A. thaliana* et *S. lycopersicum*. Plusieurs explications sont possibles. Il est probable que l'effet de l'effecteur seul ne soit pas suffisamment fort pour entrainer un phénotype fort. Cela ne signifie pas pour autant qu'il ne joue pas de rôle lors de l'infection mais complexifie grandement son étude. Une autre hypothèse serait qu'il existe chez *Ralstonia solanacearum* un ou plusieurs autres effecteurs jouant le même rôle lors de l'infection. Bien qu'aucun autre effecteur chez *Ralstonia* ne possède de motifs PPR, l'hypothèse demeure néanmoins plausible si un ou plusieurs autres effecteurs, sans forcément interagir avec un ARN, ciblent la même protéine ou la même voie chez la plante, comme le suggèrent nos données d'interactomique. Etant donné que *R. solanacearum* possède un large spectre d'hôtes, il est possible que nous ne soyons pas dans la bonne condition ou encore le bon hôte afin de voir un phénotype marqué pour le mutant *ΔripL*. Il est aussi possible que RipL ne soit pas impliqué dans la pathogénicité mais dans ce cas de figure, aucun rôle ne facilitant pas la croissance et le développement du pathogène dans la plante ne me paraît plausible pour une protéine injectée dans la cellule végétale pendant l'infection. Enfin, on peut aussi faire l'hypothèse que RipL ne joue aucun rôle dans la cellule végétale, mais cela paraît étonnant en terme de coût énergétique de maintenir un effecteur inutile pour le pathogène, qui plus est dans plusieurs souches différentes, même si la sélection ne se fait pas gène par gène mais au niveau d'un organisme entier dans son environnement.

Concernant l'injection, elle a déjà été démontrée pour la souche RS1000. Malgré tous nos efforts, aucune localisation subcellulaire n'a pu être obtenue, en raison d'extinction de gènes. Une approche basée sur les travaux de Park et al. (2017) chez *P. syringae* avec des effecteurs fusionnés au feuillet β 11 de la GFP,

infectant des lignées *A. thaliana* de surexpression de la GFP₁₋₁₀ dans différents compartiments subcellulaires pourrait être une solution envisageable. En outre, cette expérimentation permettrait en plus de valider l'injection de RipL par GMI1000 dans la plante infectée.

L'analyse de l'interactome de RipL dans *A. thaliana* a révélé de nombreux résultats, très intéressants. De nombreux interactants semblent être des cibles putatives d'effecteurs très pertinentes. Un certain nombre d'interactants de RipL semblent être localisés dans l'appareil de Golgi et/ou jouer un rôle dans la sécrétion. Je n'ai cependant pas trouvé d'information dans la littérature mettant spécifiquement en lien des défauts de l'appareil de Golgi en condition d'infection par *R. solanacearum*, alors que c'est le cas concernant *Pseudomonas syringae* (Block et Alfano, 2011). Parmi les nombreuses protéines décrites dans l'interactome de RipL, une protéine en particulier, encodée par AT5G51440, a attiré mon attention. Tout d'abord, cette petite Hsp est localisée dans la mitochondrie. RipL étant une protéine PPR et bien que sa localisation subcellulaire *in planta* ne soit pas connue, il est hautement probable que sa localisation après injection soit chloroplastique et/ou mitochondriale. L'hypothèse suggérant que RipL interagirait avec la protéine encodée par ce gène afin de promouvoir l'infection en facilitant la mort cellulaire est cependant délicate si l'on s'appuie sur les données de pathotests. En effet, si cela avait été le cas, les lignées de surexpression de RipL dans *Arabidopsis* auraient probablement montré des phénotypes de résistance ou de sensibilité aux pathogènes testés, en particulier *Pseudomonas syringae* dont l'harpine a été mise en évidence comme responsable du passage du cytochrome C de la mitochondrie vers le cytoplasme, entraînant la PCD (Krause and Durner, 2004). Cependant, en s'appuyant sur les données d'interatomique, le fait que cette protéine interagisse avec de nombreux autres effecteurs de *R. solanacearum* pourrait rappeler l'éventuelle redondance fonctionnelle expliquant pourquoi le mutant KO bactérien n'a pas montré de différence avec la souche sauvage dans nos conditions. Outre cette piste, de nombreuses autres semblent réellement prometteuses pour caractériser pleinement RipL et idéalement, les étapes suivantes consisteraient à confirmer ces interactions à l'aide d'une seconde méthode indépendante.

De manière générale concernant l'interatomique : la partie 3/3 d'ArtRipL interagit avec de nombreuses protéines, ce qui n'est pas le cas de la partie 2/2, qui contient pourtant l'intégralité de la partie 3/3. Une explication pourrait être un problème d'expression ArtRipL 2/2. Il faudrait confirmer la bonne expression de la protéine par Western Blot. Il est aussi possible que le repliement de la protéine ne soit pas le même dans les deux cas de figure, modifiant ainsi les interactions possibles. Il me semble néanmoins pertinent d'avoir travaillé avec des protéines tronquées dans ce projet puisque même en pleine longueur, la protéine va être fusionnée à un domaine AD ou DB qui risque de modifier sa structure, donc potentiellement ses interactions. Le caractère exogène de l'expression peut aussi modifier les maturations post-traductionnelles. En résumé, même sur une protéine en pleine longueur, nous n'avons pas l'assurance de retrouver la conformation exacte de la protéine *in planta* (que nous ne connaissons pas forcément de toute façon) dans la levure. Le double hybride levure reste donc une approche relativement globale, qu'il est nécessaire de confirmer par d'autres analyses, pas nécessairement de l'interatomique, pour mettre en évidence la notion de fonctionnalité dans l'interaction.

Concernant les approches en *RNAseq*, il serait intéressant parmi les nombreuses empreintes mises en évidence de voir si les segments extraits présentent un même motif, et de recouper cette information avec les prédictions données par le code PPR. Si aucun défaut d'édition, d'épissage et assez peu de variation de couverture n'ont été trouvés, la dernière étape est de tester si la traduction est différente entre la lignée de surexpression et Col0. De nombreuses variations ont été trouvées lors de l'analyse des petits ARN, en particulier sur les ARNt. Une analyse par Northern Blot de l'accumulation de ces derniers permettrait de détecter d'éventuels défauts d'accumulation. Une analyse de traduction (*ribosomal profiling*, spectrométrie de masse couplée à un enrichissement chloroplastique et mitochondrial, etc.) serait nécessaire pour compléter l'analyse. Concernant les résultats obtenus pour le séquençage des petits ARN, et en comparaison avec

d'autres jeux de données construits et analysés de manière analogue, la quantité de segments différentiels est très importante, suggérant qu'il se passe quelque chose, mais qui nous échappe encore dans les organites de la lignée de surexpression de RipL.

D'un point de vue purement technique, il est possible aussi que le niveau d'expression du gène codant pour RipL demeure trop faible pour que nous puissions trouver un phénotype associé à son expression. Bien que l'expression du gène soit quantifiable en RT-qPCR et lors du séquençage des ARN totaux, le séquençage des petits ARN montre qu'il est la cible de très nombreux ARN interférents. De plus, son expression n'a jamais pu être mise en évidence par Western Blot anti-HA sur extraits totaux. L'insertion d'un intron dans la séquence artificielle, l'utilisation d'un promoteur inductible sont autant de pistes pertinentes qui permettraient de limiter l'extinction de l'expression.

On peut aussi questionner le bien fondé d'analyser des échantillons de feuille pour un pathogène principalement racinaire, tout du moins sur les phases précoces d'infection. Il aurait été intéressant de faire les mêmes analyses sur des échantillons de racine.

Pour conclure, l'hypothèse la plus plausible reste que notre effecteur joue un rôle dans une condition particulière et/ou sur un hôte spécifique. Néanmoins, une caractérisation de ce type demeure extrêmement couteuse en termes de temps et de personne. En parallèle de la caractérisation de RipL, la thèse s'est aussi portée sur la caractérisation de la PPR d'*Erwinia amylovora*. Le paragraphe suivant va détailler et discuter les données concernant cette seconde protéine PPR bactérienne.

4. Caractérisation de la protéine PPR trouvée chez *Erwinia amylovora*

4.1. Description de l'agent pathogène *Erwinia amylovora* et de sa protéine PPR

Erwinia amylovora est une bactérie gram négative de la famille des *Enterobacteriaceae* (Norelli et al., 2003). La bactérie est responsable du feu bactérien, une maladie nécrotique pouvant affecter la sous-famille des *Spiraeoideae* (dont font partie le pommier et le poirier pour les espèces d'intérêt agronomique) mais aussi les plantes du genre *Rubus* (ronces, framboisiers, mûriers, etc.) suivant la souche (Mann et al., 2013). Pénétrant dans la plante via les ouvertures naturelles ou les blessures, *E. amylovora* induit le flétrissement et la nécrose des tissus (Paulin, 1996). Sa pathogénicité dépend de plusieurs facteurs. On peut citer en premier lieu la production d'exopolysaccharides, l'amylovoran et le levan, qui permettent à la bactérie de s'attacher aux différentes surfaces, de survivre dans des environnements difficiles et qui bloquent les faisceaux cribro-vasculaires de la plante pendant l'infection (Paulin, 1996; Koczan et al., 2011). Second facteur extrêmement important dans la pathogénicité d'*E. amylovora*, son système de sécrétion de type III (T3SS). A l'image de celui de *R. solanacearum*, il va permettre la translocation d'effecteurs bactériens (T3E) dans la cellule végétale ou la sécrétion dans l'apoplaste (Hogenhout et al., 2009; Puhar et Sansonetti, 2014). Le génome d'*E. amylovora* CFBP1430 est composé d'un plasmide, pEa29 de 28kb ainsi que d'un chromosome bactérien d'environ 3806kb. C'est sur ce dernier que se situe EAMY_0820, le gène codant pour la protéine PPR d'*Erwinia amylovora*. Il est à noter que ce gène de 1.3kb n'est pas situé sur un îlot de pathogénicité comme la plupart des effecteurs. Cela ne signifie pas pour autant que EAMY_0820 n'encode pas pour un effecteur puisque AvrRpt2_{Ea} a été décrit également comme étant hors d'un îlot de pathogénicité (Zhao et al., 2006).

EAMY_0820 est un gène codant pour une protéine putative, prédite comme possédant 9 motifs PPR purs selon la méthodologie de Cheng et al. (2016). Comme les protéines PPR sont prédites pour fixer l'ARN, nous avons, de la même façon que pour RipL, tenté d'appliquer le code PPR, mais sans succès (Tableau XXIV).

Tableau XXIV : Application du code PPR à la protéine encodée par EAMY_0820.

P : motif PPR pur. AA5 et 35 : acides aminés numéros 5 et 35 de chaque motif PPR.

	P	P	P	P	P	P	P	P	P
AA5	Gly	Thr	Lys	Asp	His	Asp	Asn	Gly	Gln
AA35	His	Asp	Asp	Lys	Asp	Asp	Asn	Asp	Leu
A	0.25	0.04	0.25	0.25	0.25	0.25	0.08	0.25	0.25
U	0.25	0.03	0.25	0.25	0.25	0.25	0.16	0.25	0.25
C	0.25	0	0.25	0.25	0.25	0.25	0.29	0.25	0.25
G	0.25	0.93	0.25	0.25	0.25	0.25	0.41	0.25	0.25
Séquence fixée putative	N	G	N	N	N	N	G	N	N

A la différence de RipL cependant, la protéine codée par EAMY_0820 n'est que prédite, suite au séquençage assez récent du génome d'*E. amylovora* CFBP1430 (Smits et al., 2010). Nous détectons cependant son ARNm par RT-qPCR et il a également été vu par *RNAseq* (Kamber et al., 2016). Cette protéine putative est de plus prédite dans de nombreuses souches d'*E. amylovora*. Comme décrit dans la littérature (Mann et al., 2013; Smits et al., 2014), certaines souches sont connues pour affecter les *Spiraeoideae* quand d'autres n'affectent que les plantes du genre *Rubus*. En utilisant BlastP avec la séquence de la protéine encodée par EAMY_0820, nous observons qu'il existe des homologues dans toutes les souches infectant les *Spiraeoideae* (dont fait partie CFBP1430), mais pas dans les souches infectant le genre *Rubus* (à l'exception de la souche Ea644 dans laquelle un homologue dont la séquence est tronquée a été trouvé). Le résultat est représenté par la figure 45.

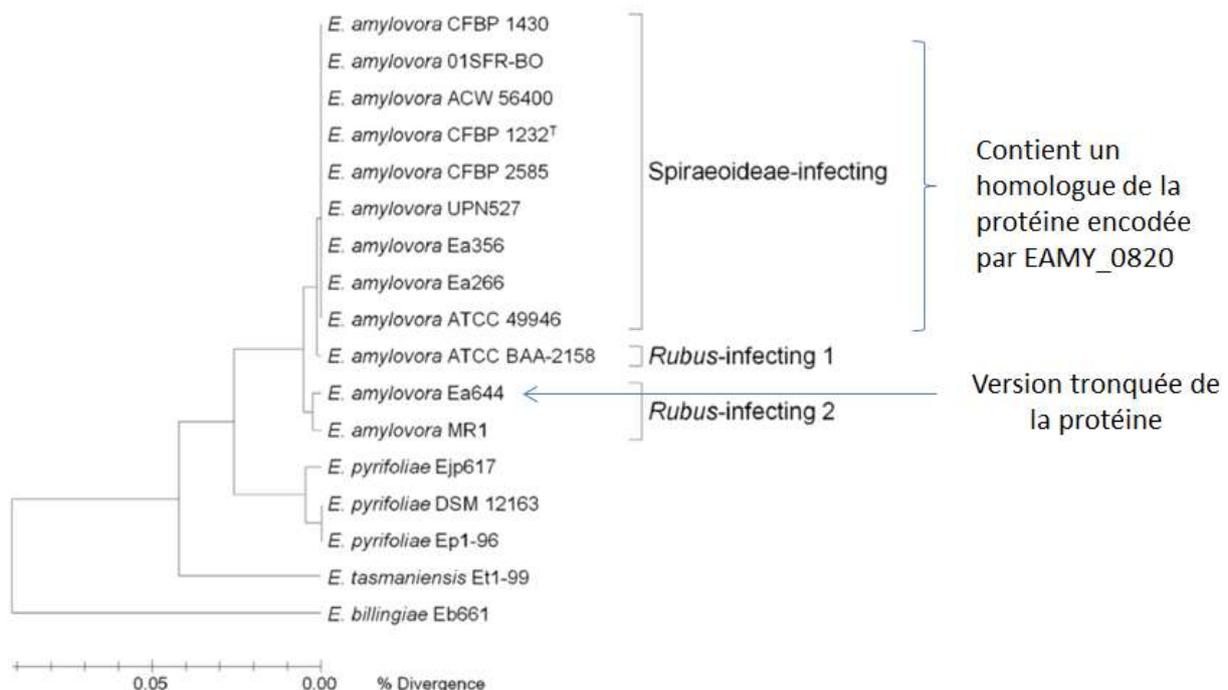


Figure 45 : Arbre phylogénétique de différentes souches et espèces d'*Erwinia* et mise en évidence des souches d'*Erwinia amylovora* contenant une protéine homologue de celle encodée par EAMY_0820 dans CFBP1430.

Adapté de Mann et al. (2013).

4.2. Injection de la protéine codée par EAMY_0820 lors de l'infection

Selon les connaissances actuelles, le sécrétome d'*E. amylovora* est composé de 12 effecteurs (Nissinen et al., 2007), loin des 72 effecteurs de *Ralstonia solanacearum* (Mukaihara et al., 2010) ou des 40 effecteurs de *P. syringae* pv. *Tomato* (Cunnac et al., 2009). En se basant sur ce constat, Zhao et al. (2006) font l'hypothèse que de nombreux effecteurs restent encore à mettre en évidence chez *E. amylovora*. En utilisant EffectiveDB (Eichinger et al., 2016), un système de prédiction d'effecteurs de Type III, la protéine encodée par EAMY_0820 est prédite comme un T3E avec un seuil de confiance élevé.

En se basant sur les informations de Oh et al. (2010), je me suis rendu compte que la composition en acides aminés de la partie N-terminale de la protéine encodée par EAMY_0820 suivait la majorité des descriptions de séquence des effecteurs de Type III.

Afin de se donner une idée précise des critères remplis, sont listés ici les 50 premiers acides aminés de la PPR d'*Erwinia* :

Met-Pro-Gln-Ser-Ala-Ile-Phe-Ser-Ser-Ser-Lys-Ile-Asn-Pro-Ile-Tyr-Ile-Asn-Asn-Thr-Ser-Gly-Ser-Thr-Ala-Thr-Ala-Tyr-Asp-Gly-Asn-Val-Ile-Thr-Arg-Met-Lys-Glu-Val-Thr-Asn-Thr-Gly-Glu-Asn-Leu-Ala-Ala-Gln-Arg-...

Les 6 règles proposées par Oh et al. (2010) sont résumées dans le tableau XXV :

Tableau XXV : Règles proposées par Oh et al. (2010) et application à la PPR d'*E. amylovora*.

Règles proposées par Oh et al. (2010)	S'applique à la PPR codée par EAMY_0820 ?
les acides aminés Ile, Leu, Val, Ala ou Pro sont trouvés en positions 3 ou 4 mais pas les deux, et sont souvent précédés par une Pro, ou un acide aminé polaire ou basique	On trouve bien de la Pro en position 2, mais pas d'Ile, Leu, Val, Ala ou Pro en positions 3 et 4
La position 5 est rarement occupée par une Met, Ile, Leu, Phe, Tyr ou Trp	Oui
Les acides aminés Asp ou Glu n'apparaissent pas dans les 12 premières positions	Oui
Les Cys sont rarement présentes après la position 5	Oui
Les 50 premiers acides aminés sont amphipathiques et riches en résidus polaires, en particulier en Ser et Gln	On trouve 6 Ser dans les 50 premiers acides aminés, mais seulement 2 Gln. L'enrichissement est net pour la Ser, mais pas pour la Gln. Pour les 50 premiers acides aminés, la proportion d'acides aminés polaires + Cys est de 58%, mais ils représentent 55% des acides aminés au total. L'enrichissement n'est donc pas net
Pas plus de 3 résidus consécutifs de Met, Ile, Leu, Val, Phe, Tyr ou Trp ne sont présents dans les 50 premiers résidus	Oui

En résumé, la plupart des règles empiriques basées sur les observations de Oh et al. (2010) sont respectées. Elles sont cependant à nuancer puisque provenant de l'étude de *Pseudomonas syringae* et non d'*E. amylovora*.

En dehors de ces informations, nous n'avons pas de preuve expérimentale que cette PPR est bien transloquée lors de l'infection. Une fusion de la PPR dans la bactérie avec une étiquette adénylate cyclase calmoduline dépendante (Mukaihara et al., 2010) a été réalisée conjointement avec notre collaboratrice Marie-Anne Barny. Cette enzyme permet de produire de l'AMP cyclique, facilement dosable et dont la concentration est faible dans la cellule végétale, à partir d'ATP. Cette enzyme n'est active qu'en présence de calmoduline, protéine présente dans la cellule végétale et absente dans la bactérie. L'expérimentation effectuée sur pommier n'a pas permis de conclure à une injection alors que les contrôles positifs et négatifs ont bien fonctionné.

Sur la fin de la thèse, j'ai effectué des fusions de la PPR d'*Erwinia* avec une étiquette HA-GFP₁₁ (Park et al., 2017) puis transformé des souches virulentes de *Pseudomonas syringae* DC3000 et *Erwinia amylovora* CFBP1430. Différentes lignées d'*A. thaliana* surexprimant la GFP₁₋₁₀ dans différents compartiments subcellulaires ont été infectées et la fluorescence a été observée à 3hpi. L'expérimentation n'a pas permis de mettre en évidence la translocation, même pour HrpW, un effecteur d'*E. amylovora* connu pour être transloqué, qui était notre contrôle positif (Gaudriault et al., 1998). Par manque de temps, l'expérimentation n'a pu être renouvelée. En se basant sur la méthodologie de la mise en évidence de HrpW, j'ai également tenté à partir de cultures liquides de *Pseudomonas* et sur boîtes avec milieu inducteur du T3SS de mettre en évidence la sécrétion de la protéine encodée par EAMY_0820 dans le milieu extérieur. Le Western Blot anti-HA n'a pas permis de mettre en évidence de sécrétion dans le milieu extérieur, et nous n'avons pas obtenu de signal non plus pour les culots bactériens, suggérant des limites techniques. Aucun contrôle positif n'était utilisé pour *Pseudomonas* et par manque de temps, l'expérimentation n'a pas encore été faite avec *Erwinia amylovora*.

Etant donné qu'il est impossible de démontrer une absence, nous ne pouvons conclure quoi que ce soit de ces résultats.

4.3. Importance d'EAMY_0820 dans la pathogénicité d'*E. amylovora*

4.3.1. Sur plante hôte

Suite à nos données préliminaires *in silico*, nos collaborateurs d'Angers ont effectué des pathotests sur *Malus domestica* « Golden delicious », plante hôte d'*Erwinia amylovora*. Elles ont été infectées avec la souche sauvage CFBP1430 et avec un mutant d'insertion d'une cassette de résistance à la kanamycine dans EAMY_0820, que l'on appellera CFBP7962. A la différence de ce que l'on a pu faire pour les pathotests *Pst* sur les lignées de surexpression de RipL par exemple, ce n'est pas un comptage de bactéries qui est effectué mais l'observation et la quantification de conséquences phénotypiques sur la plante. Les infections sont faites à 3 concentrations bactériennes différentes et les observations à 7 et 14 jours après infection. Les paramètres observés sont les suivants : le démarrage de l'infection ou non, progression de la bactérie (de la feuille à la tige), la longueur de tige nécrosée rapportée à la longueur totale de la tige. A la différence de tous les pathotests réalisés jusqu'ici avec 3 répétitions biologiques, chacun à un jour d'intervalle, le pathotest ici est effectué sur 3 blocs de 10 plantes par condition. Les figures 46 et 47 résument l'évolution des symptômes et les résultats associés à 7 et 14 jours suivant les souches et les différentes concentrations.

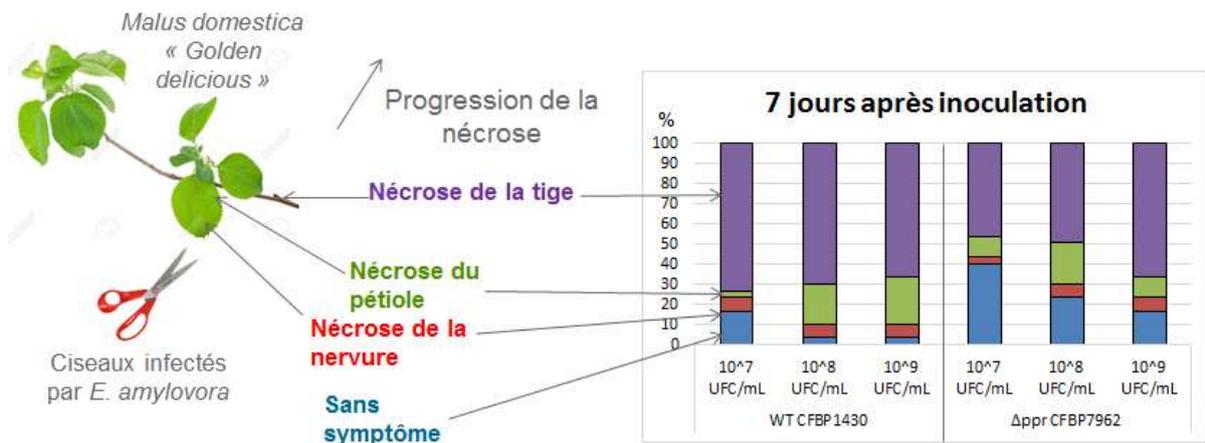


Figure 46 : Evolution des symptômes de nécrose sur le pathosytème *Erwinia amylovora*-*Malus domestica* « Golden delicious » et symptômes à 7 jours après inoculation.

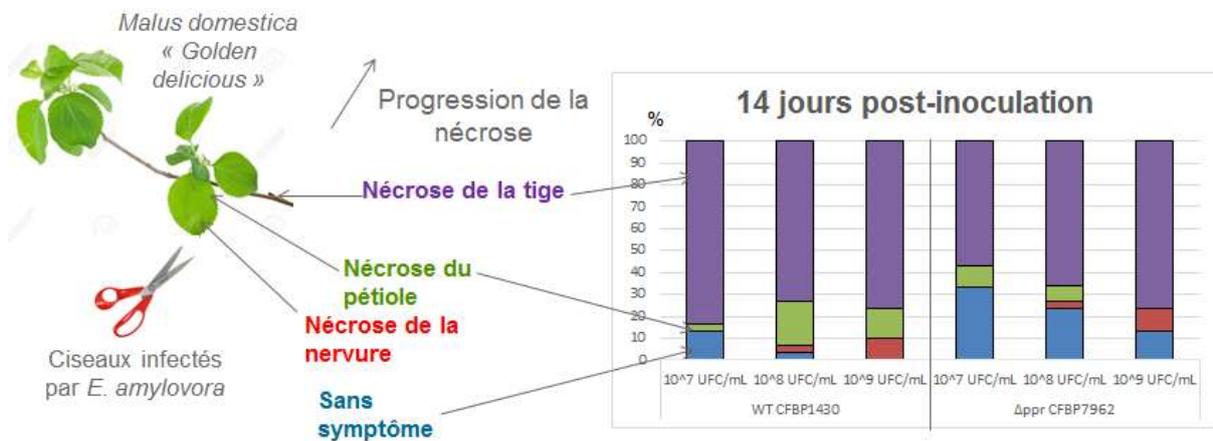


Figure 47 : Evolution des symptômes de nécrose sur le pathosytème *Erwinia amylovora*-*Malus domestica* « Golden delicious » et symptômes à 14 jours après inoculation.

Comme attendu, l'évolution des symptômes d'infection semble positivement corrélée à la concentration bactérienne initiale d'infection. Dès 7 jours d'infection, la souche mutante CFBP7962 semble présenter un certain retard dans la mise en place de l'infection et a une proportion de plantes sans symptôme nettement plus importante que lors de l'infection par la souche sauvage. C'est d'autant plus visible à 10⁷ UFC/mL mais la tendance semble s'atténuer lorsque l'on monte en concentration.

La figure 48 montre le taux d'incidence de la maladie en fonction de la concentration. On peut définir le taux d'incidence d'une maladie comme étant le nombre de nouveaux cas d'infection rapportés à la population totale sur une période donnée. Attention cependant, sont considérées comme malades les plantes pour lesquelles la bactérie a au moins atteint la tige.

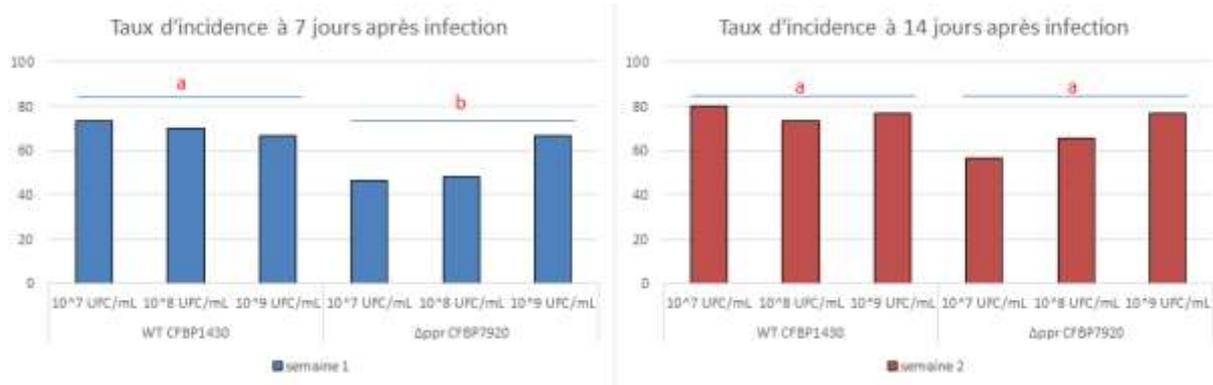


Figure 48 : Taux d'incidence de la maladie induite par *E. amylovora* sur *Malus domestica* « Golden delicious » (en pourcentage) en fonction de la concentration et de la souche à 7 et 14 jours après infection. Analyse statistique basée sur un modèle linéaire généralisé sans prise en compte de l'effet concentration et d'une ANOVA. A 7 jours, $p=0,02108$. A 14 jours, $p=0,06801$). Les analyses statistiques faites à 7 et 14 jours sont indépendantes, 30 plants par condition.

A 7 jours, le taux d'incidence toutes concentrations confondues est statistiquement différent entre la souche bactérienne sauvage et mutante. On remarque que pour les concentrations les plus faibles, le taux d'incidence est nettement inférieur pour la souche mutante comparée à la souche sauvage, et de la même façon que pour les symptômes, l'effet est beaucoup plus marqué à faible concentration. On peut noter qu'il y a plus de 20 points d'écart entre le taux d'incidence de la souche sauvage et la souche mutante à 10^7 UFC/mL à 14 jours pour l'exemple le plus marqué.

La figure 49 représente la sévérité de l'infection. Celle-ci correspond à la proportion de la tige qui est atteinte par la nécrose (en pourcentage) à 7 et 14 jours après infection.

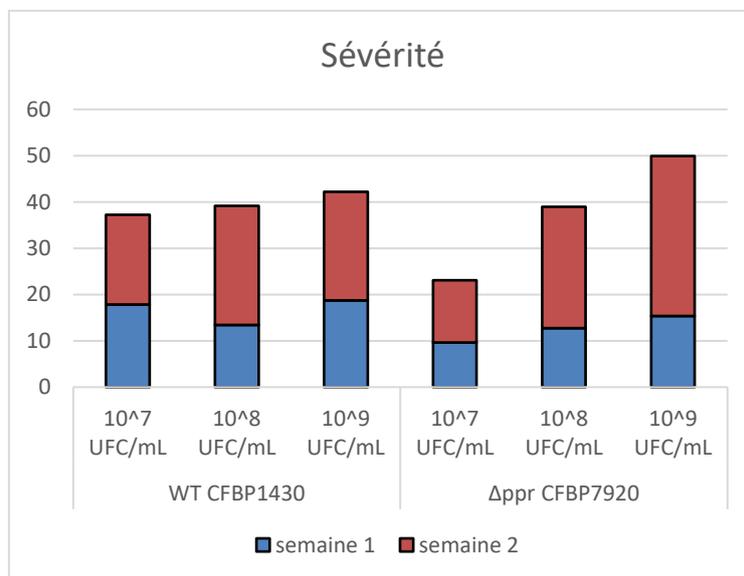


Figure 49 : Sévérité de l'infection de *Malus domestica* « Golden delicious » par *E. amylovora* (en pourcentage) en fonction de la concentration et de la souche à 7 et 14 jours après infection.

L'infection semble encore une fois moins sévère pour la souche mutante, et d'autant plus à faible concentration, que pour la souche sauvage. Ce résultat est néanmoins à mettre en perspective avec les résultats précédents, en particulier le taux d'incidence puisque les plantes ne présentant aucun symptôme tirent

énormément la moyenne de la sévérité de l'infection vers le bas. La figure suivante présente également la sévérité de l'infection, mais seulement pour les plantes dont la nécrose a atteint la tige. L'intérêt de ce graphique est de savoir si l'effet observé sur la sévérité est cohérent biologiquement parlant ou plutôt mathématique et statistique à cause d'un effet fort sur le taux d'incidence.

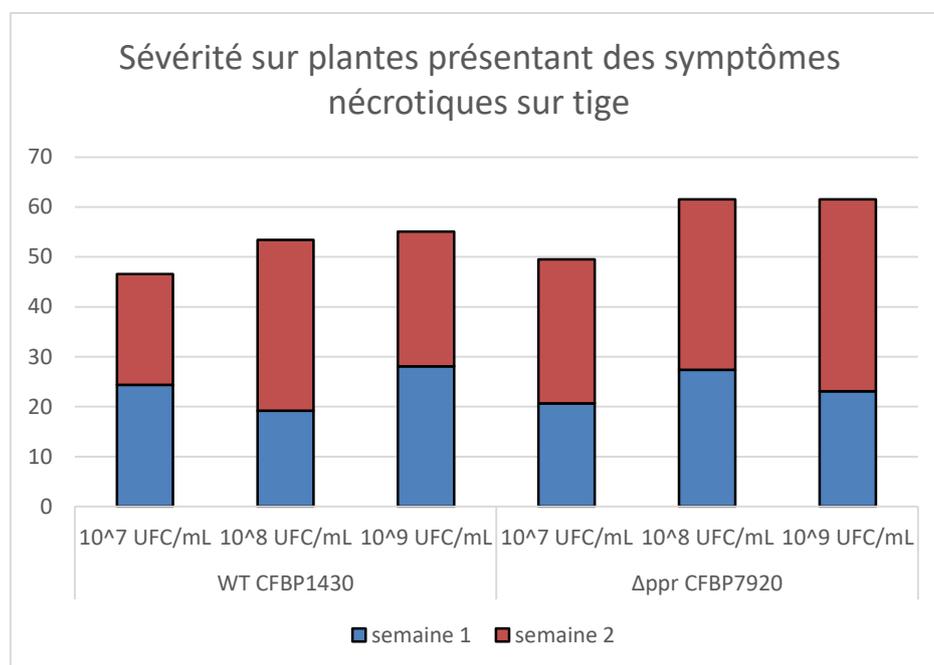


Figure 50 : Sévérité de l'infection de *Malus domestica* « Golden delicious » sur les plantes dont la nécrose a atteint la tige par *E. amylovora* (en pourcentage) en fonction de la concentration et de la souche à 7 et 14 jours après infection.

En ne tenant compte que des plantes pour lesquelles l'infection a atteint la tige, les sévérités d'infection sont nettement plus proches entre CFBP1430 et CFBP7962 pour les 3 concentrations.

Au vu de ces différents résultats, on voit que le gène EAMY_0820 est impliqué dans la pathogénicité d'*E. amylovora* contre *Malus domestica* « Golden delicious ». Après observation des taux d'incidence et de sévérité, la protéine semblerait plutôt jouer un rôle dans l'initiation de l'infection et la mise en place de la maladie plutôt que dans les phases tardives. Il est aussi possible de relever que son absence est d'autant plus handicapante pour la bactérie lorsque celle-ci est en concentration plus faible au moment de l'infection.

4.3.2. Sur plante non-hôte

Suite à la faible disponibilité d'outils et aux difficultés liés à la manipulation du pommier, le choix a été fait de travailler pour la majeure partie de la thèse sur l'interaction non-hôte entre *Erwinia amylovora* et *Arabidopsis thaliana*. Bien que n'étant pas une plante cultivée, *A. thaliana* dispose de très nombreux outils qui nous permettront de chercher le plus finement possible un éventuel mécanisme moléculaire à la protéine encodée par EAMY_0820. L'interaction *A. thaliana*-*E. amylovora* est même régulièrement utilisée dans la littérature (El-Maarouf et al., 2001; Rebutier et al., 2007; Degraeve et al., 2008; Moreau et al., 2012; Launay et al., 2016; Aksoy et al., 2017; Hamdoun et al., 2018). De façon analogue à RipL, nous avons principalement travaillé sur des lignées de surexpression de la PPR bactérienne chez *Arabidopsis*, et également avec un gène synthétique afin de limiter l'extinction de l'expression. Toutes les lignées de surexpression de la PPR d'*E. amylovora* sont construites à partir du gène de synthèse, qui sera appelé par la suite ArtEa.

La première expérimentation réalisée a été l'infection d'*A. thaliana* sauvage et des lignées de surexpression par *Erwinia amylovora* CFBP1430 et la souche mutante pour la PPR, CFBP7962.

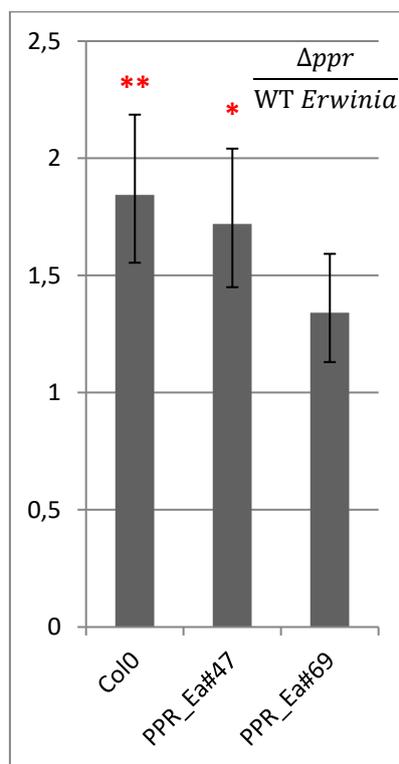


Figure 51 : Ratio du titrage bactérien in planta d'*E. amylovora* mutante par rapport à la souche sauvage à 24hpi en fonction de différents fonds génétiques d'*A. thaliana*.

Deux étoiles : p-value < 0.01 ; une étoile : p-value < 0.05. Analyse basée sur 3 répétitions biologiques.

Plusieurs informations sont nécessaires pour appréhender cette figure. Tout d'abord, j'ai représenté le ratio des titres bactériens de la souche mutante par rapport à la souche sauvage. Une valeur supérieure à 1 suggère donc que la croissance de la souche mutante est supérieure à celle de la souche sauvage dans le fond génétique végétal en question. Ensuite, l'écart-type représenté correspond à celui du ratio après correction des données. Pour déterminer si les différences observées étaient statistiquement significatives ou non, Guillem Rigail a utilisé un modèle linéaire généralisé, ce qui lui a permis de séparer les différents effets observés (réplicats techniques, dilutions lors des comptages de colonies, répétitions biologiques, etc.) pour ne calculer que celui qui nous intéresse, à savoir l'effet génotype bactérien. Il a ensuite effectué une ANOVA.

Sur Col0 tout d'abord, on observe que la croissance de la souche mutante est environ 1.8 fois plus marquée que celle de la souche sauvage, et que la différence entre les deux souches bactériennes dans ce fond génétique végétal est très significative. On retrouve un résultat allant dans le même sens pour la lignée de surexpression #47 de la PPR bactérienne dans la plante, avec une différence significative. Cette différence n'est cependant pas retrouvée dans la lignée de surexpression #69 pour laquelle aucune différence statistiquement significative ne fut mise en évidence entre les deux souches bactériennes.

4.3.3. En culture liquide

Afin de vérifier si les différents effets observés en plante hôte et en plante non-hôte sont effectivement liés à une interaction avec une plante ou s'ils proviennent d'un éventuel défaut de croissance de la bactérie, j'ai

effectué des cinétiques de croissance en culture liquide, comparatives entre *E. amylovora* sauvage et la souche mutante pour EAMY_0820. La figure 52 représente les résultats obtenus en milieu LB, un milieu riche.

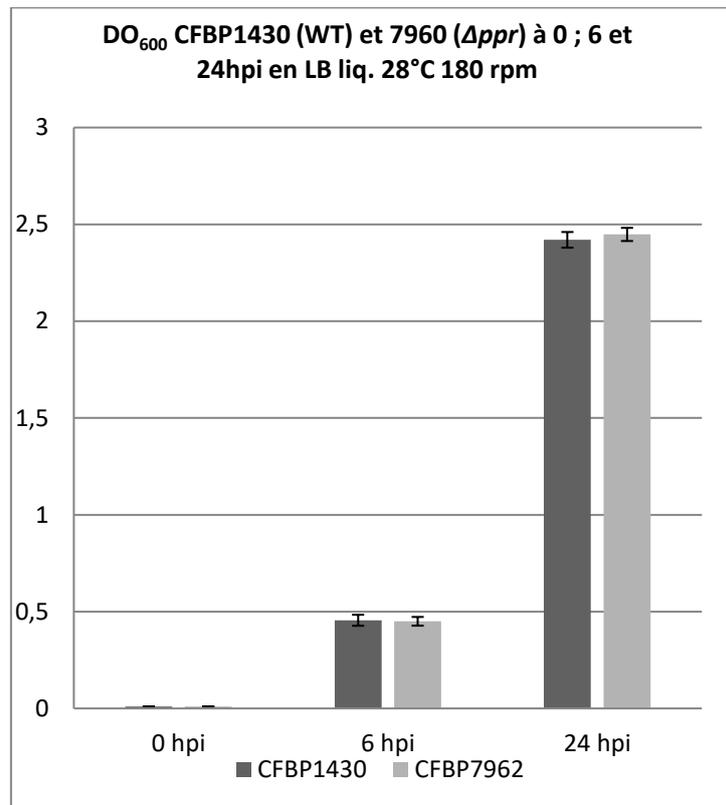


Figure 52 : Comparaison de la croissance bactérienne en milieu liquide riche pour CFBP1430 et CFBP7962 à différents temps.

Les écart-types représentent les triplicats techniques par condition. Répétition biologique représentative parmi les 2 effectuées.

Aucune différence dans les cinétiques de croissance laissant supposer qu'EAMY_0820 jouerait un rôle dans la croissance en milieu riche d'*E. amylovora* en dehors de l'infection n'a été mise en évidence.

Afin de pousser l'analyse un peu plus loin, nous avons commencé à effectuer des cinétiques de croissance en milieu minimum, se rapprochant un peu plus des conditions que la bactérie va trouver dans l'apoplaste de la plante. Par manque de temps, une seule répétition biologique a été réalisée pour le moment (Figure 53).

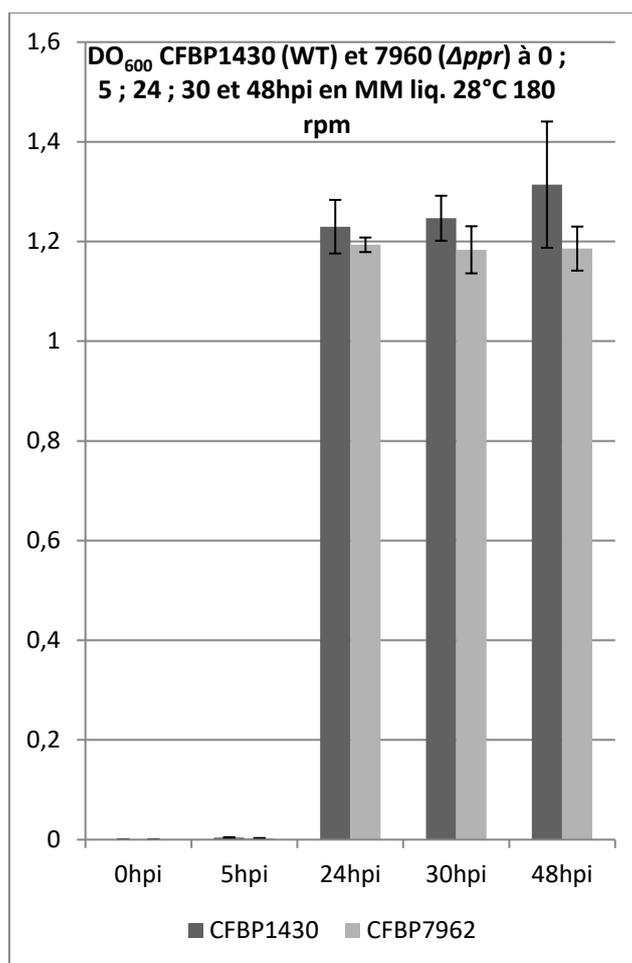


Figure 53 : Comparaison de la croissance bactérienne en milieu liquide minimum pour CFBP1430 et CFBP7962 à différents temps.

Les écart-types représentent les triplicats techniques par condition. Une seule répétition biologique.

Etant donné que nous ne disposons que d'un seul réplicat biologique, que les triplicats techniques présentent une variabilité assez forte pour ce genre d'expérimentation et que le choix des heures d'observation ne s'est pas avéré optimal pour suivre la croissance des différentes souches, nous ne pouvons rien conclure pour le moment.

4.4. Localisation subcellulaire

Les prédictions de localisation subcellulaire de Predotar (v. 1.04) pour la protéine encodée par EAMY_0820, si elle est injectée, l'envoient plutôt dans le chloroplaste (0,22) que dans la mitochondrie (0,01) ou le réticulum endoplasmique (0). TargetP (v. 1.1) prédit également la protéine dans le chloroplaste (0,625).

De manière analogue à RipL, de nombreuses expérimentations ont été réalisées afin de déterminer la localisation subcellulaire de la PPR d'*E. amylovora*, et ce sans succès jusqu'à présent. Le tableau suivant résume toutes les tentatives infructueuses réalisées durant la thèse.

Tableau XXVI : Résumé des différentes méthodes testées afin d'obtenir la localisation subcellulaire de la PPR d'*Erwinia amylovora*.

Méthode	Matériel végétal	Construction			Commentaires
		Promoteur	Insert	Etiquette	
Agroinfiltration	<i>Nicotiana benthamiana</i> sauvage	p35S	ArtEaL ₁₋₄₃₈	dsRed	Avec et sans inhibiteur de <i>silencing</i> ; observations à 3, 4 et 5 jours ; microscopie à épifluorescence et confocale ; sur feuille entière ou protoplaste
				GFP	
			ArtEa ₁₋₁₀₀	dsRed	
				GFP	
	<i>Arabidopsis thaliana</i> sauvage		ArtEa ₁₋₄₃₈	dsRed	Sans inhibiteur de <i>silencing</i> , observation à 5 jours sur feuille entière au microscope confocal
		ArtEa ₁₋₁₀₀			
Enrichissement sur gradient de Percoll et Western Blot	<i>Arabidopsis thaliana</i> surexpression		ArtEa ₁₋₄₃₈	3xHA	Sur extraits totaux ; enrichissement chloroplastes ; mitochondries et noyaux ; thylakoïdes cassés et cytosol
Infection <i>E. amylovora</i> surexprimant la PPR fusionnée à GFP ₁₁	Différentes lignées de surexpression de la GFP1-10 dans différents compartiments subcellulaires	Promoteur natif bactérien	PPR_Ea ₁₋₄₃₈ native	1xHA-GFP ₁₁	Basé sur le travail de Park et al. (2017)

4.5. Rôle *in planta*

4.5.1. Un interactant en double hybride levure chez *Arabidopsis*

De façon analogue à RipL, la PPR d'*Erwinia* a été testée en crible Y2H contre la banque d'*Arabidopsis thaliana*. Deux différences majeures par rapport à RipL. La première est que la fusion ArtEa-DB n'est pas auto-activatrice, ce qui a grandement facilité les expérimentations. La seconde, c'est qu'ArtEa ne semble interagir qu'avec très peu de protéines d'*A. thaliana* puisqu'une seule protéine a été mise en évidence lors de ce crible. Cette protéine, encodée par AT5G02150, est nommée AtFes1c (Zhang et al., 2010). On peut voir l'interaction sur la figure 54. La croissance sur milieu de sélection des diploïdes, permettant de s'assurer que l'absence de pousse sur milieu de sélection d'interaction est bien due à une absence d'interaction chez la levure, est présentée en annexe S16.

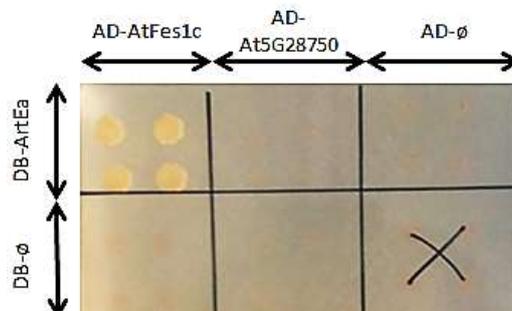


Figure 54 : Mise en évidence de l'interaction en double hybride levure entre DB-ArtEa et AD-AtFes1c (AT5G02150) sur milieu de sélection stringent.

AT5G28750 sert de contrôle négatif d'interaction, tout comme les construits AD et DB vides (\emptyset) et l'eau. Les levures déposées sur milieu de sélection d'interaction et de sélection diploïde sont issues d'une même culture.

AtFes1c est un homologue de Fes1p, une protéine décrite chez *S. cerevisiae* comme jouant le rôle d'une NEF (pour *Nucleotide Exchange Factor*) (Kabani et al., 2002). Ces protéines facilitent la libération de l'ADP dans les cycles des protéines chaperonnes Hsp70, accélérant ainsi leurs cycles de repliement des protéines. Outre AtFes1c, deux autres homologues de Fes1p existent chez *A. thaliana* : AtFes1a (AT3G09350) et AtFes1b (AT3G53800). Le gène *atfes1a* a été décrit codant pour une protéine cytosolique impliquée dans la thermotolérance, interagissant avec des Hsp70, mais aucune activité NEF n'a été mise en évidence (Zhang et al., 2010). A ma connaissance, aucune description d'AtFes1c n'a été proposée chez *A. thaliana*. Cette protéine apparaît cependant dans une publication de crible en double hybride levure comme étant très ciblée par les effecteurs de pathogènes (*Pseudomonas syringae* et *Hyaloperonospora arabidopsidis* en l'occurrence), en comparaison avec ce que l'on attendrait avec son interactome chez *A. thaliana* (Mukhtar et al., 2011b).

Afin d'obtenir une localisation subcellulaire pour AtFes1c, nous avons effectué des fusions GFP et dsred en partie C-terminale la protéine, réalisé des agroinfiltrations et des observations au microscope confocal, mais sans succès jusqu'à présent. Il est néanmoins très fortement probable qu'AtFes1c localise dans le cytoplasme à l'image de ses homologues chez la plante ou même en levure (Kabani et al., 2002; Zhang et al., 2010). Predotar prédit que la protéine est envoyée ailleurs que dans le chloroplaste, la mitochondrie et le réticulum endoplasmique (0,99), alors qu'un résultat très proche est proposé par TargetP (0,819).

Dans le but d'avoir un peu plus d'informations sur AtFes1c, nous l'avons testé contre la banque en double hybride levure. Etant donné que la fusion DB-AtFes1c est auto-activatrice, il n'a pu être testé que contre la banque DB dont les auto-activateurs ont été retirés. Des résultats existaient déjà dans la littérature concernant l'interactome d'AtFes1c chez *Arabidopsis* (Dreze et al., 2010) mais, à l'image de ce qui a été expliqué pour RipL, seules les protéines du groupe 1 avaient été testées. Le tableau XXVII résume les interactants connus et trouvés lors de notre crible pour AtFes1c chez *A. thaliana* et la figure 55 montre son réseau.

Tableau XXVII: Liste des interactants trouvés pour AtFes1c lors d'un crible contre la banque DB.

Les descriptions sont données par le site TAIR et les données de localisation subcellulaire par SUBA4. Les AGI possédant une étoile sont ceux trouvés par Dreze et al. (2010) qui n'ont pas été retrouvés dans notre analyse, ceux en gras avec une étoile correspondent à ceux trouvés par et retrouvés dans notre analyse et tous les autres à ceux trouvés dans notre analyse.

AGI interactants	Localisation subcellulaire	Description TAIR
AT1G04930*	Noyau	hydroxyproline-rich glycoprotein family protein.
AT1G05860	Noyau	Hypothetical protein
AT1G07830	Mitochondrie	ribosomal protein L29 family protein.
AT1G10850	Membrane plasmique	Leucine-rich repeat protein kinase family protein.
AT1G12620	Mitochondrie	Pentatricopeptide repeat (PPR) superfamily protein.
AT1G15385*	Cytosol	Hypothetical protein

AT1G15810	Mitochondrie	S15/NS1, RNA-binding protein.
AT1G19010	Noyau	Hypothetical protein
AT1G28135*	Noyau	Hypothetical protein
AT1G49850	Noyau	RING/U-box superfamily protein.
AT1G54200	Noyau	Hypothetical protein
AT1G56150	Cytosol	SAUR-like auxin-responsive protein family.
AT1G56170	Noyau	nuclear factor Y, subunit C2.
AT1G61870	Mitochondrie	pentatricopeptide repeat 336.
AT1G62780	Chloroplaste	Hypothetical protein
AT1G66500	Noyau	Pre-mRNA cleavage complex II.
AT1G68340	Noyau	Protein of unknown function (DUF1639).
AT1G77950	Cytosol	AGAMOUS-like 67.
AT2G17670	Chloroplaste	Tetratricopeptide repeat (TPR)-like superfamily protein.
AT2G23370	Mitochondrie	Hypothetical protein
AT2G35240	Mitochondrie	plastid developmental protein DAG, putative.
AT2G35940	Noyau	BEL1-like homeodomain 1;Encodes a member of the BEL-like homeodomain protein family.
AT2G38300	Noyau	myb-like HTH transcriptional regulator family protein.
AT2G41350	Noyau	;
AT2G42260	Noyau	uv-b-insensitive 4;Encodes a novel plant-specific protein of unknown function.
AT2G45950	Noyau	SKP1-like 20.
AT3G01180	Chloroplaste et noyau	starch synthase 2.
AT3G13160	Mitochondrie	Tetratricopeptide repeat (TPR)-like superfamily protein.
AT3G15160*	Cytosol	Hypothetical protein
AT3G17300	Mitochondrie	Hypothetical protein

AT3G23290*	Noyau	Light Sensitive Hypocotyls 4, LSH4
AT3G50330	Noyau	basic helix-loop-helix (bHLH) DNA-binding superfamily protein.
AT3G54826	Mitochondrie	Zim17-type zinc finger protein.
AT3G54850	Noyau	plant U-box 14;Encodes a protein with a typical U-box domain followed by an Armadillo repeat region.
AT3G57980	Noyau	DNA-binding bromodomain-containing protein.
AT4G02770	Chloroplaste	photosystem I subunit D-1;Encodes a protein predicted by sequence similarity with spinach PsaD to be photosystem I reaction center subunit II (PsaD1).
AT4G04640	Chloroplaste	ATPase, F1 complex, gamma subunit protein;One of two genes (with ATPC2) encoding the gamma subunit of Arabidopsis chloroplast ATP synthase.
AT4G09650	Chloroplaste	ATP synthase delta-subunit gene;Encodes the chloroplast ATPase delta-subunit.
AT4G10040*	Mitochondrie	cytochrome c-2;Encodes cytochrome c.
AT4G11830	Chloroplaste et membrane plasmique	phospholipase D gamma 2;Encodes one of three phospholipase D enzymes of the gamma class.
AT4G32470	Mitochondrie et golgi	Cytochrome bd ubiquinol oxidase, 14kDa subunit;
AT4G33520	Chloroplaste et mitochondrie	P-type ATP-ase 1;Encodes a putative metal-transporting P-type ATPase.
AT4G34700	Mitochondrie	LYR family of Fe/S cluster biogenesis protein.
AT4G36110	Noyau	SAUR-like auxin-responsive protein family.
AT4G39820*	Golgi	Tetratricopeptide repeat (TPR)-like superfamily protein.
AT5G03040	Membrane plasmique et golgi	IQ-domain 2.
AT5G03500	Noyau	Mediator complex, subunit Med7.
AT5G05790	Noyau	Duplicated homeodomain-like superfamily protein.
AT5G11060	Noyau	KNOTTED1-like homeobox gene 4.
AT5G14070	Cytosol	Thioredoxin superfamily.
AT5G14170	Noyau et cytosquelette	SWIB/MDM2 domain superfamily protein.
AT5G37055	Noyau	HIT-type Zinc finger family protein.
AT5G37550	Chloroplaste	Hypothetical protein

En jaune, AtFes1c (AT5G02150), en bleu les protéines non PPR d'*A. thaliana*, en vert, les protéines PPR d'*A. thaliana* et en rouge les effecteurs testés par Mukhtar et al. (2011) ou la plateforme InterATOME.

Il est intéressant de remarquer que 10 protéines interagissant avec AtFes1c interagissent aussi avec des effecteurs et que deux effecteurs interagissent directement avec AtFes1c.

Cependant, aucune protéine Hsp n'a été mise en évidence parmi les interactants, ce qui aurait pu appuyer un peu plus l'hypothèse de co-chaperone. Plus étonnant par contre, la proportion de protéines impliquées dans la fixation des acides nucléiques. On remarque aussi dans le tableau XXVII que de nombreuses protéines sont annotées TPR ou PPR. Après avoir testé sur TPRpred toutes les protéines annotées PPR ou TPR, il s'avère qu'AT1G61870, AT3G13160, AT2G17670 et AT1G12620 sont bien des PPR, et prédites –voire même observée pour AT1G61870- pour aller dans la mitochondrie (Uyttewaal et al., 2008; Colcombet et al., 2013). En plus de celle d'*Erwinia*, AtFes1c semble donc avoir une certaine affinité pour les protéines PPR.

Sur ces données d'interactomique, nous avons réalisé une analyse Thalemine, à l'image que ce qui a été réalisé pour RipL. La liste des interactants d'AtFes1c semble être uniquement enrichie en protéines impliquées dans les chaînes de transport des électrons (4,8^E-4), cette rubrique regroupant les chaînes chloroplastiques et mitochondriales.

Concernant les localisations subcellulaires des interactants, détaillées dans l'annexe S17, deux choses paraissent intrigantes. La première est que selon si on tient compte seulement des observations ou des observations et des prédictions, on obtient deux résultats assez différents, avec forte proportion de prédictions nucléaires. La seconde observation est la très grande proportion de protéines observées localisant dans le chloroplaste ou la mitochondrie. L'analyse Thalemine ne montre cependant pas d'enrichissement pour une localisation subcellulaire en particulier.

Enfin, et comme on peut l'observer sur la figure 55, AtFes1c interagit avec 3 protéines, qui elles-mêmes interagissent avec RipL, l'effecteur à motifs PPR de *Ralstonia solanacearum*. Ces protéines sont AT1G49850, AT2G42260 et AT1G66500. Et comme nous l'avons vu précédemment, AtFes1c interagit aussi avec deux effecteurs en plus des deux effecteurs de *Hyaloperonospora arabidopsidis*. Il s'avère très étonnant que RipL et AtFes1c partagent des interactants. L'hypothèse la plus probable serait que ces protéines soient pour la plupart impliquées dans la défense de la plante, ou que leur détournement s'avère bénéfique au pathogène dans un contexte d'infection. La littérature et les descriptions connues à ce jour pour ces protéines n'appuient cependant pas cette hypothèse. AT1G49850 encode pour une protéine à domaine RING (pour *Really Interesting New Gene finger*), un domaine assez particulier souvent rattaché aux domaines à doigts de zinc. Ces domaines permettent les interactions protéine-protéine et les protéines les possédants semblent être impliquées dans de nombreux processus biologiques (Kosarev et al., 2002). Aucune caractérisation de cette protéine n'a pour le moment été proposée. Concernant AT1G66500, je n'ai pas trouvé d'information claire le décrivant dans la littérature. Une analyse de sa séquence sur Prosite révèle cependant un motif en doigt de zinc. Pour AT2G42260, les informations trouvées ont déjà été présentées précédemment dans l'interactome de RipL.

Comme détaillé un peu plus haut, lors d'un gros projet de crible Y2H d'effecteurs de *Hyaloperonospora arabidopsidis* (*Hpa*) et *Pseudomonas syringae*, AtFes1c s'est avéré être l'une des 51 parmi les 2661 protéines d'*Arabidopsis* testées à être ciblées par les effecteurs significativement plus souvent que ce à quoi on s'attendait par rapport à son interactome *in planta* (Mukhtar et al., 2011a). Deux effecteurs de *Hpa*, HaRxL73 et ATR13, interagissent avec AtFes1c, en plus de la PPR encodée par *Erwinia amylovora* et des deux autres effecteurs trouvés par la plateforme InterATOME. Sur la base de ces données, une hypothèse plausible serait qu'AtFes1c soit une protéine régulièrement ciblée par les effecteurs, qu'elle joue un rôle dans la défense de la plante et qu'*Erwinia amylovora* et *Hpa* aient cette cible en commun.

4.5.2. Validation du rôle d'AtFes1c dans la défense d'*A. thaliana* contre les pathogènes

Afin de tester notre précédente hypothèse, nous avons travaillé avec les lignées de surexpression ArtEa, mais aussi avec une lignée homozygote d'insertion d'ADN-T dans *atfes1c*, qui sera nommée par la suite $\Delta atfes1c$. Ces lignées ont ensuite été infectées à l'aide de *Hyaloperonospora arabidopsidis* et les spores ont été comptées 7 jours après infection (Figure 56).

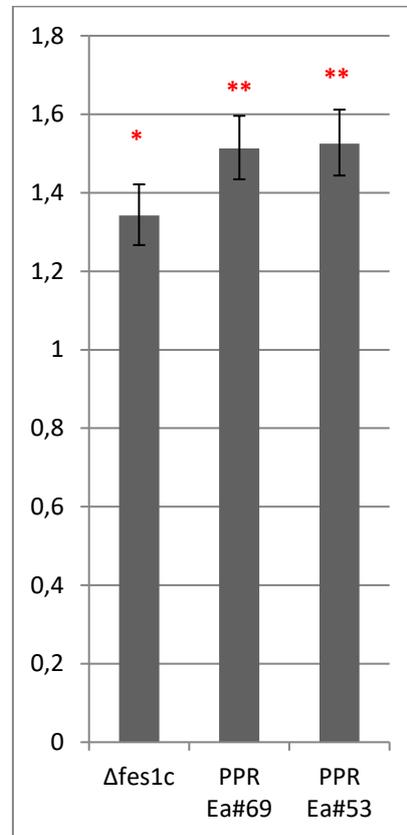


Figure 56 : Ratio de la sporulation de *Hpa* par mg de matière fraîche végétale à 7 jours entre les différents fonds génétiques d'*A. thaliana* et Col0.

Deux étoiles : p -value < 0.01 ; une étoile : p -value < 0.05. Analyse basée sur 3 répétitions biologiques.

Ces données ont été analysées de la même façon que le pathotest *Erwinia* sur *Arabidopsis*, à savoir un modèle linéaire généralisé permettant de prendre en compte l'effet des répétitions biologiques, des réplicats techniques et ici éventuellement de la masse de matière fraîche végétale. On remarque que les deux lignées de surexpression présentent une sensibilité accrue à *Hpa* en comparaison avec Col0. On retrouve une certaine sensibilité chez $\Delta atfes1c$, significative mais moins marquée. On remarquera que nous n'avons utilisé qu'un seul allèle concernant *atfes1c*. Une complémentation fonctionnelle est en cours, mais les pathotests *Hpa* nécessitant des lignées homozygotes, je n'ai pas eu le temps de mener cette expérimentation jusqu'au bout dans le cadre de la thèse.

Afin de tester si les lignées de surexpression étaient sensibles à d'autres pathogènes, des pathotests avec *Pseudomonas syringae* ont été effectués, mais n'ont pas permis de mettre de différence en évidence avec la condition contrôle (Figure 57).

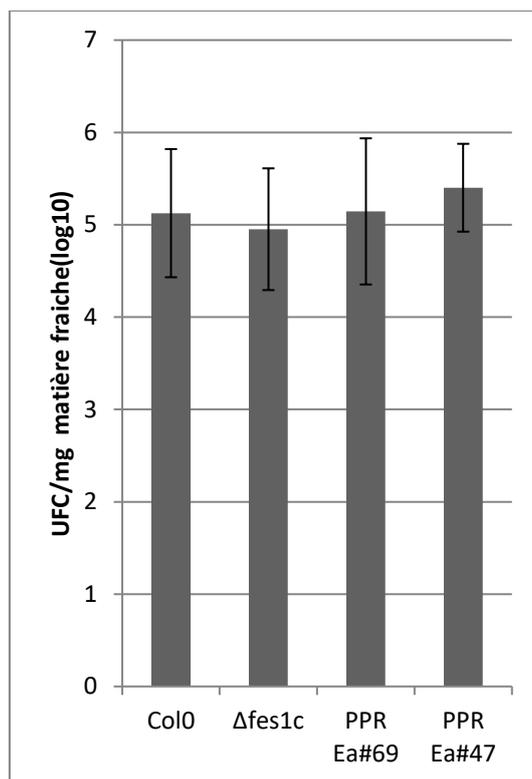


Figure 57 : Comptage de colonies de *Pseudomonas syringae* par mg de matière fraîche végétale, représentés en échelle Log10, 48h après infection.

Réplicat biologique représentatif parmi les 3 effectués.

Afin de tester si AtFes1c jouait un rôle dans la défense non-hôte contre *Erwinia amylovora*, nous avons testé la sensibilité du mutant en pathotest, avec la souche sauvage bactérienne et la souche mutante pour EAMY_0820 (Figure 58).

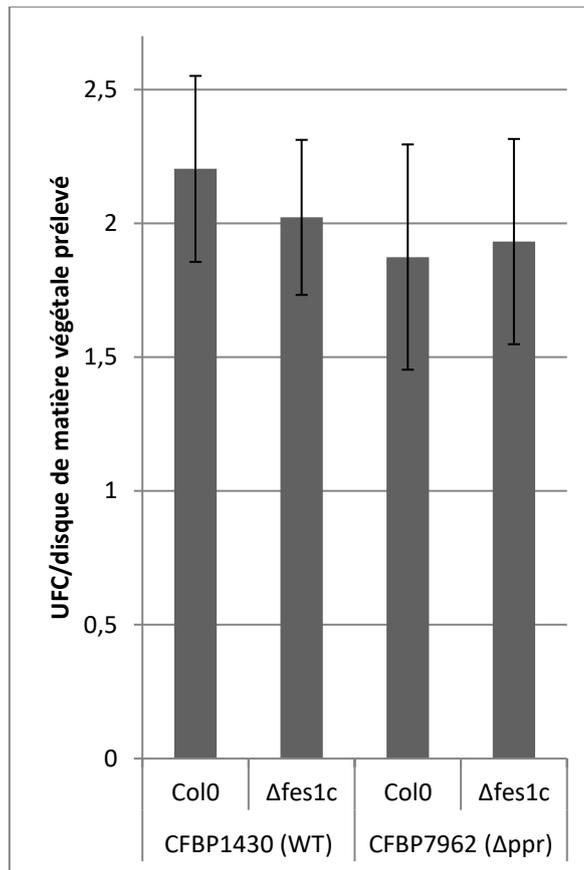


Figure 58 : Comptage de colonies d'*Erwinia amylovora* par disque de matière végétale prélevé, représentés en échelle Log10, 24h après infection.

Les données sont cependant difficilement exploitables à cause de la variabilité extrêmement marquée entre les différents échantillons. De plus, suite à un problème de croissance bactérienne, les deux répétitions biologiques initialement prévues ont été infectées le même jour. Au lieu de disposer de 12 échantillons pour deux répétitions biologiques, nous disposons d'une seule répétition à 24 échantillons. L'absence de répétition biologique couplée à une variabilité nettement plus forte que ce qui est habituellement observé nous empêche de conclure quant au rôle d'AtFes1c dans la défense d'*A. thaliana* en interaction non-hôte avec *E. amylovora*.

Pour résumer, une sensibilité du mutant Δ atfes1c et des lignées de surexpression de la PPR d'*Erwinia* à *Hpa* ont été mis en évidence. Ces mêmes lignées ne sont pas sensibles à *Pseudomonas syringae*. Les pathotests *Hpa* montrent que les niveaux de sensibilité ne sont pas les mêmes entre les lignées de surexpression et Δ atfes1c. Cela laisse supposer que la PPR bactérienne joue peut-être d'autres rôles dans la plante, peut-être un rôle plus proche de celui attendu par une protéine PPR. A l'image de ce qui a été effectué pour RipL, les paragraphes suivant vont décrire les résultats obtenus par *RNAseq* sur la lignée de surexpression d'ArtEa dans *Arabidopsis thaliana*.

4.5.3. Analyses transcriptomiques

4.5.3.1. Analyse du transcriptome des organites

Concernant les organites, aucun défaut d'édition, d'épissage ou de couverture n'a été mis en évidence après séquençage des ARN totaux.

Pour les petits ARN, après segmentation, analyse différentielle et correction de Bonferroni, nous n'avons pas vu d'empreinte putative en se basant sur la définition proposées dans le chapitre de méthodologie d'analyses de données *RNAseq*, et ce que ce soit dans le chloroplaste ou la mitochondrie.

4.5.3.2. Analyse du transcriptome nucléaire

Pour l'analyse d'expression différentielle, 89 gènes dont la PPR bactérienne ont été trouvés comme différentiellement exprimés après correction avec la méthode de Benjamini-Hochberg (Annexe S18). Parmi ces gènes, aucun n'est chloroplastique, mitochondrial ou lié à la transcription de manière générale. AT5G02150, codant pour la protéine AtFes1c, n'est pas différentiellement exprimé.

En effectuant une analyse Thalemine, on trouve nettement moins d'enrichissements passant les seuils statistiques que pour le transcriptome de la lignée de surexpression de RipL. Cependant, on retrouve la réponse aux stimuli externes (0,001), aux stimuli biotiques (0,007), aux composés dérivés de l'oxygène (0,009), aux stress (0,009) et pour terminer le métabolisme de l'acide jasmonique (0,049). Aucun enrichissement concernant la localisation subcellulaire n'a été observé.

L'analyse MapMan du métabolisme ne montre pas de défauts majeurs. Après avoir effectué une correction de p-values avec la méthode de Benjamini-Yekutieli, le test de la somme des rangs de Wilcoxon ne montre rien dans la réponse aux stress, mais suggère des différences au niveau de la photosynthèse (0.0016), du métabolisme secondaire des flavonoïdes (0.00149), des métabolites secondaires contenant du soufre (0.00157) puis pour finir dans la synthèse de l'arabinogalactane, un composé de la paroi végétale (0.00411).

Suite à ces résultats, les chlorophylles a et b ont été dosées par spectrophotométrie après extraction à l'acétone (Annexe S14). Aucune différence statistiquement significative n'a été mise en évidence sur les deux répétitions biologiques effectuées à l'aide d'un test de Student et d'une correction de tests multiples. Nous avons aussi évalué deux paramètres de la photosynthèse par PAM fluorescence : le rendement du photosystème II et le *Non-Photochemical Quenching* (pour NPQ, correspondant à la dissipation de l'excès d'énergie). Deux conditions ont été testées : la première consistait à exposer les plantes à un rayonnement de plus en plus fort et la seconde à exposer les plantes à l'obscurité, puis à une forte lumière (photoinhibition) afin de saturer les photosystèmes, puis ensuite de mesurer la récupération post-photoinhibition (Annexes S19 à S22). A nouveau, aucune différence n'a été mise en évidence.

4.5.3.3. Limites liées à la surexpression d'un gène bactérien *in planta*

La même analyse que celle faite pour ArtRipL a été effectuée pour ArtEa. En moyenne, 3168 petits ARN anti-sens de 21nt s'alignent sur ArtEa et 917 de 24nt. En comparaison, AT1G09680, une PPR pure de taille identique et dont l'expression est similaire, ne présente aucun petit ARN anti-sens de 21 ou 24nt suggérant que la PPR bactérienne est également la cible d'ARN interférents.

4.6. Exportation des résultats sur *Malus domestica* « Golden delicious »

Bien que l'interaction AtFes1c avec la PPR d'*Erwinia amylovora* puisse jouer un rôle dans l'interaction non-hôte, il serait intéressant de savoir si cette interaction existe également chez la plante hôte *Malus domestica* « Golden delicious », pour laquelle nous avons un phénotype avec le mutant KO bactérien pour EAMY_0820. En utilisant BlastP avec AtFes1c (AT5G02150), trois protéines de *Malus domestica* « Golden delicious » furent mises en évidence comme d'éventuels homologues : XP_008387795.1, renommée Fes1 α , XP_008369926.1 renommée Fes1 β et XP_017191978.1 renommée Fes1 γ . En observant les gènes codant ces protéines sur le génome du pommier, je me suis rendu compte que le gène codant pour Fes1 γ ressemblait énormément à celui codant pour Fes1 α avec des exons en plus en extrémité 5'. En regardant de plus près les

données de l'assemblage du génome, j'ai remarqué également que la couverture au niveau du codon d'initiation de *fes1 γ* était équivalente à celle que l'on pouvait avoir dans un intron, ce qui n'est pas le cas pour *fes1 α* . Toutes ces informations couplées à l'impossibilité d'amplifier l'ARNm de *fes1 γ* me laissent supposer que l'annotation correspondant à *fes1 γ* est fautive, soit à cause d'une erreur de séquençage, soit que l'on n'est pas sur le bon codon d'initiation.

Des clonages ont donc été effectués avec *Fes1 α* et *Fes1 β* en AD et en DB, et ils ont été testés en double hybride levure face à AD et DB-*ArtEa*. Aucun de ces construits ne s'est montré auto-activateur en levure et ils ont tous pu être testés en un contre un.

En contrôle positif, l'interaction AD-*AtFes1c*/DB-*ArtEa* a parfaitement fonctionné, mais aucune interaction n'a pu être mise en évidence entre *Fes1 α/β* et *ArtEa*, en AD comme en DB.

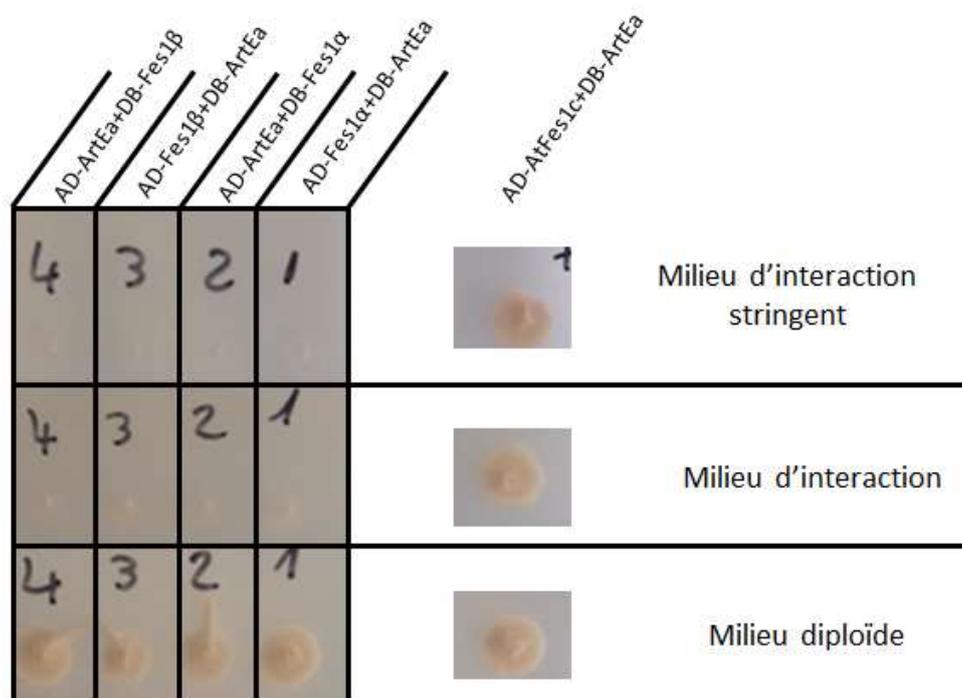


Figure 59 : Crible de double hybride levure en un contre un sur différents milieux solides, pour la PPR d'*Erwinia amylovora*, son interactant chez *A. thaliana* *AtFes1c* et ses deux homologues chez *Malus domestica* « Golden delicious » *Fes1 α* et *Fes1 β* .

4.7. Discussion

Pour résumer, *EAMY_0820* est un gène prédit pour encoder une protéine PPR chez la bactérie pathogène *Erwinia amylovora* CFBP1430. Neuf motifs sont prédits sur la protéine, et elle est également prédite pour être transloquée dans la cellule végétale via le T3SS, pour potentiellement cibler le chloroplaste. Le travail effectué est résumé par la figure 60.

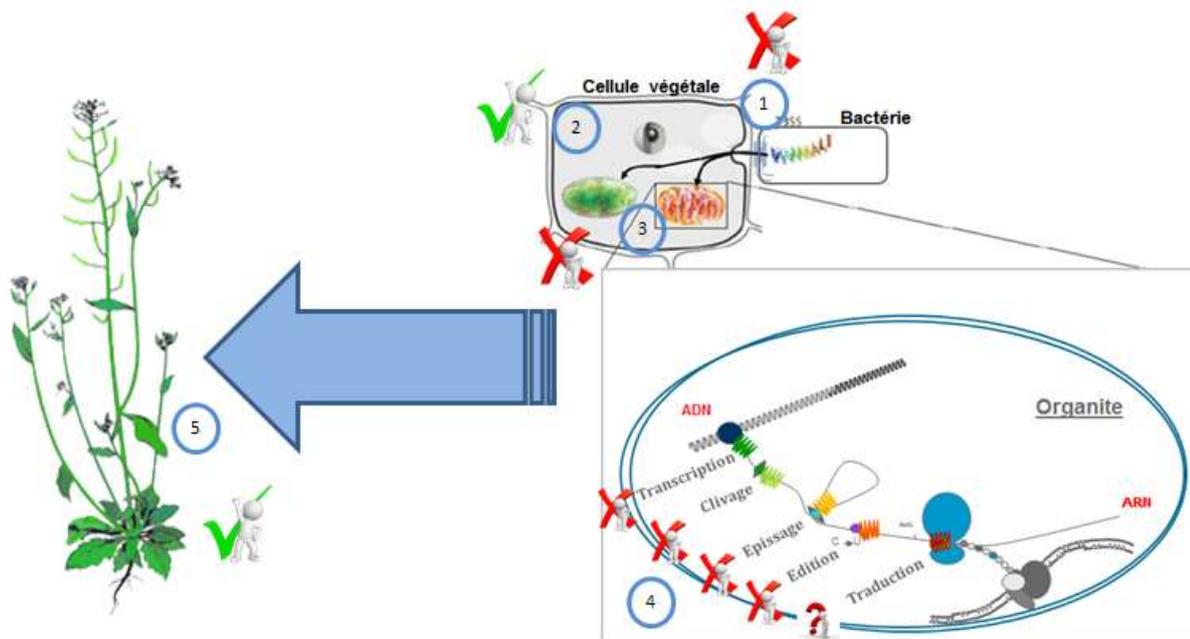


Figure 60 : Résumé des résultats obtenus lors de la thèse sur la caractérisation de la protéine encodée par EAMY_0820.

Les croix vertes font références aux expérimentations ayant mis en évidence un mécanisme concernant RipL, les points d'interrogation aux expérimentations non effectuées ou non concluante et les croix rouges à celles n'ayant pas mis en évidence de défaut ou de mécanisme. Validation de l'injection (1), localisation subcellulaire (2), protéines de la plante interagissant avec les protéines bactériennes (3), rôle potentiel dans les organites sur les différentes étapes de maturation post-transcriptionnelle (4) et conséquences sur la physiologie de la plante et sa résistance aux maladies (5).

Chez la bactérie, l'insertion d'une cassette de résistance à la kanamycine dans EAMY_0820 entraîne une diminution du taux d'incidence sur la plante hôte *Malus domestica* « Golden delicious » et à l'inverse, une meilleure survie dans la plante non-hôte *Arabidopsis thaliana*. Concernant le fait de trouver des résultats différents entre hôte et non-hôte, cela n'a rien de choquant pour plusieurs raisons. Tout d'abord, il n'est pas forcément nécessaire de mettre ces deux résultats en opposition, ils concernent des contextes d'interactions différents et apportent chacun des informations différentes. Dans un cas, l'interaction est compatible (maladie) alors que dans l'autre, elle est incompatible (résistance non-hôte). Ensuite, les méthodes utilisées dans les deux pathotests divergent. Concernant *A. thaliana*, c'est un comptage de colonies qui a été réalisé alors qu'une étude de symptômes a été faite sur pommier. Il serait intéressant de disposer d'échantillons permettant d'évaluer la croissance des deux souches bactériennes sur plante hôte, en inoculant des fruits immatures par exemple. En culture liquide, la croissance de la bactérie mutante n'est pas différente de la souche sauvage en milieu riche. Cependant, la croissance de la bactérie n'a pas été suffisamment testée dans un milieu minimum mimant l'apoplaste de la plante lors de l'infection. Il serait intéressant de voir si sa croissance est affectée dans une condition plus proche de l'infection.

Une autre explication probable de la différence observée entre *Malus domestica* et *A. thaliana* peut être l'interactant de la protéine PPR d'*Erwinia* trouvé chez *Arabidopsis*. Cette protéine, proposée par Mukhtar et al. (2011) comme étant ciblée par les effecteurs d'agents pathogènes, interagit également avec deux effecteurs de *Hyaloperonospora arabidopsidis* (*Hpa*). Un mutant d'insertion pour cet interactant ainsi que les lignées de surexpression de la PPR dans *A. thaliana* semblent être plus sensibles que Col0 à *Hpa*, suggérant qu'AtFes1c jouerait un rôle dans la défense de la plante. Ces mêmes lignées ne montrent pas de résistance ou de sensibilité

accrue à *Pseudomonas syringae*, suggérant que cette sensibilité serait spécifique à *Hpa*. Afin de confirmer un peu plus cette spécificité (ce qui demeure impossible dans l'absolu, tous les pathogènes ne pouvant être testés), il aurait été pertinent de tester par exemple *Botrytis cinerea*, ce qui aurait permis d'avoir un agent nécrotrophe en interaction hôte sur les différentes lignées, ou encore contre *Ralstonia solanacearum*, suite aux données d'interactomique obtenues. Etant donné que la PPR d'*Erwinia* interagit avec Fes1c chez *A. thaliana*, mais pas chez le pommier, ça peut en partie expliquer la différence de comportement du mutant entre les deux espèces. On peut aussi nuancer en suggérant que ne pas voir d'interaction avec les homologues de Fes1c dans le pommier ne signifie pas qu'elle n'existe pas.

Toujours concernant AtFes1c, l'étude de ses homologues chez le pommier n'a pas permis de mettre en évidence une interaction en double hybride levure avec la protéine encodée par EAMY_0820. Si cela ne signifie pas que les protéines n'interagissent pas –il est impossible de démontrer une absence- cela peut toutefois être un début d'explication des résultats obtenus sur plante hôte (taux d'incidence plus faible) et sur plante non-hôte (titre bactérien plus élevé) de notre mutant avec AtFes1c, protéine impliquée dans la défense et détournant potentiellement la protéine PPR bactérienne de son rôle lors de l'infection, si celle-ci est bien injectée. Concernant l'interactome d'AtFes1c chez *A. thaliana*, on remarque que la protéine interagit beaucoup, notamment avec d'autres protéines PPR. De nombreuses protéines interagissant avec AtFes1c ont été observées dans les organites. En couplant les informations obtenues sur les homologues d'AtFes1c chez la levure par exemple et les données de son interactome, on peut supposer qu'AtFes1c joue le rôle d'une protéine NEF cytosolique avec une chaperonne impliquée dans l'adressage de protéines vers les organites. Une seconde hypothèse, en lien avec les données des pathotests et des homologues d'AtFes1c connus chez *Arabidopsis*, serait qu'AtFes1c soit issue d'une duplication de gènes à partir d'AtFes1a, protéine impliquée dans la résistance de la plante au stress chaud, et qu'AtFes1c joue le rôle « d'appât à effecteurs » en condition d'infection. On peut donc imaginer un modèle de leurre où AtFes1c protégerait AtFes1a en condition d'infection (Block et Alfano, 2011). Il serait intéressant, comme dit précédemment, de tester les mutant *Afes1a* et *c* contre de nombreux autres pathogènes, et de regarder si cette distribution AtFes1a, b et c est conservées dans d'autres plantes, notamment des plantes hôtes de *Hpa*.

Pour revenir à la caractérisation de la PPR d'*Erwinia*, aucun mécanisme moléculaire évident n'a été mis en évidence dans les organites. Si la protéine est effectivement transloquée et localisée dans les organites, il est alors fort probable qu'elle soit impliquée dans la traduction, ce que nos approches transcriptomiques ne nous ont pas permis de tester. Concernant les approches *RNAseq*, les mêmes doutes levés par l'analyse des lignées de surexpression d'ArtRipL demeurent. Le niveau d'expression de la PPR a-t-il été suffisant ? Le nombre de *reads* dénombré lors du séquençage des ARN totaux suggère que oui mais l'analyse des petits ARN montre que de nombreux ARN interférents semblent cibler le gène codant pour la PPR bactérienne. A travers cette étude, il a été démontré que cette PPR joue un rôle lors de l'infection. Cependant, mis à part les prédictions, absolument aucune preuve n'a été apportée concernant sa translocation et rien ne nous permet de réfuter le fait que cette PPR puisse jouer un rôle dans l'infection, mais dans la bactérie. Un certain nombre d'expérimentations reste cependant faisable afin de valider la translocation. Nous disposons du matériel pour en réaliser, comme les Western Blot anti-HA sur le surnageant de la bactérie surexprimant une version avec étiquette de la PPR, ou les fusions GFP₁₁ détaillées précédemment. Ces expérimentations, lourdes, n'ont été tentées qu'à une seule reprise par manque de temps, et n'ont pas fonctionné pour le contrôle positif. Il est donc primordial de les refaire, afin au moins de faire fonctionner ces différents systèmes pour un effecteur connu d'*Erwinia*, et dans l'idéal pour la PPR aussi.

Discussion générale.

L'hypothèse générale de la thèse est qu'il existe des mécanismes de régulation de l'expression génétique des organites, en lien avec les protéines PPR, en réponse au stress biotique.

Nous avons choisi de tester cette hypothèse pour de nombreuses raisons. Tout d'abord, on sait que les organites sont impliqués dans la défense face aux organismes pathogènes. Le chloroplaste est connu comme étant pour partie impliqué dans la synthèse d'acide jasmonique et salicylique (Creelman et Mullet, 1997; Strawn et al., 2007) et il peut également amplifier la signalisation via la production de ROS en condition de stress (Kangasjärvi et al., 2012). Ensuite, il existe de nombreuses données montrant que l'activité des organites est modifiée en condition de stress biotique. Du côté du chloroplaste, l'activité photosynthétique diminue après infection par des organismes pathogènes (Bonfig et al., 2006; Kangasjärvi et al., 2012). Du côté de la mitochondrie, le métabolisme respiratoire augmente, probablement suite à la diminution de l'activité photosynthétique afin d'assurer un apport continu en énergie à la cellule (Kangasjärvi et al., 2012). Il est donc évident que l'activité des organites en condition de stress est largement modulée. Reste à savoir comment cette activité est modulée. De nombreuses publications démontrent que l'expression des gènes nucléaires codant pour des protéines chloroplastiques et mitochondriales est largement remodelée en condition de stress (Meyer et al., 2001; Bilgin et al., 2010; de Torres Zabala et al., 2015). Les deux organites possèdent un génome, suggérant que les variations d'activités en condition de stress biotique pourraient provenir de la régulation de l'expression de leurs gènes.

En soutien de cette hypothèse, on trouve des exemples de protéines régulées par les stress et susceptibles de jouer un rôle régulateur aux différentes étapes de l'expression des gènes des organites. La première est la transcription. Comme détaillé en introduction, de nombreux facteurs peuvent être impliqués dans la régulation de la transcription des organites, que ce soit les différentes ARN polymérases PEP (d'origine plastidiale) et NEP (d'origine nucléaire) (Liebers et al., 2017) ou les facteurs de transcription Sigma dans le chloroplaste (Lerbs-Mache, 2011). Les facteurs Sigma sont d'autant plus intéressants que nous disposons d'informations sur l'expression d'un facteur en particulier, SIG3, en condition de stress biotique. La protéine SIG3 est décrite comme étant impliquée dans la reconnaissance de nombreux promoteurs dans le chloroplaste, mais très peu de défauts transcriptomiques ont été mis en évidence chez le mutant d'insertion d'ADN-T (Lerbs-Mache, 2011). Il est supposé que SIG3 complète l'action des autres facteurs Sigma. Des données de transcriptomiques montrent que l'expression du gène codant pour SIG3 est induite 12h après infection par *Hyaloperonospora arabidopsidis*, tout comme *psbN*, l'un des deux gènes (avec *atpH*) sous sa dépendance (Wang et al., 2011). Des données de l'équipe montrent que la lignée d'insertion d'ADN-T $\Delta sig3$ est presque deux fois plus sensible au même organisme pathogène que la plante sauvage (communication personnelle). On est donc dans un cas de figure où l'on a en condition de stress biotique une augmentation de la transcription de l'élément régulateur et de l'élément régulé.

La seconde étape est la maturation des ARN. Elle implique, comme nous l'avons vu, de nombreuses protéines interagissant plus ou moins directement avec l'ARN. Parmi ces protéines, les protéines PPR sont, pour certaines, induites en réponse aux stress, aussi bien biotiques qu'abiotiques (Chen et al., 2018) et sont également des cibles de micro ARN en condition de stress (Khraiwesh et al., 2012). PPRL, une protéine PPR présentée en introduction, dont l'expression est régulée par ARN interférence, a attiré mon attention et mériterait que l'on étudie plus en profondeur son mécanisme d'action (Katiyar-Agarwal et al., 2006). Il a été démontré que l'expression de *pprl* diminuait après infection par *Pst* possédant le facteur d'avirulence *AvrRpt2*, via un mécanisme d'ARN interférence de la plante. Sa surexpression entraîne un retard lors de la

mise en place de la réponse hypersensible, l'hypothèse étant que PPRL agit comme régulateur négatif de la voie de résistance RPS2, reconnaissant le facteur d'avirulence AvrRpt2.

Les dernières étapes sont la traduction, dans laquelle certaines protéines PPR peuvent aussi être impliquées, et la maturation post-traductionnelle des protéines, où peuvent aussi se jouer des étapes de régulation. Si l'on étend la définition de l'expression des gènes des organites à l'obtention d'une protéine fonctionnelle, on peut alors inclure les étapes de repliement nécessaires à la plupart des protéines afin d'être fonctionnelles comme une éventuelle régulation de l'expression génétique. Dans ce cas, les protéines Hsp (*Heat Shock Proteins*) sont de très bonnes candidates en tant qu'élément régulateur putatif. Je ne dispose pas de données sur les Hsp en condition de stress biotique, mais leur rôle a été très détaillé lors de stress chaud. La plus connue dans le chloroplaste d'*A. thaliana* est AtHsp101 (AT1G74310). Si la protéine est quasiment indétectable en condition de culture optimale, elle s'accumule lors de fortes températures (Lee et al., 2007). Le mutant d'insertion d'ADN-T dans le gène codant pour AtHsp101 rend la plante très sensible au stress chaud. On peut alors imaginer que les chaperonnes viennent non pas induire l'expression des gènes des organites mais la maintenir à un niveau suffisant, une forme d'homéostasie de l'expression qui est aussi une régulation suite à une variation environnementale.

Toutes ces données préliminaires vont donc dans le sens de notre hypothèse de départ, à savoir qu'il existe une régulation de l'expression des gènes des organites en réponse au stress biotique.

Ce n'est pourtant pas ce que l'on a observé lors de la thèse. Dans le cas de l'étude de la protéine PGN, nous n'avons pas observé de régulation de l'édiction ou de variation l'accumulation de transcrits *nad6* et *cox2* en condition de stress biotique avec *Botrytis cinerea* (Laluk et al., 2011). Au sujet des deux PPR bactériennes, même si nous avons obtenu des résultats très intéressants, nous n'avons pas mis en évidence de modifications majeures du transcriptome des organites après surexpression chez *A. thaliana* et analyse *RNAseq* des ARN totaux et des petits ARN. Nos données ne mettent donc pas en évidence de régulation en condition d'infection, et la littérature non plus. A ma connaissance, aucun effecteur n'a été mis en évidence comme modulant directement l'expression génétique des organites, à l'image de ce que les effecteurs TAL peuvent faire dans le noyau (Scholze et Boch, 2011). Il n'existe presque aucune donnée de variation de l'expression des gènes des organites chez les plantes en condition de stress biotique. Cette absence de publication ne signifie pas nécessairement que cela n'a pas été regardé mais peut aussi vouloir dire qu'il n'y a rien de spécifique à voir au niveau de l'ARN. La très longue demi-vie des ARN des organites, qui peut aller jusqu'à 40h, représente également un frein à l'hypothèse de régulation transcriptionnelle. En effet, on s'attend à une réponse rapide en condition fluctuante et dans l'hypothèse d'une régulation transcriptionnelle, les données de demi-vie suggèrent que la réponse ne serait effective qu'après 40h dans certains cas. De plus, le transcriptome mitochondrial est connu pour être d'une très grande stabilité, même en condition de stress (communication personnelle). Même si le transcriptome chloroplastique est connu pour être modulé selon les étapes de développement, ces variations peuvent s'expliquer par la pluralité des appareils transcriptionnels en présence. Dans le chloroplaste, la transcription va être effectuée pour partie par la PEP, codée par le génome plastidial, et par deux NEP, codée par le noyau, à savoir RPOTMP et RPOTP. Ces différentes enzymes reconnaissent différents promoteurs et donc peuvent être spécifiques de certains groupes de gènes du chloroplaste. N'étant pas toutes actives lors des mêmes étapes de développement, cela peut expliquer les variations de transcription observées sans que ça ne corresponde à une régulation fine. Enfin, un autre argument en faveur de l'absence de régulation transcriptionnelle est la diminution du nombre de copies du génome chloroplastique chez les feuilles matures (Rowan et Bendich, 2009). Si la quantité d'ADN chloroplastique devient indétectable, comment évoquer la régulation de la transcription ? Du côté de la maturation post-transcriptionnelle, à l'image de PGN, de nombreuses protéines PPR, impliquées dans la maturation des ARN ont été étudiées en condition de stress biotique et ont un mutant d'insertion plus sensible à un pathogène. On peut citer les cas de SLO2 et PPR40 par exemple (Zsigmond et al., 2008; Zhu et al., 2012b). Toutefois, peut-on parler de régulation si

l'expression de la PPR et de sa cible ne sont pas modifiées en réponse au pathogène ? Comme PGN, je pense que ces mutants ont un défaut mineur dans la mitochondrie, visible seulement lorsque la plante est en condition de stress, en l'occurrence biotique. Cela n'en fait pas pour autant des protéines de régulation si l'activité de la PPR n'est pas modifiée en réponse au pathogène.

En résumé, ni lors de la thèse, ni dans la littérature nous n'avons beaucoup de preuves démontrant une régulation de l'expression génétique des organites en condition de stress biotique. Plusieurs hypothèses peuvent expliquer ce constat.

La première serait que nous ne sommes pas dans les bonnes conditions (biologiques, techniques) pour voir cette régulation. En condition de stress chaud, un rare papier met en évidence chez le maïs une variation d'édition (Nakajima et Mulligan, 2001). On peut supposer alors que nous n'avons pas testé assez de conditions plante/stress à la recherche de variations de transcription/maturation des ARN afin de mettre en évidence une éventuelle régulation. Il est possible aussi que les techniques utilisées jusqu'à présent ne suffisent pas à donner une vue suffisamment exhaustive. L'approche par RNAseq s'est pourtant démocratisée et en théorie, il serait possible d'accéder à ces données sans réaliser les expérimentations à la paille. Pourtant, en regardant de plus près les bases de données NCBI de séquençage des ARN, sur les 21094 échantillons concernant *A. thaliana* et couvrant de nombreuses conditions de stress, moins de 500 correspondent à des banques avec déplétion des ARN ribosomiaux, prérequis à une analyse des transcriptomes chloroplastiques et mitochondriaux. Ce manque peut s'expliquer par le surcote qu'engendre la déplétion si l'on ne s'intéresse pas aux organites d'une part, et à la très forte proportion de reads chloroplastiques d'autre part. On peut aussi supposer que la faible proportion de ces jeux de données peut tenir du fait que les données restent difficiles à analyser, les pipelines proposés dans la littérature étant assez récents (Castandet et al., 2016b; Malbert et al., 2018).

On peut aussi imaginer que la régulation ne se fait pas gène par gène mais sur le génome complet. Cette hypothèse est logique car à la différence du noyau, la majorité du génome des organites sont transcrits. L'exemple décrit plus haut concernant les ARN polymérase PEP et NEP est intéressant car s'il existe une régulation de la transcription via ces enzymes, elle pourrait se faire sur de grands groupes de gènes. L'exemple des Hsp est également intéressant dans ce contexte, car la régulation pourrait se faire sur de nombreuses protéines indépendamment de la position du gène sur le génome, de la dépendance d'une ARN polymérase ou d'une autre, etc. La stabilité du transcriptome mitochondrial est aussi un argument intéressant. En effet, il est tout à fait possible que cette stabilité soit entretenue par un ou plusieurs éléments régulateurs en condition de stress, malgré un environnement cellulaire modifié. Kangasjärvi et al. (2012) proposent que cette stabilité du transcriptome permette d'assurer la continuité dans l'apport en énergie suite à la diminution de l'activité photosynthétique. A ce moment-là, ce serait le reste de la cellule, en particulier le noyau, qui modifierait l'environnement cellulaire afin de permettre cette stabilité. On peut citer ici l'exemple de PGN. En condition de stress salin, nous voyons que l'édition sur le 5'UTR de *nad6* reste stable par rapport à la condition contrôle, avec une augmentation de la transcription de *pgn* et de *nad6*. Même si cela reste à confirmer au niveau protéique pour PGN, cela suggère une induction du gène codant pour PGN pour répondre à une demande accrue de l'édition de *nad6* puisqu'il y a un peu plus d'ARN à éditer. Cela permettrait en théorie de conserver un niveau d'édition stable, participant ainsi à la régulation de l'expression génétique de la mitochondrie. De manière générale, on peut même supposer que les variations d'expression des protéines PPR impliquées dans la maturation des ARN permettraient de maintenir une certaine homéostasie des ARN. La quantité d'ARN peut dépendre du nombre de chloroplastes/mitochondries et/ou de copies des génomes des organites, mais cela ne me paraît pas pertinent dans une logique de régulation nécessitant une réponse rapide. Une autre hypothèse qu'il est possible d'avancer est que l'état global de l'organite régule l'expression de son génome. Par état global, j'entends niveau de pH, absence d'effecteur détecté ou encore état d'oxydo-réduction de ses composants par exemple.

Une régulation de l'ensemble du génome ne permettant pas une régulation individuelle, on peut aussi supposer que la régulation n'a pas lieu lors de la transcription ou de la maturation des ARN, mais lors de la traduction. Les protéines PPR pouvant aussi être impliquées dans la traduction, la modulation de l'expression d'un certain nombre d'entre elles en condition de stress biotique est un argument pour continuer les recherches de ce côté. Cependant, les outils techniques permettant d'accéder à la traduction sont nettement plus lourds. On pense par exemple au *ribosome profiling*, c'est-à-dire au séquençage des ARN en cours de traduction, protégés par les ribosomes (Ingolia et al., 2009; Mustroph et al., 2009). Un autre argument en faveur de cette hypothèse est que dans le chloroplaste de *Chlamydomonas reinhardtii*, la traduction de certaines protéines est contrôlée par leur assemblage dans un complexe (Choquet et al., 2001). On parle alors de contrôle par épistasie de synthèse. L'idée générale est que lorsqu'une protéine particulière d'un complexe fait défaut (on parle alors de protéine dominante), les autres protéines du complexe contrôlées par épistasie de synthèse vont voir leur taux de traduction diminuer. Le mécanisme a été particulièrement détaillé pour le cytochrome b6f mais peut être élargi puisque selon les auteurs, chaque complexe de la membrane des thylakoïdes aurait au moins une protéine dont la traduction est contrôlée par épistasie (Choquet et Vallon, 2000; Choquet et al., 2001). Les deux avantages de cette stratégie sont qu'elle réduit les pertes d'énergie dans la synthèse de protéines inutilisées tout en permettant une certaine flexibilité dans l'adaptation face à un environnement changeant (Pogson et al., 2008). Bien que cela soit démontré chez *C. reinhardtii*, aucune donnée chez *A. thaliana* concernant l'épistasie de synthèse via des protéines codées par les organites n'est connue à ce jour. L'épistasie de synthèse peut s'imaginer intra-organite, avec une régulation de la traduction des gènes des organites par des protéines codées dans les organites. Elle peut aussi s'imaginer avec des protéines codées par le noyau et envoyées aux organites, régulant la traduction des protéines codées par les organites.

En résumé, s'il existe une forme de régulation de l'expression génétique des organites en condition de stress, elle peut se faire à différents niveaux, de la transcription à la maturation des protéines, en passant par la maturation des ARN et la traduction (Figure 61).

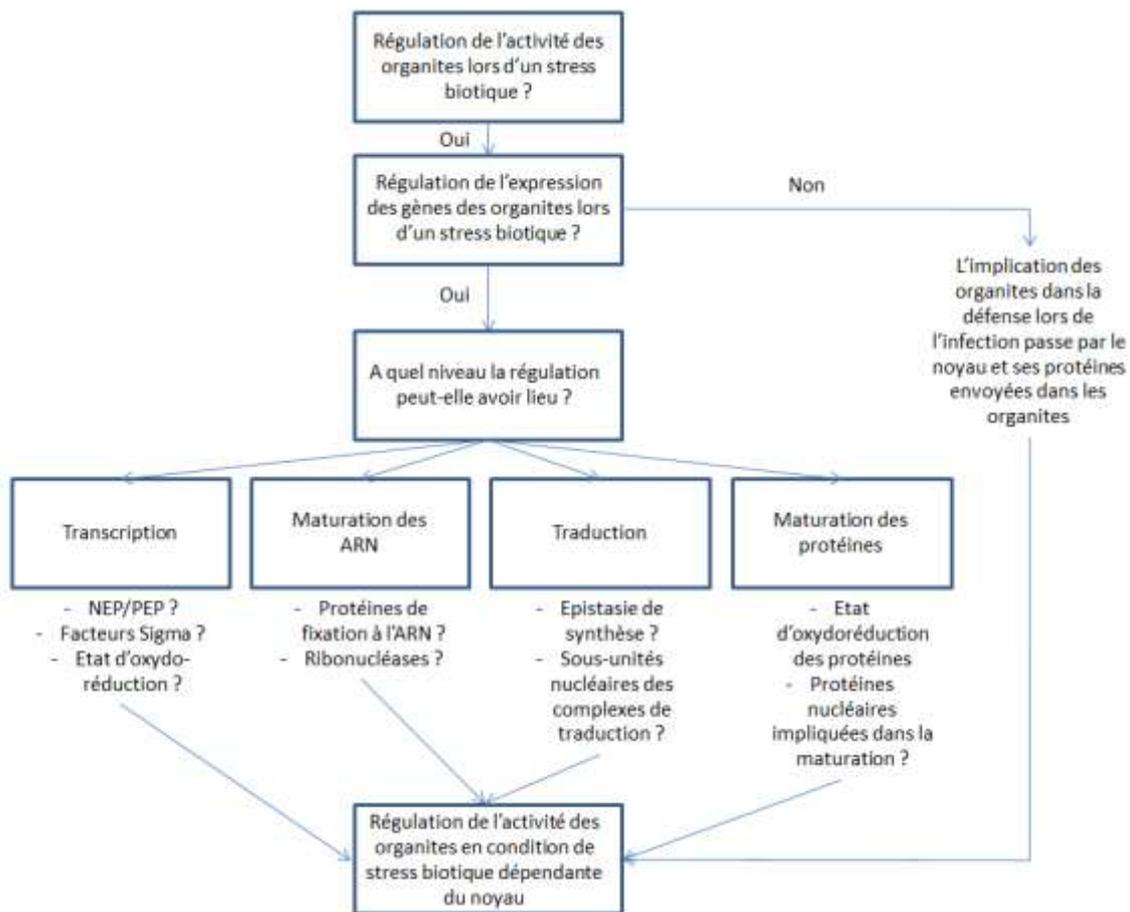


Figure 61 : Résumé des différentes hypothèses de régulation ou d'absence de régulation de l'expression des gènes des organites en condition de stress biotique.

Compte-tenu de la longue demi-vie des ARN des organites, les hypothèses de régulation de la transcription ne m'apparaissent pas être les plus pertinentes en réponse à un stress. L'hypothèse traductionnelle me semble être la plus pertinente à creuser. Parmi les possibilités mentionnées, on remarquera toutefois qu'elles comportent toutes (à l'exception de l'hypothèse d'épistasie de synthèse intra-organite) une protéine régulatrice codée par le noyau. Que l'on parle des facteurs Sigma, des NEP, des protéines PPR, de la plupart des sous-unités des chaînes de transfert des électrons ou des protéines senseurs phosphorylées suite aux variations d'oxydo-réduction, toutes ces protéines sont codées par le noyau. Avec les très nombreux systèmes de régulation de l'expression génétique dans le noyau –méthylation, interférence par ARN, pléiade de facteurs de transcription, etc.-, on imagine facilement une régulation fine de l'expression des gènes codant pour des protéines adressées aux organites qui permettrait une régulation fine de l'activité des organites, et même éventuellement une forme de régulation par épistasie de synthèse via ses protéines si elles venaient à être en sous-nombre. En théorie, la centralisation nucléaire permettrait de coordonner les différentes réponses de la cellule en condition de stress, ce qui présente de nombreux avantages. On peut même émettre l'hypothèse que tous les gènes nécessitant une régulation transcriptionnelle ont été transférés des organites vers le noyau au cours de l'évolution.

Pour rappel, le chloroplaste et la mitochondrie sont le fruit d'évènements d'endosymbiose, c'est-à-dire de l'incorporation de bactéries dans des cellules eucaryotes primitives. Il existe plusieurs hypothèses sur les forces évolutives contrôlant les transferts des gènes des organites vers le noyau. On peut imaginer qu'à travers l'accumulation de mutations et en l'absence de pression de sélection pour conserver un gène donné, celui-ci

aura tendance à disparaître. C'est l'hypothèse du cliquet de Muller présentée en introduction (Muller, 1931). En restant à l'intérieur d'une cellule, l'ancêtre bactérien est passé d'un être vivant se reproduisant de façon sexuée à un être vivant asexué. Il n'est alors plus possible pour lui de se débarrasser par recombinaison des mutations délétères qui risquent alors de s'accumuler et de le voir disparaître. Le noyau possédant de nombreux systèmes de correction des mutations, il est logique d'imaginer un transfert de l'information génétique vers celui-ci afin d'assurer sa pérennité. Les organites étant en plus des compartiments cellulaires chargés en radicaux libres, ça augmente d'autant plus le risque de mutations délétères. Chez les plantes pourtant, les prédictions de Muller sont à nuancer. La polyploïdie des organites à génome couplée aux recombinaisons inter-organites permettent de diminuer le risque d'accumulation de mutations délétères (Martin et Herrmann, 1998). Concernant les gènes restés dans les organites, on peut aussi supposer que dans l'éventualité d'un transfert du gène vers le noyau, les caractéristiques physico-chimiques de la protéine (composition, taille, structure, hydrophobicité, etc.) ne permettent pas son ré-import vers les organites (De Grey, 2005). Une autre hypothèse est celle proposée par Allen (2015). Allen suggère que l'expression des gènes des organites passe par une perception rapide des fluctuations environnementales, et propose que l'état d'oxydo-réduction soit, via des protéines senseurs, à l'origine d'une régulation de la transcription chez les organites. D'un point de vue évolutif, Allen fait une hypothèse suggérant que les gènes conservés dans les organites sont ceux nécessitant une régulation de la transcription, via l'état d'oxydo-réduction de l'organite. Donc nombreuses hypothèses existent quant à la dynamique de transfert de gènes des organites vers le noyau. Bien qu'aucun modèle proposé n'explique toutes les observations listées précédemment, l'hypothèse d'un transfert vers le noyau des gènes des organites nécessitant une régulation transcriptionnelle est celle qui me semble tenir compte du plus grand nombre d'observations. Elle demeure toutefois perfectible, la réponse se trouvant potentiellement à la croisée de plusieurs hypothèses.

Lors de ce travail de thèse, j'ai étudié la régulation de l'expression des gènes des organites en condition de stress biotique. Les mécanismes potentiels de régulation ne manquent pas. Mais avant de chercher à démontrer leur existence, il me semble primordial de trouver d'abord des variations nettes de l'expression des gènes en condition de stress biotique. En ce sens, l'accentuation des efforts de recherche, l'utilisation d'approches globales –*RNAseq* ou *Ribosome Profiling* par exemple-, de la sueur et des larmes permettront peut-être d'affiner notre compréhension du rôle des organites en condition de stress biotique. Cependant, compte-tenu de la longue demi-vie des ARN des organites, l'hypothèse d'une régulation transcriptionnelle ne me paraît pas être la plus pertinente. Je pencherai donc vers une régulation traductionnelle et post-traductionnelle.

Matériels et méthodes.

1. Matériel biologique utilisé lors de la thèse

1.1. Espèces végétales utilisées et conditions de cultures

La majorité du travail de thèse a été réalisée sur la plante modèle *Arabidopsis thaliana* écotype *Columbia* (Col0). Des pathotests ont été effectués sur *Solanum lycopersicum* « Super Marmande », *Malus domestica* « Golden delicious » issus de pollinisation libre et *A. thaliana*. Enfin, les agroinfiltrations effectuées pour déterminer les localisations subcellulaires des protéines fusionnées à des protéines fluorescentes ont été faites sur *Nicotiana benthamiana*.

A. thaliana est cultivée en condition de jours courts (jour : 8h à 20°C ; nuit : 16h à 18°C avec une hygrométrie de 70%) ou jours longs (jour : 16h à 20°C ; nuit : 6h à 18°C avec une hygrométrie de 70%). *Malus domestica* « Golden delicious » est cultivé à 25°C jour, 17°C nuit en éclairage naturel. *Nicotiana benthamiana* est cultivé en condition de jours longs (jour : 16h à 23°C ; nuit : 6h à 18°C avec une hygrométrie de 70%). *Solanum lycopersicum* « Super Marmande » est cultivé selon les conditions précisées par Morel et al. (2018). *Malus domestica* « Golden delicious » est cultivé selon les conditions de Laurent et al. (1987).

1.2. Bactéries utilisées

La souche d'*Escherichia coli* utilisée pour les différents clonages est la souche DH5 α , sensible à la cassette suicide *ccdb* des vecteurs Gateway. Les vecteurs vides Gateway, contenant encore la cassette *ccdb* sont quant à eux multipliés dans la souche DB3.1, résistante.

Afin de transformer de façon stable ou transitoire *Arabidopsis thaliana* et *Nicotiana benthamiana*, nous avons utilisé la bactérie *Agrobacterium tumefaciens*. Responsable d'une maladie chez les plantes appelée galle du collet, cette bactérie est connue pour insérer de l'ADN dans le génome de la plante qu'elle infecte. Nous avons travaillé avec la souche C58C1 (pMP90) avec ou sans le plasmide pSOUP, celui-ci étant nécessaire pour l'utilisation des plasmides pGREEN utilisés pour la localisation subcellulaire.

Pour les pathotests, nous travaillons avec les souches sauvages *Erwinia amylovora* CFBP1430, *Ralstonia solanacearum* GMI1000 et *Pseudomonas syringae* pv. tomato DC3000. Les mutants $\Delta ripL$ de *Ralstonia solanacearum* et $\Delta eamy_0820$ d'*Erwinia amylovora* ont été obtenus respectivement par nos collaborateurs Nemo Peeters et Marie-Anne Barny.

1.3. Champignons utilisés

Pour les pathotests, nous travaillons avec *Botrytis cinerea* B05.10 et *Hyaloperonospora arabidopsidis* Noco2.

1.4. Souches de levure

Pour le double hybride levure, deux souches ont été utilisées : la souche Y8800 (*MAT α*) et la souche Y8930 (*MAT α*).

2. Méthodes

2.1. Préparation de bactéries compétentes

Les bactéries DH5 α utilisées pour les transformations par électroporation sont préparées de la façon suivante :

- Inoculation de 1L de milieu LB (TableauXXVIII) par 3 à 10mL d'une culture de nuit de DH5 α saturée, puis incubation à 37°C avec agitation jusqu'à obtenir une DO₆₀₀ comprise entre 0,5 et 0,6.
- A partir de cette étape, les bactéries sont systématiquement conservées en glace et le matériel a été stocké à -20°C avant utilisation.
- Centrifugation à 4000rpm, 4°C pendant 15min, puis resuspension dans un volume d'eau stérile.
- Centrifugation à 4000rpm, 4°C pendant 15min, puis resuspension dans un demi-volume d'eau stérile.
- Centrifugation à 4000rpm, 4°C pendant 15min, puis resuspension dans 20mL de glycérol 10% stérile.
- Centrifugation à 4000rpm, 4°C pendant 15min, puis resuspension dans 2mL de glycérol 10% stérile.
- Stocker les bactéries en aliquots de 100 μ L à -80°C après «snap-freeze» dans de l'azote liquide.

Les agrobactéries compétentes utilisées sont préparées de la façon suivante :

- Inoculer 48mL de milieu LB par 2mL d'une culture de nuit C58C1 pMP90 (+/-pSOUP) saturée, puis incubation à 30°C avec agitation jusqu'à obtenir une DO₆₀₀ de 0,5.
- A partir de cette étape, les bactéries sont systématiquement conservées en glace et le matériel a été stocké à -20°C avant utilisation.
- Centrifuger à 4000rpm, 4°C pendant 15min, puis resuspension dans 10mL Tris 10mM pH8 préalablement refroidi.
- Centrifuger à 4000rpm, 4°C pendant 10min, puis resuspension dans 10mL LB froid.
- Stocker les bactéries en aliquots de 100 μ L à -80°C après «snap-freeze» dans de l'azote liquide.

Tableau XXVIII : Composition du milieu LB.

Bactotryptone	10g
Extraits de levure	5g
NaCl	5g
Ajuster le pH à 7,5	
Agar	12g
Eau milliQ	qsp. 1L
Autoclaver	

2.2. Génotypage

2.2.1. Extraction d'ADN génomique

Le protocole suivant permet d'extraire de l'ADN génomique d'*Arabidopsis thaliana* :

- Broyer de 50 à 100mg de feuille en plaque 96 puits avec une bille, à l'aide du TissueLyserII de Qiagen, deux fois 30s à 20Hz. Centrifugation 30s à 1000rpm pour culoter la poudre.
- Ajouter 400 μ L tampon AP1 préchauffé à 65°C, vortexer 15s et centrifuger 1min à 3000rpm.
- Ajouter 130 μ L tampon AP2, vortexer 15s puis centrifuger 1min à 3000g.
- Incuber 10min à -20°C pour précipiter les protéines puis centrifuger 10min à 4000rpm.
- Transférer 400 μ L de surnageant dans une nouvelle plaque.

- Ajouter 600µL de tampon AP3, vortexer 15s et centrifuger 1min à 3000rpm.
- Transférer 750µL de surnageant dans la plaque Whatman GF/B puis sceller avec un film laissant passer les gaz.
- Centrifuger 5min à 4000rpm.
- Ajouter 750µL de tampon AW/E, centrifuger 15min à 4000rpm, puis transférer la plaque Whatman GF/B sur plaque 96 pour l'élution.
- Ajouter 100µL de tampon AE, incubé 1min à température ambiante puis centrifuger 1min à 4000rpm.
- Stocker l'ADN génomique extrait à -20°C.

Les tampons sont préparés de la façon suivante :

Pour la RNase A :

- Dissoudre dans l'eau distillée stérile la RNase A (sigma R5125) à 100mg/mL.
- Bain marie d'eau bouillante : 20min.
- Conserver à -20°C.

Pour les différents tampons, les compositions sont données dans les tableaux XXIX à XXXIII.

Tableau XXIX: Composition d'AP1 : tampon de lyse, dénaturation des protéines et lyse des parois cellulaires.

AP1 (500mL)		
	Vol. (mL)	Masse (g)
Tris 1M pH8.0	50	
EDTA 0.5M	50	
NaCl 2.5M	100	
SDS 20%	31.5	
PVP40		5
Bisulfite de sodium		5
H ₂ O	qsp. 500mL	

Ajouter 1mL de RNase A pour 400mL de tampon AP1, stocker à -4°C.

Tableau XXX : Composition d'AP2 : précipitation des protéines et des polysaccharides par acidification du milieu.

AP2 (200mL)	
	Vol. (mL)
Acétate de Potassium 5M	120
Acide Acétique Glacial	23
H ₂ O	qsp. 200mL

Tableau XXXI : Composition d'AP3 : constitution des sels d'ADN.

AP3 (150mL)		
	Vol. (mL)	Masse (g)
Guanidine Chloride		37.25
Ethanol 96%	100	
H ₂ O	qsp. 150mL	

Dissoudre le chlorure de guanidium dans l'H₂O jusqu'à un volume final de 50mL, puis ajouter l'éthanol.

Tableau XXXII: Composition d'AW : dessaler et laver l'ADN.

AW (1L)	
	Vol. (mL)
Tris 1M pH8.0	22.5
EDTA 0.5M	0.2
Acétate de Potassium 5M	32
H ₂ O	qsp 1L

Le tampon AW/E = 100mL AW + 170mL Ethanol 96%.

Tableau XXXIII: Composition d'AE : tampon d'éluion.

AE (200mL) Tris 10mM ; EDTA 0.5mM	
	Vol. (mL)
Tris 1M pH8.0	2
EDTA 0.5M	0.2
H ₂ O	qsp. 200mL

2.2.2. PCR de géotypage

L'ADN génomique extrait à l'aide des plaques Whatman GF/B est suffisamment pur pour que l'on puisse utiliser avec réussite une Taq polymérase que nous avons produite nous-même. On suit le protocole de PCR détaillé dans le tableau XXXIV :

Tableau XXXIV : Protocole PCR et composition de la réaction pour le géotypage

ADN génomique	1µL
Amorce sens 10µM	0,6µL
Amorce antisens 10µM	0,6µL
Tampon Taq 10x	1µL
Taq maison	0,25µL
10mM dNTP	0,2µL
Betaine	3µL
BSA 100mg/mL	1µL
H ₂ O qsp 10µL	2,35µL
Total	10µL

Programme

95°C	2min	} 35 cycles
95°C	15s	
50-65°C	15s	
68°C	1min/kb	
68°C	5min	
12°C	∞	

2.2.3. Migration sur gel d'agarose

La séparation des molécules d'ADN par taille, après PCR, se fait par électrophorèse sur gel d'agarose. Les gels sont préparés d'après les informations des tableaux XXXV et XXXVI :

Tableau XXXV : Préparation d'un gel d'agarose.

Agarose	0,8 à 2%
TAE 1X	de 20 à 200mL
Faire fondre l'agarose au micro-onde	
Laisser refroidir	
Bromure d'éthidium	1 μ L/100mL
Couler le gel avec le peigne	

Tableau XXXVI : Composition du tampon TAE.

Tris	89mM
Acide acétique	89mM
EDTA	2mM
Ajuster à pH8	

Le gel est ensuite placé dans une cuve à électrophorèse remplie de tampon TAE 1X. L'ADN est mélangé à un tampon de charge, le rouge de crésol, afin de faciliter le dépôt et la visualisation des échantillons. On utilise également un marqueur de poids moléculaire afin de servir d'échelle et d'estimer la taille des fragments d'ADN (1kb DNA Ladder, New England Biolabs). On applique ensuite un champ électrique, entre 50 et 150V afin d'effectuer la migration. La révélation se fait sur table UV.

2.3. Clonage et transformation de plantes

2.3.1. PCR de clonage

On utilise pour les PCR de clonage une Taq haute-fidélité afin de limiter au maximum les mutations. On utilise la Taq iProof High-Fidelity de chez Biorad et les informations nécessaires sont résumées dans le tableau XXXVII :

Tableau XXXVII : Programme PCR et composition de la réaction pour le clonage.

ADN	1 μ L	Programme	
Amorce sens 10 μ M	2,5 μ L	98°C	30s
Amorce antisens 10 μ M	2,5 μ L	98°C	10s
Tampon 5X Biorad	10 μ L	50-65°C	30s
Taq	0,5 μ L	72°C	1min/kb
dNTP	1 μ L	72°C	7'
H ₂ O qsp 10 μ L	32,5 μ L	12°C	∞
Total	50 μ L		

} 35 cycles

Dans certains cas de figure, on souhaite pour un même insert avoir seulement un gène codant pour une protéine, ou une fusion, avec une étiquette poly-histidine ou une protéine fluorescente. Il faut donc faire une PCR avec un insert dont la séquence codante se termine avec un codon stop dans le premier cas, et sans codon stop dans le second. En pratique, cela peut se faire en modifiant une seule paire de bases. Dans ce cas de figure, ce qui a été fait par soucis de simplicité est de réaliser seulement l'un des deux clonages, et d'effectuer une mutagenèse ponctuelle dans le second cas. Les amorces de PCR sont construites de façon à ce que la base mutée se situe au centre des deux amorces, qui vont être complémentaires. On effectue ensuite une PCR à

l'identique de celle que l'on vient de présenter, mais sur seulement 20 cycles. On effectue ensuite une digestion de 5µL du produit PCR avec 0,5µL d'enzyme Dpn1 en incubant 1h à 37°C. Cette enzyme digère l'ADN méthylé donc va permettre la digestion de l'ADN plasmidique qui n'est pas muté. Après transformation dans des bactéries *E. coli* DH5α, celles-ci vont recombiner l'ADN linéaire obtenu par PCR avec les parties chevauchantes afin d'obtenir un plasmide muté.

2.3.2. Précipitation PEG

Afin d'éliminer les dimères d'amorces et purifier les produits PCR de clonage, on effectue une précipitation PEG (pour Polyéthylène Glycol). On mélange 45µL du produit PCR avec 135µL de tampon TE (Tris EDTA), 90µL de PEG/Mg (30% PEG 8000, 30mM MgCl₂) et 1µL de glycogène azur afin de faciliter la précipitation et colorer le culot. On centrifuge à 20000rpm pendant 15min, puis rapidement on élimine le surnageant et on resuspend dans 20µL de TE.

2.3.3. Réaction BP

La recombinaison BP (Invitrogen) permet de recombiner les séquences *attB1/attB2* présentes aux extrémités des inserts obtenus par PCR avec les séquences des sites *attP1/attP2* présents dans le plasmide pDONR207. Entre *attP1* et 2 se trouve un gène codant pour CCDB, protéine perturbant l'activité de l'ADN gyrase des *E. coli* souche DH5α et empêchant la réplication bactérienne. Le pDONR207 possède un autre marqueur de sélection, un gène permettant la résistance à la gentamycine. En sélectionnant sur boîte avec de la gentamycine, on s'assure en théorie de ne permettre la pousse que des bactéries DH5α contenant un pDONR207 sans le gène codant pour CCDB, donc potentiellement un insert. Cette réaction catalysée par l'enzyme BP permet d'obtenir un plasmide pDONR207 contenant l'insert, flanqué des séquences *attL1* et 2. La composition de la réaction est donnée dans le tableau XXXVIII.

Tableau XXXVIII : Composition de la réaction BP.

50ng de pDONR207 vide
2µL de précipitation PEG
1µL de tampon BP
0,4µL d'enzyme BP
TE qsp. 5µL

Cette réaction s'effectue à 25°C pendant minimum une heure, dans l'idéal sur la nuit.

2.3.4. Traitement à la protéinase K

La protéinase K est une enzyme qui va avoir pour rôle de digérer les protéines présentes dans la réaction BP une fois terminée. Cette étape permet d'augmenter significativement le rendement de transformation. On utilise 5µL de réaction BP et 0,5µL de protéinase K. On incube 10min à 37°C puis on place le tube en glace avant la transformation en *E. coli* DH5α.

2.3.5. Transformation avec *E. coli*

Comme vu précédemment, nous travaillons avec des bactéries *E. coli* DH5α électrocompétentes. Les bactéries sont sorties du congélateur -80°C et placées en glace 30min environ jusqu'à décongélation. Les cuves d'électroporation de 2mm de diamètre sont également placées en glace. On ajoute les 5µL de réaction BP à 100µL de bactéries compétentes, et on mélange en tournant l'extrémité du cône dans le tube. On incube en glace 1min, on transfère dans la cuve et on électropore à l'aide du Biorad MicroPulser (paramètre Ec2, avec dans l'idéal un temps de passage compris entre 4 et 5ms). On ajoute ensuite 900µL de milieu LB. La vitesse

de cette étape est extrêmement critique, la moindre seconde de perdue entraîne une perte en efficacité de transformation assez importante. Les bactéries sont ensuite incubées pendant une heure à 37°C pour ensuite être étalées sur boîte LB avec de la gentamycine toute la nuit à 37°C.

2.3.6. PCR et cultures liquides de bactéries

On réalise ensuite une PCR sur les colonies ainsi obtenues. Les colonies sont prélevées à l'aide d'un cône 10µL stérile. Ce cône est trempé quelques secondes dans 20µL d'eau stérile puis dans 5mL de milieu LB liquide avec de la gentamycine. Ce mélange eau et bactéries servira à la PCR sur colonies avec les amorces DNR5 et 3 sur le pDNOR207 et bordant la zone d'insertion. Le protocole de PCR est le même que celui vu précédemment pour le génotypage de plantes, à la seule différence que le temps de dénaturation initial est de 10min afin de permettre la lyse des bactéries. La culture liquide est incubée sur la journée ou la nuit avec agitation à 37°C. Si la pousse est bonne, le volume de 5mL est idéal, 4mL serviront à l'extraction d'ADN plasmidique, et le dernier mL à faire un stock glycérol.

2.3.7. Stock glycérol

Les stocks glycérols sont faits en mélangeant 1,2mL de culture liquide bactérienne dans du milieu LB du jour, avec 600µL de glycérol 50% stérile. Après vortexage, les tubes sont stockés à -80°C.

2.3.8. Extraction d'ADN plasmidique (Miniprep)

On culotte 2 à 4mL de culture liquide DH5α, on élimine le surnageant et l'on resuspend dans 600µL de TE. On utilise ensuite le kit PureYield Plasmid Miniprep System de Promega, en suivant le protocole fournisseur.

2.3.9. Séquençage des plasmides

Le séquençage des plasmides se fait sur 15µL de la miniprep à environ 100ng/µL, après dosage au NanoDrop 1000 (Thermo Scientific). On ajoute 2µL d'amorce diluée à 10µM et le séquençage est effectué par la société Eurofins Genomics.

2.3.10. Réaction LR

Une fois les résultats de séquençage obtenus, on effectue une recombinaison LR. Lors de cette réaction, l'insert présent dans le pDONR207 va être transféré dans un plasmide de destination ou pDEST. Les séquences *attL1* et 2 vont être recombinées avec les sites *attB1* et 2 des plasmides de destination. La composition de la réaction de LR est résumée dans le tableau XXXIX.

Tableau XXXIX : Composition de la réaction LR.

100ng de pDONR207 avec l'insert
100ng de pDEST
1µL de tampon LR
0,4µL d'enzyme LR
TE qsp. 5µL

La réaction se fait à 25°C, pendant une heure minimum, dans l'idéal sur la nuit.

On effectue ensuite une transformation *E. coli* comme vu précédemment, avec ensuite génotypage, miniprep mais sans séquençage, puisqu'il paraît assez improbable qu'une mutation puisse apparaître entre le séquençage du pDONR207 et la LR.

2.3.11. Transformation avec *A. tumefaciens*

Afin d'insérer des transgènes dans *Arabidopsis thaliana* de façon stable ou pour réaliser des agroinfiltrations, on transforme *Agrobacterium tumefaciens* avec les plasmides de destination. A l'aide d'agrobactéries compétentes, on réalise le protocole suivant : 10µL de miniprep sont mélangés à 100µL de bactéries. Après passage dans l'azote liquide, on incube 5min à 37°C, on ajoute 900µL de milieu LB liquide puis on incube 3h à 30°C avec agitation. On étale ensuite sur boîte avec les antibiotiques nécessaires et on incube à 30°C pendant 3 jours. Les colonies sont testées de la même façon que les colonies DH5α.

2.3.12. Transformation d'*A. thaliana* par la méthode goutte

Une fois les agrobactéries transformées sélectionnées, on va transformer *Arabidopsis thaliana* par la méthode de la goutte, avec le protocole suivant :

- Strier les agrobactéries sur boîte LB avec antibiotiques et incuber pendant deux jours à 30°C.
- Effectuer une culture, à partir d'une colonie isolée de la boîte, dans 5mL de milieu LB avec antibiotiques à 30°C sur la nuit avec agitation.
- Doser la culture le lendemain et centrifuger pendant 10min à 4000rpm le volume nécessaire afin d'avoir 2mL à $DO_{600}=1$ après resuspension.
- Eliminer le surnageant et resuspendre dans 2mL de sucrose 5% fraîchement préparé.
- Ajouter du Silwett L77 à 0.05% final.
- En serre, on transforme des au stade boutons floraux, soit entre 5 et 8 semaines. On élimine les siliques déjà formées et les fleurs.
- A l'aide d'une pipette 200, déposer une goutte sur chaque bouton floral.
- Couvrir pendant 24 à 48h, puis recommencer l'opération une semaine plus tard, cette fois-ci sans couper les siliques et les fleurs.
- Récolter les graines.

2.3.13. Sélection BASTA

Les construits pGREEN sont sélectionnés grâce à une résistance herbicide (BASTA, glufosinate d'ammonium). On sème sur terreau les graines récoltées après la méthode goutte et à 10 jours, on traite 3 fois au BASTA à 300µM tous les deux jours. On teste en PCR de génotypage les plantes résistantes après extraction d'ADN génomique.

2.3.14. Sélection hygromycine

D'autres construits, comme les pGWB13 et 14, permettent à la plante d'être résistante à un antibiotique, l'hygromycine. On stérilise 500µL de graines récoltées après la méthode goutte en les incubant 5min dans une solution de 8 volumes d'éthanol 70% et 2 volumes de javel. On effectue ensuite 2 lavages rapides dans de l'éthanol 70% puis un dernier dans de l'éthanol 100%. On laisse l'éthanol s'évaporer sous hotte à flux laminaire et on sème sur boîte ½ MS 1% sucrose 25µg/mL d'hygromycine (Tableau XL). On conserve les boîtes 2 jours à 4°C et à l'obscurité, puis on les expose 6h à la lumière en chambre de culture *in vitro*. Les 5 jours qui suivent, les boîtes sont conservées en chambre *in vitro* mais à l'obscurité. Les transformants seront les plantes étiolées. On les transfère ensuite en terre, avec hygrométrie saturante pendant 2 à 3 jours. Les plantes ayant survécu seront testées comme vu précédemment.

Tableau XL : Composition du milieu ½ MS.

MS (Sigma)	0.66g
Sucrose	3g
H ₂ O qsp. 300mL	
Ajuster le pH à 5,7 KOH	
Bacto agar	2,4g
Autoclaver	

2.3.15. Agroinfiltration de *Nicotiana benthamiana*

Les agroinfiltrations se font sur la plante modèle *Nicotiana benthamiana*. Elle doit être âgée de 3 à 4 semaines et doit être agroinfiltrée avant qu'elle n'ait de fleurs afin de limiter les risques d'interférence par ARN. La veille de l'agroinfiltration, on lance une culture d'agrobactéries transformées avec un plasmide contenant notre gène d'intérêt, dans 10mL de milieu LB liquide à 30°C avec agitation. Le matin, on dose la culture, on culote, élimine le surnageant et on resuspend dans du milieu MMA afin d'obtenir une DO₆₀₀ comprise entre 0,2 et 0,3. On incube à l'obscurité pendant 2 à 3h et on infiltre la face subapicale d'une feuille avec la solution, à l'aide d'une seringue sans aiguille. Les observations peuvent être faites 3 à 5 jours après agroinfiltration.

2.4. Microscopie

Les observations en microscopie à épifluorescence ont été effectuées sur un Nikon Eclipse TE-2000-S. Les observations en microscopie confocale ont été effectuées sur un Zeiss LSM 880 par Séverine Domenichini, responsable du plateau imagerie de l'IPS2.

2.5. Transcriptomique

2.5.1. Extraction d'ARN

Les extractions d'ARN sont effectuées avec du Nucléozol (Macherey-Nagel), et le protocole suivi est celui du fournisseur (version du 02/16 rev. 02). A noter que l'étape facultative au 4-bromoanisole est systématiquement faite car elle permet entre autres d'éliminer une part de l'ADN génomique restant. Les ARN sont ensuite stockés à -20°C ou conservés en glace.

2.5.2. Traitement DNase

Afin d'éliminer ce qu'il peut rester d'ADN génomique avec les ARN, on effectue un traitement DNase à l'aide du kit DNase Max (Mo Bio), en suivant le protocole fournisseur. Les ARN traités sont ensuite stockés à -20°C ou conservés en glace. On vérifie l'efficacité du traitement DNase par PCR, de la même façon que l'on réalise les PCR d'éditotype, à la seule différence qu'elle est faite sur 10µL de volume total avec 1µL d'ARN traité. La PCR d'éditotype sera détaillée un peu plus bas.

2.5.3. Construction de banques de séquençage *RNAseq*

La qualité des ARN est testée à l'aide de puces Agilent RNA Nano Chip et de l'Agilent 2100 Bioanalyser, en suivant le protocole fournisseur. La construction des différentes banques de séquençage a été réalisée par la plateforme transcriptomique POPS, à l'IPS2.

2.5.3.1. Pour les ARN totaux

La construction de la banque de séquençage a été réalisée à l'aide du kit et en suivant le protocole fournisseur Illumina Truseq Stranded Total RNA avec le kit de déplétion Ribo-Zero Plant. Le séquençage est effectué sur NextSeq 500 (Illumina) avec des reads single-end de 75 paires de bases en moyenne.

2.5.3.2. Pour les petits ARN

La construction de la banque de séquençage a été réalisée à l'aide du kit et en suivant le protocole fournisseur Illumina Truseq Small.

2.5.4. Rétro-transcription

Les rétro-transcriptions sont effectuées de la façon suivante :

- Mélanger 5µL d'ARN à 100ng/µL à 0,5µl de dNTP à 40mM et à 0,5µl d'oligo dT si l'on souhaite regarder par la suite seulement l'expression de gènes nucléaires, sinon 0,5µL de *random primers*.
- Incuber 5min à 65°C puis 2min à 4°C.
- Rajouter au précédent mélange 2µL de tampon 5X FS, 1µL de DTT 0,1M, 0,5µL de RNase OUT et 0,5µL de SS II.
- Incuber 50min à 42°C, puis 15min à 70°C et ensuite placer les tubes en glace.

2.5.5. PCR d'éditotype

Les mesures de niveau d'édition des ARN des organites se font sur de l'ADNc obtenu avec des *random primers*. Suivant le niveau d'expression des gènes étudiés, l'ADNc pourra être plus ou moins dilué pour la PCR. Le protocole est résumé dans le tableau XLI.

Tableau XLI: Programme et composition de la réaction pour la PCR d'éditotype.

ADN	2µL	98°C	2min	} 35 cycles
Amorce sens 10µM	0,6µL	98°C	10s	
Amorce antisens 10µM	0,6µL	50-65°C	15s	
Tampon Terra 2x	10µL	68°C	1min/kb	
Taq Terra	0,13µL	68°C	5'	
H ₂ O qsp 10µL	6.7µL	12°C	∞	
Total	20µL			

Il est à noter que le volume de 20µL est idéal pour cette PCR. En effet, 2µL vont être déposés sur gel afin de vérifier si la PCR a bien fonctionné, et le volume restant servira au séquençage, effectué également par Eurofins Genomics, sans étape de purification supplémentaire.

2.5.6. PCR quantitative (qPCR)

Les qPCR sont effectuées, comme les PCR d'éditotype, sur de l'ADNc, obtenu à l'aide d'oligo dT ou de *random primers*. On suit le protocole suivant (Tableau XLII), en plaque 384 :

Tableau XLII: Programme et composition de la réaction de qPCR.

ADN	2 μ L	95°C	1min	} 40 cycles
Amorce sens 10 μ M	0,125 μ L	95°C	3s	
Amorce antisens 10 μ M	0,125 μ L	50-65°C	20s	
SyBR Premix Ex Taq	2.5 μ L	65°C	5s	
H ₂ O qsp 10 μ L	0.25 μ L	95°C	+0,5°C/cycle	
Total	5 μ L			

On travaille en triplicats techniques pour chaque échantillon, avec analyse de la courbe de *melting* post-amplification. Les données aberrantes (courbe de *melting* ne correspondant pas à un produit d'amplification, valeur de Ct très éloignée des autres répliquats techniques) sont éliminées. On utilise le SyBR Premix Ex Taq de chez TaKaRa, avec la machine CFX384 real-Time System de chez Biorad.

2.6. Pathotest

2.6.1. *Ralstonia solanacearum*

Toute manipulation du pathogène *Ralstonia solanacearum* doit se faire dans un laboratoire avec un niveau de sécurité biologique de 3 minimum (NSB3). Les infections ont été faites à Toulouse au LIPM sur *Solanum lycopersicum* et *Arabidopsis thaliana* en suivant le protocole suivant :

- Lancer des cultures de *Ralstonia* de 100mL environ sur la nuit en milieu minimum, à 28°C avec agitation.
- Mesurer la DO₆₀₀ le lendemain. Il faut 1L de culture à 0.05 de DO (c'est-à-dire 5.10⁷ ufc/mL, une DO₆₀₀ de 1 correspondant à 10⁹ ufc/mL), avec dilution dans de l'eau du robinet conservée à température ambiante. Faire les dilutions dans une éprouvette. A l'aide d'une seconde éprouvette, homogénéiser la solution.

2.6.1.1. Pour *Solanum lycopersicum*

- Arroser le plateau carré de tomates avec 500mL de culture en suivant une ligne sur chaque bord, afin que chaque angle du plateau soit imbibé de la solution. Faire cela sur les 4 bords et garder les 100 derniers mL environ pour les mettre au milieu du plateau. Placer les plantes en chambre de culture.

2.6.1.2. Pour *Arabidopsis thaliana*

- Eliminer les *Arabidopsis* les plus abimées.
- En placer 3x4 sur un plateau carré (les *Arabidopsis* sont cultivées dans des Jiffy-7).
- Remplir le plateau avec 500mL de culture bactérienne sans trop arroser les plantes directement.
- Laisser imbiber 10min. Sortir les plantes du plateau et les placer sur du papier aluminium à côté.
- Prendre du terreau et placer une dose par plateau carré directement sur la solution bactérienne.
- Etaler le terreau et aplanir au maximum, puis poser proprement les plants d'*Arabidopsis* directement dessus.
- Placer les plantes en chambre de culture.

Pour les deux plantes, une lecture des symptômes est réalisée tous les jours ou deux jours.

2.6.2. *Erwinia amylovora*

De la même façon que *Ralstonia*, *Erwinia amylovora* doit être manipulé au minimum dans un laboratoire NSB3, à Versailles ou à Angers.

2.6.2.1. Sur *A. thaliana*

Les infections sont réalisées sur des plantes *Arabidopsis* de 5 semaines cultivées en condition jours courts, dans du terreau avec des copeaux de liège pour que les feuilles ne soient pas en contact directement avec le terreau. Le protocole suivi est le suivant :

- Lancer une préculture d'*Erwinia amylovora* en milieu liquide LB le matin, à 28°C avec agitation, puis le soir, étaler la préculture sur 4 boîtes LB et incuber à 28°C sur la nuit.
- Le lendemain matin, ajouter 500µL d'eau stérile par boîte et racler avec une étaleur stérile afin de collecter les bactéries.
- Doser une dilution au 1/10^{ème} des bactéries collectées et préparer environ 50mL d'une solution d'eau stérile avec les bactéries à une DO₆₀₀ de 1.
- Infiltrer la face subapicale des *Arabidopsis* à l'aide d'une seringue sans aiguille.
- Prélever 24h après infiltration les feuilles infectées dans un tube eppendorf 2mL avec une bille et ajouter 500µl d'eau stérile. Broyer à l'aide d'un mélangeur à peinture jusqu'à obtenir un broyage homogène pour tous les tubes. Pour faire de la transcriptomique, ne pas ajouter les 500µL d'eau stérile et placer directement les tubes dans l'azote liquide puis à -80°C.
- Effectuer des dilutions en série d'un facteur 10 et déposer sur boîte LB avec antibiotique.
- Incuber 2 jours à 28°C puis effectuer les comptages.

2.6.2.2. Sur *Malus domestica* « Golden delicious »

Les pathotests sur pommier ont été effectués par nos collaborateurs Marie-Noëlle Brisset et Matthieu Gaucher à l'IRHS à Angers. Le protocole suivi est celui de Laurent et al. (1987), les seules différences étant les conditions de culture des pommiers. Celles employées ont été décrites précédemment.

2.6.3. *Botrytis cinerea*

On infecte des plantes *Arabidopsis* saines de 4 semaines en jours courts, en suivant le protocole suivant :

- Cultiver *Botrytis cinerea* souche B05.10 sur boîte PDA pendant 10 à 15 jours, à température ambiante et à la lumière
- Déposer 5mL de milieu liquide ½ PDB sur boîte et racler avec un étaleur stérile avec de collecter les spores.
- Filtrer avec du Miracloth et compter les spores sur cellule de Thoma.
- Ajuster la concentration à 5.10⁴ spores/mL avec du ½ PDB.
- Déposer 5µL au centre de la face apicale de la feuille d'*Arabidopsis* et incuber en chambre de culture pendant 2 à 3 jours.
- Effectuer la mesure du diamètre de la lésion.

La préparation du milieu PDA consiste à faire bouillir 200g de pommes de terre coupées en morceaux dans de l'eau. L'ensemble est ensuite filtré sur coton cardé, puis on rajoute 20g de glucose, 20g d'agar, on ajuste le volume à 1L et on autoclave 20min à 120°C. Le milieu ½ PDB est réalisé de la même façon, en ajustant le volume à 2L et sans ajout d'agar.

2.6.4. *Pseudomonas syringae*

On infecte des plantes *Arabidopsis* saines de 4 semaines en jours courts, en suivant le protocole suivant :

- Lancer une culture liquide sur la nuit de *Pseudomonas syringae* dans du LB liquide à 30°C avec agitation.
- Le lendemain, doser la culture jusqu'à atteindre une DO₆₀₀ comprise entre 0,4 et 0,6.
- Centrifuger à température ambiante 1min à 4000rpm puis resuspendre dans une solution de MgCl₂ 10mM afin d'atteindre une DO₆₀₀ de 0,01.
- Infiltrer la face subapicale des feuilles d'*Arabidopsis* à l'aide d'une seringue sans aiguille.
- 48h plus tard, prélever les feuilles infectées à l'aide d'un emporte-pièce. Afin de lisser la variabilité biologique, un échantillon est constitué de 3 disques de 3 feuilles de 3 plantes différentes de même génotype.
- Placer les disques 10s dans de l'éthanol 70%, puis 10s dans de l'eau stérile, sécher sur sopalin et placer dans un tube eppendorf 1,5mL préalablement pesé
- Peser les échantillons
- Broyer les échantillons à l'aide d'un pilon stérile et ajouter 150µL de MgCl₂ 10mM.
- Réaliser 8 dilutions en série d'un facteur 10 et déposer ces dilutions sur boîte LB
- Incuber 3 jours à 28°C et effectuer les comptages d'UFC (Unités Formant des Colonies).

2.6.5. *Hyaloperonospora arabidopsidis*

On suit le protocole suivant afin d'effectuer le pathotest :

- Semer 10 pots d'*Arabidopsis* Col0 sur terreau en jours courts, dans des pots ronds de 5cm de diamètre, 5mg de graines par pot.
- 10 jours après le semis, on effectue la première infection. Afin d'activer les spores, deux cycles d'infection sur Col0 sont nécessaires avant d'infecter les lignées à tester.
- Décongeler des cotylédons infectés, stockés à -80°C dans un tube Falcon 50mL, sur glace pendant 2 à 3h.
- Ajouter 5mL d'eau stérile et vortexer pendant 1min.
- Filtrer les cotylédons sous hotte à flux laminaire avec compresses et pinces stériles.
- Compter les spores sur cellule de Nageotte et diluer dans de l'eau stérile afin d'atteindre une concentration de 5.10⁴ spores/mL.
- Pulvériser 2mL par pot de façon homogène.
- Incuber les pots en chambre de culture jours courts, sous cloche avec hygrométrie saturante, pendant 24h.
- Entre-ouvrir la cloche pendant 3 jours puis sprayer de l'eau stérile à l'intérieur de la cloche et directement sur les plantes le 4^{ème} jour. Refermer la cloche.
- Au 7^{ème} jour, récolter les spores des plantes infectées. A l'aide de ciseaux stérilisés à la flamme, couper les cotylédons à la base sans prendre de terre, les placer dans un falcon 50mL, ajouter 5mL d'eau stérile et vortexer 1min. Filtrer, doser et diluer comme vu précédemment et sprayer à nouveau des pots de Col0 de 10 jours préparés de la même façon. Réaliser un second cycle d'infection sur Col0.
- Après récolte des spores du second cycle d'infection, stocker quelques pots à -80°C dans un falcon 50mL puis infecter les génotypes à tester de la même façon.
- A 7 jours, récolter les cotylédons à raison de un falcon 50mL par pot infecté et peser la matière fraîche. Ajouter 5mL d'eau stérile, vortexer 1min et filtrer comme vu précédemment.

- Dénombrer le nombre de spores par tube et rapporter cette valeur par rapport à la masse de matière fraîche.

2.7. Double Hybride levure

2.7.1. Préparation de levures compétentes

- Inoculer 20mL de YEPD (Tableau XLIII) dans un erlenmeyer de 2L avec une oese d'un étalement de levure sur boîte YEPD solide. Incuber à 30°C avec agitation sur la nuit.
- Le lendemain matin, ajouter 80mL de YEPD et incuber à 30°C avec agitation pendant une à deux heures.
- Faire chauffer le Salmon Sperm DNA (Sigma) pendant 5min dans un bain-marie à 95°C puis le conserver en glace.
- Centrifuger 10mL de la culture de levures 5min à 3000rpm et jeter le surnageant.
- Resuspendre délicatement le culot dans 1mL d'eau stérile.
- Centrifuger 5min à 3000rpm et jeter le surnageant.
- Resuspendre le culot dans 1mL de TE/LiAc, puis centrifuger 5 min à 3000rpm. Jeter le surnageant.
- Resuspendre le culot dans 200µL de TE/LiAc, puis ajouter 1 mL de TE/LiAc/PEG et 20µL de Salmon Sperm DNA et mélanger par inversion.

Tableau XLIII : Compositions respectives des solutions de TE/LiAc ; TE/LiAc/PEG et YEPD.

TE/LiAc	8 volumes d'eau stérile	
	1 volume de TE 10X	
	1 volume LiAc 1M	
TE/LiAc/PEG	8 volumes PEG 3350 44% stérile	
	1 volume TE 10X	
	1 volume LiAc 1M	
YEPD	Extraits de levure	10g
	Bactopeptone	20g
	Glucose	20g
	Agar pour milieu solide	20g
	H ₂ O milliQ	qsp. 1L
	Autoclaver	

2.7.2. Transformation de levures

Ces transformations se font après LR, avec comme plasmides de destination les pDEST-AD (pour la souche Y8800, *MAT_a*) et pDEST-DB (pour la souche Y8930, *MAT_α*). On suit le protocole suivant :

- Mélanger par pipetage 120µL de levures compétentes avec 10µL de plasmide à 50ng/µL.
- Incuber 30min à 30°C.
- Effectuer un choc thermique de 15min à 42°C en bain-marie.
- Centrifuger 5min à 3000rpm et jeter délicatement le surnageant en pipetant.
- Resuspendre les culots dans 100µL d'eau stérile puis centrifuger à nouveau 5min à 3000rpm.

- Jeter 90µL du surnageant à la pipette et resuspendre par agitation rapide le culot avec les 10µL restants.
- Déposer 5µL par transformation sur boîte Sc-L (Tableau XLIV) pour sélection les *MATα* et Sc-W pour les *MATα*. Incuber les boîtes 72°C à 30°C.
-

Tableau XLIV: Composition du milieu Sc.

Pour 1L	
Mélange d'acides aminés	2.6g
Yeast Nitrogen Base	3.4g
Sulfate d'amonium	10g
Glucose	20g
Ajuster pH à 5.9 avec NaOH 10N	
Agar pour milieu solide	20g
H ₂ O milliQ	qsp. 1L
Autoclaver	

Dans un litre de milieu Sc, ajouter 8mL de Trp (40mM), 8mL d'Ura (20mM), 8mL d'His (100mM) et 15mL d'Ade (65mM) pour obtenir du Sc-L.

2.7.3. Lyse des levures et PCR

Avant d'effectuer une PCR de génotypage sur les levures, il est nécessaire de les lyser. Pour cela, on suit le protocole suivant :

- Piquer une colonie de levure et la resuspendre dans 15µL de tampon de lyse (Tableau XLV).
- Incuber 15min à 37°C puis 5min à 95°C.
- Ajouter 100µL d'eau stérile, centrifuger 10min à 3000rpm et effectuer une PCR de génotypage en suivant le protocole des PCR sur colonies bactériennes.
-

Tableau XLV: Composition du tampon de lyse.

Tampon de lyse	Zymolase 20T	2.5mg/mL
	Tampon sodium phosphate pH7,4	0,1M

2.7.4. Cribles contre les banques AD et DB

Les cribles sont effectués par Dario Monachello, sur la plateforme InterATOME. Le protocole suivi est celui de Dreze et al. (2010) mais adapté pour pouvoir réaliser les cribles en cultures liquides (Guillaumot et al., 2017).

2.8. Analyses bio-informatiques et logiciels utilisés.

Les cartes de plasmides sont analysées avec le logiciel ApE V.2.0.50. Les différents logiciels utilisés dans l'analyse de données *RNAseq* sont détaillés dans le chapitre consacré. On peut cependant ajouter que les données qui en ressortent ont été analysées avec le logiciel MapMan V.3.6.0RC1. Les analyses d'enrichissement GO (*Gene Ontology*) ont été faites sur ThaleMine V1.10.4 avec Araport11. Les analyses de

séquences de protéines PPR ont été effectuées à l'aide de TPRpred V.11.0. avec les paramètres standards d'exécution. Les analyses statistiques ont été effectuées avec RStudio V1.1.456. et R V3.5.1. Les génomes chloroplastiques et mitochondriaux sont visualisés à l'aide du logiciel Artemis (Carver et al., 2012). Les alignements multiples de séquences ont été effectués avec MUSCLE (McWilliam et al., 2013) et les matrices HMM utilisées pour la recherche de PPR bactériennes par HMMSearch V.2.26.0. Les analyses d'interactomique ont été effectuées sur Cytoscape V.3.6.1.

Table des matières.

Remerciements.....	2
Abréviations.....	5
Sommaire	6
Avant-propos.....	9
Définitions	10
Introduction	11
1. Interaction plante-pathogène vue par le pathogène	11
1.1. Origines et importance de la phytopathologie	11
1.2. Différents types de pathogènes	12
1.3. Qu'est-ce qu'un effecteur ?.....	12
1.3.1. Définition	12
1.3.2. Caractéristiques des effecteurs.....	12
1.3.3. Sécrétion et translocation	13
1.3.4. Redondance fonctionnelle des effecteurs.....	14
1.3.5. Cibles connues d'effecteurs de type III	14
1.3.6. Effecteurs à motif eucaryote	16
1.3.6.1. F-Box	16
1.3.6.2. TAL	17
1.3.6.3. PPR.....	18
1.3.6.4. Hypothèse du transfert horizontal de gènes.....	18
2. Interaction plante-pathogène vue par la plante.....	19
2.1. Reconnaissance du pathogène.....	19
2.1.1. Défenses préformées	20
2.1.2. Mise en place de la défense non spécifique.....	21
2.1.3. Mise en place de la défense spécifique.....	22
2.1.4. Modèle en zigzag	23
2.2. Principaux acteurs de la défense des plantes	24
2.2.1. Les gènes R	24
2.2.2. Les protéines kinases.....	25
2.2.3. Les hormones végétales	25
2.2.4. Les ROS	26
2.3. Rôle des organites dans la défense	27
2.3.1. Cas du chloroplaste	27
2.3.2. Cas de la mitochondrie	27
3. Régulation de l'expression génétique des organites pendant un stress biotique.....	28

3.1.	Génomique des organites chez <i>Arabidopsis thaliana</i>	28
3.1.1.	Génome chloroplastique	28
3.1.2.	Génome mitochondrial.....	29
3.2.	Expression génétique chez les organites d' <i>Arabidopsis thaliana</i>	31
3.2.1.	Transcription.....	31
3.2.2.	Maturation 5' et 3' UTR des ARN	31
3.2.3.	Edition.....	32
3.2.4.	Epissage	32
3.3.	Acteurs de la régulation de l'expression génétique des organites	32
3.3.1.	NEP	32
3.3.2.	Facteurs sigma	32
3.3.3.	RIP/MORF	33
3.3.4.	ORRM.....	34
3.3.5.	Protéines PPR	35
3.4.	Les protéines PPR	35
3.4.1.	Caractéristiques.....	35
3.4.1.1.	Motif PPR et fixation à l'ARN	35
3.4.1.2.	Différentes familles	36
3.4.1.3.	Distribution des protéines PPR dans l'arbre du vivant.....	36
3.4.1.4.	Code PPR.....	37
3.4.2.	Rôles des protéines PPR dans la maturation des ARN des organites.....	38
3.4.2.1.	Transcription.....	38
3.4.2.2.	Maturation 5' et 3' UTR des ARN	38
3.4.2.3.	Epissage	39
3.4.2.4.	Edition C vers U.....	39
3.4.2.5.	Traduction.....	40
3.5.	Organites et stimuli environnementaux.....	42
3.5.1.	Modifications de l'expression génétique des organites en réponse à un stress biotique	42
3.5.2.	Protéines PPR impliquées dans la réponse aux stress biotiques.....	43
3.5.2.1.	PPRL	43
3.5.2.2.	PPR1/PPR2	44
3.5.2.3.	SLO2	44
3.5.2.4.	PPR40.....	44
3.5.2.5.	PGN	45
4.	Projet de thèse	45

Chapitre 1 : Proposition d'un pipeline d'analyse de données <i>RNAseq</i> pour le chloroplaste et la mitochondrie	47
1. Présentation du pipeline d'analyse de données <i>RNAseq</i>	47
1.1. Pipeline appliqué à l'étude du transcriptome du chloroplaste.....	47
1.2. Pipeline appliqué à l'étude du transcriptome de la mitochondrie	48
2. Discussion	48
2.1. Génération des données	48
2.2. Analyse de l'expression des gènes	49
2.3. Analyse de couverture.....	52
2.4. Analyse d'épissage.....	53
2.5. Analyse d'édition	54
Chapitre 2 : Amélioration du système de prédiction de sites d'édition sous dépendance de PPR E+	55
1. Mise en contexte	55
1.1. Etude de CLB19, DYW2 et NUWA.....	55
1.2. Analyse transcriptomique des mutants $\Delta nuwa$ et $\Delta dyw2$	56
1.3. Hypothèse et projet de recherche	57
2. Lien fonctionnel entre DYW2 et les PPR E+	58
2.1. Validation du lien fonctionnel entre DYW2 et les PPR E+	58
2.1.1. Analyse <i>RNAseq</i> du transcriptome d'AT1G56570 (Salk_141937)	59
2.1.2. Analyse <i>RNAseq</i> du transcriptome d'AT5G08305 (Salk_058773)	59
2.1.3. Résumé des résultats obtenus et approche probabiliste.....	61
2.2. Application du critère DYW2 et rangs des sites prédits	63
2.2.1. Construction du système de prédiction	63
2.2.2. Exemple de CLB19	64
2.2.3. Analyse des rangs avec et sans le critère E+	66
3. Extension à l'ensemble des PPR E+	68
4. Discussion	70
Chapitre 3 : Caractérisation fonctionnelle de la protéine PPR PGN en condition de stress	74
1. Etat des lieux	74
1.1. Une PPR mitochondriale à l'expression induite par <i>Botrytis cinerea</i>	74
1.2. Un mutant hypersensible au glucose, au NaCl et à l'acide abscissique	75
1.3. L'expression de gènes impliqués dans le fonctionnement de la mitochondrie perturbée chez le mutant Δpgn	76
1.4. Hypothèse de travail.....	78
2. PGN, une PPR E+ impliquée dans l'édition	78
2.1. Etude <i>in silico</i>	78
2.2. Utilisation de la nouvelle méthode de prédiction pour PPR E+	78

2.3.	Analyse du transcriptome	79
2.4.	Complémentation fonctionnelle du mutant <i>Δpgn</i>	80
3.	PGN, une PPR dont la transcription est induite en condition de stress	81
3.1.	En condition de stress biotique	81
3.2.	En condition de stress salin	82
4.	Discussion	83
Chapitre 4 : Mise en évidence et caractérisation fonctionnelle de protéines bactériennes à motifs PPR		
PPR		
1.	Mise en évidence de protéines à motifs PPR chez les bactéries	85
1.1.	Initiation du projet.....	85
1.2.	Recherche de motifs PPR chez les bactéries	85
1.2.1.	Méthodologie	85
1.2.2.	Résultats	86
1.2.3.	Discussion	89
1.3.	Choix des objets d'étude	90
2.	Hypothèse de travail	91
3.	Caractérisation de la protéine PPR trouvée chez <i>Ralstonia solanacearum</i>	92
3.1.	Présentation du pathogène <i>Ralstonia solanacearum</i> et de sa protéine PPR.....	92
3.2.	Importance de RipL dans la pathogénicité de <i>Ralstonia solanacearum</i>	93
3.3.	Injection.....	94
3.4.	Localisation subcellulaire.....	94
3.5.	Rôle <i>in planta</i>	95
3.5.1.	Double Hybride levure.....	95
3.5.2.	Analyses transcriptomiques	99
3.5.2.1.	Analyse du transcriptome des organites	100
3.5.2.2.	Analyse du transcriptome nucléaire.....	101
3.5.2.3.	Limites liées à la surexpression d'un gène bactérien <i>in planta</i>	101
3.5.3.	Pathotests.....	102
3.6.	Discussion	103
4.	Caractérisation de la protéine PPR trouvée chez <i>Erwinia amylovora</i>	106
4.1.	Description de l'agent pathogène <i>Erwinia amylovora</i> et de sa protéine PPR.....	106
4.2.	Injection de la protéine codée par EAMY_0820 lors de l'infection.....	108
4.3.	Importance d'EAMY_0820 dans la pathogénicité d' <i>E. amylovora</i>	109
4.3.1.	Sur plante hôte	109
4.3.2.	Sur plante non-hôte.....	112
4.3.3.	En culture liquide.....	113
4.4.	Localisation subcellulaire.....	115

4.5.	Rôle <i>in planta</i>	116
4.5.1.	Un interactant en double hybride levure chez <i>Arabidopsis</i>	116
4.5.2.	Validation du rôle d'AtFes1c dans la défense d' <i>A. thaliana</i> contre les pathogènes.....	122
4.5.3.	Analyses transcriptomiques.....	124
4.5.3.1.	Analyse du transcriptome des organites.....	124
4.5.3.2.	Analyse du transcriptome nucléaire.....	125
4.5.3.3.	Limites liées à la surexpression d'un gène bactérien <i>in planta</i>	125
4.6.	Exportation des résultats sur <i>Malus domestica</i> « Golden delicious ».....	125
4.7.	Discussion.....	126
Discussion générale.		129
Matériels et methodes.		135
1.	Matériel biologique utilisé lors de la thèse.	135
1.1.	Espèces végétales utilisées et conditions de cultures.....	135
1.2.	Bactéries utilisées.....	135
1.3.	Champignons utilisés.....	135
1.4.	Souches de levure.....	135
2.	Méthodes	136
2.1.	Préparation de bactéries compétentes.....	136
2.2.	Génotypage.....	136
2.2.1.	Extraction d'ADN génomique.....	136
2.2.2.	PCR de génotypage.....	138
2.2.3.	Migration sur gel d'agarose.....	138
2.3.	Clonage et transformation de plantes.....	139
2.3.1.	PCR de clonage.....	139
2.3.2.	Précipitation PEG.....	140
2.3.3.	Réaction BP.....	140
2.3.4.	Traitement à la protéinase K.....	140
2.3.5.	Transformation avec <i>E. coli</i>	140
2.3.6.	PCR et cultures liquides de bactéries.....	141
2.3.7.	Stock glycérol.....	141
2.3.8.	Extraction d'ADN plasmidique (Miniprep).....	141
2.3.9.	Séquençage des plasmides.....	141
2.3.10.	Réaction LR.....	141
2.3.11.	Transformation avec <i>A. tumefaciens</i>	142
2.3.12.	Transformation d' <i>A. thaliana</i> par la méthode goutte.....	142
2.3.13.	Sélection BASTA.....	142
2.3.14.	Sélection hygromycine.....	142

2.3.15.	Agroinfiltration de <i>Nicotiana benthamiana</i>	143
2.4.	Microscopie	143
2.5.	Transcriptomique	143
2.5.1.	Extraction d'ARN.....	143
2.5.2.	Traitement DNase	143
2.5.3.	Construction de banques de séquençage <i>RNAseq</i>	143
2.5.3.1.	Pour les ARN totaux.....	144
2.5.3.2.	Pour les petits ARN	144
2.5.5.	PCR d'éditotype	144
2.5.6.	PCR quantitative (qPCR)	144
2.6.1.	<i>Ralstonia solanacearum</i>	145
2.6.1.1.	Pour <i>Solanum lycopersicum</i>	145
2.6.1.2.	Pour <i>Arabidopsis thaliana</i>	145
2.6.2.	<i>Erwinia amylovora</i>	146
2.6.2.1.	Sur <i>A. thaliana</i>	146
2.6.2.2.	Sur <i>Malus domestica</i> « Golden delicious ».....	146
2.6.3.	<i>Botrytis cinerea</i>	146
2.6.4.	<i>Pseudomonas syringae</i>	147
2.6.5.	<i>Hyaloperonospora arabidopsidis</i>	147
2.7.	Double Hybride levure.....	148
2.7.1.	Préparation de levures compétentes	148
2.7.2.	Transformation de levures	148
2.7.3.	Lyse des levures et PCR	149
2.7.4.	Cribles contre les banques AD et DB	149
2.8.	Analyses bio-informatiques et logiciels utilisés.....	149
	Table des matières.	151
	Table des figures.....	157
	Table des tableaux.....	161
	Annexes.	163
	Liste des amorces PCR utilisées pour le clonage.....	189
	Liste des amorces PCR utilisées pour mesurer l'édition.....	190
	Liste des amorces utilisées pour le génotypage des lignées d'insertion d'ADN-T	192
	Références.....	196

Table des figures.

Figure 1: Jules Breton, La bénédiction des blés en Artois (1857).....	12
Figure 2: Observation par microscopie électronique après purification (B) et représentation schématique (C) du système de sécrétion de type III de <i>Salmonella enterica</i>	14
Figure 3: Cibles connues d'effecteurs de type III de <i>Pseudomonas syringae</i> DC3000 chez <i>Arabidopsis thaliana</i>	15
Figure 4: Domaines fonctionnels des effecteurs TAL de <i>Xanthomonas</i>	17
Figure 5: Représentation schématique des différents types d'interactions plante-pathogène.	20
Figure 6: Représentation schématique du modèle proposé par Van der Hoorn et Kamoun (2008).	23
Figure 7 : Représentation du modèle en « zigzag » proposé par Jones et Dangl (2006).	24
Figure 8 : Représentation schématique de la chaîne de transport des électrons de la membrane thylakoïdienne.	29
Figure 9 : Représentation schématique de la chaîne de transport des électrons de la mitochondrie.	30
Figure 10 : Structure 3D d'une protéine PPR.....	35
Figure 11 : Représentation schématique des différentes familles de protéines PPR.	36
Figure 12 : Proportion de protéines PPR suivant une distribution phylogénétique.....	37
Figure 13 : Représentation schématique de l'organisation de l'expression des gènes des organites de la transcription à la traduction.	42
Figure 14 : Représentation schématique des principales étapes de l'analyse différentielle.	51
Figure 15 : Représentation schématique d'un cas hypothétique et extrême démontrant l'importance d'une analyse de couverture en complément d'une analyse d'expression différentielle.....	52
Figure 16 : Représentation schématique des différents types de <i>reads</i> à proximité d'un évènement d'épissage.	53
Figure 17 : Représentation schématique de la protéine de fusion CLB19-GS tag.	55
Figure 18 : Représentation schématique des PPR interagissant avec CLB19, par la méthode TAP.....	55
Figure 19 : Résumé des défauts d'édition trouvés chez les mutants d'insertion $\Delta nuwa$ et $\Delta dyw2$	56
Figure 20 : Modèle de l'éditosome des PPR E+ chez <i>Arabidopsis thaliana</i> proposé par Guillaumot et al.	57
Figure 21 : Représentation schématique de la PPR E+ CLB19.....	64
Figure 22 : Partie 3' de la séquence codant pour Nad1 avec ses sites d'édition.	71
Figure 23 : Expression du gène <i>pgn</i> chez le sauvage (Wt) et le mutant $\Delta bik1$ (<i>bik1</i>) à 0 et 48h après infection par <i>Bc</i>	74
Figure 24 : Symptômes observés après inoculation avec <i>Bc</i> par la méthode goutte, après 5 jours (D) et pulvérisation après 7 jours (E), chez le sauvage (Wt) et le mutant Δpgn (<i>pgn</i>).	75
Figure 25 : Observations au microscope à épifluorescence d'épiderme de <i>N. benthamiana</i> exprimant transitoirement PGN-GFP.	75
Figure 26 : Hypersensibilité du mutant Δpgn au NaCl (A), au glucose (B) et à l'ABA (C) après 10 jours.....	76

Figure 27 : Profils d'expression de gènes associés à la mitochondrie à 0 et 12h après traitement NaCl chez le sauvage (Wt) et le mutant <i>Δpgn</i> (<i>pgn</i>). Gènes associé au transport d'électrons (A), à la transcription (B), à l'oxydation/réduction (C), à la régulation du cytochrome c (D) et aux sous-unités de la NAD(P)H déshydrogénase (E).	77
Figure 28 : Observation de l'accumulation de peroxyde d'hydrogène par coloration DAB chez le sauvage (Wt) et le mutant <i>Δpgn</i> (<i>pgn</i>) avec ou sans traitement au NaCl.....	77
Figure 29 : Résultat affiché pour PGN (AT1G56570) par l'outil TPRpred.	78
Figure 30 : Chromatogrammes du site <i>nad6_leader_73</i> chez <i>A. thaliana</i> Col0, <i>Δpgn</i> , et les lignées <i>Δpgn</i> complémentées avec PGN avec étiquette 3xHA et sans étiquette.....	80
Figure 31 : Représentations graphiques des données de RT-qPCR effectuées sur Col0 et <i>Δpgn</i> avant infection (T0) et à 48h après infection (T48) par <i>Botrytis cinerea</i>	82
Figure 32 : Représentations graphiques des données de RT-qPCR effectuées sur plantules de 10 jours Col0 et <i>Δpgn</i> , sur 1/2 MS 1% sucrose avec et sans 150mM NaCl	83
Figure 33 : Diagramme de Venn représentant les protéines PPR bactériennes selon leur méthode de mise en évidence.....	87
Figure 34 : Consensus du motif pur chez <i>A. thaliana</i> basé sur les données de Cheng et al. (2016) (en haut) et sur les motifs trouvés chez les protéines PPR bactériennes quelle que soit leur méthode de mise en évidence (en bas).	88
Figure 35 : Proportion de protéines PPR bactériennes trouvées selon leur annotation et en fonction de leur méthode d'obtention.....	89
Figure 36 : Représentation schématique des différentes étapes de caractérisation du rôle des PPR bactériennes de <i>Ralstonia solanacearum</i> et <i>Erwinia amylovora</i>	91
Figure 37 : Pourcentage de survie en fonction du temps d' <i>Arabidopsis thaliana</i> infecté par <i>Ralstonia solanacearum</i> sauvage (GMI1000A, en rouge) ou <i>ΔripL</i> (GRS176A, en noir).....	93
Figure 38 : Pourcentage de survie en fonction du temps de <i>Solanum lycopersicum</i> infecté par <i>Ralstonia solanacearum</i> sauvage (GMI1000A, en rouge) ou <i>ΔripL</i> (GRS176A, en noir).....	93
Figure 39 : Représentation schématique des différents construits utilisés en double-hybride levure.	95
Figure 40 : Interactome d'ArtRipL chez <i>Arabidopsis thaliana</i>	98
Figure 41 : Pourcentage de survie en fonction du temps d' <i>Arabidopsis thaliana</i> Col0 (en rouge) ou surexprimant ArtRipL, infecté par <i>Ralstonia solanacearum</i> sauvage.....	102
Figure 42 : Comptage de spores de <i>Hyaloperonospora arabidopsidis</i> par mg de matière fraîche végétale 7 jours après infection chez <i>A. thaliana</i>	103
Figure 43 : Comptage de colonies de <i>Pseudomonas syringae</i> par mg de matière fraîche végétale, représentés en échelle Log10, 48h après infection.	103
Figure 44 : Résumé des résultats obtenus lors de la thèse sur la caractérisation de RipL.	104
Figure 45 : Arbre phylogénique de différentes souches et espèces d' <i>Erwinia</i> et mise en évidence des souches d' <i>Erwinia amylovora</i> contenant une protéine homologue de celle encodée par EAMY_0820 dans CFBP1430.	

.....	107
Figure 46 : Evolution des symptômes de nécrose sur le pathosytème <i>Erwinia amylovora-Malus domestica</i> « Golden delicious » et symptômes à 7 jours après inoculation.....	110
Figure 47 : Evolution des symptômes de nécrose sur le pathosytème <i>Erwinia amylovora-Malus domestica</i> « Golden delicious » et symptômes à 14 jours après inoculation.....	110
Figure 48 : Taux d'incidence de la maladie induite par <i>E. amylovora</i> sur <i>Malus domestica</i> « Golden delicious » (en pourcentage) en fonction de la concentration et de la souche à 7 et 14 jours après infection.	111
Figure 49 : Sévérité de l'infection de <i>Malus domestica</i> « Golden delicious » par <i>E. amylovora</i> (en pourcentage) en fonction de la concentration et de la souche à 7 et 14 jours après infection.	111
Figure 50 : Sévérité de l'infection de <i>Malus domestica</i> « Golden delicious » sur les plantes dont la nécrose a atteint la tige par <i>E. amylovora</i> (en pourcentage) en fonction de la concentration et de la souche à 7 et 14 jours après infection.	112
Figure 51 : Ratio du titrage bactérien in planta d' <i>E. amylovora</i> mutante par rapport à la souche sauvage à 24hpi en fonction de différents fonds génétiques d' <i>A. thaliana</i>	113
Figure 52 : Comparaison de la croissance bactérienne en milieu liquide riche pour CFBP1430 et CFBP7962 à différents temps.	114
Figure 53 : Comparaison de la croissance bactérienne en milieu liquide minimum pour CFBP1430 et CFBP7962 à différents temps.....	115
Figure 54 : Mise en évidence de l'interaction en double hybride levure entre DB-ArtEa et AD-AtFes1c (AT5G02150) sur milieu de sélection stringent.	116
Figure 55 : Réseau des interactants d'AtFes1c en double hybride chez la levure contenant les effecteurs ayant des interactants en commun avec AtFes1c.	120
Figure 56 : Ratio de la sporulation de <i>Hpa</i> par mg de matière fraîche végétale à 7 jours entre les différents fonds génétiques d' <i>A. thaliana</i> et Col0.	122
Figure 57 : Comptage de colonies de <i>Pseudomonas syringae</i> par mg de matière fraîche végétale, représentés en échelle Log10, 48h après infection.	123
Figure 58 : Comptage de colonies d' <i>Erwinia amylovora</i> par disque de matière végétale prélevé, représentés en échelle Log10, 24h après infection.....	124
Figure 59 : Crible de double hybride levure en un contre un sur différents milieux solides, pour la PPR d' <i>Erwinia amylovora</i> , son interactant chez <i>A. thaliana</i> AtFes1c et ses deux homologues chez <i>Malus domestica</i> « Golden delicious » Fes1 α et Fes1 β	126
Figure 60 : Résumé des résultats obtenus lors de la thèse sur la caractérisation de la protéine encodée par EAMY_0820.	127
Figure 61 : Résumé des différentes hypothèses de régulation ou d'absence de régulation de l'expression des gènes des organites en condition de stress biotique.....	133

Table des tableaux.

Tableau I : Résumé des différents facteurs Sigma et de leurs spécificités supposées.	32
Tableau II : Résumé des différentes protéines RIP connues chez <i>A. thaliana</i> et de la proportion de sites d'édition impactés par leur mutation.	33
Tableau III : Sites différentiellement édités chez le mutant Salk_141937 (AT1G56570) en comparaison avec Col0	59
Tableau IV : Sites différentiellement édités chez le mutant SK_141937 (AT1G56570) en comparaison avec Col0 avec ajout du critère DYW2.	59
Tableau V : Sites différentiellement édités chez le mutant Salk_058773 (AT5G08305) en comparaison avec Col0	59
Tableau VI : Sites différentiellement édités chez le mutant SK_058773 (AT5G08305) en comparaison avec Col0 avec le critère DYW2.	60
Tableau VII : Résumé des sites connus et trouvés comme dépendants de PPR E+ et dépendance à DYW2.	62
Tableau VIII : Répartition des sites d'édition observés par Guillaumot et al. puis ajustés par une nouvelle analyse et répartition par dépendance.	63
Tableau IX : Probabilités de fixation d'une base azotée pour chaque motif de CLB19.	64
Tableau X : Premiers sites prédits pour être fixés par CLB19 en appliquant le code PPR.	65
Tableau XI : Premiers sites prédits pour être fixés par CLB19 en appliquant le code PPR et le critère E+	65
Tableau XII : Comparaison des rangs moyens des sites d'édition connus de PPR E+ avec et sans le critère E+	66
Tableau XIII : Classement des premiers sites prédits pour MEF37 en se basant sur le code PPR et le critère E+	67
Tableau XIV : Résumé des données obtenues sur les lignées d'insertion d'ADN-T sur des gènes codants pour des PPR E+, pour lesquelles nous disposons d'individus homozygotes.	69
Tableau XV : Sites différentiellement édités chez le mutant Salk_141937 (AT1G56570, Δ pgn) en comparaison avec Col0 avec ajout du critère DYW2.	79
Tableau XVI: Analyse d'expression différentielle de nad6 et cox2.	79
Tableau XVII : Espèces bactériennes contenant des gènes codants pour des protéines à motifs PPR obtenues par la matrice HMM d' <i>A. thaliana</i> et leur phylogénie.	86
Tableau XVIII : Espèces bactériennes contenant des gènes codants pour des protéines à motifs PPR obtenues par la matrice HMM bactérienne et leur phylogénie.	87
Tableau XIX: Application du code PPR à RipL.	93
Tableau XX : Liste des différentes approches testées pour la localisation subcellulaire de RipL.	94
Tableau XXI : Liste des interactants obtenus après cribles en double hybride levure en fonction de la banque et du construit testé.	96
Tableau XXII : Nombre d'interactants trouvés chez <i>A. thaliana</i> par Dreze et al. (2010) et la plateforme InterATOME pour les interactants d'ArtRipL.	97
Tableau XXIII: Epissage de l'intron 2 de nad7.	100
Tableau XXIV : Application du code PPR à la protéine encodée par EAMY_0820.	107
Tableau XXV : Règles proposées par Oh et al. (2010) et application à la PPR d' <i>E. amylovora</i>	108
Tableau XXVI : Résumé des différentes méthodes testées afin d'obtenir la localisation subcellulaire de la PPR d' <i>Erwinia amylovora</i>	116
Tableau XXVII: Liste des interactants trouvés pour AtFes1c lors d'un crible contre la banque DB.	117
Tableau XXVIII : Composition du milieu LB.	136
Tableau XXIX: Composition d'AP1 : tampon de lyse, dénaturation des protéines et lyse des parois	

cellulaires.	137
Tableau XXX : Composition d'AP2 : précipitation des protéines et des polysaccharides par acidification du milieu.	137
Tableau XXXI : Composition d'AP3 : constitution des sels d'ADN.	137
Tableau XXXII: Composition d'AW : dessaler et laver l'ADN.....	138
Tableau XXXIII: Composition d'AE : tampon d'élution.	138
Tableau XXXIV : Protocole PCR et composition de la réaction pour le génotypage.....	138
Tableau XXXV : Préparation d'un gel d'agarose.....	139
Tableau XXXVI : Composition du tampon TAE.	139
Tableau XXXVII : Programme PCR et composition de la réaction pour le clonage.	139
Tableau XXXVIII : Composition de la réaction BP.....	140
Tableau XXXIX : Composition de la réaction LR.	141
Tableau XL : Composition du milieu 1/2 MS.....	143
Tableau XLI: Programme et composition de la réaction pour la PCR d'éditotype.	144
Tableau XLII: Programme et composition de la réaction de qPCR.	145
Tableau XLIII : Compositions respectives des solutions de TE/LiAc ; TE/LiAc/PEG et YEPD.	148
Tableau XLIV: Composition du milieu Sc.	149
Tableau XLV: Composition du tampon de lyse.	149

Annexes.

Annexe S1 : Script de l'analyse par segmentation.

Extraction des débuts et fins de chaque *read* aligné sur le génome chloroplastique.

```
<script>
bedtools bamtobed -i Input_primary.bam|awk '{print $1 "\t" $2 "\t" $3 "\t" $6}'|grep "Name_of_plastid_chromosome" >coverage.bed
</script>
```

Utilisation de R pour calculer la couverture par base.

```
<script>
rm(list=ls()); gc();
library(data.table)
genome_size <- 154478
file<-system("ls *_chloro.bed",T)
data <- fread(file[1], sep="\t", showProgress = FALSE)
colnames(data) <- c("Pt", "start", "end", "strand")
i_plus <- which(data$strand == "+")
i_minus <- which(data$strand == "-")
coverage <- data.frame(
  pos = 0:genome_size,
  Minus = tabulate(data$end[i_minus]+1, nbins=genome_size+1),
  Plus = tabulate(data$start[i_plus]+1, nbins=genome_size+1)
)
rm(data);gc();
colnames(coverage)<-c("pos",paste(strsplit(file[1],"_chloro.bed")[[1]][1],"_minus",sep=""),paste(strsplit(file[1],"_chloro.bed")[[1]][1],"_plus",sep=""))
for (i in 2:length(file)){
data <- fread(file[i], sep="\t", showProgress = FALSE)
colnames(data) <- c("Pt", "start", "end", "strand")
i_plus <- which(data$strand == "+")
i_minus <- which(data$strand == "-")
tmp <- data.frame(
  pos = 0:genome_size,
  Minus = tabulate(data$end[i_minus]+1, nbins=genome_size+1),
  Plus = tabulate(data$start[i_plus]+1, nbins=genome_size+1)
)
rm(data);gc();
colnames(tmp)<-c("pos",paste(strsplit(file[i],"_chloro.bed")[[1]][1],"_minus",sep=""),paste(strsplit(file[i],"_chloro.bed")[[1]][1],"_plus",sep=""))
coverage=merge(coverage,tmp, by="pos")
}

write.table(coverage, file="chloro_read_starts.txt",sep="\t",quote=F,row.names=FALSE)

data <- fread(file[1], sep="\t", showProgress = FALSE)
colnames(data) <- c("Pt", "start", "end", "strand")
i_plus <- which(data$strand == "+")
i_minus <- which(data$strand == "-")
coverage <- data.frame(
  pos = 0:genome_size,
  Minus = tabulate(data$start[i_minus]+1, nbins=genome_size+1),
  Plus = tabulate(data$end[i_plus]+1, nbins=genome_size+1)
)
rm(data);gc();
colnames(coverage)<-c("pos",paste(strsplit(file[1],"_chloro.bed")[[1]][1],"_minus",sep=""),paste(strsplit(file[1],"_chloro.bed")[[1]][1],"_plus",sep=""))
for (i in 2:length(file)){
data <- fread(file[i], sep="\t", showProgress = FALSE)
colnames(data) <- c("Pt", "start", "end", "strand")
i_plus <- which(data$strand == "+")
i_minus <- which(data$strand == "-")
tmp <- data.frame(
  pos = 0:genome_size,
  Minus = tabulate(data$start[i_minus]+1, nbins=genome_size+1),
  Plus = tabulate(data$end[i_plus]+1, nbins=genome_size+1)
)
rm(data);gc();
colnames(tmp)<-c("pos",paste(strsplit(file[i],"_chloro.bed")[[1]][1],"_minus",sep=""),paste(strsplit(file[i],"_chloro.bed")[[1]][1],"_plus",sep=""))
coverage=merge(coverage,tmp, by="pos")
}

write.table(coverage, file="chloro_read_ends.txt",sep="\t",quote=F,row.names=FALSE)
</script>
```

Segmentation des débuts de *reads*.

```
<script>
chloro=read.table(file="chloro_read_starts.txt",sep="\t",h=T)[-1]
min <- grep("_minus", colnames(chloro))
min_chloro=chloro[,min]
plus_chloro=chloro[,-min]
library(robseg)

varDiff <- function(x, method='MAD'){
  n = length(x)
  if(method == "MAD"){
    return(mad(diff(x)/sqrt(2)))
  }
  if(method == "HALL"){
    wei <- c(0.1942, 0.2809, 0.3832, -0.8582)
    mat <- wei %*% t(x)
    mat[2, -n] = mat[2, -1]
    mat[3, -c(n-1, n)] = mat[3, -c(1, 2)]
    mat[4, -c(n-2, n-1, n)] = mat[4, -c(1, 2, 3)]
    return(sqrt(sum(apply(mat[, -c(n-2, n-1, n)], 2, sum)^2) / (n-3)))
  }
}

wrapRobSeg <- function(x){
  std.dev <- std.dev <- varDiff(x, method="HALL")
  x_ <- std.dev
  beta <- 2*log(length(x))*std.dev
  threshold <- param*std.dev
  return(Rob_seg.std(x, "Outlier", lambda = beta, lthreshold = threshold))
}

param = 5
res.min <- wrapRobSeg(sqrt(log(min_chloro[,1]+1)))
split.min=res.min$est
for (i in 2:dim(min_chloro)[2]){
  x<-sqrt(log(min_chloro[,i]+1))
  tmp <- wrapRobSeg(x)
  split.min=sort(union(split.min,tmp$est))
}
res.plus <- wrapRobSeg(sqrt(log(plus_chloro[,1]+1)))
split.plus=res.plus$est
for (i in 2:dim(plus_chloro)[2]){
  x<-sqrt(log(plus_chloro[,i]+1))
  tmp <- wrapRobSeg(x)
  split.plus=sort(union(split.plus,tmp$est))
}
</script>
```

Comptage par segment.

```
<script>
start=1
stop=split.min[1]-1
count_min=colSums(min_chloro[start:stop,])
names(count_min)=names(min_chloro)
for (i in 1:(length(split.min)-1)){
  start=as.numeric(split.min[i])
  stop=split.min[i+1]-1
  tmp=colSums(min_chloro[start:stop,])
  names(tmp)=names(min_chloro)
  count_min=rbind(count_min,tmp)
}
start=1
stop=split.min[1]-1
nom=paste("Minus_segment",as.character(start),as.character(stop),sep="_")
for (i in 1:(length(split.min)-1)){
  start=as.numeric(split.min[i])
  stop=split.min[i+1]-1
  tmp=paste("Minus_segment",as.character(start),as.character(stop),sep="_")
  nom=c(nom,tmp)
}
row.names(count_min)=nom
```

```

start=1
stop=split.plus[1]-1
count_plus=colSums(plus_chloro[start:stop,])
names(count_plus)=names(plus_chloro)
for (i in 1:(length(split.plus)-1)){
#la dernière rupture correspond à la dernière base du génome
start=as.numeric(split.plus[i])
stop=split.plus[i+1]-1
tmp=colSums(plus_chloro[start:stop,])
names(tmp)=names(plus_chloro)
count_plus=rbind(count_plus,tmp)
}

start=1
stop=split.plus[1]-1
nom=paste("Plus_segment",as.character(start),as.character(stop),sep="_")
for (i in 1:(length(split.plus)-1)){
start=as.numeric(split.plus[i])
stop=split.plus[i+1]-1
tmp=paste("Plus_segment",as.character(start),as.character(stop),sep="_")
nom=c(nom,tmp)
}
row.names(count_plus)=nom
noms_min=unlist(strsplit(colnames(count_min),"_minus"))
colnames(count_min)=colnames(count_plus)=noms_min
count=rbind(count_plus,count_min)
write.table(count,file="chloro_starts_count.txt",sep="\t",quote=FALSE)
</script>

```

Analyse différentielle.

```

<script>
require(edgeR)
conds<-c("WT","EA")
rep<-3
ordre.ech<-c(grep(conds[1],colnames(count)),grep(conds[2],colnames(count)))
countTable<-count[,ordre.ech]
cat("Library size \n")
print(apply( countTable,2,sum))
cat("Number of analyzed segments", nrow(countTable),"n")

geno<-rep(conds,each=rep)
batch<-rep(paste("repbio",1:rep,sep="_"),length(conds))
geno <- as.factor (geno)
batch <- as.factor(batch)
geno <- relevel(geno, ref=conds[1])
design <- model.matrix( ~geno+batch)
m <- 1e6 * t(t(countTable) / colSums(countTable))
ridx <- rowSums(m > 1) >= rep
id.filtres<-rownames(countTable)[!ridx]
filteredCountTable <- countTable[!ridx,]
countTable <- countTable[ridx,]
cat("Number of analyzed segments after edgeR filtering: ", nrow(countTable),"n")

cds <- DGEList( countTable , group = paste(geno,batch))
cds <- calcNormFactors( cds )
cds <- estimateDisp( cds, design)
plotBCV(cds)
fit<-glmFit(cds,design)
lrt <- glmLRT(fit, coef=2)
N <- cds$samples$lib.size
f <- cds$samples$norm.factors
TMM <- N*f / mean(N*f) ## normalized total counts
norm <- scale(cds$count, center=FALSE, scale=TMM)
baseMeanA <- rowMeans(scale(cds$counts[, which(geno == conds[1])], center = FALSE, scale = TMM[which(geno == conds[1])]))
baseMeanB <- rowMeans(scale(cds$counts[, which(geno == conds[2])], center = FALSE, scale = TMM[which(geno == conds[2])]))
foldChange <- baseMeanB/baseMeanA
log2FoldChange <- log2(foldChange)
norm<-data.frame(id=rownames(norm),baseMeanA,baseMeanB,log2FoldChange)
colnames(norm)<-cbind("id",conds[1],conds[2],"log2FoldChange")
final<-data.frame(norm,PValue=lrt$table$PValue)
final<-cbind(final,FDR=p.adjust(final$PValue,"BH"))
baseMeanAFiltered <- rowMeans(scale(filteredCountTable[, which(geno == conds[1])],

```

```

center = FALSE, scale = TMM[which(geno == conds[1])])
baseMeanBFiltered <- rowMeans(scale(filteredCountTable[, which(geno == conds[2])],
center = FALSE, scale = TMM[which(geno == conds[2])]))
foldChangeFiltered <- baseMeanBFiltered/baseMeanAFiltered
log2FoldChangeFiltered <- log2(foldChangeFiltered)
colfiltre <- rep("filtered",length(id.filtres))
filtres <- data.frame(id=rownames(filteredCountTable),baseMeanAFiltered,baseMeanBFiltered,log2FoldChangeFiltered,colfiltre,colfiltre)
names(filtres)<-names(final)
DE.Bonf<-which(final$PValue*length(final$PValue)<0.05)
cat("nbre DE (Bonf) :",length(DE.Bonf),"n")
DE.BH<-which(final$FDR<0.05)
cat("nbre DE (BH) :",length(DE.BH),"n")
cat("Gene division \n")
print(table(decideTestsDGE(lrt , p.value = 0.05)))
nom<-"chloro_starts_CompleteListe.txt"

complete<-rbind(final,filtres)
write.table(complete,nom,row.names=FALSE,col.names=TRUE,sep="\t",quote=FALSE)
nom<-paste("filter_chloro_starts_Bonferroni-",0.05,".txt",sep="")
write.table(final[DE.Bonf,],nom,row.names=FALSE,col.names=TRUE,sep="\t",quote=FALSE)
nom<-paste("filter_chloro_starts_BH-",0.05,".txt",sep="")
write.table(final[DE.BH,],nom,row.names=FALSE,col.names=TRUE,sep="\t",quote=FALSE)
</script>

```

Segmentation des *reads* de fin.

```

<script>
chloro=read.table(file="chloro_read_ends.txt",sep="\t",h=T)[-1,-1]
min <- grep("_minus", colnames(chloro))
min_chloro=chloro[,min]
plus_chloro=chloro[,-min]
library(robseg)

varDiff <- function(x, method='MAD'){
  n = length(x)
  if(method == "MAD"){
    return(mad(diff(x)/sqrt(2)))
  }
  if(method == "HALL"){
    wei <- c(0.1942, 0.2809, 0.3832, -0.8582)
    mat <- wei %*% t(x)
    mat[2, -n] = mat[2, -1]
    mat[3, -c(n-1, n)] = mat[3, -c(1, 2)]
    mat[4, -c(n-2, n-1, n)] = mat[4, -c(1, 2, 3)]
    return(sqrt(sum(apply(mat[, -c(n-2, n-1, n)], 2, sum)^2) / (n-3)))
  }
}

wrapRobSeg <- function(x){
  std.dev <- std.dev <- varDiff(x, method="HALL")
  x_ <- std.dev
  beta <- 2*log(length(x))*std.dev
  threshold <- param*std.dev
  return(Rob_seg.std(x, "Outlier", lambda = beta, lthreshold = threshold))
}

param = 5
res.min <- wrapRobSeg(sqrt(log(min_chloro[,1]+1)))
split.min=res.min$st.est
for (i in 2:dim(min_chloro)[2]){
  x<-sqrt(log(min_chloro[,i]+1))
  tmp <- wrapRobSeg(x)
  split.min=sort(union(split.min,tmp$st.test))
}
res.plus <- wrapRobSeg(sqrt(log(plus_chloro[,1]+1)))
split.plus=res.plus$st.est
for (i in 2:dim(plus_chloro)[2]){
  x<-sqrt(log(plus_chloro[,i]+1))
  tmp <- wrapRobSeg(x)
  split.plus=sort(union(split.plus,tmp$st.test))
}
</script>

```

Comptage par segment.

```

<script>
start=1
stop=split.min[1]-1
count_min=colSums(min_chloro[start:stop,])
names(count_min)=names(min_chloro)
for (i in 1:(length(split.min)-1)){
start=as.numeric(split.min[i])
stop=split.min[i+1]-1
tmp=colSums(min_chloro[start:stop,])
names(tmp)=names(min_chloro)
count_min=rbind(count_min,tmp)
}
start=1
stop=split.min[1]-1
nom=paste("Minus_segment",as.character(start),as.character(stop),sep="_")
for (i in 1:(length(split.min)-1)){
start=as.numeric(split.min[i])
stop=split.min[i+1]-1
tmp=paste("Minus_segment",as.character(start),as.character(stop),sep="_")
nom=c(nom,tmp)
}
row.names(count_min)=nom

start=1
stop=split.plus[1]-1
count_plus=colSums(plus_chloro[start:stop,])
names(count_plus)=names(plus_chloro)
for (i in 1:(length(split.plus)-1)){
#la dernière rupture correspond à la dernière base du génome
start=as.numeric(split.plus[i])
stop=split.plus[i+1]-1
tmp=colSums(plus_chloro[start:stop,])
names(tmp)=names(plus_chloro)
count_plus=rbind(count_plus,tmp)
}

start=1
stop=split.plus[1]-1
nom=paste("Plus_segment",as.character(start),as.character(stop),sep="_")
for (i in 1:(length(split.plus)-1)){
start=as.numeric(split.plus[i])
stop=split.plus[i+1]-1
tmp=paste("Plus_segment",as.character(start),as.character(stop),sep="_")
nom=c(nom,tmp)
}
row.names(count_plus)=nom
noms_min=unlist(strsplit(colnames(count_min),"_minus"))
colnames(count_min)=colnames(count_plus)=noms_min
count=rbind(count_plus,count_min)
write.table(count,file="chloro_ends_count.txt",sep="\t",quote=FALSE)
</script>

```

Analyse différentielle.

```

<script>
require(edgeR)
conds<-c("WT","EA")
rep<-3
ordre.ech<-c(grep(conds[1],colnames(count)),grep(conds[2],colnames(count)))
countTable<-count[,ordre.ech]
cat("Library size \n")
print(apply( countTable,2,sum))
cat("Number of analyzed segments", nrow(countTable),"n")

geno<-rep(conds,each=rep)
batch<-rep(paste("repbio",1:rep,sep="_"),length(conds))
geno <- as.factor (geno)
batch <- as.factor(batch)
geno <- relevel(geno, ref=conds[1])
design <- model.matrix( ~geno+batch)
m <- 1e6 * t((countTable) / colSums(countTable))
ridx <- rowSums(m > 1) >= rep
id.filtres<-rownames(countTable)[!ridx]

```

```

filteredCountTable <- countTable[!ridx,]
countTable <- countTable[ridx,]
cat("Number of analyzed segments after edgeR filtering: ", nrow(countTable), "\n")

cds <- DGEList( countTable , group = paste(geno,batch))
cds <- calcNormFactors( cds )
cds <- estimateDisp( cds, design)
plotBCV(cds)
fit<-glmFit(cds,design)
lrt <- glmLRT(fit, coef=2)
N <- cds$samples$lib.size
f <- cds$samples$norm.factors
TMM <- N*f / mean(N*f)  ## normalized total counts
norm <- scale(cds$count, center=FALSE, scale=TMM)
baseMeanA <- rowMeans(scale(cds$count[, which(geno == conds[1])], center = FALSE, scale = TMM[which(geno == conds[1])]))
baseMeanB <- rowMeans(scale(cds$count[, which(geno == conds[2])], center = FALSE, scale = TMM[which(geno == conds[2])]))
foldChange <- baseMeanB/baseMeanA
log2FoldChange <- log2(foldChange)
norm<-data.frame(id=rownames(norm),baseMeanA,baseMeanB,log2FoldChange)
colnames(norm)<-cbind("id",conds[1],conds[2],"log2FoldChange")
final<-data.frame(norm,PValue=lrt$stale$PValue)
final<-cbind(final,FDR=p.adjust(final$PValue,"BH"))
baseMeanAFiltered <- rowMeans(scale(filteredCountTable[, which(geno == conds[1])],
                                center = FALSE, scale = TMM[which(geno == conds[1])]))
baseMeanBFiltered <- rowMeans(scale(filteredCountTable[, which(geno == conds[2])],
                                center = FALSE, scale = TMM[which(geno == conds[2])]))
foldChangeFiltered <- baseMeanBFiltered/baseMeanAFiltered
log2FoldChangeFiltered <- log2(foldChangeFiltered)
colfiltre <- rep("filtered",length(id.filtres))
filtres <- data.frame(id=rownames(filteredCountTable),baseMeanAFiltered,baseMeanBFiltered,log2FoldChangeFiltered,colfiltre,colfiltre)
names(filtres)<-names(final)
DE.Bonf<-which(final$PValue*length(final$PValue)<0.05)
cat("nbre DE (Bonf) :",length(DE.Bonf),"\n")
DE.BH<-which(final$FDR<0.05)
cat("nbre DE (BH) :",length(DE.BH),"\n")
cat("Gene division \n")
print(table(decideTestsDGE(lrt , p.value = 0.05)))
nom<-"chloro_ends_CompleteListe.txt"

complete<-rbind(final,filtres)
write.table(complete,nom,row.names=FALSE,col.names=TRUE,sep="\t",quote=FALSE)
nom<-paste("filter_chloro_ends_Bonferroni-",0.05,".txt",sep="")
write.table(final[DE.Bonf,],nom,row.names=FALSE,col.names=TRUE,sep="\t",quote=FALSE)
nom<-paste("filter_chloro_ends_BH-",0.05,".txt",sep="")
write.table(final[DE.BH,],nom,row.names=FALSE,col.names=TRUE,sep="\t",quote=FALSE)
</script>

```

Annexe S2 : Lignées de mutants d'insertion pour lesquelles nous n'avons pas trouvé d'homozygote.

Ho : Homozygote pour le gène codant la PPR E+ ; HT : Hétérozygote pour le gène codant la PPR E+.

Lignées sans HO disponible		
Lignée mutante utilisée	AGI	Remarques
Salk_122927	AT3G25970	Pas d'HO trouvé sur la première génération. HT disponible pour re-semer.
SAIL_549_E07	AT1G06150	
Salk_039979	AT3G01580	
Salk_027870	AT3G02330	
Salk_122936	AT3G25970	
GB_300H08	AT4G32430	
Salk_130949	AT3G15930	
Salk_055676	AT3G50420	Erreur de Salk sur la zone d'insertion du T-DNA.
Salk_025267	AT3G63370	
Sail_1210_A07	AT1G50270	Pas d'amplification du T-DNA par PCR.
Salk_045339	AT1G71490	
SAIL_377_B04	AT2G33680	
Salk_091381	AT3G25060	
Salk_090615	AT3G47840	
Salk_115308	AT2G35030	Pas d'HO sur la première et la seconde génération issue d'un HT.
Salk_081270	AT4G21300	
Salk_130945	AT5G08510	

AXI97796.1 cytochrome c oxidase subunit 2 [Gracilaria vermicu... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_009138187.1 cytochrome c oxidase subunit 2 [Gracilaria ver... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

ART65149.1 cytochrome c oxidase subunit 2 [Gracilaria changii] SIFIKREGIYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

AOX48918.1 cytochrome c oxidase subunit 2 [Gelidium crinale f... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

AT281798.1 cytochrome c oxidase subunit 2 [Picochlorum sp. 's... PLYIRREGVYGGQCEICGVNHGFMPIVIEAVPLNPNYI W AQRL

YP_008994795.1 cytochrome c oxidase subunit 2 [Bathycoccus pr... SIFLNREGVYGGQCEICGVNHGFMPIVIEAVPLNPNYI W SNKL

YP_006665885.1 cytochrome c oxidase subunit 2 [Porphyra umbil... SIFVKKREGIFYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_009445900.1 cytochrome c oxidase subunit 2 [Betaphycus gel... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009365198.1 cytochrome c oxidase subunit 2 [Gracilaria cha... SIFIKREGIYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_009040899.1 cytochrome c oxidase subunit 2 (mitochondrion)... SIFVKKREGIFYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_009114043.1 cytochrome c oxidase subunit 2 [Ahnfeltia plic... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009132701.1 cytochrome c oxidase subunit 2 [Bangia fuscocu... SIFIKREGIFYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_006280878.1 cytochrome c oxidase subunit 2 [Pyropia yezoen... SIFVKKREGIFYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AAR99501.1 cytochrome oxidase subunit II [Chlorella vulgaris] FVFIKREGVYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_007890507.1 cytochrome c oxidase subunit II [Andalucia god... DVFIKREGVYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AHX02474.1 cytochrome c oxidase subunit 2 [Riquetophycus sp. ... SIFIKREGVYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

AGH24453.1 cytochrome c oxidase subunit 2 [Reclinomonas ameri... SVFIKREGVYGGQCEICGVNHGFMPIVIEAVPLKDYV W TSSQI

YP_009441347.1 cytochrome c oxidase subunit 2 [Lympia mucosa] SMFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

AGH24386.1 cytochrome c oxidase subunit 2 [Reclinomonas ameri... SVFIKREGIFYGGQCEICGVNHGFMPIVIEAVPLKDYV W TSSQI

YP_009295522.1 cytochrome c oxidase subunit 2 [Mastocarpus pa... SIFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009159734.1 cytochrome c oxidase subunit 2 [Pyrococcus nitida] SIFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AML60643.1 cytochrome c oxidase subunit 2 [Moromonas marocensis] NIFIKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AGH24319.1 cytochrome c oxidase subunit 2 [Reclinomonas ameri... SVFIKREGIFYGGQCEICGVNHGFMPIVIEAVPLKDYV W TSSQI

AVK39572.1 cytochrome c oxidase subunit 2 [Sheathia arcuata] SMFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_009048890.1 cytochrome c oxidase subunit 2 [Wildemanian sch... SIFVKKREGIFYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_009348173.1 cytochrome c oxidase subunit 2 [Corallina offi... SIFIKREGVYGGQCEICGINHGFMPIVIEAVSLPNYI W TANKL

YP_009002113.1 cytochrome c oxidase subunit 2 [Sporolithon du... SIFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009040512.1 cytochrome c oxidase subunit 2 [Kappaphycus st... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

AXI97771.1 cytochrome c oxidase subunit 2 [Gracilaria tenuist... SIFLNREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_007890776.1 cytochrome c oxidase subunit II [Seculamonas e... TMFIKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TSSQI

YP_009490426.1 cytochrome c oxidase subunit 2 [Gracilaria ten... SIFLNREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_004062208.1 cox2 gene product [Gracilariopsis andersonii] SIFVKKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_009504387.1 cytochrome c oxidase subunit 2 [Gracilaria cho... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

AVK39596.1 cytochrome c oxidase subunit 2 [Sirodotia delicatula] SMFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W TANKL

YP_009511794.1 cytochrome c oxidase subunit 2 [Gracilaria cau... SIFIKREGIYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_009502225.1 cytochrome c oxidase subunit 2 [Porolithon onk... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009511819.1 cytochrome c oxidase subunit 2 [Gracilaria ferox] SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009490451.1 cytochrome c oxidase subunit 2 [Gracilaria tex... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_009316517.1 cytochrome c oxidase subunit 2 [Kappaphycus al... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009130770.1 cytochrome c oxidase subunit 2 [Gracilaria chi... SIFLNREGIYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_006234144.1 cytochrome c oxidase subunit 2 [Pyropia haitan... SIFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AXF36245.1 cytochrome c oxidase subunit 2 [Gracilariopsis ten... SMFVKKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

AHX02545.1 cytochrome c oxidase subunit 2 [Sebdenia flabellata] SIFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009019507.1 cytochrome c oxidase subunit 2 [Gracilaria sal... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

YP_009138127.1 cytochrome c oxidase subunit 2 [Calliarthron t... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W TANKL

APC24895.1 cytochrome c oxidase subunit 2 [Gracilariopsis and... SIFVKKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

AIB08178.1 cytochrome c oxidase subunit 2 (mitochondrion) [Py... SIFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_009317627.1 cytochrome c oxidase subunit 2 [Pterocladia mu... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009511894.1 cytochrome c oxidase subunit 2 [Gracilariopsis... SMFVKKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

AHX02521.1 cytochrome c oxidase subunit 2 [Schizymenia dubyi] SMFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

AHB34952.1 cytochrome c oxidase subunit 2 (mitochondrion) [Py... SIFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AHB35429.1 cytochrome c oxidase subunit 2 [Pyropia perforata] SIFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

AHX02424.1 cytochrome c oxidase subunit 2 [Ceranium japonicum] SVFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W SNKL

AVK39546.1 cytochrome c oxidase subunit 2 [Paralemanea sp.] SMFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W TANKL

YP_009317535.1 cytochrome c oxidase subunit 2 [Gelidium scler... SIFIKREGVYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009402605.1 cytochrome c oxidase subunit 2 [Sheathia arcuata] SMFVKKREGIYGGQCEICGVNHGFMPIVIEAVSLPNYI W TANKL

YP_009511945.1 cytochrome c oxidase subunit 2 [Melanthalia in... SMFVKKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

AIY34288.1 cytochrome c oxidase subunit 2 [Grateloupia taiwan... SIFIKREGIYGGQCEICGINHGFMPIVIEAVSLPNYI W SNKL

YP_009511844.1 cytochrome c oxidase subunit 2 [Gracilaria gra... SIFIKREGVYGGQCEICGINHGFMPIVIEAVKLSNYI W SNKL

Annexe S3 : Alignement multiple de séquences protéiques de Cox2 issues de différentes espèces non végétales sur BlastN, en partant de la séquence protéique de Cox2 d'*A. thaliana*. Exclusion faite sur le taxon 3193 avec les paramètres standards.

YP_003875498.1	cytochrome c oxidase subunit 2 [Silene latifolia]	VEAVSRKDYGSWYSTQLIPQTGEA	259
YP_009270661.1	cytochrome c oxidase subunit 2 [Nelumbo nucifera]	VEAVSRKDYGSWYSNQLIPQTGEA	260
BAD83476.2	cytochrome oxidase subunit 2 [Nicotiana tabacum]	VEAVSRKDYGSWYSNQLIPQTGEA	260
AXU98939.1	cytochrome c oxidase subunit 2 [Brassica juncea]	VEAVPRKDYGSRYSNQLIPQTGEA	260
CAA69761.3	cytochrome c oxidase subunit 2 [Arabidopsis thaliana]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AXU99057.1	cytochrome c oxidase subunit 2 [Brassica juncea]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AXS77402.1	Cox2 [Fallopia multiflora]	VEAVSRKDYGSRYANQLIPQAGEA	260
YP_009243659.1	cytochrome c oxidase subunit 2 [Cannabis sativa]	VEAVPRKDYGSRYSNQLIPQTGEA	259
YP_005090502.1	cytochrome c oxidase subunit 2 [Lotus japonicus]	VEAVSSKDYGSWYSNQLIPQ	255
YP_004927497.1	cox2-1 [Brassica oleracea]	VEAVPRKDYGSRYSNQLIPQTGEA	260
YP_009472098.1	cytochrome c oxidase subunit 2 [Arabidopsis th...	VEAVPRKDYGSRYSNQLIPQTGEA	260
ARM39101.1	cytochrome c oxidase subunit 2 [Isatis tinctoria]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AXU98876.1	cytochrome c oxidase subunit 2 [Brassica juncea]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AAB92366.1	cytochrome c oxidase subunit II [Brassica rapa sub...	VEAVPRKDYGSRYSNQLIPQTGEA	260
CBJ20691.1	cytochrome c oxidase subunit 2 [Beta vulgaris subs...	VEAVSRKDYGSWYSNQ	252
AXH81406.1	cytochrome c oxidase subunit 2 [Parinari campestris]	VEAVPRKDYGSWYSNQFIPQTEEA	260
CAD20888.1	cytochrome oxidase subunit 2 [Ruscus aculeatus]	VEAVSLKDYGSWYSNQLIPQTN	258
YP_009466051.1	cytochrome c oxidase subunit 2 [Arabis alpina]	VEAVPRKDYGSRYSNQLIPQTGEA	260
ANA04434.1	cytochrome c oxidase subunit 2 [Schrenkiella parvula]	VEAVPRKDYGSRYSNQLIPQTGEA	260
YP_004927539.1	cox2-1 [Brassica carinata]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AHH81707.1	cox2-1 [Brassica juncea]	VEAVPRKDYGSRYSNQLIPQTGEA	259
AXH81400.1	cytochrome c oxidase subunit 2 [Hirtella physophora]	VEAVPRKDYASWYSNQFIPQTEEA	260
AAB88866.1	cytochrome c oxidase subunit II [Raphanus sativus]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AFI44282.1	cytochrome c oxidase subunit 2 [Silene vulgaris]	VEAVSRKDYGSRYSTQLIPQTGEA	265
AIE42542.1	cytochrome c oxidase subunit 2 [Raphanus sativus]	VEAVPRKDYGSRYSNQLIPQTGEA	265
AGC81698.1	cytochrome c oxidase subunit 2 [Raphanus sativus]	VEAVPRKDYGSRYSNQLIPQTGEA	262
AUD56993.1	cytochrome oxidase subunit 2 [Luculia grandifolia]	VEAVSRKDYGSRYENQLIPQTGEA	260
YP_009477628.1	cytochrome c oxidase subunit 2 [Nymphaea color...	VEAVSLKDYVSWYSNQFIPQTEEA	258
CAA37047.2	cytochrome-c oxidase [Pisum sativum]	VEAVSSKDYGSWYSNQLXPQTG	257
BAD66739.1	cytochrome oxidase subunit 2 [Beta vulgaris subsp....	VEAVSRKDYGSRYSNQLIPQTGEA	263
CAA40875.1	NcoxII [Beta vulgaris subsp. vulgaris]	VEAVSRKDYGSRYSNQLIPQTGEA	260
AUD57000.1	cytochrome oxidase subunit 2 [Schizocolea linderi]	VEAVDRKDYGSRYENQLIPQTGEA	260
AAF43631.1	cytochrome oxidase subunit 2 [Pseudeminia comosa]	VEAVSTKDYGSWYFNQF	252
AEM46191.1	cytochrome c oxidase subunit 2 [Silene vulgaris]	VEAVSRKDYGSRYSTQLIPQTGEA	265
BAI82399.1	cytochrome c oxidase subunit 2 [Beta vulgaris subs...	VEAVSRKDYGSWYSNQ	252
ABB43241.1	cytochrome c oxidase subunit II [Solanum tuberosum]	VEAVSRKDDGSWYSNQLIPQTN	258
P98012.1	RecName: Full=Cytochrome c oxidase subunit 2; AltNam...	VEAVSRKDYGSRYSNQLIPQTGEA	260
AUD57032.1	cytochrome oxidase subunit 2 [Paederia foetida]	VEAVDRKDYGSRYENQLIPQTGEA	260
AUD57003.1	cytochrome oxidase subunit 2 [Palicourea guianensis]	VEAVDRKDYGSRYENQLIPQTGKA	260
AAF27821.1	cytochrome oxidase subunit 2 [Lespedeza thunbergii...	VEAVSTKDYGSWYFNQF	252
AUD57046.1	cytochrome oxidase subunit 2 [Theligonum cynocrambel]	VEAVDRKDYGSRYENELIPQTGEA	260
YP_009045795.1	cytochrome c oxidase subunit 2 [Batis maritima]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AUD56998.1	cytochrome oxidase subunit 2 [Ophiorrhiza mungos]	VEAVDRKDYGSRYENQLIPQTGEA	260
AUD57008.1	cytochrome oxidase subunit 2 [Pentas lanceolata]	VEAVDRKDYGSRYENQLIPQTGEA	260
AAF43639.1	cytochrome oxidase subunit 2 [Amphicarpaea bracteata]	VEAVSTKDYGSWYFSQF	252
AUD56999.1	cytochrome oxidase subunit 2 [Faramea multiflora]	VEAVDRKDYGSRYENQLIPQTGEA	260
AUD56996.1	cytochrome oxidase subunit 2 [Colletocema dewevrei]	VEAVDRKDYGSRYENQLIPQTGEA	260
AUD57012.1	cytochrome oxidase subunit 2 [Argostemma yappii]	VEAVDRKDYGSRYENQLIPQTGEA	260
AUD57005.1	cytochrome oxidase subunit 2 [Gaertnera phyllostac...	VEAVDRKDYGSRYENQLIPQTGPA	260
AUD57007.1	cytochrome oxidase subunit 2 [Danais xanthorrhoea]	VEAVDRKDYGSRYENQLIPQTGKA	260
AAF25663.1	cytochrome oxidase subunit 2 [Cologania lemmonii]	VEAVSTKDYGSWYFNQF	252
AUD57035.1	cytochrome oxidase subunit 2 [Plocama pendula]	VEAVDRKDYGSRYENELIPQTGEA	260
AUD57040.1	cytochrome oxidase subunit 2 [Rubia horrida]	VEAVDRKDYGSRYENKLIIPQTGEA	260
AAF43635.1	cytochrome oxidase subunit 2 [Teramnus uncinatus]	VEAVSTKDYGSWYFNQF	252
AAF45144.1	cytochrome oxidase subunit 2 [Dumasia villosa]	VEAVSTKDSGSWYFNQF	252
AUD57009.1	cytochrome oxidase subunit 2 [Kohautia caespitosa]	VEAVDRKDYGSRYENQLIPQTEQA	260
AUD57010.1	cytochrome oxidase subunit 2 [Anthospermum spathul...	VEAVDRKDYGSRYENQLIPQSGEA	260
YP_002608184.1	cytochrome c oxidase subunit 2 [Carica papaya]	VEAVPRKDYGSRYSNQLIPQTGEA	260
AUD57024.1	cytochrome oxidase subunit 2 [Guettarda scabra]	VEAVSRKDYASRYSNQLIPQTGEA	260
YP_009153933.1	cytochrome c oxidase subunit 2 [Gossypium hirs...	VEAVPRKDYGSRYSNQLIPQTGEA	260
YP_009412620.1	cytochrome c oxidase subunit 2 [Hibiscus canna...	VEAVPRKDYGSRYSNQLIPQTGEA	260
AUD57039.1	cytochrome oxidase subunit 2 [Rondeletia odorata]	VEAVSRKDYASRYSNQLIPQTREA	260

Annexe S4 : Alignement multiple de séquences protéiques de Cox2 issues de différentes espèces végétales sur BlastN, en partant de la séquence protéique de Cox2 d'*A. thaliana*.
Sélection faite sur le taxon 3193 avec les paramètres standards.

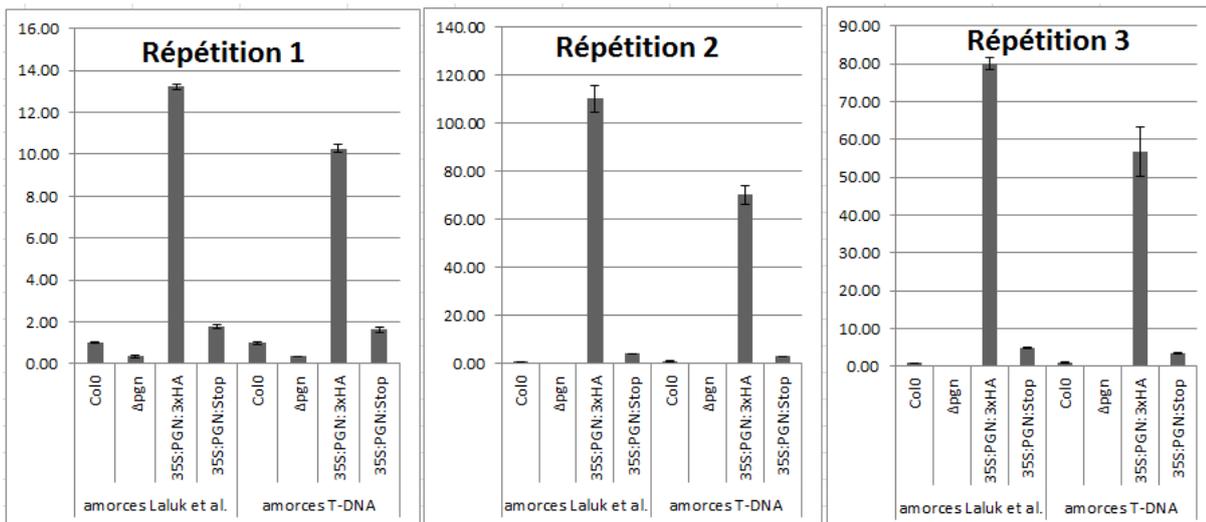
KJ485850.1	Rhazya stricta mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KF386162.1	Vaccinium macrocarpon mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
EU431224.1	Carica papaya mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
NC 037463.1	Citrus sinensis mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MG736621.1	Citrus sinensis mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MH004292.1	Chrysanthemum boreale mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KX063855.1	Diplostephium hartwegii voucher TEX:Vargas 456 mit...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KT959112.1	Castilleja paramensis mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
JN098455.1	Mimulus guttatus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KF177345.1	Salvia miltiorrhiza mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
JN107812.1	Boea hygrometrica mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
NC 026515.1	Hyoscyamus niger mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF579563.1	Nicotiana attenuata mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
LT855979.1	Betula pendula genome assembly, organelle: mitocho...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
XM 019402612.1	PREDICTED: Nicotiana attenuata NADH-ubiquinone...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
XM 016645336.1	PREDICTED: Nicotiana tabacum ribosomal protein...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KT997964.1	Nicotiana sylvestris cultivar TW 137 mitochondrion...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KT997964.1	Nicotiana sylvestris cultivar TW 137 mitochondrion...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KR780036.1	Nicotiana tabacum cultivar zhongyan 100 mitochondr...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KR780036.1	Nicotiana tabacum cultivar zhongyan 100 mitochondr...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KM207679.1	Nicotiana tabacum/Hyoscyamus niger hybrid clone co...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
BA000042.1	Nicotiana tabacum mitochondrial DNA, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF989958.1	Solanum tuberosum chromosome 1 mitochondrion, comp...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF989958.1	Solanum tuberosum chromosome 1 mitochondrion, comp...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF989954.1	Solanum tuberosum chromosome 2 mitochondrion, comp...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF989953.1	Solanum tuberosum chromosome 1 mitochondrion, comp...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF989961.1	Solanum commersonii strain Lz3.2 chromosome 2 mito...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
CP023756.1	Solanum lycopersicum cultivar I-3 chromosome 00	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF034194.1	Solanum pennellii bio-material TGRC:LA0716 mitoch...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF034193.1	Solanum lycopersicum bio-material TGRC:LA1479 mito...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MF034192.1	Solanum lycopersicum bio-material TGRC:LA1421 mito...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KU363907.1	Cannabis sativa cultivar Sievers Infinity mitochon...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KU310670.1	Cannabis sativa mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KR059940.1	Cannabis sativa cultivar Carmagnola mitochondrion,...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KRS559021.1	Heuchera parviflora var. saurensis voucher Folk 97...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
HG975450.1	Solanum pennellii chromosome chl1, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
HG975450.1	Solanum pennellii chromosome chl1, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
AK325645.1	Solanum lycopersicum cDNA, clone: LEFL2004G18, HTC...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KR610474.1	Nelumbo nucifera mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KJ865410.1	Capsicum annuum cultivar Jeju mitochondrion, compl...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KJ865410.1	Capsicum annuum cultivar Jeju mitochondrion, compl...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KJ865410.1	Capsicum annuum cultivar Jeju mitochondrion, compl...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KJ865409.1	Capsicum annuum cultivar CMS line FS4401 mitochond...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KJ865409.1	Capsicum annuum cultivar CMS line FS4401 mitochond...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KJ865409.1	Capsicum annuum cultivar CMS line FS4401 mitochond...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KC821969.1	Liriodendron tulipifera mitochondrion, complete ge...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KC821969.1	Liriodendron tulipifera mitochondrion, complete ge...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MG770607.2	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MG770607.2	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MG735191.1	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MG735191.1	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
MG735191.1	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KF815390.1	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KF815390.1	Helianthus annuus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
KP940485.1	Neobartsia pedicularoides NADH dehydrogenase subun...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
XM 006364193.2	PREDICTED: Solanum tuberosum NADH-ubiquinone o...	GAACAACCGCGCTGGTCTAATAGATCGACTTT
AP014526.1	Hevea brasiliensis mitochondrial DNA, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
AP014526.1	Hevea brasiliensis mitochondrial DNA, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT
GQ856147.1	Citrullus lanatus mitochondrion, complete genome	GAACAACCGCGCTGGTCTAATAGATCGACTTT

Annexe S5 : Aligement multiple de séquences génomiques de nad6 issues de différentes espèces végétales sur BlastN, en partant de la séquence génomique de nad6 d'*A. thaliana*.
Sélection faite sur le taxon 3193 avec les paramètres standards.

Annexe S6 : Gènes différentiellement exprimés dans la lignée mutante Δpgn en comparaison avec Col0, après correction des p-values par la méthode de Benjamini-Hochberg.

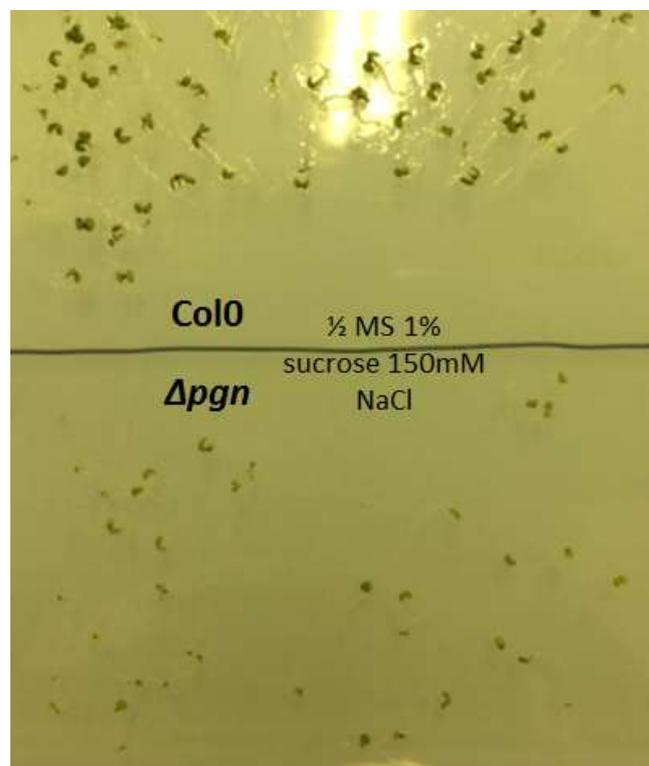
AGI	Comptages normalisés pour Col0	Comptages normalisés pour Δpgn	log2FoldChange	BH
AT1G02310	31.03525167	3.107024072	-3.320302521	0.00678076
AT1G02470	82.77568348	25.29090854	-1.710588153	0.000425529
AT1G04580	14.43182016	0.404014686	-5.158701718	0.001176803
AT1G07430	58.22547729	26.33855756	-1.144474222	0.024073059
AT1G12845	154.2668825	251.6675617	0.706090892	0.009613001
AT1G13830	18.74055378	3.133326592	-2.580396528	0.020593179
AT1G14250	1125.714924	2114.597034	0.909541238	2.59E-05
AT1G22550	170.6368941	81.29099449	-1.069762168	0.001488257
AT1G27030	240.9491274	380.1370671	0.657791132	0.035173789
AT1G29460	40.08184231	78.88466213	0.976795997	0.028525814
AT1G29510	40.93926815	81.10231736	0.986257825	0.011366176
AT1G34060	120.0950876	30.837291	-1.961429202	1.31E-05
AT1G45616	16.19669469	0.404014686	-5.32514788	0.000388288
AT1G49500	18.38005148	124.4419869	2.759260623	0.005635098
AT1G51780	1.408098473	49.69380643	5.141245918	3.01E-06
AT1G53480	237.3976827	2.180374039	-6.766586395	6.83E-70
AT1G53490	31.70228593	10.35132339	-1.614771648	0.007498848
AT1G53830	99.3268331	5.261095591	-4.238748346	1.19E-10
AT1G66100	3243.432702	5711.652778	0.816386773	0.013518663
AT1G66760	256.9040453	450.467181	0.810192394	0.010082083
AT1G67810	175.9631674	79.75001622	-1.141716758	0.00073013
AT1G73260	677.510628	227.4691731	-1.574572535	1.27E-05
AT1G76790	334.7220176	726.469613	1.117939001	0.008711027
AT1G76960	1188.068503	503.6096527	-1.238240183	2.42E-06
AT1G77960	73.73454054	134.1514949	0.863450625	0.009613001
AT1G79900	18.25055097	3.267324679	-2.481758289	0.045339999
AT2G02990	71.889119	21.00876369	-1.774782158	0.00395636
AT2G19970	24.93904152	1.624963413	-3.939926879	0.016667169
AT2G22810	8.421158324	28.87567012	1.777763834	0.009780201
AT2G23010	152.4451439	263.0311094	0.786943246	0.001734143
AT2G30770	59.96903074	160.090985	1.41660251	0.003912284
AT2G30830	8.437881776	58.98614421	2.805423328	0.00010094
AT2G33020	39.47313345	7.272264355	-2.440394498	0.00395636
AT2G37760	217.3984722	130.3200488	-0.738282752	0.014795147
AT2G38240	30.32881822	129.2846996	2.091790359	0.000136401
AT2G38530	101.7410617	2.303447033	-5.464963819	2.02E-13
AT2G39330	394.7512679	1000.201525	1.341274905	0.00395636
AT2G41480	22.22152013	4.736439819	-2.230082553	0.048038422

AT2G41800	152.5527179	6.609158162	-4.528697552	4.70E-17
AT2G43550	429.4733085	770.2070503	0.842677858	8.70E-05
AT3G09260	398.0291754	749.6787812	0.913398385	0.034918553
AT3G10320	38.3528445	10.77885078	-1.831130213	0.034693743
AT3G21500	147.8194138	49.69380643	-1.5726978	0.003359154
AT3G25900	32.41473377	9.106092565	-1.831745694	0.009520898
AT3G26330	0.277100461	10.37483609	5.226535621	0.00395636
AT3G29250	50.00295098	10.77885078	-2.213809871	0.007498848
AT3G46980	195.0888089	332.5698284	0.769526268	0.013049471
AT3G47340	759.3105815	1604.918516	1.079738033	1.00E-05
AT3G49210	140.3148607	60.89151409	-1.20435472	0.000425529
AT3G58770	0	4.2291819	Inf	0.038401788
AT3G63380	403.1034306	196.2630409	-1.038361543	4.87E-07
AT4G03320	29.94467174	7.193303217	-2.057572994	0.018410537
AT4G11910	66.33402134	20.71031699	-1.679399353	0.001247967
AT4G12480	211.5048357	41.71348447	-2.342104913	0.007625026
AT4G12490	282.4735767	43.14615797	-2.710811918	3.74E-13
AT4G16260	214.8859814	118.2491357	-0.86174173	0.002239227
AT4G21830	102.0978377	218.3282313	1.09654638	0.032916975
AT4G22880	29.17402271	4.655046772	-2.64781676	0.000397323
AT4G23680	114.6066043	40.1512818	-1.513172233	0.001944609
AT4G24000	186.7371625	53.13235681	-1.813346453	1.27E-05
AT4G24150	449.1910693	202.9177132	-1.146434438	0.000151924
AT4G30280	5.814341263	27.50618544	2.242068425	0.025541499
AT4G32810	41.03346445	2.424088118	-4.081286917	1.01E-06
AT4G34230	469.7810037	243.1279784	-0.950272453	0.007498848
AT4G35160	4.73330401	53.73840562	3.505034027	3.14E-06
AT4G37150	59.33992264	15.51083826	-1.935726389	0.047454793
AT5G02760	216.1237305	382.6000122	0.82397943	0.001176803
AT5G07100	106.8391881	53.22616014	-1.005233522	0.003238382
AT5G08330	289.3947148	421.6986204	0.543173727	0.048038422
AT5G16030	921.5874314	1499.090453	0.701894489	0.006732128
AT5G18030	34.85736328	69.88020386	1.003420374	0.049983091
AT5G23820	297.1658127	689.1958389	1.21364584	0.000666024
AT5G26260	6.221078085	37.98628519	2.610242113	0.001176803
AT5G39520	179.2257061	25.87464211	-2.792166763	3.16E-05
AT5G42800	34.4996013	6.007561678	-2.521728229	0.00010094
AT5G50950	733.1015725	1199.757983	0.710658408	0.018410537
AT5G51060	4.456203548	30.30110148	2.7654832	0.014159327
AT5G53320	21.70422065	1.890527679	-3.521114743	0.001250153
AT5G59320	26.58975136	0.548937866	-5.598083615	3.27E-08
AT5G65080	2.963661048	24.77515743	3.063441863	8.49E-06

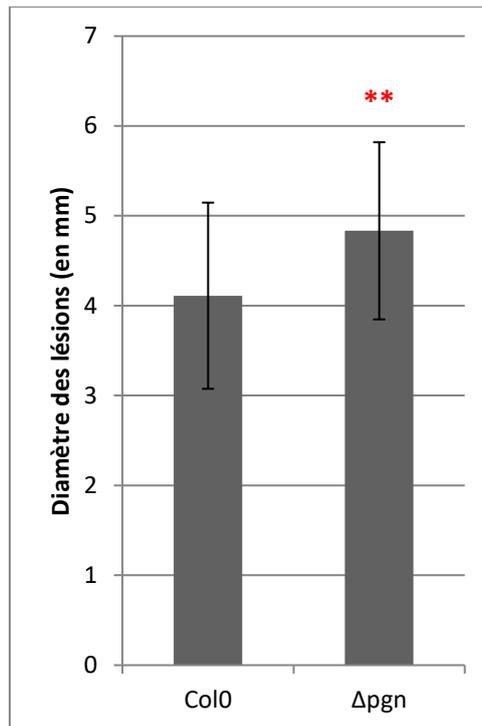


Annexe S7 : Représentations graphiques des données de RT-qPCR effectuées sur feuilles de plantes de 4 semaines en jours longs de différents génotypes

Chaque condition est composée d'une feuille d'une seule plante, les écart-types représentent la variabilité technique. Les répétitions sont biologiques et effectuées à un jour d'écart. Chaque mesure a été réalisée en triplicats techniques avec analyse de courbe de fusion. Les données sont corrigées par la méthode du $\Delta\Delta C_t$ sur l'actine de la plante et comparé à Col0.

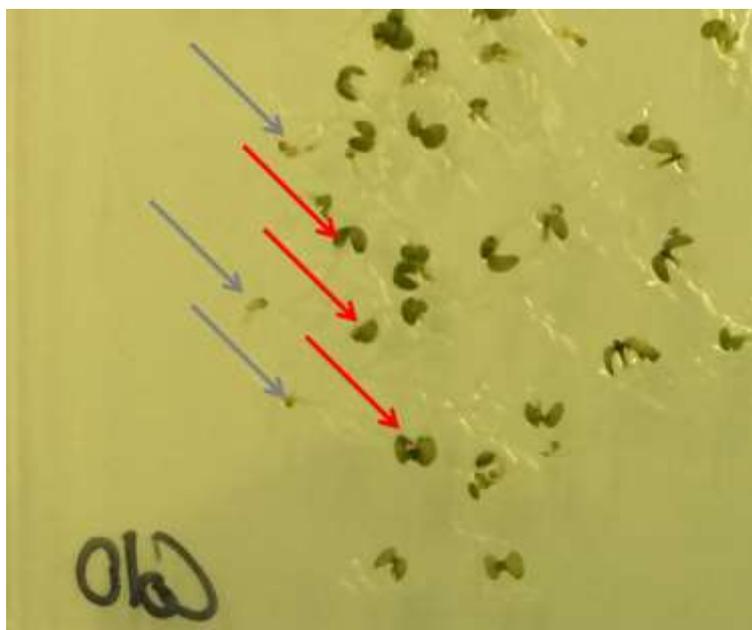


Annexe S8 : Comparaison entre plantules Col0 et Δ pgn sur milieu 1/2MS 1% sucrose supplémenté avec 150mM NaCl.



Annexe S9 : Mesure de tailles de lésions nécrotiques causées par *Botrytis cinerea* à 72h après infection par la méthode goutte de Col0 et Δpgn.

Les données sont basées sur 12 à 18 points de mesure par condition et deux répétitions biologiques. Le test statistique utilisé est un test de Student, les deux étoiles signifiant que la *p-value* est située entre 0.01 et 0.001.



Annexe S10 : Comparaison entre plantules non affectées (flèches rouges) et affectées (flèches grises) d'*A. thaliana* sur milieu 1/2MS 1% sucrose supplémenté avec 150mM NaCl.

Annexe S11 : Analyse par segmentation pour le chloroplaste des données issues du séquençage des petits ARN de la lignée de surexpression d'ArtRipL, comparées à Col0.

extrémité	Orientation	Segment	Col0	ArtRipL	log2FoldChange	Bonferroni	Annotation
fin	-	36812-36845	6175.60	12307.43	0.994877875	2.54E-02	En 3' du précurseur psaA-psaB-rps14
début	-	36842-36874	6136.93	11987.94	0.965994034	3.71E-02	
fin	-	48391-48414	55.47	97.41	0.812494786	0.02848934	dans le 3' du précurseur ndhK-ndhJ
début	-	48430-48438	37.73	73.39	0.959718091	7.99E-03	
fin	-	89448-89463	103.29	273.73	1.406024782	3.79E-06	anti-sens ycf2
début	-	89481-89488	93.54	252.62	1.433350888	1.70E-06	
fin	-	117186-117199	268.59	418.19	0.638742615	0.01982067	5' ndhD
début	-	117204-117234	272.00	419.94	0.626597033	1.46E-02	
fin	-	143794-143872	383.19	727.93	0.92574219	0.0001165	intergénique
début	-	143868-143890	311.45	612.24	0.975114156	4.62E-05	
fin	-	149120-149146	133.83	234.89	0.811579512	1.64E-02	anti-sens ycf2
début	-	149154-149169	133.15	228.99	0.782161069	1.72E-02	
fin	+	6873-6892	52.11	110.94	1.090160274	4.64E-03	5' UTR psbK
début	+	6893-6920	51.64	115.64	1.163214769	1.03E-03	
fin	+	8266-8299	491.57	750.33	0.610127851	2.32E-02	intergénique
début	+	8297-8351	517.48	790.05	0.610461004	0.02415806	
fin	+	32586-32610	1212.04	4321.85	1.834213321	3.6643E-11	dans le 5' du précurseur psbD-psbC-psbZ
début	+	32617-32655	1195.22	4311.86	1.851036826	1.06E-10	
fin	+	47200-47227	289.15	510.18	0.819187229	4.66E-04	intron trnL
début	+	47235-47268	263.29	474.71	0.850414494	5.39E-04	
fin	+	76612-76640	1105.69	2452.88	1.149532888	3.6827E-05	intron petD
début	+	76639-76690	1125.04	2481.00	1.140944164	2.60E-05	
fin	+	89479-89494	126.57	219.60	0.794973251	1.46E-02	interne ycf2
début	+	89502-89528	133.66	235.90	0.819638594	0.01640339	
fin	+	94758-94780	309.32	613.16	0.987181203	3.6827E-05	intergénique
début	+	94776-94854	383.55	728.61	0.925753793	0.0001165	
fin	+	149160-149167	92.47	255.24	1.464779414	7.605E-07	interne ycf2
début	+	149185-149200	102.88	267.07	1.376255385	4.58E-06	

Annexe S12 : Analyse par segmentation pour la mitochondrie des données issues du séquençage des petits ARN de la lignée de surexpression d'ArtRipL, comparées à Col0.

extrémité	Orientation	Segment	Col0	ArtRipL	log2FoldChange	Bonferroni	Annotation
fin	-	116580-116620	682.70	1315.60	0.946398897	1.24E-02	5' UTR cox1
début	-	116622-116636	617.42	1276.78	1.048191945	4.98E-03	
fin	-	124634-124665	3050.37	6550.61	1.102646256	0.008958775	trnH
début	-	124675-124677	558.99	1295.99	1.213153185	4.88E-04	
fin	-	180653-180680	1189.63	2097.55	0.818191761	4.21E-02	trnS
début	-	180695-180714	329.03	616.90	0.90683009	1.33E-02	
fin	-	321362-321400	238.79	447.38	0.905754736	0.014988158	trnE
début	-	321395-321439	560.27	1006.74	0.845502008	4.29E-02	
fin	-	321874-321907	51.78	146.40	1.499499299	5.25339E-05	trnS
début	-	321899-321942	33.87	63.59	0.908788674	4.33E-02	
fin	-	321908-321919	126.43	255.35	1.014083876	1.50E-02	trnS
début	-	321946-322250	176.74	373.68	1.080151606	2.57E-03	
fin	+	41450-41516	205.45	415.06	1.014533393	5.84E-03	trnS
début	+	41511-41522	126.13	246.24	0.965200503	2.41E-02	
fin	+	43542-43547	18157.24	130929.81	2.850176385	1.11E-08	trnN
début	+	43556-43578	17917.50	124501.82	2.796725544	7.11793E-11	
fin	+	49221-49228	167.24	447.92	1.421300995	1.77559E-05	5' UTR atp6
début	+	49243-49275	165.69	438.15	1.402912665	4.05E-05	
fin	+	308965-308976	949.68	1953.80	1.040776774	1.20E-02	3' ccmC
début	+	308985-309037	1127.68	2170.59	0.944727322	2.32E-02	

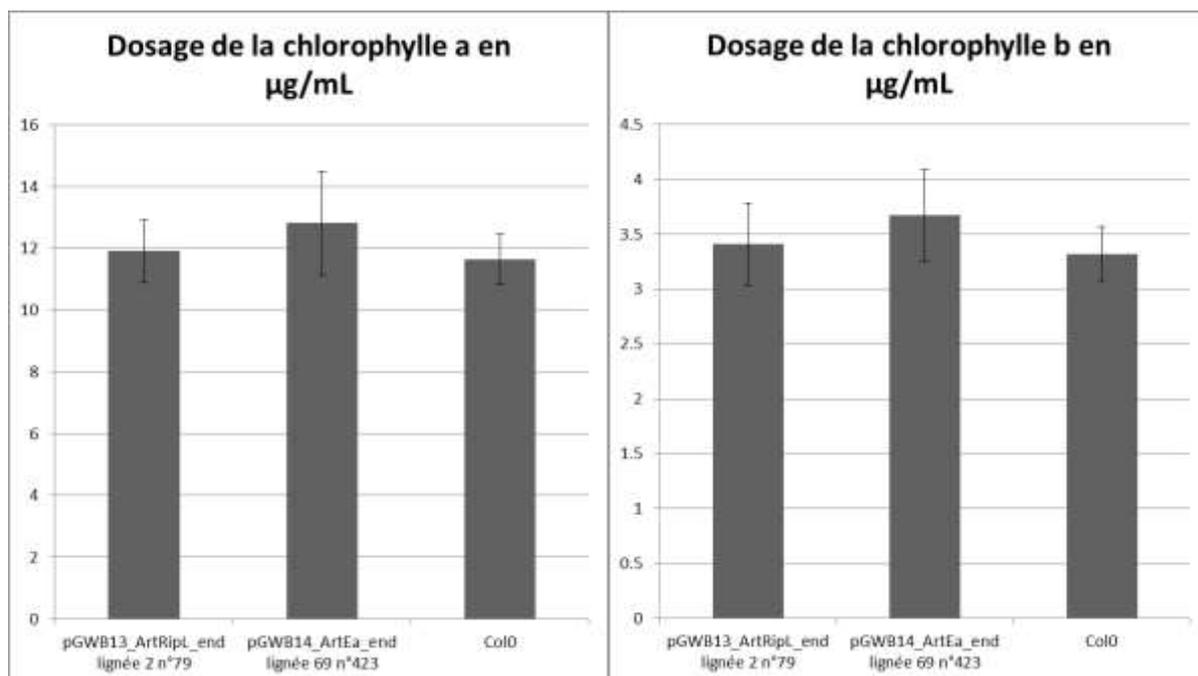
Annexe S13 : Gènes différentiellement exprimés dans la lignée de surexpression d'ArtRipL en comparaison avec Col0, après correction des *p-values* par la méthode de Benjamini-Hochberg.

AGI	Comptages normalisés pour Col0	Comptages normalisés pour ArtRipL	log2FoldChange	BH
AT1G01060	48.670128	99.475512	1.031305	0.046178
AT1G02920	330.976602	124.246488	-1.413524	0.039945
AT1G03940	7.297745	41.759958	2.516598	0.000212
AT1G05675	61.791251	6.382488	-3.275212	0.005785
AT1G07050	100.607466	42.707692	-1.236170	0.012717
AT1G07135	63.083708	5.460332	-3.530207	0.000010
AT1G08930	1506.836175	850.529719	-0.825089	0.038242
AT1G13340	72.980357	163.483729	1.163567	0.001082
AT1G15010	93.064063	13.968444	-2.736053	0.000212
AT1G16850	169.817086	68.451809	-1.310821	0.006227
AT1G20440	3539.207050	1550.430063	-1.190758	0.002356
AT1G22810	9.871083	0.875918	-3.494341	0.010498
AT1G25400	60.517789	22.788980	-1.409023	0.011581
AT1G27730	365.345085	46.962539	-2.959677	0.004807
AT1G06463	622.271121	1102.116147	0.824661	0.006867
AT1G32920	263.637691	102.093568	-1.368665	0.019806
AT1G32960	12.003858	1.193055	-3.330766	0.007042
AT1G35210	181.705643	18.988916	-3.258374	0.000148

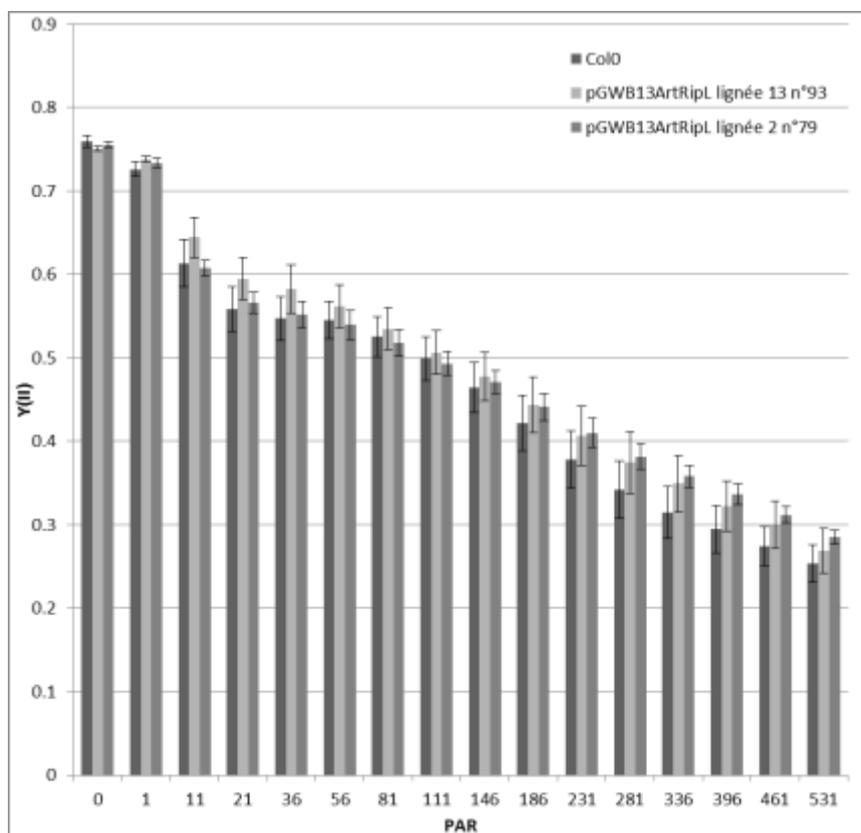
AT1G43160	90.768033	8.260807	-3.457830	0.000004
AT1G44800	39.552022	143.880392	1.863047	0.000200
AT1G07457	19.397508	3.060051	-2.664244	0.001469
AT1G49320	5.730334	43.675094	2.930120	0.000000
AT1G51090	221.906085	116.262990	-0.932557	0.020739
AT1G54870	56.166213	0.654108	-6.424031	0.000000
AT1G56240	51.639172	2.939229	-4.134956	0.000000
AT1G63750	289.940546	127.793625	-1.181941	0.002927
AT1G63820	12.661756	1.207227	-3.390709	0.004653
AT1G65390	17.253329	3.881219	-2.152293	0.044641
AT1G66090	383.627936	45.688369	-3.069809	0.002393
AT1G66160	103.914279	45.872452	-1.179694	0.009854
AT1G67105	12.321635	53.417908	2.116130	0.000110
AT1G69840	822.438111	458.279072	-0.843681	0.031095
AT1G70640	26.706002	7.348071	-1.861726	0.014702
AT1G72416	553.270335	327.309525	-0.757329	0.008132
AT1G73260	12.380590	0.377548	-5.035277	0.001823
AT1G75830	82.849531	0.437959	-7.563554	0.000000
AT1G78410	49.509259	7.987097	-2.631955	0.001435
AT1G78440	95.963977	192.982613	1.007906	0.019806
AT1G78970	131.124993	56.514936	-1.214239	0.005950
AT1G80760	37.897439	89.445527	1.238909	0.017767
AT1G80840	183.819635	24.275578	-2.920713	0.000017
AT5G10340	146.116739	263.001438	0.847949	0.018983
AT5G10420	1.498047	9.566211	2.674865	0.039112
AT5G12030	15.007498	0.377548	-5.312880	0.002626
AT5G13930	2465.846889	4935.857476	1.001218	0.005747
AT5G17220	121.042404	415.088368	1.777906	0.000000
AT5G17460	338.277462	178.410654	-0.923005	0.000506
AT5G23660	199.068400	65.593818	-1.601632	0.000003
AT5G23730	85.221201	161.155097	0.919166	0.023087
AT5G24470	426.384817	258.843659	-0.720075	0.026355
AT5G26920	434.126501	89.051949	-2.285396	0.021727
AT5G28030	188.774667	323.613843	0.777608	0.021268
AT5G36925	23.822624	3.221450	-2.886550	0.000172
AT5G37770	573.025966	297.534473	-0.945544	0.018518
AT5G39410	358.881485	192.230130	-0.900673	0.011201
AT5G42380	9.295480	0.714519	-3.701486	0.013920
AT5G42800	44.185548	329.363288	2.898033	0.000000
AT5G44120	126.373979	1.962323	-6.008993	0.000000
AT5G44420	90.881196	16.161998	-2.491376	0.017009
AT5G44430	25.727316	1.267638	-4.343086	0.000042
AT5G45340	188.424585	11.291642	-4.060660	0.000000
AT5G46600	175.386985	97.910449	-0.841007	0.037301
AT5G46710	310.539862	155.692786	-0.996076	0.003616
AT5G50600	76.352122	0.377548	-7.659865	0.000000
AT5G50700	76.352122	0.377548	-7.659865	0.000000
AT5G51190	124.298679	7.794542	-3.995203	0.000000
AT5G52310	4711.738634	1892.554466	-1.315925	0.000169
AT5G54060	28.212443	178.516305	2.661652	0.000000
AT5G57220	130.000648	23.214707	-2.485408	0.002468
AT5G57560	507.454226	100.181404	-2.340663	0.000000
AT5G61600	308.073095	27.500787	-3.485728	0.003214
AT5G62165	98.332742	40.345176	-1.285276	0.003051
AT5G62470	539.289902	276.789674	-0.962271	0.019806
AT5G64750	66.409428	10.511012	-2.659486	0.000986
AT5G65080	16.937544	217.399602	3.682053	0.000000
AT5G65920	97.469253	45.092931	-1.112046	0.017060
AT5G66150	10.418300	37.028400	1.829512	0.001466

AT5G67480	594.378170	334.980433	-0.827304	0.012768
AT3G05880	2090.291250	1188.361393	-0.814730	0.006227
AT3G09870	51.433920	3.774569	-3.768336	0.000060
AT3G10720	1447.870323	781.192960	-0.890182	0.007748
AT3G10930	25.692358	3.990718	-2.686619	0.027308
AT3G15670	49.413671	0.714519	-6.111795	0.000004
AT3G22142	213.177940	90.633202	-1.233947	0.000272
AT3G22640	115.496142	1.132644	-6.672007	0.000000
AT3G22930	17.456884	3.673581	-2.248537	0.021809
AT3G24750	12.177861	75.062717	2.623836	0.000000
AT3G25180	153.232212	46.000760	-1.735990	0.035904
AT3G26830	36.660928	5.884118	-2.639345	0.021809
AT3G26910	199.596072	114.086620	-0.806954	0.049487
AT3G27660	31.597862	0.377548	-6.387024	0.000001
AT3G28290	495.261864	155.447499	-1.671764	0.000000
AT3G28300	495.261864	155.447499	-1.671764	0.000000
AT3G28340	143.863512	54.995164	-1.387324	0.003616
AT3G28500	48.844640	19.432574	-1.329723	0.046725
AT3G29590	23.160338	98.668190	2.090929	0.000003
AT3G30460	18.954931	45.183467	1.253222	0.039112
AT3G30775	97.571433	34.613907	-1.495107	0.023867
AT3G44260	652.494491	199.351408	-1.710652	0.030097
AT3G45960	9.286746	1.106239	-3.069510	0.046477
AT3G46230	17.205270	0.437959	-5.295911	0.005343
AT3G48520	238.269911	43.795884	-2.443730	0.000002
AT3G48740	567.163111	325.307413	-0.801960	0.007748
AT3G49300	1.936257	14.707466	2.925206	0.003616
AT3G50930	196.461400	44.085700	-2.155863	0.000025
AT3G50970	253.467378	95.833875	-1.403192	0.013470
AT3G51240	224.897601	432.291052	0.942735	0.000257
AT3G51860	42.800726	11.279288	-1.923959	0.014507
AT3G52450	30.250669	10.185403	-1.570464	0.021727
AT3G55980	1106.058587	255.341493	-2.114928	0.000598
AT3G56880	370.603309	175.227258	-1.080649	0.001291
AT3G57450	286.565260	136.885974	-1.065889	0.006227
AT3G57530	943.586687	434.015336	-1.120409	0.013920
AT3G60160	309.731808	167.820633	-0.884099	0.002052
AT3G61190	38.402661	9.511461	-2.013467	0.045528
AT3G63380	18.032169	1.313877	-3.778671	0.002703
AT2G02100	658.542640	368.856710	-0.836216	0.017009
AT2G02450	288.446378	162.178159	-0.830724	0.021577
AT2G05380	606.315653	284.903241	-1.089597	0.010121
AT2G18660	60.333356	12.358213	-2.287486	0.000118
AT2G20750	319.495899	524.454746	0.715021	0.020539
AT2G21660	1928.553316	1177.792985	-0.711433	0.045748
AT2G22500	697.614989	303.866239	-1.198995	0.044641
AT2G22880	62.998931	14.528135	-2.116478	0.005554
AT2G24540	135.012304	231.871406	0.780234	0.035904
AT2G25735	224.572744	41.797691	-2.425688	0.003532
AT2G26530	487.161367	240.987150	-1.015444	0.034793
AT2G29460	26.235229	6.221089	-2.076266	0.038399
AT2G30300	8.131004	30.547735	1.909560	0.002453
AT2G32030	32.484204	3.855723	-3.074665	0.007089
AT2G32200	24.931495	2.783491	-3.163002	0.000350
AT2G32210	48.244680	10.286391	-2.229633	0.023504
AT2G33850	326.051585	84.964124	-1.940175	0.000000
AT2G35710	79.854308	26.603255	-1.585767	0.029102
AT2G38240	192.884718	22.647376	-3.090323	0.008725
AT2G39200	33.506603	7.410423	-2.176818	0.006758

AT2G39650	158.053206	59.807998	-1.402000	0.001633
AT2G41100	1926.831178	578.865905	-1.734929	0.000145
AT2G43150	216.183934	99.019217	-1.126479	0.005950
AT2G43570	53.264562	12.788570	-2.058321	0.005950
AT2G44840	55.938321	15.818456	-1.822228	0.010498
AT2G46400	152.808292	69.130464	-1.144329	0.038399
AT4G00970	555.535333	286.425012	-0.955721	0.012822
AT4G01390	98.434690	22.043465	-2.158816	0.000000
AT4G02330	221.904729	75.144947	-1.562192	0.000012
AT4G08040	45.692520	5.811475	-2.974982	0.000000
AT4G09820	1.676757	12.325116	2.877856	0.039112
AT4G11280	775.457444	248.460828	-1.642029	0.000900
AT4G11290	23.119868	2.184133	-3.404001	0.000005
AT4G11310	11.586635	0.437959	-4.725522	0.021727
AT4G12480	9.585720	0.654108	-3.873287	0.021268
AT4G12720	823.920112	433.803425	-0.925463	0.013007
AT4G13577	559.040504	919.820142	0.718399	0.021727
AT4G14090	31.653468	134.800595	2.090391	0.001633
AT4G14370	63.734059	12.929226	-2.301429	0.005343
AT4G17860	17.589740	3.742503	-2.232659	0.013881
AT4G21680	18.538725	0.815507	-4.506702	0.000040
AT4G22880	59.947867	354.128457	2.562492	0.000000
AT4G23190	173.593484	55.389613	-1.648025	0.000049
AT4G23810	275.565932	37.037782	-2.895328	0.000000
AT4G24570	507.692870	50.919411	-3.317668	0.000000
AT4G27280	487.227107	117.791318	-2.048361	0.000208
AT4G29020	302.165829	512.740842	0.762889	0.008725
AT4G29780	151.523871	44.503825	-1.767544	0.011996
AT4G30280	65.285251	6.462733	-3.336541	0.000000
AT4G30290	33.534310	7.255595	-2.208472	0.021727
AT4G30350	327.207857	181.191782	-0.852690	0.002468
AT4G31800	266.547077	151.091034	-0.818972	0.013279
AT4G33550	24.157530	67.036867	1.472482	0.045359
AT4G34480	771.549119	1455.537776	0.915722	0.001566
AT4G36600	8.780338	0.377548	-4.539545	0.040714
AT4G37260	224.401771	118.009408	-0.927182	0.016909
AT4G37290	15.319573	0.276560	-5.791641	0.000016
AT4G37990	26.278391	3.734901	-2.814735	0.041538
AT4G39670	20.745062	4.253105	-2.286180	0.008633
ATCG00950	16486.923593	32477.284963	0.978109	0.003616
ATCG01180	16486.923593	32477.722922	0.978128	0.003616
PPR_Rs	0.000000	1189.279630	Inf	0.000000

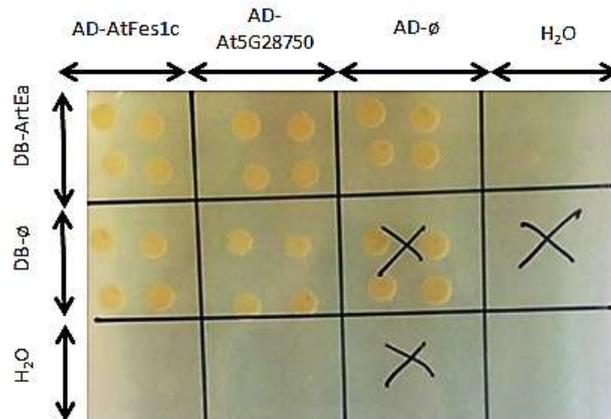


Annexe S14 : Dosage des chlorophylles a et b dans différents fonds génétiques.
 Réplicat biologique représentatif parmi les deux effectués. Les écarts-types représentent la variabilité parmi les 10 plantes testées par condition.



Annexe S15 : Rendement du photosystème II en fonction d'un rayonnement photosynthétiquement actif (RPA) croissant pour 3 génotypes d'*A. thaliana*.

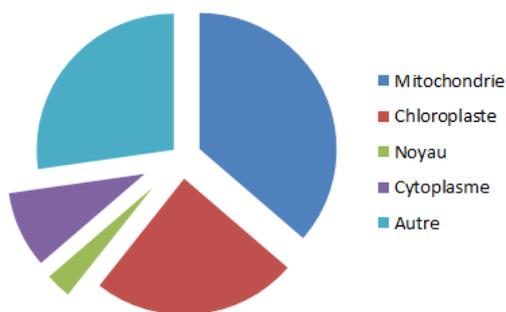
Les écarts-types représentent la variabilité observée entre les 10 plantes testées par condition.



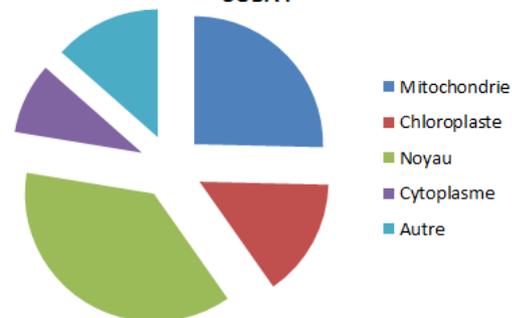
Annexe S16 : Mise en évidence de la pousse sur milieu de sélection des diploïdes des interactions testées précédemment.

AT5G28750 sert de contrôle négatif d'interaction, tout comme les construits AD et DB vides (\emptyset) et l'eau.

Localisations subcellulaires observées pour les interactants d'AtFes1c selon SUBA4



Localisations subcellulaires observées et prédites pour les interactants d'AtFes1c selon SUBA4



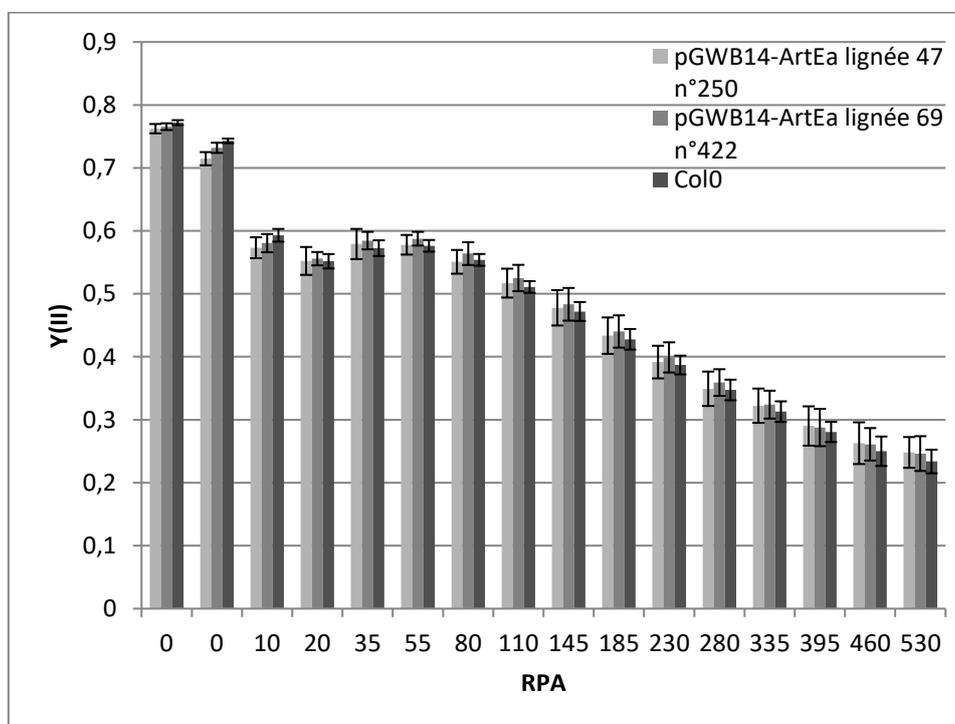
Annexe S17 : Localisations subcellulaires observées selon SUBA4 (à gauche) ; observée selon SUBA4 et prédites par SUBAcon (à droite) pour les interactants d'AtFes1c chez *A. thaliana*.

Annexe S18 : Gènes différemment exprimés dans la lignée de surexpression d'ArtEa en comparaison avec Col0, après correction des *p-values* par la méthode de Benjamini-Hochberg.

AGI	Comptages normalisés pour Col0	Comptages normalisés pour ArtEa	log2FoldChange	BH
AT1G22570	375.541899	699.819553	0.898009	0.000131
AT1G30900	136.722659	237.911398	0.799172	0.017931
AT1G44350	882.675626	1275.680848	0.531312	0.014432

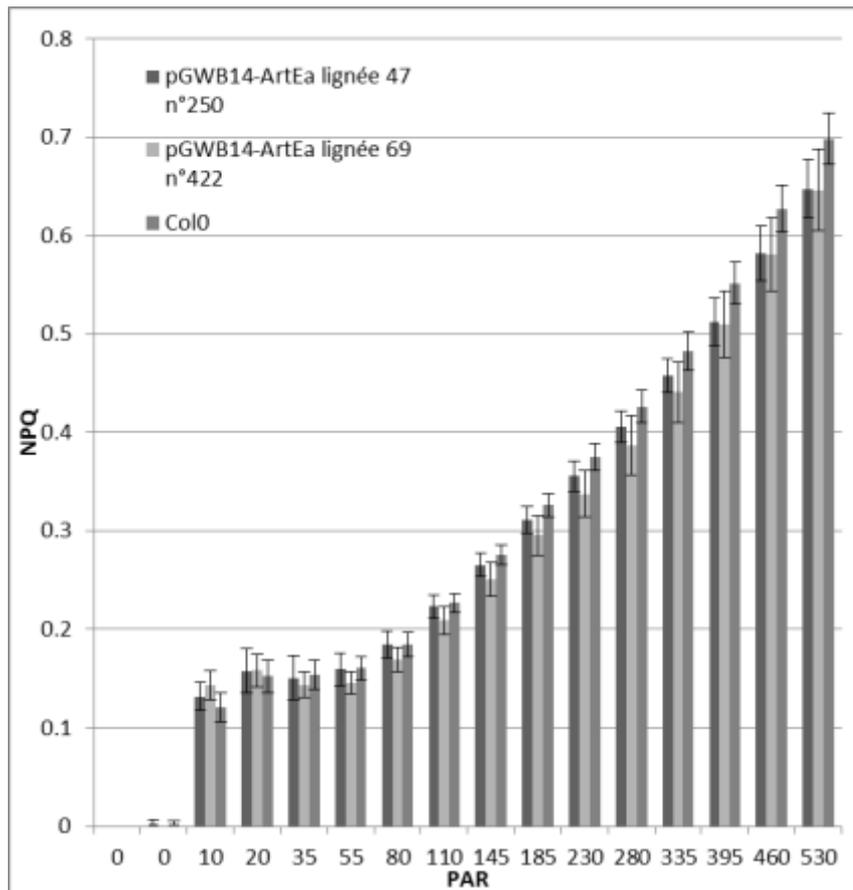
AT1G44800	46.881254	129.077660	1.461156	0.000099
AT1G44830	18.066688	2.203385	-3.035538	0.007118
AT1G45145	354.086809	786.817668	1.151926	0.004855
AT1G53480	542.361737	2.081918	-8.025198	0.000000
AT1G56120	178.759087	308.138599	0.785563	0.036770
AT1G56240	60.980674	5.580397	-3.449912	0.000068
AT1G58190	12.329030	38.935290	1.659019	0.027824
AT1G73750	734.299320	1187.480509	0.693464	0.026056
AT1G75830	98.376854	0.443052	-7.794700	0.000000
AT1G75945	3.556536	208.549888	5.873776	0.000000
AT1G78970	155.241500	65.262600	-1.250186	0.000954
AT1G80830	115.097903	213.424493	0.890864	0.024063
AT5G01380	1.880147	20.601128	3.453806	0.036770
AT5G01600	4424.578305	2647.074994	-0.741141	0.013353
AT5G05600	1266.098451	2083.941126	0.718925	0.001922
AT5G07010	145.612814	55.002079	-1.404579	0.000461
AT5G09220	277.472082	852.443007	1.619261	0.001922
AT5G10760	963.818591	1678.063197	0.799964	0.033368
AT5G17220	143.339425	535.417270	1.901228	0.000000
AT5G22580	1651.163676	1070.985247	-0.624545	0.029820
AT5G23240	380.353891	716.790130	0.914208	0.001922
AT5G35760	2.298977	23.460235	3.351154	0.041958
AT5G43780	545.051125	933.517142	0.776285	0.001617
AT5G45340	222.668795	29.585442	-2.911940	0.008863
AT5G45380	160.287317	259.119579	0.692958	0.041958
AT5G47230	314.993499	78.229946	-2.009529	0.046900
AT5G50600	90.664729	1.084662	-6.385224	0.000000
AT5G50700	90.664729	1.084662	-6.385224	0.000000
AT5G54060	33.411757	214.996483	2.685885	0.000002
AT5G57220	153.549964	39.359486	-1.963925	0.041958
AT5G57760	85.105728	30.772288	-1.467624	0.014432
AT5G62165	116.458031	48.016047	-1.278222	0.014146
AT5G63450	57.864815	298.827331	2.368554	0.000001
AT5G65070	38.858374	115.966023	1.577405	0.000099
AT5G65080	20.052006	264.946672	3.723884	0.000000
AT3G02020	953.793169	1733.873633	0.862250	0.041958
AT3G04720	437.810863	744.842782	0.766628	0.001215
AT3G09940	39.175846	150.402079	1.940788	0.007898
AT3G10930	30.339465	4.174151	-2.861641	0.018503
AT3G11010	564.574244	916.630796	0.699177	0.029847
AT3G21370	47.416578	0.443052	-6.741772	0.000001
AT3G22275	38.648488	126.966270	1.715961	0.012203
AT3G22740	97.622922	185.314371	0.924683	0.022719
AT3G25760	1153.677422	2234.887030	0.953962	0.000533
AT3G26210	106.656194	203.998834	0.935593	0.002595

AT3G26910	236.306437	142.583548	-0.728851	0.046900
AT3G27660	37.517826	0.774951	-5.597328	0.000915
AT3G28270	102.971715	199.602898	0.954885	0.048930
AT3G28540	184.629141	380.355485	1.042718	0.000043
AT3G29590	27.447715	136.547168	2.314641	0.029847
AT3G33528	6.582810	0.331899	-4.309887	0.033472
AT3G44970	26.967280	74.019825	1.456702	0.004942
AT3G48520	282.959238	1103.739700	1.963734	0.000001
AT3G53040	16.256864	0.542331	-4.905732	0.029501
AT3G55920	267.946107	130.366024	-1.039375	0.000147
AT3G55980	1307.702590	446.556516	-1.550120	0.018731
AT3G57260	289.413214	657.894750	1.184726	0.004855
AT3G61280	337.979618	569.883815	0.753732	0.013324
AT2G23170	56.731694	22.961640	-1.304929	0.040394
AT2G26560	196.573403	383.293896	0.963383	0.000051
AT2G30766	150.569408	270.791662	0.846755	0.017293
AT2G32290	200.900060	351.314778	0.806286	0.004760
AT2G41240	1.149030	18.152969	3.981717	0.004584
AT4G02330	262.714081	122.428513	-1.101554	0.000439
AT4G04223	85.675427	211.905388	1.306467	0.000406
AT4G08300	21.395014	95.270784	2.154759	0.015951
AT4G10120	730.897806	1251.810870	0.776275	0.003601
AT4G10500	4.913452	31.136476	2.663797	0.003601
AT4G11280	917.126571	339.766004	-1.432579	0.002595
AT4G11310	13.758711	0.443052	-4.956726	0.013682
AT4G14090	37.488466	152.029847	2.019836	0.007118
AT4G15233	209.623473	369.424153	0.817478	0.028796
AT4G15530	655.904996	1272.191486	0.955757	0.004974
AT4G16740	99.442667	247.023240	1.312710	0.040715
AT4G20110	182.912227	301.580400	0.721391	0.042315
AT4G21870	185.921802	67.041820	-1.471563	0.018731
AT4G23190	205.364396	83.193865	-1.303637	0.001614
AT4G23260	325.142501	540.304954	0.732702	0.010081
AT4G23810	325.768736	90.935006	-1.840940	0.013619
AT4G24570	599.780206	116.272817	-2.366920	0.006566
AT4G28520	195.916048	2.148695	-6.510631	0.000000
AT4G32280	124.272251	52.675531	-1.238299	0.004814
AT4G32460	329.388551	197.643300	-0.736891	0.026998
AT4G36670	299.985178	529.119684	0.818703	0.002076
AT4G37290	18.092217	0.443052	-5.351750	0.000102
PPR_Ea	0.732187	1929.296951	11.363575	0.000000



Annexe S19 : Rendement du photosystème II en fonction d'un rayonnement photosynthétiquement actif (RPA) croissant pour 3 géotypes d'*A. thaliana*.

Expérimentation représentative parmi les deux répétitions biologiques effectuées. Les écarts-types représentent la variabilité observée entre les 10 plantes testées par condition.



Annexe S20 : Observation du NPQ (*Non-Photochemical Quenching*) en fonction d'un rayonnement photosynthétiquement actif (RPA) croissant pour 3 géotypes d'*A. thaliana*.
 Expérimentation représentative parmi les deux répétitions biologiques effectuées. Les écarts-types représentent la variabilité observée entre les 10 plantes testées par condition.

Liste des amorces PCR utilisées pour le clonage.

Projet	Objectif	Nom	Séquence 5' vers 3'
Amorces générales méthode Gateway	Clonage	U5_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAAC CATG
		U3endstop_R	AGATTGGGGACCCTTTGTACAAGAAAGCTGGGTCTCCACCTCCGG ATC
	Génotypage clones	DNR5 _F	CTGGCAGTTCCTACTCTCG
		DNR3 _R	GATGGTCGGAAGAGGCATAA
		AD_F	CGCGTTTGAATCACTACAGGG
		DB_F	GGCTTCAGTGGAGACTGATATGCCTC
		B2_R	CCATCAAACCCTTTGTACAAGAAAGTTGGG
p35S-B_F	AGTGGAAAAGGAAGGTGGCT		
PGN	Clonage	PGN_end_F	GAAGTCGGAAGTGGAGATCCGGAGGTGG
		PGN_end_R	CCACCTCCGGATCTCCAGTTCGGACTTC
		PGN_Stop_F	GGAGATAGAACCATGAGCATAACCAAACCTGCACGCTCC
		PGN_Stop_R	TCCACCTCCGGATCTCAAGTTCGGACTTCTTGATCATTAAAC
	Séquençage	PGN_seq1_F	GCAAGCAAATTCATGCGTCTGTG
	Génotypage et qPCR	qPCR_PGN_F	ACCGGGTTTACGCATTTGGGTG
		qPCR_PGN_R	CGCATGAATTTGCTTGCCGGTTGTAAC
PPR bactériennes : RipL	Clonage mutation ponctuelle	ArtRs_Stop_Muta_F	GGGTAAGGGGAAAAGCTAAGTGAGACCCAGCTTTC
		ArtRs_Stop_Muta_R	GAAAGCTGGGTCTCACTTAGCTTTTCCCTTACCC
	Clonage double hybride levure	ArtRs_Y2H_F	GGAGATAGAACCCTGATGCTTCTTCACAGTTGGGAGAT
		ArtRs_Y2H_R	TCCACCTCCGGATCTCACTTAGCTTTTCCCTTACCC
		ArtRs_Y2H_P1_2_P1_3_F	GGAGATAGAACCATGGATGCTTCTTCACAGTTGGGAGAT
		ArtRs_Y2H_P1_2_R	TCCACCTCCGGATCACTCAGTGAGAGCAACAGAAACAGC
		ArtRs_Y2H_P2_2_F	GGAGATAGAACCATGCTTATCGCTCATCTCACACCTGGA
		ArtRs_Y2H_P2_2_P3_3_R	TCCACCTCCGGATCACTTAGCTTTTCCCTTACCCCTAGTTCC
		ArtRs_Y2H_P1_3_R	TCCACCTCCGGATCAAGCAGCCTCAACAAGAGCAAG
		ArtRs_Y2H_P2_3_F	GGAGATAGAACCATGGGTAGACTTAGAGCTGAAGCTGCT
		ArtRs_Y2H_P2_3_R	TCCACCTCCGGATCAAGTAGGAAGAAGCTGCTCTGGC
	ArtRs_Y2H_P3_3_F	GGAGATAGAACCATGGCTGCTCATCTTCTTCAGGTGG	
	Clonage pré-séquence	ArtRs_preseq_R	TCCACCTCCGGATCTGAGGAGTGAAGTGGTGAGCTT
	Génotypage et qPCR	ArtRs_geno_F	GACACTTGCTGATCTCGCTG
		ArtRs_geno_R	CTCCGATAGCAGGCTTAGCT
	Séquençage	ArtRs_seq1_F	CCAGATGCTGGATACGATGC
		ArtRs_seq2_F	ATACCCTGGTGCATCTGCTG
ArtRs_seq3_F		CTTATCGTCTCATCTCACACCTG	
PPR bactériennes : EAMY_0820	Clonage double hybride levure	ArtEa_Y2H_F	GGAGATAGAACCCTGGATCTACCCTACCCTTACGATG
		ArtEa_Y2H_R	TCCACCTCCGGATCAAACAGCAGGAGAGTAAGC
		Fes1_Alpha_Beta_F	GGAGATAGAACCATGGCGGACGGCGGAC
		Fes1_Alpha_R	TCCACCTCCGGATCAATTCTGAATGTGAGCTA
	Fes1_Beta_R	TCCACCTCCGGATCAATTCTGTGTGAGCTG	
	Clonage pré-séquence	ArtEa_preseq_R	CTCCACCTCCGGATCTTCCCTGCTTGCAGAGC
	Clonage injection et loc. subcellulaire	GFP11_PPR_Ea_F	GATAGAACCATGGGTATCGCTAACGCGACATCG
		GFP11_PPR_Ea_R	CTCCACCTCCGGATCCAACGGCAGGTGAATATG
		GFP11_hrpW_F	GATAGAACCATGGGCTGCTGCTGCTGCTGCCCT
	GFP11_hrpW_R	CCTCCGGATCTTCAGCCACCTTCAGGTTGGC	
	Génotypage et qPCR	ArtEa_Geno2_F	TAAGGCTCCTGAGAAGGCTG
		ArtEa_Geno2_R	GCAAAGACTGGCACACATCA
		qGFP11_hrpW_F	CCTGAAGGTGGCTGAAGGAT
		qGFP11_R	GTACTCGTGCAGCACCATGT
		qGFP11_PPR_Ea_F	ATATACGCCCTGGACACCTG
PPREa_qPCR_F		CGCGCGATGAAAAGAGCTAA	
PPREa_qPCR_R	CAGGTGTCCAGGGCGTATAT		

Liste des amorces PCR utilisées pour mesurer l'édition.

Projet	Nom	Séquence 5' vers 3'
Validation méthodologie prédiction PPR E+	ATP4_F	TTGTGCATTAAGTTCGAAGAAGA
	ATP4_R	GCAAATTGCTTCCCCACTAA
	CCMB_3_F	CCACACCAATCACGAGTTTTT
	CCMB_3_R	GAGGGTCCAGAAGAGAATTGG
	CCMB_5_F	TACGTACGGTGGAGGTTGGT
	CCMB_5_R	TGGGAAAACCACAGAAAACA
	CCMC_3_F	ACCGGTATAGAAATGGGTGCT
	CCMC_3_R	TGGGCGCTTTAACCATTAG
	CCMC_5_F	AGCTACGCGCAAATTCTCAT
	CCMC_5_R	TCCGGAAGAGCGAAGATAAA
	CCMFC_3_F	GATTGCCTTTGGGTTACGAA
	CCMFC_3_R	CCGCAGCCACTACTTTGACT
	CCMFC_5_F	TGGTCCAACACTACATAACTTTTTCTTTT
	CCMFC_5_R	ATTCGTCGTTGCTCATTCC
	COX2_3_F	TCCGATGAGCAGTCACTCAC
	COX2_3_R	CCCCTCCCTCACCTACTCT
	COX2_5_F	GGGAGCAGAAAATGAAGGAA
	COX2_5_R	GTGACTGCTCATCGGAACTG
	COX3_F	AACCACCGTAGGAGGTGTGA
	COX3_R	AAACCATGAAAGCCTGTTGC
	MATR_3_F	ATCAAGGGGGATCTCGTTCT
	MATR_3_R	GCTTCCCAAGCTCTATGCTG
	MATR_5_F	TTACGATCCCGAGTTCCAG
	MATR_5_R	CCACAATGATGGCTCTGTTG
	MTTB_3_F	GCCTTTTCGACATTTGTTGC
	MTTB_3_R	TCTCCCTCATTCCACTCGTC
	MTTB_5_F	TCCTTCACTTTTAGCTTTGAATTACT
	MTTB_5_R	ACGAGTCAAAAGGCAGGGTA
	NAD1_3_F	CTCCGTTTGATCTCCAGAA
	NAD1_3_R	AGGAAGCCATTGAAAGGTGA
	NAD1_5_F	TCGAGAAAAGGTCCCATCC
	NAD1_5_R	TGAGCTGCAGATCGTAATGC
	NAD3_F	AGCAAGGAGCGAGAAAACAA
	NAD3_R	CCCATTTTGTGCCCTATCAC
	NAD4_3_F	ACAGGGAATTGGAGGTAGCA
	NAD4_3_R	TGAAATTGCCATGTTGCAC
	NAD4_5_F	TTTAGCGCCCATCTTTGAC
	NAD4_5_R	TGTCTCGAACCCATACTCC
	NAD4_MID_F	CAAACAGGAACCACCGATTT
	NAD4_MID_R	TGCTACCTCCAATTCCTGT
	NAD4L_F	CGTTCCCGAAATGGATCTTA
	NAD4L_R	ACGAGCGACTGAAAGGAGAG

NAD5_3_F	TCAATTCCAACGAGCCTTTC
NAD5_3_R	CAAGTCCTCCAGTTCGATT
NAD5_5_F	GCTCGGTAGTTCGTCAGCAG
NAD5_5_R	CCCCAAGAAGCATCAAACAT
NAD5_INTRON_F	GGCTCGTAGTACCCCTTTT
NAD5_INTRON_R	CCTCCCATCCCAGGAATAAT
NAD5_MID1_F	GGATGGGAGGGAGTAGGTCT
NAD5_MID1_R	AGTTAGAGATGCCGCAAGCA
NAD5_MID2_F	CCATCAGTGGGAACTTTGCT
NAD5_MID2_R	CAATCGTCGGAATGTGTACG
NAD6_3_F	AAAGTTCCATTCAGGGAAGG
NAD6_3_R	GCCCCAATCATGGCTACTAA
NAD6_5_F	GAAGGGCTTGGAAGAAGAAAA
NAD6_5_R	TCCCACCAAAAAGATCAGTCC
RPL5_3_F	AAGGGGTTTCGACAGGAAAGT
RPL5_3_R	ATTTAGCCGCGAGCAATCTA
RPL5_5_F	TTCGAAAGGAGGAGTAGGA
RPL5_5_R	TGTGTTGGCCGAAGTGATAA
RPS3_3_F	TCGGGAACAACCTCTCAATC
RPS3_3_R	TTTCGGATATAGCACGTCTCC
RPS3_5_F	TGAGGGGAAGGTTGGTCATA
RPS3_5_R	CGCTTGAATGAAGACTCG

Liste des amorces utilisées pour le génotypage des lignées d'insertion d'ADN-T.

Projet	Nom	Séquence 5' vers 3'
Amorces générales	LBb1.3_SK	ATTTTGCCGATTTCGGAAC
	LBa1_SK	TGGTTCACGTAGTGGGCCATCG
	O8474_GB	ATAATAACGCTGCGGACATCTACATTTT
	LB1_SAIL	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC
Validation méthodologie prédiction PPR E+	GB_300H08_LP	ACGCTTGCGGTACATATCAAG
	GB_300H08_RP	TCTTAGCTTCTGTGTGGGCTC
	SK_081321_RP	TGCCAAAAGGTTCTTTGCCAG
	SK_081321_LP	TTCAATTCGTTCCCCATTCCC
	SAIL_107_G04_LP	TGGTTGCAGAACAAAATGATG
	SAIL_1210_A07_LP	AATGGCTCGTAGCAAAATCTG
	SAIL_15_E07C1_LP	TTGGAATGCAGTTATTGGAGG
	SAIL_205_G08_LP	CTGATTAGAGCTCATGCCAG
	SAIL_377_B04_LP	ATCGTGTATTCCTGGGCTTG
	SAIL_405_A09_LP	AAGCCAGAACGCTCTAACTCC
	SAIL_649_C02_LP	TCAAATGCTATTCGATTTCGG
	SAIL_549_E07_LP	GCGCTTCTGGTGTCTTCAA
	SALK_000371.54.75_LP	TCATGCTTTGGAGTAATTCCG
	SALK_001587.53.85_LP	ATTATGCGAGGCTTGTGTTTG
	SALK_024585.22.40_LP	TCTTGCGTGGTGGACTTACT
	SALK_025267.28.55_LP	TTGAGGACAGAAGCGTGGAT
	SALK_025427.37.80_LP	ATGTCCCCAAGTATTGGATC
	SALK_027870.30.15_LP	TTGACTGGTGGTCTTTAGG
	SALK_027874.48.35_LP	ACAGGTGTTTGTATGCTATGCC
	SALK_045339.36.30_LP	TCATTTCCGGTTATGCGCAG
	SALK_055676.49.60_LP	TTACAAGATCCGCCACAAGTC
	SALK_058773.55.75_LP	CTTGAATCCACCATTTCAGTG
	SALK_064675.54.25_LP	GAGAAGCTTTGATGTGATGCC
	SALK_066720.47.85_LP	TCCCATCATCAGCTGCTCAT
	SALK_085377.43.30_LP	AACGAGAGCGAGAGGAAAAAC
	SALK_088120.56.00_LP	ACATTACGGTGGATCTTGCGAG
	SALK_090615.52.30_LP	AAGGAGAAAAGTAGCGGTGAGG
	SALK_091381.55.00_LP	GACGATCAAAGCGTGCTTAAG
	SALK_104072.31.20_LP	ATTGCTCAATTTGTGTCCCTG
	SALK_115308.56.00_LP	CATGGACGCATGTGATTACTG
	SALK_122936_LP	GTCAGGATAGTGAGCGTCGAG
	SALK_124160.17.10_LP	TGTCGATTAACAGCGTTTAACC
	SALK_130945.42.45_LP	AGGTGGGTCAGTTGATACGG
SALK_132959.51.00_LP	CACAGCCGTCTCTATGCAAC	
SALK_201330_LP	TTAGAAACCACATGCCCACTC	
SALK_202435_LP	GACTTTAGAACCGGTTCTGGG	
SALK_204996_LP	AGTCGCTTCTCTGGCCTTAG	

SALKseq_036259.101_LP	GCTCAGGCTGTTTTTGATCAG
SALKseq_039979.1_LP	GTGTTTCTCCCCTGAAGAAGC
SALKseq_047811.1_LP	ACCACAAACGGTAGCAGAGTG
SALKseq_071391.1_LP	ACCTCCTAGCATCTCTGCTCC
SALKseq_081270.3_LP	AACACATTCATGGCAGCTTTC
SALKseq_130949.2_LP	ATTGTTGGCAAGACCAACAAC
SALK_141937_LP	CATCATCTTCCTCACTCTCGC
SALK_123752_LP	TGGAGGAAGCGACAATGGTT
SAIL_107_G04_RP	AATCGTCAAAGTCGCATGATC
SAIL_1210_A07_RP	TTTCATCGAACACTTTCTGGG
SAIL_15_E07C1_RP	TGCGGGCCTAATATAACATTG
SAIL_205_G08_RP	TGTTTGACAACCTTATCGC
SAIL_377_B04_RP	CTGGGTTGGAAGAACAACCTCTG
SAIL_405_A09_RP	GACGATCTTATCCTTCTCCGG
SAIL_649_C02_RP	GAGAGACACAGCTTCGTGGAC
SAIL_549_E07_RP	CAGCAAGCAAACCTCATCTGA
SALK_000371.54.75_RP	ACGTTGCAACACAGTGAAGC
SALK_001587.53.85_RP	GAACACTTGCTAGGTTGCCAG
SALK_024585.22.40_RP	CCCATCCCGCTTGCTGTAG
SALK_025267.28.55_RP	ACTCTCCAGTCGAAGCCTTTG
SALK_025427.37.80_RP	TTCGTGGAATGCGATTATCTC
SALK_027870.30.15_RP	AAACCTTTTCCCACATTCCTG
SALK_027874.48.35_RP	TAAGGCTTTATCCGCTAAGCC
SALK_045339.36.30_RP	CATGGATGTGACACGCGTTC
SALK_055676.49.60_RP	CCGGAATACTCCGAATAGAG
SALK_058773.55.75_RP	AGTCTTTCCTCGATTTAGCGC
SALK_064675.54.25_RP	GATCAACTCTTCAAGCGCAAC
SALK_066720.47.85_RP	CGGTGCAGATGGGAGAAGAG
SALK_085377.43.30_RP	TCCAAACAGTAACATCCCGAG
SALK_088120.56.00_RP	ACCTTGATTCAATCTGCATCG
SALK_090615.52.30_RP	CTTGACCGATTGCTTGTAAG
SALK_091381.55.00_RP	TAATGCTTTTCACTCGGTTGG
SALK_104072.31.20_RP	GTCAAGAAGCGCACTCAAAG
SALK_115308.56.00_RP	TTTCAATCTACCAGCACGACC
SALK_122936_RP	CTTCTCAGATCATGCTCGGAC
SALK_124160.17.10_RP	TATTTGCCGGAACAAGTGTTT
SALK_130945.42.45_RP	ATCACCTCGTCGTTGGTAGC
SALK_132959.51.00_RP	ATGGCTGATTTTGATTGGCTGAAA
SALK_201330_RP	CACACTCAAACACATCCAACG
SALK_202435_RP	GGAACTCCATGTGATCAATGAG
SALK_204996_RP	TTTCTGATGGATTCTGCATCC
SALKseq_036259.101_RP	ATCAACACACCGTAAGCCAAC
SALKseq_039979.1_RP	CATGTATTGTTGAGAACCGCC
SALKseq_047811.1_RP	TCATGCGTCTGTGATTAACG
SALKseq_071391.1_RP	TGACTCAAACCACCTCAATCC

	SALKseq_081270.3_RP	ATCGATGTGCCGAGTAAATTG
	SALKseq_130949.2_RP	GGATATGGCTCGTGGGGTTT
	SALK_141937_RP	TGTACAAGCAAATGTTGCTGG
	SALK_123752_RP	GCTGCTCTTAACGCTGCAC
PPR bactériennes : EAMY_0820	SK_208486_LP	TCGGCTAAGTATTGTTGCACC
	SK_208486_RP	CATTCTCGAACACAGGCTCTC

Séquence artificielle de RipL utilisée dans les différents clonages (ArtRipL).

ATGTGGAGGTCTAACATCGGAGATCAGAGAGTGAGACACGCTCCTGATGCTCCAGATATGCAACAACAACCTGCTTCTGGT
CAGACCTCTGAACCTAGACCTGAAGGATCTAGAGCTGTGAGGGCTTCTAATGAGGCTCTTGCTCCTCTTGCTGCTAACAGA
TGGTCTGTTCTCGGATCTGCTTCTAACGCTGGTTTTAGAAAGGTCTACCATGCCCTCCCAAAGAGGCCCTGAACCTGCTGAT
GGTGGTGTGATGCTTCTTACAGGTGGGAGATCCTAGGTCTAATGCTGCTGCTATCGCTACCTATGCTGATGAGGGATGT
CCAGCTGATGAAGTGCTCAAGCTCACCACTTCACTCCTCAGAATCGGTGGTGTATTCGCTAGATCAATGGCTCTCTTCGCT
GCTGTGAAGGATGCTGGAATCTGTCTAACGTTGGTACCTACTCAGCTGCTATCTCTGCTTGTGAGAAGGCTGGAAGAATG
GATGAGGCTTTGGTTCTCCTCAACGAGCTTAAGGCTATCGGATCTGATGATCCTATGATGAGGCCAAACGTTGTGACTTAC
AGTGTGCAATCAGTGCATGCGAAAAGGCTGGTAGAGCTGATCATGCTGTTGCTTTGCTCGGAGAGTTGAAGGCTTTGGCT
GCTCACGATCCTACCATGAGGCCAATGTGGTTACTTACTCTGCTACCATCTCAGCATGCGAGAAAGCAGGATGGGTTGAC
GAAGCTCTCGCTCTCCTTGATGAGTTGAGAGCTTTGGGACAGTCTGATCCTGCTCTCAGACCTAACGTGATCACATACTCT
GCAGCTATTTAGCTTGCATGGGAGCTGGACAAGCTGATAGAGCTAGATGGCTTGTGCTCAGGCTATCAGAGATGAGGTT
GTGCTTCCAGATGCTGGATACGATGCTAGAGCTAACACCCTCGATTTCCACGCTAACCTCGTTTTTCAGAGAGCCAGCTCCA
TCTTCTAGACCTGTTGGAGTTGCTGCTCCAGTTGCTCTCGCTTTGATGTCTTACCATGGACACAGGGGTAACCTCAACGGT
GAGACTCATTACATCGTGGGACAGCATGGTGGTGTATCAAGTCTGCTGTTCTCGAGACTCTCAACAGAGTGAGGCCCT
AGAGCTTACGTGGTGTCTCCTACTAATGCTGGTATGTTGATCGCTACCGCTCCAGATCCTTACGCTGTTTGGTCTGCTGAG
GATGATCTTCCACCTAGTCTTAGGGCTGCTTCTTCTGCTTTGCCAGATGCTCTTGCTGATAGGCTCAGAGCTATGTGGCCT
CAAGCTGGTGTCTTGAAGCTCTTGCAGCTGGTCAATCTGATGCTATGTCTATCGATGATGCTGAGGCTATGAGTTCTGAT
GAAGAGGATGATGGACTCATCAGGTTCCACGCTGATCCAGCAGCTAGAGCTGCACAGCTTAGAGATCAAGGCTAGAGTT
AGAGAGGCTGCTACCGCTGAAGATGTTCTTCTGCTTGTGAGGCTGCTGGTAGACTTAGAGCTGAAGCTGCTGAAAGATAC
CCTGGTGCATCTGCTGAGCAGCTCAAGGATGCTTTTTCTCCCTGGATCTGAAAATGTGGATTCTAAGCTTTAGAAGGCTCAAC
GATATGAGGCTCGATTGTGATGATGCAGCTCAGGCTTTGGCAAGCTGCTTCTGATCATTACAGGTCTGCTAGGGATGCTGCT
CTTAAGCTCAACCTCCCTGGTGTATTTGTGACCCCTCTTGGACAGCTGAGGCTGAATGTGAAGCTCAAGCTACTAGACTC
GAGATCACTCACGGACTCCTCGATGGAATTTGGAGGTTCAAGGGTTGGAGGCAGTCTTTTCGATGCTGGACACCCTAAGTCT
CTCGATGTGGTTCTTAAGCCTCTTACAGAGGCTCTTGCACCTGGAAACGCTGAGGCTCTCGCTACACTTCTTTCTGTGAT
GTTCTATCCAGGCTGAGTTCCACGATCAGCTCAACATCATTACGGTGTGCTGCACAGTTGGTGTGCTGCTGGTGGTGGT
GTGCTTCTCTTCTGCTCTCCTTAACAGACTTAGACCTGAGGCTCAGGCTGCTAGAAGGGCTATGTCAGCTGCTGCTTACGCT
GAGATTGTGAAGGCAATCGCTCTTTTGGGACCAGCTCTTAAGGGTGAATGCCTGCTGCTGCACCTGAGCCAATGCTTATT
GATGAGCCTGGTGTGTCTGAGAGGGCTGTTTCTGTTGCTCTCACTGAGCTTATCGCTCATCTCACACCTGGAATCGAGTCA
GCTAGAAGAGTTGCAGCTCTCGCACAGTCCACATGAGACTGGTGTCTTTGTCTGTTGGAGATCAGGTTAGGGCTTTTCGGACTC
TGGCATAGAAGTGTGAGGTGTTTCGGAAGGGCTAGACAAGCTTTTTGAAGCTGTGGCTGCATCTGGATGGCCTTTCAGAGGCT
TTGTGAGCTCCAGATGTTCAAGCTCTCCACGCAGAGATCAGACAGGGATATGATGCTGCTGTGGGAAAGATGCAAGCAACC
GCTGGAAGAGTTTTTCGCTGCTATGTGGGAGCAGGTTAGAGGATCTCTTGGACAGGTGAGAGATGCTCAGACAGTTGCTCAG
CTCATGGCAAGAATGCATGTGCCTGATGAGCTTGGAAATCCACGTTGCAGATGGATGCCAGCACATGAGATCTTGCCTTGG
GATGCTCCTGTGGCTTTGAGAGCTGGTGGAAATCGATGCTGCAAGAATGGTGGCTCCTTTCACTGCTCTCCAGGCTTTCTTG
AGAATCACAGACAACCCTTGTCTCAGCCAGAGCAGCTTCTTCTACTGCTGCTCATCTTCTTACAGGTGGTGTGCTCTCGCT
AAGAGGCTGGTGTGCTGGTGCAGCTTCTGCACTTAAGCCTGCACTTACGCTTCTTCTACTTTCTGCGCTGATCTCAGAATG
CAGCTCGTGCATACTGCTCACAGAATCGAGAGAAGAGAGGTTGCAACCCTCTGCACTTCTGTGTACCTGATATTTGGGCT
GCTGCAACCAGAGTTGAGCATAGACTCAGGGCTCTCGTTGATCCTAAGGCTGTTACTGGACAGGGAATCACTTCTGCTGGA
ACTGCAGAGGAACCTTACAGATGCTGCTATGGCTGCTGCTGCTGATATGGCAGCTCTTAGACTTGAGCAGAGAAGGGAAGGA
CTCCTTGGCTTTGCCTCCTGGTGTTCCTGCTACTCTCACTAGAAAAGCAAGCTAGGTCTGCTCTCTTGTCTCAGAGGTATTGCT
GCAGCTGGAAGATTGACCGCAGAGCTTATCAGATGTACGAAAAGGCGAGCTCCAGCTATGGGAGAAGCTAACCTTGTCTCAA
AGGGCTAAGGCTCTCGGAAAGATTGCTGCTGAAGTTCAAGCTGCTGGTGTGCTAAGGCTGGTGGGCTTTTCGTTGATACTGTT
ACCCTTGGAGATGCTCCTGAGTCTGCTTCTCCTAATCTGAGGCTGCACTCTTGTCTGCTGCTGTTTCTAGATGCGAGTTC
CTCGCTGAGTACATGAGATTCTGCGTGCAGACACTTGTGATCTCGCTGCAGGTAGAAGAAACGCTCAGTCTAGAATCAGA
CAGCCAGCTGGTGTGCTTCTGCTTACGTGAGAGAAGCTGATGCAATCCAGGGAATCCATGATGATATCGTGGATAGGCTC
TACGAGGCTTGCAGAGAGATTGTTGGACTTATCGAAGCTAAGCCTGCTATCGGAGATACCAGGGAAGCTATCGATGCAGAG
ACTAACTTTCGTGCATGAGTTGGTGGATTCTCTCGAGTGGAGAGCAATCGTGGAAAACCGATATGACACAGAGGGCTTGGATG
AGAGCTACCCTTGTGTTGCTGCAGTGTCAATCGATCACGCTTCAAGAGATGTGCTCCAGGGTATGAGTACCATGCTCGCT
AAGATCGAGGATGATTTGAGAGAGTCTAAGGGAAGGCTCGTGGAGTACGCTGAAAAGACTGGTGTGATCAGGCTTACAGGGCT
TCACTCAAGCTCGAGGATGTTACTCTCGGAGTTGGTGTGCTAGGGCTCAGACTAGGCTTATCGATAGGCTTAAACGCTCTC
GCTGGTGAAGTAGGGCTCAACCTTCTAAGTCAAAGGGAAAGGGAAGCTAAGGGTAAGGGAAAAGCTAAGTGA

Séquence artificielle d'EMY 0820 utilisée dans les différents clonages (ArtEa).

ATGCCTCAGTCTGCTATCTTCTCATCTTCTAAGATCAACCCATCTACATCAACAACACCTCTGGATCTACCGCTACCGCT
TACGATGGAAACGTGATCACCAGGATGAAGGAAGTGACTAACACCGGAGAGAACCCTCGCTGCTCAAAGACAGTCTGATCAC
AAGTCTGATGCTGGTTCTACCTACAACGGAAACATCACCCCATCTCTGAGAAGAACAACGATCAGGTGAACACCCCTGGA
CTCGTTACTGGTGGTGAGGCTATTAAGTACAACCACGAGAACAAGAAAGATGATCTCAAGAACGCTGAGATCGGAAGAGAG
ATCAAGAAGCTCTGCAAGCAGGGAAAGATTGGAGAGGCTGAGAAGCTCGTGAACGCTATGTCTGATCCTCATGTGTGGGTT
CTCACCGCTCTTATGGAAGGATGCTTCTCTAAGCACAAAGTGCGGAGATGCTGAGCAGGTTTTCAAGTCTATGGAAGATAAG
AACCTCCCTGTGGATCTCTGGGTGTTCAAGTCTCTCATCAAGTACTACACCAGGACCGTGAATGCTGAAAGGGCTGAGTGG
GTGATGACCAAGATGCAAAAAGGCTGGAATCAAGCCTGATGAGGGAATCCTCGATAAGCTCCTCGAGGTGTACAGAAAGGCT
GGTAACCATGCTGGTGGTGGATGCACGTTTTCAAAGCAGCTCCATCAGCTCAAGTGCACCTCCTAAGCTTGCTGCTATGCACAAA
GAGATGAACGATCACGTTGAGACTGGTGATATCAAGGGTGCTGAGAGGGTGTCTTCAAGATTAACCAGTACGGACTCACC
TCAGATCAAAAAGTCTCTCGATATCATGCTCAAGGGATACGGATACACCGGAAATATCGCTGAGGCTAAGTCTATCTTTGAG
AAGCCAGCTATCAACGATATCACCCCTTACAACAGGCTCATGAAGGTTTGCCTCAAGCTAAGGCTCCTGAGAAGGCTGAA
CAAGTGCTCGCTAGAATCAAAGGTGCTAACCTCAAGCCTAACGTGCAGTCTTACGGAAATGCTCATCAAGGCTTACATCGTG
GCTTCAGATTGGCCTGGTGCTGAAAGAGCTATGAAGAGGGCTAAAAGAAGCTGAGTGCAGGACCTGATTCTGATGTGTGCCAG
TCTTTGCTCAAGACTTGCGGAGCTACTGGACAGATCGATAGGGCTCAATGGGTTTTGAGGGAAGCTTCTGTTGAGGTGAGG
GATCCACTCTTCAAGGTGCCAAGATCTGAGTTGCACAACCTACCTCATCTACGCTCTCGATACTTGGAGGAAGCAGAACCCT
GCTTACTCTCCTGCTGTTTGA

Références.

- Aksoy, H.M., Kaya, Y., et Tengku Abdul Hamid, T.H.** (2017). Expression of the *dspA/E* gene of *Erwinia amylovora* in non-host plant *Arabidopsis thaliana*. *Biotechnol. Biotechnol. Equip.* **31**: 85-90.
- Alfano, J.R. et Collmer, A.** (2004). TYPE III SECRETION SYSTEM EFFECTOR PROTEINS: Double Agents in Bacterial Disease and Plant Defense. *Annu. Rev. Phytopathol.* **42**: 385-414.
- Allen, J.F.** (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci.* **112**: 10231-10238.
- Alvim Kamei, C.L., Boruc, J., Vandepoele, K., Van den Daele, H., Maes, S., Russinova, E., Inze, D., et De Veylder, L.** (2008). The PRA1 Gene Family in *Arabidopsis*. *Plant Physiol.* **147**: 1735-1749.
- Amirsadeghi, S., Robson, C. a., et Vanlerberghe, G.C.** (2007). The role of the mitochondrion in plant responses to biotic stress. *Physiol. Plant.* **129**: 253-266.
- Anderson, S. et al.** (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457-465.
- André, C., Levy, a, et Walbot, V.** (1992). Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* **8**: 128-32.
- Andrés, C., Lurin, C., et Small, I.D.** (2007). The multifarious roles of PPR proteins in plant mitochondrial gene expression. *Physiol. Plant.* **129**: 14-22.
- Andrés-Colás, N., Zhu, Q., Takenaka, M., De Rybel, B., Weijers, D., et Van Der Straeten, D.** (2017). Multiple PPR protein interactions are involved in the RNA editing system in *Arabidopsis* mitochondria and plastids. *Proc. Natl. Acad. Sci.*: 201705815.
- Apel, K. et Hirt, H.** (2004). REACTIVE OXYGEN SPECIES: Metabolism, Oxidative Stress, and Signal Transduction. *Annu. Rev. Plant Biol.* **55**: 373-399.
- Arenas-M, A., Takenaka, M., Moreno, S., Gómez, I., et Jordana, X.** (2013). Contiguous RNA editing sites in the mitochondrial *nad1* transcript of *Arabidopsis thaliana* are recognized by different proteins. *FEBS Lett.* **587**: 887-891.
- Bari, R. et Jones, J.D.G.** (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* **69**: 473-488.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., et Small, I.** (2012). A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genet.* **8**: 4-11.
- Barkan, A. et Small, I.** (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **65**: 415-42.
- Beardslee, T. a., Roy-Chowdhury, S., Jaiswal, P., Buhot, L., Lerbs-Mache, S., Stern, D.B., et Allison, L. a.** (2002). A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. *Plant J.* **31**: 199-209.
- Becerra, J.X.** (2015). Macroevolutionary and geographical intensification of chemical defense in plants driven by insect herbivore selection pressure. *Curr. Opin. Insect Sci.* **8**: 15-21.
- Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., et Barkan, A.** (2008). The Pentatricopeptide Repeat Protein PPR5 Stabilizes a Specific tRNA Precursor in Maize Chloroplasts. *Mol. Cell. Biol.* **28**: 5337-5347.
- Bendich, a. J.** (2004). Circular Chloroplast Chromosomes: The Grand Illusion. *Plant Cell Online* **16**: 1661-1666.
- Benne, R., Van Den Burg, J., Brakenhoff, J.P.J., Sloof, P., Van Boom, J.H., et Tromp, M.C.** (1986). Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**: 819-826.

- Bentolila, S., Oh, J., Hanson, M.R., et Bukowski, R.** (2013). Comprehensive High-Resolution Analysis of the Role of an Arabidopsis Gene Family in RNA Editing. *PLoS Genet.* **9**.
- Bertolla, F., Frostegård, Å., Brito, B., Nesme, X., et Simonet, P.** (1999). During Infection of Its Host, the Plant Pathogen *Ralstonia solanacearum* Naturally Develops a State of Competence and Exchanges Genetic Material. *Mol. Plant-Microbe Interact.* **12**: 467-472.
- Bertolla, F., Van Gijsegem, F., Nesme, X., et Simonet, P.** (1997). Conditions for natural transformation of *Ralstonia solanacearum*. *Appl. Environ. Microbiol.* **63**: 4965-4968.
- Bertolla, F. et Simonet, P.** (1999). Horizontal gene transfers in the environment: Natural transformation as a putative process for gene transfers between transgenic plants and microorganisms. *Res. Microbiol.* **150**: 375-384.
- Van Der Biezen, E. a. et Jones, J.D.G.** (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* **23**: 454-456.
- Bilgin, D.D., Zavala, J. a., Zhu, J., Clough, S.J., Ort, D.R., et Delucia, E.H.** (2010). Biotic stress globally downregulates photosynthesis genes. *Plant, Cell Environ.* **33**: 1597-1613.
- Blatch, G.L. et Lässle, M.** (1999). The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. *BioEssays* **21**: 932-939.
- Block, A. et Alfano, J.R.** (2011). Plant targets for *Pseudomonas syringae* type III effectors: Virulence targets or guarded decoys? *Curr. Opin. Microbiol.* **14**: 39-46.
- Block, A., Guo, M., Li, G., Elowsky, C., Clemente, T.E., et Alfano, J.R.** (2010). The *Pseudomonas syringae* type III effector HopG1 targets mitochondria, alters plant development and suppresses plant innate immunity. *Cell. Microbiol.* **12**: 318-330.
- Boch, J. et Bonas, U.** (2010). *Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function. *Annu. Rev. Phytopathol.* **48**: 419-436.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., et Bonas, U.** (2009). Breaking the Code of DNA Binding. *Science* (80-.). **1509**: 1509-12.
- Bock, R.** (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* **15**: 11-22.
- Bonas, U., Stall, R.E., et Staskawicz, B.** (1989). Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet.* **218**: 127-136.
- Bonfig, K.B., Schreiber, U., Gabler, A., Roitsch, T., et Berger, S.** (2006). Infection with virulent and avirulent *P. syringae* strains differentially affects photosynthesis and sink metabolism in *Arabidopsis* leaves. *Planta* **225**: 1-12.
- Boussardon, C., Avon, A., Kindgren, P., Bond, C.S., Challenor, M., Lurin, C., et Small, I.** (2014). The cytidine deaminase signature HxE(x)nCxxC of DYW1 binds zinc and is necessary for RNA editing of *ndhD-1*. *New Phytol.* **203**: 1090-1095.
- Boussardon, C., Salone, V., Avon, a., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., et Lurin, C.** (2012). Two Interacting Proteins Are Necessary for the Editing of the *NdhD-1* Site in *Arabidopsis* Plastids. *Plant Cell* **24**: 3684-3694.
- Buchanan, B., Gruissem, W., et Jones, R.** (2015). *BIOCHEMISTRY and MOLECULAR BIOLOGY of PLANTS* (Wiley).
- Carver, T., Harris, S.R., Berriman, M., Parkhill, J., et McQuillan, J.A.** (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**: 464-469.
- Castandet, B., Choury, D., Bégu, D., Jordana, X., et Araya, A.** (2010). Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res.* **38**: 7112-7121.
- Castandet, B., Hotto, A.M., Strickler, S.R., et Stern, D.B.** (2016a). ChloroSeq, an Optimized Chloroplast RNA-Seq Bioinformatic Pipeline, Reveals Remodeling of the Organellar Transcriptome Under Heat Stress. *Genes|Genomes|Genetics* **6**: 2817-2827.
- Castandet, B., Hotto, A.M., Strickler, S.R., et Stern, D.B.** (2016b). ChloroSeq, an Optimized

- Chloroplast RNA-Seq Bioinformatic Pipeline, Reveals Remodeling of the Organellar Transcriptome Under Heat Stress. *G3 Genes|Genomes|Genetics* **6**: 2817-2827.
- Catanzariti, A.-M.** (2006). Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors. *Plant Cell Online* **18**: 243-256.
- Chateigner-Boutin, A.L., Colas Des Francs-Small, C., Fujii, S., Okuda, K., Tanz, S.K., et Small, I.** (2013). The e domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J.* **74**: 935-945.
- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-García, A., Andrés, C., Gutiérrez-Nava, M.D.L.L., Cantero, A., Delannoy, E., Jiménez, L.F., Lurin, C., Small, I., et Leo, P.** (2008). CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. *Plant J.* **56**: 590-602.
- Chekmenev, D.S., Haid, C., et Kel, a. E.** (2005). P-Match: Transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Res.* **33**: 432-437.
- Chen, G., Zou, Y., Hu, J., et Ding, Y.** (2018). Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genomics* **19**: 720.
- Cheng, S. et al.** (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* **85**: 532-547.
- Chevalier, F., Perazza, D., Laporte, F., Le Henanff, G., Hornitschek, P., Bonneville, J.-M., Herzog, M., et Vachon, G.** (2008). GeBP and GeBP-Like Proteins Are Noncanonical Leucine-Zipper Transcription Factors That Regulate Cytokinin Response in Arabidopsis. *Plant Physiol.* **146**: 1142-1154.
- Choquet, Y. et Vallon, O.** (2000). Synthesis, assembly and degradation of thylakoid membrane proteins. *Biochimie* **82**: 615-634.
- Choquet, Y., Wostrikoff, K., Rimbault, B., Zito, F., Girard-Bascou, J., Drapier, D., et Wollman, F.A.** (2001). Assembly-controlled regulation of chloroplast gene translation. *Biochem. Soc.* **29**: 421-6.
- Colas Des Francs-Small, C. et Small, I.** (2014). Surrogate mutants for studying mitochondrially encoded functions. *Biochimie* **100**: 234-242.
- Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthomé, R., et Lurin, C.** (2013). Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. *RNA Biol.* **10**.
- Collmer, A., Schneider, D.J., et Lindeberg, M.** (2009). Lifestyles of the Effector Rich: Genome-Enabled Characterization of Bacterial Plant Pathogens. *Plant Physiol.* **150**: 1623-1630.
- Creelman, R. a. et Mullet, J.E.** (1997). Biosynthesis and Action of Jasmonates in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**: 355-381.
- Cunnac, S., Lindeberg, M., et Collmer, A.** (2009). Pseudomonas syringae type III secretion system effectors: repertoires in search of functions. *Curr. Opin. Microbiol.* **12**: 53-60.
- Dahan, J., Tcherkez, G., Macherel, D., Benamar, a., Belcram, K., Quadrado, M., Arnal, N., et Mireau, H.** (2014). Disruption of the CYTOCHROME C OXIDASE DEFICIENT1 Gene Leads to Cytochrome c Oxidase Depletion and Reorchestrated Respiratory Metabolism in Arabidopsis.
- Davila, J.I., Arrieta-Montiel, M.P., Wamboldt, Y., Cao, J., Hagmann, J., Shedge, V., Xu, Y.-Z., Weigel, D., et Mackenzie, S.A.** (2011). Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. *BMC Biol.* **9**: 64.
- Degrave, A., Fagard, M., Perino, C., Brisset, M.N., Gaubert, S., Laroche, S., Patrit, O., et Barny, M. -a.** (2008). Erwinia amylovora Type Three-Secreted Proteins Trigger Cell Death and Defense Responses in Arabidopsis thaliana. *Mol. Plant-Microbe Interact.* **21**: 1076-1086.
- Delannoy, E., Stanley, W. a., Bond, C.S., et Small, I.D.** (2007). Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem. Soc. Trans.* **35**: 1643-1647.

- Dhawan, R., Luo, H., Foerster, A.M., AbuQamar, S., Du, H.-N., Briggs, S.D., Scheid, O.M., et Mengiste, T.** (2009). HISTONE MONOUBIQUITINATION1 Interacts with a Subunit of the Mediator Complex and Regulates Defense against Necrotrophic Fungal Pathogens in Arabidopsis. *Plant Cell Online* **21**: 1000-1019.
- Diaz, M.F., Bentolila, S., Hayes, M.L., Hanson, M.R., et Mulligan, R.M.** (2017). A protein with an unusually short PPR domain, MEF8, affects editing at over 60 Arabidopsis mitochondrial C targets of RNA editing. *Plant J.* **92**: 638-649.
- Dietz, K.-J. et Pfannschmidt, T.** (2011). Novel Regulators in Photosynthetic Redox Control of Plant Metabolism and Gene Expression. *Plant Physiol.* **155**: 1477-1485.
- Dobin, A., Davis, C. a., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., et Gingeras, T.R.** (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.
- Doke, N.** (1985). NADPH-dependent O₂ generation in membrane fractions isolated from wounded potato tubers inoculated with *Phytophthora infestans*. *Physiol. Plant Pathol.* **27**: 311-322.
- Dreze, M. et al.** (2011). Evidence for Network Evolution in an Arabidopsis Interactome Map. *Science* (80-.). **333**: 601-607.
- Dreze, M., Monachello, D., Lurin, C., Cusick, M.E., Hill, D.E., Vidal, M., et Braun, P.** (2010). High-quality binary interactome mapping 2^e éd. (Elsevier Inc.).
- Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M.A., Arnold, R., et Rattei, T.** (2016). EffectiveDB - Updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* **44**: D669-D674.
- El-Maarouf, H., Barny, M.A., Rona, J.P., et Bouteau, F.** (2001). Harpin, a hypersensitive response elicitor from *Erwinia amylovora*, regulates ion channel activities in Arabidopsis thaliana suspension cells. *FEBS Lett.* **497**: 82-84.
- Eubel, H., Braun, H.P., et Jansch, L.** (2003). New Insights into the Respiratory Chain of Plant Mitochondria. Supercomplexes and a Unique Composition of Complex II. *PLANT Physiol.* **133**: 274-286.
- Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S. a., Law, T.F., Grant, S.R., Dangel, J.L., et Carrington, J.C.** (2007). High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS One* **2**.
- Fisk, D.G., Walker, M.B., et 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.
- Flor, H.H.** (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.*: 275-296.
- Fukui, K. et Kuramitsu, S.** (2011). Structure and Function of the Small MutS-Related Domain. *Mol. Biol. Int.* **2011**: 1-9.
- Gaudriault, S., Brisset, M.N., et Barny, M.A.** (1998). HrpW of *Erwinia amylovora*, a new Hrp-secreted protein. *FEBS Lett.* **428**: 224-228.
- Giegé, P., Hoffmann, M., Binder, S., et Brennicke, A.** (2000). RNA degradation buffers asymmetries of transcription in Arabidopsis mitochondria. *EMBO Rep.* **1**: 164-170.
- Giovannoni, J.** (2004). Genetic regulation of fruit development and ripening. *Plant Cell Online* **16**: 170-181.
- Glass, F., Härtel, B., Zehrmann, A., Verbitskiy, D., et Takenaka, M.** (2015). MEF13 requires MORF3 and MORF8 for RNA editing at eight targets in mitochondrial mRNAs in Arabidopsis thaliana. *Mol. Plant* **8**: 1466-1477.
- De Grey, a. D.N.J.** (2005). Forces maintaining organellar genomes: Is any as strong as genetic code disparity or hydrophobicity? *BioEssays* **27**: 436-446.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., et Grienenberger, J.M.** (1989). RNA

editing in wheat mitochondria results in the conservation of protein sequences. *Nature* **341**: 660-662.

- Guillaumot, D., Lopez-Obando, M., Baudry, K., Avon, A., Rigai, G., Falcon de Longevialle, A., Broche, B., Takenaka, M., Berthomé, R., De Jaeger, G., Delannoy, E., et Lurin, C.** (2017). Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc. Natl. Acad. Sci.* **114**: 8877-8882.
- Hackett, J.B., Shi, X., Kobylarz, A.T., Lucas, M.K., Wessendorf, R.L., Hines, K.M., Bentolila, S., Hanson, M.R., et Lu, Y.** (2017). An Organelle RNA Recognition Motif Protein is Required for Photosynthetic Subunit psbF Transcript Editing. *Plant Physiol.* **173**: pp.01623.2016.
- Hagemann, R.** (2010). The foundation of extranuclear inheritance: Plastid and mitochondrial genetics. *Mol. Genet. Genomics* **283**: 199-209.
- Hamdoun, S., Gao, M., Gill, M., Kwon, A., Norelli, J.L., et Lu, H.** (2018). Signalling requirements for *Erwinia amylovora* -induced disease resistance, callose deposition and cell growth in the non-host *Arabidopsis thaliana*. *Mol. Plant Pathol.* **19**: 1090-1103.
- Hammani, K. et Giegé, P.** (2014). RNA metabolism in plant mitochondria. *Trends Plant Sci.* **19**: 380-389.
- Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, a.-L., Shikanai, T., et Small, I.** (2009). A Study of New Arabidopsis Chloroplast RNA Editing Mutants Reveals General Features of Editing Factors and Their Target Sites. *Plant Cell* **21**: 3686-3699.
- Härtel, B., Zehrmann, A., Verbitskiy, D., et Takenaka, M.** (2013). The longest mitochondrial RNA editing PPR protein MEF12 in *Arabidopsis thaliana* requires the full-length E domain. *RNA Biol.* **10**: 1337-1342.
- Hase, Y., Trung, K.H., Matsunaga, T., et Tanaka, A.** (2006). A mutation in the *uvi4* gene promotes progression of endo-reduplication and confers increased tolerance towards ultraviolet B light. *Plant J.* **46**: 317-326.
- Hauler, A., Jonietz, C., Stoll, B., Stoll, K., Braun, H.-P., et Binder, S.** (2013). RNA PROCESSING FACTOR 5 is required for efficient 5' cleavage at a processing site conserved in RNAs of three different mitochondrial genes in *Arabidopsis thaliana*. *Plant J.* **74**: 593-604.
- He, Y., Wang, J., Gou, L., Shen, C., Chen, L., Yi, C., Wei, X., et Yang, J.** (2013). Comprehensive analysis of expression profile reveals the ubiquitous distribution of PPPDE peptidase domain 1, a Golgi apparatus component, and its implications in clinical cancer. *Biochimie* **95**: 1466-1475.
- Herbers, K., Conrads-Strauch, J., et Bonas, U.** (1992). Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature* **356**: 172-174.
- Heuer, H., Yin, Y.N., Xue, Q.Y., Smalla, K., et Guo, J.H.** (2007). Repeat domain diversity of *avrBs3*-like genes in *Ralstonia solanacearum* strains and association with host preferences in the field. *Appl. Environ. Microbiol.* **73**: 4379-4384.
- Hoeberichts, F. a. et Woltering, E.J.** (2003). Multiple mediators of plant programmed cell death: Interplay of conserved cell death mechanisms and plant-specific regulators. *BioEssays* **25**: 47-57.
- Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R., et Kamoun, S.** (2009). Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Mol. Plant-Microbe Interact.* **22**: 115-122.
- Van der Hoorn, R. a. L. et Kamoun, S.** (2008). From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *Plant Cell Online* **20**: 2009-2017.
- Hu, Z. et al.** (2016). Mitochondrial Defects Confer Tolerance against Cellulose Deficiency. *Plant Cell* **28**: 2276-2290.
- Huang, Z., Andrianov, V.M., Han, Y., et Howell, S.H.** (2001). Identification of Arabidopsis proteins that interact with the cauliflower mosaic virus (CaMV) movement protein. *Plant Mol. Biol.* **47**: 663-675.

- Ikeda, T.M. et Gray, M.W.** (1999). Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol. Cell. Biol.* **19**: 8113-22.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., et Weissman, J.S.** (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* (80-.). **324**: 218-223.
- Jamir, Y., Guo, M., Oh, H.S., Petnicki-Ocwieja, T., Chen, S., Tang, X., Dickman, M.B., Collmer, A., et Alfano, J.R.** (2004). Identification of *Pseudomonas syringae* type III effectors that can suppress programmed cell death in plants and yeast. *Plant J.* **37**: 554-565.
- Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., et Valent, B.** (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**: 4004-4014.
- Jixin Dong, Chunhong Chen, Z.C.** (2003). Expression profiles of Arabidopsis WRKY superfamily during plant defense response.pdf. *Plant Mol. Biol.* **51**: 21-37.
- Jones, J.D.G. et Dangl, J.L.** (2006). The plant immune system. *Nature* **444**: 323-329.
- Joo, J.H., Wang, S., Chen, J.G., Jones, a M., et Fedoroff, N. V** (2005). Different Signaling and Cell Death Roles of Heterotrimeric G Protein {alpha} and {beta} Subunits in the Arabidopsis Oxidative Stress Response to Ozone\r10.1105/tpc.104.029603. *Plant Cell* **17**: 957-970.
- Kabani, M., Beckerich, J.-M., et Brodsky, J.L.** (2002). Nucleotide exchange factor for the yeast Hsp70 molecular chaperone Ssa1p. *Mol. Cell. Biol.* **22**: 4677-89.
- Kamber, T., Buchmann, J.P., Pothier, J.F., Smits, T.H.M., Wicker, T., et Duffy, B.** (2016). Fire blight disease reactome: RNA-seq transcriptional profile of apple host plant defense responses to *Erwinia amylovora* pathogen infection. *Sci. Rep.* **6**: 1-12.
- Kangasjärvi, S., Neukermans, J., Li, S., Aro, E.M., et Noctor, G.** (2012). Photosynthesis, photorespiration, and light signalling in defence responses. *J. Exp. Bot.* **63**: 1619-1636.
- Karpenahalli, M.R., Lupas, A.N., et Söding, J.** (2007). TPRpred: A tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* **8**: 1-8.
- Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, a., Zhu, J.-K., Staskawicz, B.J., et Jin, H.** (2006). A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci.* **103**: 18002-18007.
- Khraiwesh, B., Zhu, J.K., et Zhu, J.** (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**: 137-148.
- Kipreos, E.T. et Pagano, M.** (2000). The F-box protein family. *Genome Biol.* **1**: REVIEWS3002.
- Klaff, P.** (1991). Changes in Chloroplast mRNA Stability during Leaf Development. *Plant Cell Online* **3**: 517-529.
- Kleine, T. et Leister, D.** (2016). Retrograde signaling: Organelles go networking. *Biochim. Biophys. Acta - Bioenerg.* **1857**: 1313-1325.
- Klutstein, M., Shaked, H., Sherman, A., Avivi-Ragolsky, N., Shema, E., Zenvirth, D., Levy, A. a., et Simchen, G.** (2008). Functional conservation of the yeast and Arabidopsis RAD54-like genes. *Genetics* **178**: 2389-2397.
- Koczan, J.M., Lenneman, B.R., McGrath, M.J., et Sundin, G.W.** (2011). Cell Surface Attachment Structures Contribute to Biofilm Formation and Xylem Colonization by *Erwinia amylovora*. *Appl. Environ. Microbiol.* **77**: 7031-7039.
- Koonin, E. V, Makarova, K.S., et Aravind, L.** (2001). Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annu. Rev. Microbiol.* **55**: 709-742.
- Kosarev, P., Mayer, K.F.X., et Hardtke, C.S.** (2002). Evaluation and classification of RING-finger domains encoded by the Arabidopsis genome. *Genome Biol.* **3**: RESEARCH0016.
- Kotera, E., Tasaka, M., et Shikanai, T.** (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**: 326-330.
- Krause, M. et Durner, J.** (2004). Harpin Inactivates Mitochondria in Arabidopsis Suspension Cells. *Mol. Plant-Microbe Interact.* **17**: 131-139.

- Kubori, T.** (1998). Supramolecular Structure of the Salmonella typhimurium Type III Protein Secretion System. *Science* (80-.). **280**: 602-605.
- Kuhn, K., Richter, U., Meyer, E.H., Delannoy, E., Falcon de Longevialle, a., O'Toole, N., Borner, T., Millar, a. H., Small, I.D., et Whelan, J.** (2009). Phage-Type RNA Polymerase RPOTmp Performs Gene-Specific Transcription in Mitochondria of Arabidopsis thaliana. *Plant Cell Online* **21**: 2762-2779.
- Kuźniak, E. et Skłodowska, M.** (2004). The effect of Botrytis cinerea infection on the antioxidant profile of mitochondria from tomato leaves. *J. Exp. Bot.* **55**: 605-612.
- Lahaye, T. et Bonas, U.** (2001). Molecular secrets of bacterial type III effector proteins. *Trends Plant Sci.* **6**: 479-485.
- Laluk, K., AbuQamar, S., et Mengiste, T.** (2011). The Arabidopsis Mitochondria-Localized Pentatricopeptide Repeat Protein PGN Functions in Defense against Necrotrophic Fungi and Abiotic Stress Tolerance. *Plant Physiol.* **156**: 2053-2068.
- Launay, A., Patrit, O., Wénès, E., et Fagard, M.** (2016). DspA/E Contributes to Apoplastic Accumulation of ROS in Non-host A. thaliana. *Front. Plant Sci.* **7**: 1-12.
- Laurent, J., Paulin, J.P., et Zucca, J.** (1987). Ultrastructural study of Erwinia amylovora strains: Effect of culture conditions and fixation procedures. *Protoplasma* **139**: 1-8.
- Law, S.R., Narsai, R., Taylor, N.L., Delannoy, E., Carrie, C., Giraud, E., Millar, a. H., Small, I., et Whelan, J.** (2012). Nucleotide and RNA Metabolism Prime Translational Initiation in the Earliest Events of Mitochondrial Biogenesis during Arabidopsis Germination. *Plant Physiol.* **158**: 1610-1627.
- Lechner, E., Achard, P., Vansiri, A., Potuschak, T., et Genschik, P.** (2006). F-box proteins everywhere. *Curr. Opin. Plant Biol.* **9**: 631-638.
- Lee, U., Rioflorido, I., Hong, S.W., Larkindale, J., Waters, E.R., et Vierling, E.** (2007). The Arabidopsis ClpB/Hsp100 family of proteins: Chaperones for stress and chloroplast development. *Plant J.* **49**: 115-127.
- Lerbs-Mache, S.** (2011). Function of plastid sigma factors in higher plants: Regulation of gene expression or just preservation of constitutive transcription? *Plant Mol. Biol.* **76**: 235-249.
- Leu, K.C., Hsieh, M.H., Wang, H.J., Hsieh, H.L., et Jauh, G.Y.** (2016). Distinct role of Arabidopsis mitochondrial P-type pentatricopeptide repeat protein-modulating editing protein, PPME, in nad1 RNA editing. *RNA Biol.* **13**: 593-604.
- Levine, a, Tenhaken, R., Dixon, R. a, et Lamb, C.** (1994). H₂O₂ from the oxidative burst orchestrates the plant hypersensitive response. *Cell* **79**: 583-593.
- Liao, Y., Smyth, G.K., et Shi, W.** (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923-930.
- Libbrecht, M.W. et Noble, W.S.** (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **16**: 321-332.
- Liebers, M., Grübler, B., Chevalier, F., Lerbs-Mache, S., Merendino, L., Blanvillain, R., et Pfannschmidt, T.** (2017). Regulatory Shifts in Plastid Transcription Play a Key Role in Morphological Conversions of Plastids during Plant Development. *Front. Plant Sci.* **8**: 1-8.
- Liere, K., Weihe, A., et Börner, T.** (2011). The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J. Plant Physiol.* **168**: 1345-1360.
- Lightowers, R.N. et Chrzanowska-Lightowers, Z.M.A.** (2008). PPR (pentatricopeptide repeat) proteins in mammals: important aids to mitochondrial gene expression. *Biochem. J.* **416**: e5-e6.
- Lipinski, K. a., Puchta, O., Surendranath, V., Kudla, M., et Golik, P.** (2011). Revisiting the Yeast PPR proteins-Application of an iterative hidden markov model algorithm reveals new members of the rapidly evolving family. *Mol. Biol. Evol.* **28**: 2935-2948.
- Loake, G. et Grant, M.** (2007). Salicylic acid in plant defence-the players and protagonists. *Curr. Opin. Plant Biol.* **10**: 466-472.
- Loiselay, C., Gumpel, N.J., Girard-Bascou, J., Watson, a. T., Purton, S., Wollman, F. -a., et**

- Choquet, Y.** (2008). Molecular Identification and Function of cis- and trans-Acting Determinants for petA Transcript Stability in *Chlamydomonas reinhardtii* Chloroplasts. *Mol. Cell. Biol.* **28**: 5529-5542.
- De Longevialle, A.F., Meyer, E.H., Andrés, C., Taylor, N.L., Lurin, C., Millar, a. H., et Small, I.D.** (2007). The Pentatricopeptide Repeat Gene *OTP43* Is Required for *trans* -Splicing of the Mitochondrial *nad1* Intron 1 in *Arabidopsis thaliana*. *Plant Cell* **19**: 3256-3265.
- De Longevialle, A.F., Small, I.D., et Lurin, C.** (2010). Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol. Plant* **3**: 691-705.
- López-Solanilla, E., Bronstein, P. a., Schneider, A.R., et Collmer, A.** (2004). HopPtoN is a *Pseudomonas syringae* Hrp (type III secretion system) cystein protease effector that supresses pathogen-induced necrosis associated with both compatible and incompatible plant interactions. *Mol. Microbiol.* **54**: 353-365.
- Lurin, C. et al.** (2004). Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis^W. *Plant Cell* **16**: 2089-2103.
- Malbert, B., Rigail, G., Brunaud, V., Lurin, C., et Delannoy, E.** (2018). Bioinformatic Analysis of Chloroplast Gene Expression and RNA Posttranscriptional Maturations Using RNA Sequencing. In *Methods in molecular biology*, p. 279-294.
- Manavski, N., Guyon, V., Meurer, J., Wienand, U., et Brettschneider, R.** (2012). An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria. *Plant Cell* **24**: 3087-3105.
- Mann, R.A., Smits, T.H.M., Bühlmann, A., Blom, J., Goesmann, A., Frey, J.E., Plummer, K.M., Beer, S. V., Luck, J., Duffy, B., et Rodoni, B.** (2013). Comparative Genomics of 12 Strains of *Erwinia amylovora* Identifies a Pan-Genome with a Large Conserved Core. *PLoS One* **8**.
- Manners, J.M., Penninckx, I. a. M. a., Vermaere, K., Kazan, K., Brown, R.L., Morgan, A., Maclean, D.J., Curtis, M.D., Cammue, B.P. a., et Broekaert, W.F.** (1998). The promoter of the plant defensin gene PDF1.2 from *Arabidopsis* is systemically activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid. *Plant Mol. Biol.* **38**: 1071-1080.
- Margulis, L., Knoll, A.H., et Merezchkovsky, K.** (2012). Lynn Margulis, 1938 – 2011. **109**: 2138.
- Maris, C., Dominguez, C., et Allain, F.H.T.** (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**: 2118-2131.
- Marone, D., Russo, M. a., Laidò, G., De Leonardis, A.M., et Mastrangelo, A.M.** (2013). Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: Active guardians in host defense responses. *Int. J. Mol. Sci.* **14**: 7302-7326.
- Martin, W. et Herrmann, R.G.** (1998). Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why? *Plant Physiol.* **118**: 9-17.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., et Lopez, R.** (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**: 597-600.
- Meyer, S., Saccardy-Adji, K., Rizza, F., et Genty, B.** (2001). Inhibition of photosynthesis by *Colletotrichum lindemuthianum* in bean leaves determined by chlorophyll fluorescence imaging. *Plant, Cell Environ.* **24**: 947-956.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H., et Michelmore, R.W.** (2003). Genome-wide analysis of NBS-LRR – encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809-834.
- Miranda, R.G., Rojas, M., Montgomery, M.P., Gribbin, K.P., et Barkan, A.** (2017). RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *Rna* **23**: 586-599.
- Monaghan, J. et Zipfel, C.** (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant Biol.* **15**: 349-357.
- Moore, D., Robson, G., et Trinci, T.** (2011). 21st Century Guidebook to Fungi.
- Moreau, M., Degrave, A., Vedel, R., Bitton, F., Patrit, O., Renou, J.P., Barny, M.A., et Fagard, M.** (2012). EDS1 Contributes to Nonhost Resistance of *Arabidopsis thaliana* Against *Erwinia amylovora*. *Mol. Plant-Microbe Interact.* **25**: 421-430.

- Morel, A., Peeters, N., Vaillau, F., Barberis, P., Jiang, G., Berthomé, R., et Guidot, A.** (2018). Plant pathogenicity phenotyping of *Ralstonia solanacearum* strains. In *Methods in Molecular Biology*, p. 223-239.
- Moscou, M.J. et Bogdanove, A.J.** (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* (80-.). **326**: 1501.
- Mukaihara, T., Tamura, N., et Iwabuchi, M.** (2010). Genome-wide identification of a large repertoire of *Ralstonia solanacearum* type III effector proteins by a new functional screen. *Mol. Plant-Microbe Interact.* **23**: 251-262.
- Mukhtar, M.S. et al.** (2011a). Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science* (80-.). **333**: 596-601.
- Mukhtar, M.S. et al.** (2011b). Plant Immune System Network. *Science* (80-.).: 596-601.
- Muller, H.J.** (1931). *Some genetic aspects of sex.* Univ. Chicago Press.
- Mullet, J.E.** (1993). Dynamic regulation of chloroplast transcription. *Plant Physiol.* **103**: 309-313.
- Murayama, M., Hayashi, S., Nishimura, N., Ishide, M., Kobayashi, K., Yagi, Yusuke, Asami, T., Nakamura, T., Shinozaki, K., et Hirayama, T.** (2012). Isolation of *Arabidopsis* ahg11, a weak ABA hypersensitive mutant defective in nad4 RNA editing. *J. Exp. Bot.* **63**: 5301-5310.
- Mustroph, a., Zanetti, M.E., Jang, C.J.H., Holtan, H.E., Repetti, P.P., Galbraith, D.W., Girke, T., et Bailey-Serres, J.** (2009). Profiling transcriptomes of discrete cell populations resolves altered cellular priorities during hypoxia in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **106**: 18843-18848.
- Nakajima, Y. et Mulligan, R.M.** (2001). Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. *Curr. Genet.* **40**: 209-213.
- Newman, M.-A., Sundelin, T., Nielsen, J.T., et Erbs, G.** (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Front. Plant Sci.* **4**: 1-14.
- Nissinen, R.M., Ytterberg, A.J., Bogdanove, A.J., VAN Wijk, K.J., et Beer, S. V** (2007). Analyses of the secretomes of *Erwinia amylovora* and selected hrp mutants reveal novel type III secreted proteins and an effect of HrpJ on extracellular harpin levels. *Mol. Plant Pathol.* **8**: 55-67.
- Norelli, J.L., Jones, A.L., et Aldwinckle, H.S.** (2003). Fire Blight Management in the Twenty-first Century: Using New Technologies that Enhance Host Resistance in Apple. *Plant Dis.* **87**: 756-765.
- Occhialini, A., Cunnac, S., Reymond, N., Genin, S., et Boucher, C.** (2005). Genome-wide analysis of gene expression in *Ralstonia solanacearum* reveals that the hrpB gene acts as a regulatory switch controlling multiple virulence pathways. *Mol. Plant. Microbe. Interact.* **18**: 938-949.
- Oh, C.S., Carpenter, S.C.D., Hayes, M.L., et Beer, S. V.** (2010). Secretion and translocation signals and DspB/F-binding domains in the type III effector DspA/E of *Erwinia amylovora*. *Microbiology* **156**: 1211-1220.
- Okuda, K., Shoki, H., Arai, M., Shikanai, T., Small, I., et Nakamura, T.** (2014). Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *Plant J.* **80**: 870-882.
- Orlob, G.** (1971). History of Plant Pathology in the Middle Ages. In *Introduction to Plant Diseases* (Springer US: Boston, MA), p. 7-20.
- Panstruga, R.** (2003). Establishing compatibility between plants and obligate biotrophic pathogens. *Curr. Opin. Plant Biol.* **6**: 320-326.
- Park, E., Lee, H.-Y., Woo, J., Choi, D., et Dinesh-Kumar, S.P.** (2017). Spatiotemporal Monitoring of *Pseudomonas syringae* Effectors via Type III Secretion Using Split Fluorescent Protein Fragments. *Plant Cell* **29**: 1571-1584.
- Park, Y.J., Lee, H.J., Kwak, K.J., Lee, K., Hong, S.W., et Kang, H.** (2014). MicroRNA400-

- guided cleavage of pentatricopeptide repeat protein mRNAs renders *Arabidopsis thaliana* more susceptible to pathogenic bacteria and fungi. *Plant Cell Physiol.* **55**: 1660-1668.
- Paulin, J.P.** (1996). Control of Fireblight in European Pome Fruits. *Outlook Agric.* **25**: 49-55.
- Pazhouhandeh, M., Dieterle, M., Marrocco, K., Lechner, E., Berry, B., Brault, V. r., Hemmer, O., Kretsch, T., Richards, K.E., Genschik, P., et Ziegler-Graff, V. r.** (2006). F-box-like domain in the polerovirus protein P0 is required for silencing suppressor function. *Proc. Natl. Acad. Sci.* **103**: 1994-1999.
- Pedras, M.S.C. et Adio, A.M.** (2008). Phytoalexins and phytoanticipins from the wild crucifers *Thellungiella halophila* and *Arabidopsis thaliana*: Rapalexin A, wasalexins and camalexin. *Phytochemistry* **69**: 889-893.
- Peeters, N., Carrère, S., Anisimova, M., Plener, L., Cazalé, A., et Genin, S.** (2013). Repertoire, unified nomenclature and evolution of the Type III effector gene set in the *Ralstonia solanacearum* species complex. *BMC Genomics* **14**: 859.
- Petnicki-Ocwieja, T., Schneider, D.J., Tam, V.C., Chancey, S.T., Shan, L., Jamir, Y., Schechter, L.M., Janes, M.D., Buell, C.R., Tang, X., Collmer, A., et Alfano, J.R.** (2002). Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 7652-7.
- Pfalz, J.** (2006). pTAC2, -6, and -12 Are Components of the Transcriptionally Active Plastid Chromosome That Are Required for Plastid Gene Expression. *Plant Cell Online* **18**: 176-197.
- Pfalz, J., Bayraktar, O.A., Prikryl, J., et Barkan, A.** (2009). Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J.* **28**: 2042-2052.
- Pfund, C., Tans-Kersten, J., Dunning, F.M., Alonso, J.M., Ecker, J.R., Allen, C., et Bent, A.F.** (2004). Flagellin Is Not a Major Defense Elicitor in *Ralstonia solanacearum* Cells or Extracts Applied to *Arabidopsis thaliana*. *Mol. Plant-Microbe Interact.* **17**: 696-706.
- Piqué, N., Miñana-Galbis, D., Merino, S., et Tomás, J.** (2015). Virulence Factors of *Erwinia amylovora*: A Review. *Int. J. Mol. Sci.* **16**: 12836-12854.
- Pitzschke, A., Schikora, A., et Hirt, H.** (2009). MAPK cascade signalling networks in plant defence. *Curr. Opin. Plant Biol.* **12**: 421-426.
- Pogson, B.J., Woo, N.S., Förster, B., et Small, I.D.** (2008). Plastid signalling to the nucleus and beyond. *Trends Plant Sci.* **13**: 602-609.
- Popova, L.P., Tsonev, T.D., et Vaklinova, S.G.** (1988). Changes in Some Photosynthetic and Photorespiratory Properties in Barley Leaves after Treatment with Jasmonic Acid. *J. Plant Physiol.* **132**: 257-261.
- Prell, H.H. et Day, P.** (2001). *Plant-Fungal Pathogen Interaction* (Springer Berlin Heidelberg: Berlin, Heidelberg).
- Prikryl, J., Rojas, M., Schuster, G., et Barkan, a.** (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci.* **108**: 415-420.
- Puhar, A. et Sansonetti, P.J.** (2014). Type III secretion system. *Curr. Biol.* **24**: R784-R791.
- Ramamurthi, K.S. et Schneewind, O.** (2005). A Synonymous Mutation in *Yersinia enterocolitica* yopE Affects the Function of the YopE Type III Secretion Signal. *J. Bacteriol.* **187**: 707-715.
- Ranjan, R. et Lewak, S.** (1992). Jasmonic acid promotes germination and lipase activity in non-stratified apple embryos. *Physiol. Plant.* **86**: 335-339.
- Reboutier, D., Frankart, C., Briand, J., Biligui, B., Laroche, S., Rona, J.-P., Barny, M.-A., et Bouteau, F.** (2007). The HrpN(ea) harpin from *Erwinia amylovora* triggers differential responses on the nonhost *Arabidopsis thaliana* cells and on the host apple cells. *Mol. Plant. Microbe. Interact.* **20**: 94-100.
- Rigail, G. et al.** (2016). Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief. Bioinform.:* bbw092.
- Robert-Seilantantz, a, Shan, L., Zhou, J.M., et Tang, X.** (2006). The *Pseudomonas syringae* pv.

- tomato DC3000 type III effector HopF2 has a putative myristoylation site required for its avirulence and virulence functions. *Mol Plant Microbe Interact* **19**: 130-138.
- Robinson, M.D. et Oshlack, A.** (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**.
- Rodríguez-Herva, J.J. et al.** (2012). A bacterial cysteine protease effector protein interferes with photosynthesis to suppress plant innate immune responses. *Cell. Microbiol.* **14**: 669-681.
- Rojas, M., Ruwe, H., Miranda, R.G., Zoschke, R., Hase, N., Schmitz-Linneweber, C., et Barkan, A.** (2018). Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5 and PPR10. *Nucleic Acids Res.:* 1-12.
- Romer, P., Hahn, S., Jordan, T., Strauss, T., Bonas, U., et Lahaye, T.** (2007). Plant Pathogen Recognition Mediated by Promoter Activation of the Pepper Bs3 Resistance Gene. *Science* (80-). **318**: 645-648.
- Rowan, B. a. et Bendich, A.J.** (2009). The loss of DNA from chloroplasts as leaves mature: Fact or artefact? *J. Exp. Bot.* **60**: 3005-3010.
- Sagan, L.** (1967). On the origin of mitosing cells. *J. Theor. Biol.* **14**: 225-IN6.
- Salanoubat, M. et al.** (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**: 497-502.
- Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., et Lurin, C.** (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* **581**: 4132-4138.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., et Tabata, S.** (1999). Complete Structure of the Chloroplast Genome of *thaliana*. *DNA Res.* **290**: 283-290.
- Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J.M., et Knoop, V.** (2013). A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol.* **10**: 1549-1556.
- Scharte, J., Schön, H., et Weis, E.** (2005). Photosynthesis and carbohydrate metabolism in tobacco leaves during an incompatible interaction with *Phytophthora nicotianae*. *Plant, Cell Environ.* **28**: 1421-1435.
- Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas, a., et Barkan, a.** (2006). A Pentatricopeptide Repeat Protein Facilitates the trans-Splicing of the Maize Chloroplast rps12 Pre-mRNA. *Plant Cell Online* **18**: 2650-2663.
- Scholze, H. et Boch, J.** (2011). TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.* **14**: 47-53.
- Sharma, M. et Pandey, G.K.** (2016). Expansion and Function of Repeat Domain Proteins During Stress and Development in Plants. *Front. Plant Sci.* **6**: 1-15.
- Shi, X., Bentolila, S., et Hanson, M.R.** (2016). Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in *Arabidopsis*. *Plant Signal. Behav.* **11**: 1-4.
- Shi, X., Castandet, B., Germain, A., Hanson, M.R., et Bentolila, S.** (2017). ORRM5, an RNA recognition motif-containing protein, has a unique effect on mitochondrial RNA editing. *J. Exp. Bot.* **68**: 2833-2847.
- Shi, X., Hanson, M.R., et Bentolila, S.** (2015). Two RNA recognition motif-containing proteins are plant mitochondrial editing factors. *Nucleic Acids Res.* **43**: 3814-3825.
- Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T., et Hashimoto, T.** (2001). The chloroplast clpP gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. *Plant Cell Physiol.* **42**: 264-273.
- Sloan, D.B., Wu, Z., et Sharbrough, J.** (2018). Correction of Persistent Errors in *Arabidopsis* Reference Mitochondrial Genomes. *Plant Cell* **30**: 525-527.
- Small, I.D. et Peeters, N.** (2000). The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**: 45-47.

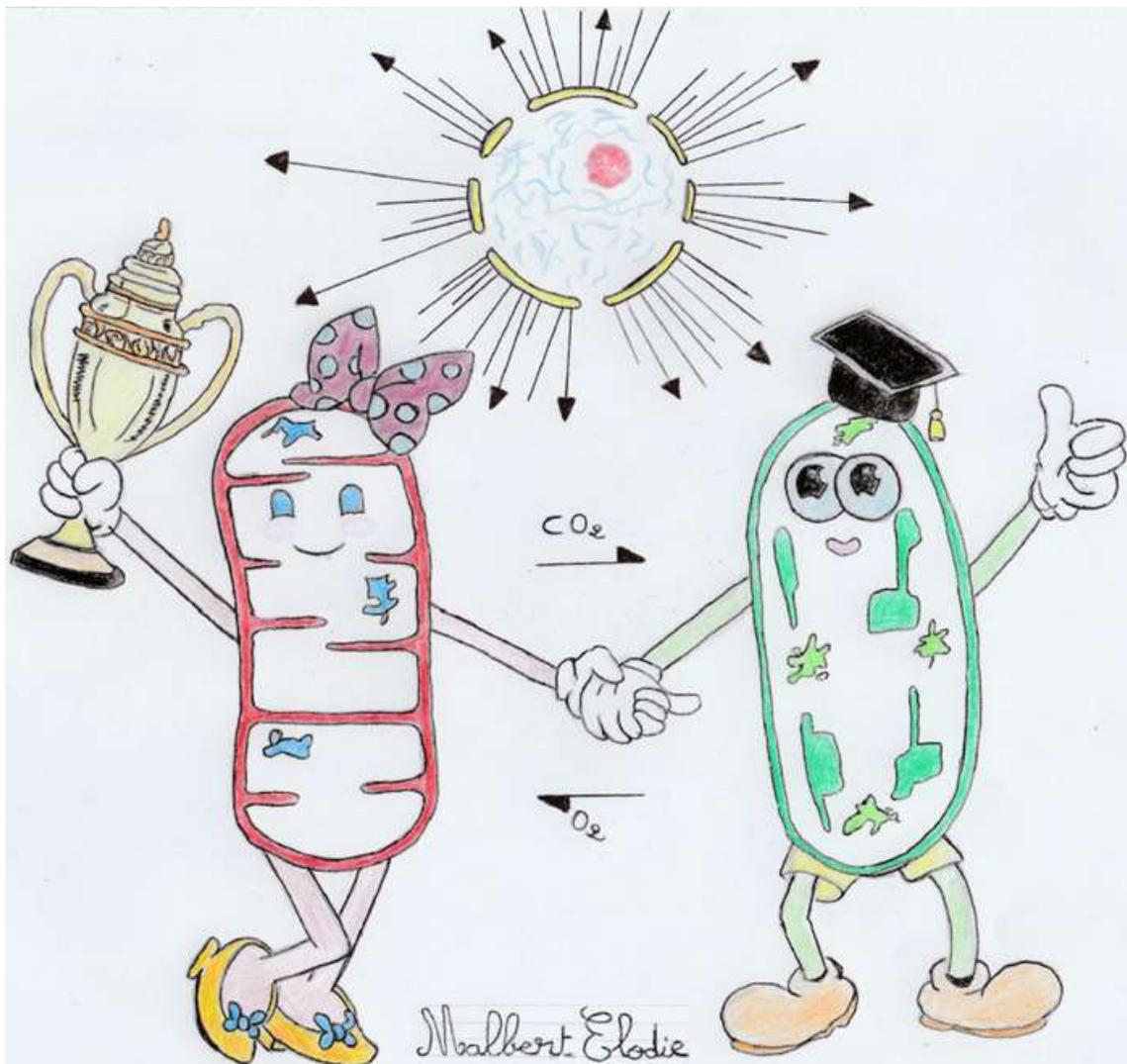
- Small, I.D., Rackham, O., et Filipovska, A.** (2013). Organelle transcriptomes: Products of a deconstructed genome. *Curr. Opin. Microbiol.* **16**: 652-658.
- Smits, T.H.M., Guerrero-Prieto, V., Hernández-Escarcega, G., Blom, J., Goesmann, a, Rezzonico, F., Duffy, B., et Stockwell, V.O.** (2014). Whole-genome sequencing of *Erwinia amylovora* strains from Mexico detects SNPs in *rpsL* conferring streptomycin resistance and in the *avrRpt2* effector altering host interactions. *Genome Announc.* **2**: e01229-13.
- Smits, T.H.M., Rezzonico, F., Kamber, T., Blom, J., Goesmann, A., Frey, J.E., et Duffy, B.** (2010). Complete Genome Sequence of the Fire Blight Pathogen *Erwinia amylovora* CFBP 1430 and Comparison to Other *Erwinia* spp. *Mol. Plant-Microbe Interact.* **23**: 384-393.
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O., et Knoop, V.** (1999). Plant mitochondrial RNA editing. *J. Mol. Evol.* **48**: 303-12.
- Stern, D.B., Goldschmidt-Clermont, M., et Hanson, M.R.** (2010). Chloroplast RNA Metabolism. *Annu. Rev. Plant Biol.* **61**: 125-155.
- Strawn, M. a., Marr, S.K., Inoue, K., Inada, N., Zubieta, C., et Wildermuth, M.C.** (2007). Arabidopsis isochorismate synthase functional in pathogen-induced salicylate biosynthesis exhibits properties consistent with a role in diverse stress responses. *J. Biol. Chem.* **282**: 5919-5933.
- Suarez-Rodriguez, M.C., Adams-Phillips, L., Liu, Y., Wang, H., Su, S.-H., Jester, P.J., Zhang, S., Bent, a. F., et Krysan, P.J.** (2006). MEKK1 Is Required for flg22-Induced MPK4 Activation in Arabidopsis Plants. *Plant Physiol.* **143**: 661-669.
- Sun, T., Bentolila, S., et Hanson, M.R.** (2016). The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci.* **21**: 962-973.
- Sung, T.Y., Tseng, C.C., et Hsieh, M.H.** (2010). The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria. *Plant J.* **63**: 499-511.
- Szurek, B., Marois, E., Bonas, U., et Van Ackerveken, G. Den** (2001). Eukaryotic features of the Xanthomonas type III effector AvrBs3: Protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. *Plant J.* **26**: 523-534.
- Takenaka, M., Verbitskiy, D., Zehrmann, A., et Brennicke, A.** (2010). Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of Arabidopsis thaliana. *J. Biol. Chem.* **285**: 27122-27129.
- Takenaka, M., Zehrmann, a., Verbitskiy, D., Kugelman, M., Hartel, B., et Brennicke, a.** (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci.* **109**: 5104-5109.
- Takenaka, M., Zehrmann, A., Brennicke, A., et Graichen, K.** (2013). Improved Computational Target Site Prediction for Pentatricopeptide Repeat RNA Editing Factors. *PLoS One* **8**: 1-12.
- Taylor, N.L., Tan, Y.F., Jacoby, R.P., et Millar, a. H.** (2009). Abiotic environmental stress induced changes in the Arabidopsis thaliana chloroplast, mitochondria and peroxisome proteomes. *J. Proteomics* **72**: 367-378.
- Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S.Y., Howe, G. a., et Browse, J.** (2007). JAZ repressor proteins are targets of the SCFCOII complex during jasmonate signalling. *Nature* **448**: 661-665.
- Timmis, J.N., Ayliff, M. a., Huang, C.Y., et Martin, W.** (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**: 123-135.
- Torres, M.A., Jones, J.D.G., et Dangl, J.L.** (2006). Reactive oxygen species signaling in response to pathogens. *Plant Physiol.* **141**: 373-378.
- De Torres Zabala, M. et al.** (2015). Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nat. Plants* **1**: 15074.
- Unseld, M., Marienfeld, J.R., Brandt, P., et Brennicke, A.** (1997). The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15**: 57-61.

- Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., et Giegé, P.** (2008). PPR336 is Associated with Polysomes in Plant Mitochondria. *J. Mol. Biol.* **375**: 626-636.
- Vanlerberghe, G.C.** (2013). Alternative oxidase: A mitochondrial respiratory pathway to maintain metabolic and signaling homeostasis during abiotic and biotic stress in plants. *Int. J. Mol. Sci.* **14**: 6805-6847.
- Veronese, P., Nakagami, H., Bluhm, B., Abuqamar, S., Chen, X., Salmeron, J., Dietrich, R.A., Hirt, H., et Mengiste, T.** (2006). The membrane-anchored BOTRYTIS-INDUCED KINASE1 plays distinct roles in Arabidopsis resistance to necrotrophic and biotrophic pathogens. *Plant Cell* **18**: 257-73.
- Vidal, G., Ribas-Carbo, M., Garmier, M., Dubertret, G., Rasmusson, a. G., Mathieu, C., Foyer, C.H., et De Paepe, R.** (2007). Lack of Respiratory Chain Complex I Impairs Alternative Oxidase Engagement and Modulates Redox Signaling during Elicitor-Induced Cell Death in Tobacco. *Plant Cell Online* **19**: 640-655.
- Wagoner, J. a., Sun, T., Lin, L., et Hanson, M.R.** (2015). Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. *J. Biol. Chem.* **290**: 2957-2968.
- Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D.U., Fu, X.D., et Dong, X.** (2011). Timing of plant immune responses by a central circadian regulator. *Nature* **470**: 110-115.
- Waszczak, C., Carmody, M., et Kangasjärvi, J.** (2018). Reactive Oxygen Species in Plant Signaling. *Annu. Rev. Plant Biol.* **69**: 209-236.
- Weihe, A., Liere, K., et Borner, T.** (2012). *Organelle Genetics*. C.E. Bullerwell, éd (Springer Berlin Heidelberg: Berlin, Heidelberg).
- Wilton, M., Subramaniam, R., Elmore, J., Felsensteiner, C., Coaker, G., et Desveaux, D.** (2010). The type III effector HopF2 Pto targets Arabidopsis RIN4 protein to promote *Pseudomonas syringae* virulence. *Proc. Natl. Acad. Sci.* **107**: 2349-2354.
- Xie, T., Chen, D., Wu, J., Huang, X., Wang, Y., Tang, K., Li, J., Sun, M., et Peng, X.** (2016). Growing Slowly 1 locus encodes a PLS-type PPR protein required for RNA editing and plant development in Arabidopsis. *J. Exp. Bot.* **67**: 5687-5698.
- Xin, X.-F. et He, S.Y.** (2013). *Pseudomonas syringae* pv. *tomato* DC3000: A Model Pathogen for Probing Disease Susceptibility and Hormone Signaling in Plants. *Annu. Rev. Phytopathol.* **51**: 473-498.
- Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., et Nakamura, T.** (2013). Elucidation of the RNA Recognition Code for Pentatricopeptide Repeat Proteins Involved in Organelle RNA Editing in Plants. *PLoS One* **8**.
- Yang, J., Zhang, M., et Wang, X.** (2017). Crystal structure of the chloroplast RNA editing factor MORF2. *Biochem. Biophys. Res. Commun.* **495**: 2038-2043.
- Yang, L., Peng, X., et Sun, M.X.** (2011). AtNG1 encodes a protein that is required for seed germination. *Plant Sci.* **181**: 457-464.
- Yang, S.F. et Hoffman, N.E.** (1984). Ethylene biosynthesis and its regulation in higher plants. *Plant Physiol.* **35**: 155-189.
- Yap, A., Kindgren, P., Colas Des Francs-Small, C., Kazama, T., Tanz, S.K., Toriyama, K., et Small, I.** (2015). AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5, and also promotes atpF splicing in Arabidopsis and rice. *Plant J.* **81**: 661-669.
- Yin, P. et al.** (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**: 168-171.
- Yuan, H. et Liu, D.** (2012). Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. *Plant J.* **70**: 432-444.

- Zghidi-Abouid, O., Merendino, L., Buhr, F., Malik Ghulam, M., et Lerbs-Mache, S.** (2011). Characterization of plastid psbT sense and antisense RNAs. *Nucleic Acids Res.* **39**: 5379-5387.
- Zhang, J.X., Wang, C., Yang, C.Y., Wang, J.Y., Chen, L., Bao, X.M., Zhao, Y.X., Zhang, H., et Liu, J.** (2010). The role of arabidopsis AtFes1A in cytosolic Hsp70 stability and abiotic stress tolerance. *Plant J.* **62**: 539-548.
- Zhao, Y., He, S.-Y., et Sundin, G.W.** (2006). The *Erwinia amylovora* avrRpt2EA gene contributes to virulence on pear and AvrRpt2EA is recognized by Arabidopsis RPS2 when expressed in *Pseudomonas syringae*. *Mol. Plant. Microbe. Interact.* **19**: 644-654.
- Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J., et Börner, T.** (2012). The Primary Transcriptome of Barley Chloroplasts: Numerous Noncoding RNAs and the Dominating Role of the Plastid-Encoded RNA Polymerase. *Plant Cell* **24**: 123-136.
- Zhu, Q. et al.** (2012a). SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J.* **71**: 836-849.
- Zhu, Q., Dugardeyn, J., Zhang, C., Mühlenbock, P., Eastmond, P.J., Valcke, R., De Coninck, B., Öden, S., Karampelias, M., Cammue, B.P. a., Prinsen, E., et Van Der Straeten, D.** (2014). The arabidopsis thaliana RNA editing factor SLO2, which affects the mitochondrial electron transport chain, participates in multiple stress and hormone responses. *Mol. Plant* **7**: 290-310.
- Zhu, Q., Meyer, E.H., et Van Der Straeten, D.** (2012b). Functional analysis of SLO2 provides new insight into the role of plant PPR proteins. *Plant Signal. Behav.* **7**: 1209-11.
- Zsigmond, L., Rigo, G., Szarka, a., Szekely, G., Otvos, K., Darula, Z., Medzihradzky, K.F., Koncz, C., Koncz, Z., et Szabados, L.** (2008). Arabidopsis PPR40 Connects Abiotic Stress Responses to Mitochondrial Electron Transport. *Plant Physiol.* **146**: 1721-1737.
- Zubo, Y.O., Potapova, T. V., Yamburenko, M. V., Tarasenko, V.I., Konstantinov, Y.M., et Börner, T.** (2014). Inhibition of the electron transport strongly affects transcription and transcript levels in Arabidopsis mitochondria. *Mitochondrion* **19**: 222-230.

Résumé de thèse vulgarisé :

Les eucaryotes sont composés de cellules, elles-mêmes possédant de petites structures possédant leur propre ADN, assurant entre autres l'apport et la conversion de l'énergie. Ces structures sont les organites à génome. Il en existe deux principaux types chez la plante, le chloroplaste et la mitochondrie. L'objectif de la thèse a été de regarder si l'expression des gènes des organites de la plante modèle *A. thaliana* était influencée par des protéines d'une famille particulière, les protéines PPR, lorsque la plante est attaquée par un pathogène. Ces protéines PPR (pour *pentatricopeptide repeat*) sont nécessaires au bon fonctionnement de l'expression des gènes des organites quand elles proviennent de la plante, mais peuvent aussi venir du pathogène, probablement pour faciliter l'attaque. Ce travail de thèse a pour objectif de mieux comprendre le rôle de ces protéines, du côté de la plante comme du pathogène, en condition d'infection.



Lay summary:

Eukaryotes are made of cells, themselves having little structures possessing their own DNA. These structures, known as genomic organelles, are mainly involved in energy supply and conversion. There are two main types in a plant cell: chloroplasts and mitochondria. The aim of my PhD studies was to test if organellar gene expression of the model plant *A. thaliana* was modified by a specific protein family, the PPR proteins, during a pathogen attack. These PPR proteins (for *pentatricopeptide repeat*) are needed for the proper functioning of organellar gene expression when they are native from the plant cell, but can also come from the pathogen, probably to facilitate the infection. This work aims to understand the role of these proteins during infection.

Titre : Caractérisation de protéines PPR impliquées dans le stress biotique chez *A. thaliana*.

Mots clés : PPR ; stress biotique ; organites ; transcription ; *A. thaliana* ; *Erwinia amylovora* ; *Ralstonia solanacearum*

Résumé : A la différence des mammifères, les plantes ne possèdent pas de cellules spécialisées dans la défense face aux pathogènes. Chaque cellule végétale peut déclencher une réponse immunitaire. Pour interagir avec succès avec la plante, les pathogènes doivent alors supprimer ou contourner les défenses de l'hôte. Afin d'y parvenir, les bactéries pathogènes disposent des effecteurs, des protéines qui peuvent être injectées dans la cellule végétale. De nombreux effecteurs sont connus pour cibler les organites lors de l'infection, confirmant l'importance du chloroplaste et de la mitochondrie dans les mécanismes de défense des cellules végétales. Dans ces conditions, il demeure vital pour la plante de garder la main sur l'expression des gènes des organites afin d'assurer une réponse proportionnée au risque encouru sans pénaliser la croissance de façon disproportionnée. A la différence de l'expression des gènes nucléaires, la régulation de l'expression des gènes des organites se fait principalement lors d'étapes très complexes de maturation post-transcriptionnelle. Parmi les protéines impliquées dans ces étapes de maturation, on trouve les protéines *pentatricopeptide repeat* (PPR). Les protéines PPR sont impliquées dans de nombreuses étapes de maturation des ARN des organites, comme l'édition C vers U ou l'épissage. Elles sont également présentes chez d'autres eucaryotes, mais n'ont jamais été étudiées chez les bactéries. L'hypothèse testée dans le cadre de la thèse est que ces protéines PPR, qu'elles soient d'origine exogène ou endogène, sont impliquées dans des modifications de l'expression des gènes des organites en condition de stress biotique.

Afin de tester notre hypothèse, nous nous sommes intéressés à PGN (*Pentatricopeptide repeat protein for Germination on NaCl*) chez la plante modèle *A. thaliana*.

Caractérisée par Laluk *et al.* (2011), le mutant KO montre une sensibilité accrue au nécrotrophe *Botrytis cinerea*, et l'expression du gène codant pour cette protéine est induite après infection. Nous avons mis en évidence des défauts d'édition dans la séquence non codante en amont de *nad6* et dans *cox2*, deux gènes mitochondriaux. Leur édition ne varie cependant pas en condition d'infection par *Botrytis cinerea*.

Dans la même optique, à la suite d'un crible bioinformatique, nous nous sommes intéressés à deux protéines PPR bactériennes que nous avons trouvées chez les phytopathogènes *Erwinia amylovora* et *Ralstonia solanacearum*. Probablement obtenues par les bactéries par transfert horizontal de gènes, il s'agit de la première caractérisation de PPR bactériennes à notre connaissance. Ces protéines possèdent des caractéristiques d'effecteurs, c'est-à-dire des protéines injectées par la bactérie dans la plante durant l'infection. Si nous n'avons pas vu de modification du transcriptome des organites de la plante provoquée par la surexpression de ces protéines PPR exogènes, nous avons cependant mis en évidence une baisse significative du taux d'incidence de la maladie provoquée par *E. amylovora* en l'absence d'un gène fonctionnel codant pour sa PPR chez la plante hôte *Malus domestica* « Golden delicious ». Pour la PPR d'*E. amylovora* comme celle de *R. solanacearum*, nous avons également trouvés plusieurs interactants en double hybride levure chez *A. thaliana*, représentant de nombreuses cibles putatives à étudier. Afin de réaliser ces expérimentations et d'obtenir ces résultats, nous avons eu besoin d'outils particuliers. Nous avons donc développé un pipeline spécifique d'analyse de données de séquençage d'ARN ainsi qu'une méthode améliorée de prédiction des zones de fixation des protéines PPR, ouvrant la voie à une caractérisation simplifiée de nombreuses protéines.

Titre : Characterization of PPR proteins involved in biotic stress in *A. thaliana*.

Keywords : PPR ; biotic stress ; organelles ; transcription ; *A. thaliana* ; *Erwinia amylovora* ; *Ralstonia solanacearum*

Abstract :

Compared to mammals, plants do not have highly specialized cells involved in defense against pathogens. Each plant cell is able to start an immune response. To interact successfully with plants, pathogens have to block or bypass host defenses. To do so, phytopathogenic bacteria can use effectors, which are basically bacterial proteins injected in the plant cell during infection. Several effectors are known to target organelles during infection, supporting the idea that chloroplasts and mitochondria are key players in plant cell defense. As a reason, it remains necessary for the plant to keep organellar gene expression under control in order to ensure a response in proportion to the risk, without penalizing growth. Unlike nuclear gene expression, organellar gene expression regulation goes through highly complex post-transcriptional maturation steps. Among proteins involved in these events, PPR proteins (for *pentatricopeptide repeat*) are known to be very important. PPR proteins are involved in several RNA maturation steps in organelles, like C to U editing or splicing. They are studied in several eukaryotes, but not in bacteria. During my PhD studies, the hypothesis is exogenous or endogenous PPR proteins are involved in organellar gene expression modifications during biotic stress.

To test our hypothesis, we work on PGN (*Pentatricopeptide repeat protein for Germination on NaCl*) in the plant model *A. thaliana*.

Characterized by Laluk *et al.* (2011), the KO mutant displays an enhanced sensitivity to the necrotrophic pathogen *Botrytis cinerea*, and PGN gene expression is induced after infection. We find two editing defects for the KO mutant, in *nad6* 5' non coding sequence and in *cox2* coding sequence. However, editing at these two sites does not vary in wild type plants during *Botrytis cinerea* infection.

Using a bioinformatic screen, we find several bacterial PPR proteins. Two of them are encoded by bacterial plant pathogens: *Erwinia amylovora* and *Ralstonia solanacearum*. To our knowledge, these proteins, putatively obtained through horizontal gene transfer, are the first bacterial PPR proteins to be characterized. They also share similarities with bacterial effectors. If overexpression of these bacterial PPR proteins in *A. thaliana* does not unveil organellar transcriptome modifications, we show a decrease of the incidence rate of the disease caused by *E. amylovora* in the host plant *Malus domestica* "Golden delicious" without a functional gene coding for the PPR protein. For both *Erwinia* and *Ralstonia* PPR, we find several interactants in *A. thaliana* using Yeast Two Hybrid, each of them representing a potential target that could be studied.

In order to perform these experiments and obtain these results, we needed very specific tools. During the PhD studies, we develop an *RNAseq* analysis pipeline and an enhanced method to predict PPR binding sites, opening the way to an easier characterization of several PPR proteins.