



**HAL**  
open science

## Les multiples facettes du chloroplaste: approches interdisciplinaires

Etienne Delannoy

► **To cite this version:**

Etienne Delannoy. Les multiples facettes du chloroplaste: approches interdisciplinaires. Sciences de l'environnement. Université Paris Saclay, 2022. <tel-04665377>

**HAL Id: tel-04665377**

**<https://hal.inrae.fr/tel-04665377v1>**

Submitted on 31 Jul 2024

**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.



HAL Authorization

# Les multiples facettes du chloroplaste des plantes

Approches interdisciplinaires

**Habilitation à diriger des recherches  
de l'Université Paris-Saclay**

présentée et soutenue à Gif-sur-Yvette le 6 avril 2022, par

**Etienne DELANNOY**

## Composition du jury

<b>Christophe AMBROISE</b> Professeur, Université Evry-Val d'Essonne	Rapporteur
<b>Ingrid LAFONTAINE</b> Professeure, Sorbonne Université	Rapporteuse
<b>Philippe GIEGE</b> Directeur de recherche, CNRS	Rapporteur
<b>Mathilde FAGARD</b> Directrice de recherche, INRAE	Examinatrice
<b>Laurence MARECHAL-DROUARD</b> Directrice de recherche, CNRS	Examinatrice
<b>Olivier VALLON</b> Directeur de recherche, CNRS	Examinateur

**Titre :** Les multiples facettes du chloroplaste des plantes. Approches interdisciplinaires

**Mots clés :** plastome, transcriptome, réseau de gènes, stress, protéines à domaine pentatricopeptide, corégulation

**Résumé :** le chloroplaste est un des organites des cellules végétales mais il partage avec les mitochondries la particularité d'avoir son propre génome. Il occupe une place centrale dans la physiologie des plantes en tant que lieu de la photosynthèse et également en raison de son implication dans de nombreuses voies métaboliques. Pour mieux en comprendre le fonctionnement, j'ai étudié le chloroplaste sous plusieurs angles.

Tout d'abord, l'étude de l'évolution des génomes chloroplastiques et du contenu en gènes en lien avec la perte de la photosynthèse chez plusieurs espèces d'orchidées (notamment *Rhizanthella gardneri*) permet d'identifier les fonctions non-photosynthétiques des plantes. Des analyses transcriptomiques et métabolomiques sur ces mêmes plantes indiquent que la perte partielle ou totale de l'activité photosynthétique s'accompagne de changements métaboliques majeurs sans pour autant nécessiter de nouvelles fonctions métaboliques.

Ensuite, la caractérisation fonctionnelle de gènes codant notamment pour les protéines à domaines

pentatricopeptides répétés chez *Arabidopsis thaliana* m'ont permis de mieux comprendre les mécanismes contrôlant l'expression génétique du chloroplaste lors du développement ou lors des réponses aux stress. Pour cela j'ai développé des approches d'analyse du transcriptome au niveau moléculaire, bio-informatique et statistique.

En condition de stress, l'activité du chloroplaste semble systématiquement impactée. Cette impression a été confirmée par une méta-analyse de données transcriptomiques que j'ai co-pilotée. Le réseau de gènes corégulés obtenu représente la réponse commune à de multiples stress et nous y observons la place prépondérante des activités associées au chloroplaste. La caractérisation de quelques gènes de fonction inconnue dans ce réseau confirme son fort pouvoir prédictif. Il constitue donc une base solide pour mes futurs travaux sur le rôle du chloroplaste dans la réponse aux stress simples et multiples. Compte-tenu de mon expérience dans le travail à l'interface et du fait que le chloroplaste est un carrefour de la physiologie végétale, ces projets seront nécessairement interdisciplinaires.

**Title : The multifaceted chloroplast of plants. Interdisciplinary approaches.**

**Keywords :** transcriptome, gene network, coregulation, stress, pentatricopeptide repeat proteins

**Abstract :** The chloroplast is one of the organelles of plant cells, but it shares with the mitochondria the peculiarity of having its own genome. It occupies a central place in the physiology of plants as a site of photosynthesis and also because of its involvement in many metabolic pathways. To better understand how it works, I studied the chloroplast from several angles.

First, the study of the evolution of chloroplast genomes and of the gene content linked to the loss of photosynthesis in several species of orchids (in particular *Rhizanthella gardneri*) makes it possible to identify the non-photosynthetic functions of plants. . Transcriptomic and metabolomic analyzes on these same plants indicate that the partial or total loss of photosynthetic activity is accompanied by major metabolic changes without requiring new metabolic functions.

Then, the functional characterization of genes coding in particular for pentatricopeptide repeat proteins in *Arabidopsis thaliana* allowed me to better understand the mechanisms controlling the genetic expression of the plastid during development or during stress responses.

For this I have developed approaches to transcriptome analysis at the molecular, bioinformatics and statistical levels.

Under stress conditions, the activity of the chloroplast seems to be systematically affected. This impression was confirmed by a meta-analysis of transcriptomic data that I co-piloted. The network of co-regulated genes obtained represents the common response to multiple stresses and we observe the preponderant place of activities associated with the chloroplast. The characterization of a few genes of unknown function in this network confirms its strong predictive power. It therefore constitutes a solid basis for my future work on the role of the chloroplast in the response to single and multiple stresses. Given my experience working at the interface and the fact that the chloroplast is a crossroads of plant physiology, these projects will necessarily be interdisciplinary.

# Contenu

Préambule .....	5
Parcours de recherche.....	5
Introduction.....	6
Approches évolutives : étude d'orchidées non autotrophes .....	9
Comparaison de plastomes .....	10
Comparaison de génomes.....	13
Comparaison de transcriptomes .....	16
Approches de génomique fonctionnelle : étude des plantes modèles.....	22
Régulation de l'expression génétique du chloroplaste .....	22
Le chloroplaste lors du développement.....	29
Réponses aux stress chez les plantes : le chloroplaste en filigrane .....	32
Approches de méta-analyse : la place centrale du chloroplaste dans la réponse au stress.....	44
Projet : Vers le réseau de régulation du chloroplaste en réponse au stress.....	48
Bibliographie.....	52
Curriculum vitae .....	59
Liste des publications .....	62

# Préambule

Je pourrais vous présenter mes travaux suivant leur ordre chronologique et prétendre que ce parcours a été linéaire et planifié pour arriver là où j'en suis aujourd'hui mais ça n'est évidemment pas le cas, en tous cas pas pour moi. Comme pour beaucoup d'entre nous je crois, mon parcours et mes travaux de recherche ont été faits de tours et de détours au gré des opportunités, des rencontres et des envies. Néanmoins, ce sont ces rencontres, ces collaborations avec des chimistes, des biochimistes, des biophysiciens, des génomiciens, des écologues, des bio-informaticiens, des biostatisticiens... qui ont nourri les compétences du biologiste moléculaire que je suis pour devenir un analyste de données moléculaires, fan d'interdisciplinarité. Je vous présenterai donc mes travaux de recherche non pas dans l'ordre chronologique mais selon les grands axes qui me permettront, je l'espère, de mettre en lumière les multiples facettes du chloroplaste et l'intérêt d'une approche interdisciplinaire pour son étude. Ce plan m'amènera cependant à mettre en regard des résultats qui s'étalent sur 20 ans durant lesquels les techniques (et mes compétences) ont largement évolué. Je commencerai donc par un bref rappel de mon parcours.

## Parcours de recherche

De 1999 à 2003, lors de mon DEA puis de ma thèse, j'ai caractérisé moléculairement les peroxydases de classe III associées à la résistance du cotonnier à *Xanthomonas campestris* pv. *malvacearum*, agent causal de la bactériose. Le travail sur cette plante de grande culture et de forte importance économique a été effectué dans le laboratoire de Résistance des Plantes à l'IRD de Montpellier sous la direction de Michel Nicole. L'équipe déployait des outils performants de biochimie, de génétique et de microscopie pour étudier tant le côté de la plante que celui du pathogène. L'analyse moléculaire était en retrait mais en cours de développement lorsque je suis arrivé. Il va sans dire que les ressources génomiques étaient extrêmement limitées.

En 2004 et 2005, j'ai effectué un premier stage post-doctoral dans le laboratoire de Génétique et Biophysique des Plantes de Christophe Robaglia à l'université de la Méditerranée à Marseille. J'y ai travaillé sur le lien entre le stress lié au cadmium et les petits ARNs chez *Arabidopsis thaliana*. Cela m'a permis de découvrir les outils de recherche disponibles sur le modèle *Arabidopsis* (génome disponible, banques de mutants, outils de transcriptomiques). De plus, les petits ARNs étaient (et sont toujours) un domaine de recherche extrêmement actif établissant un lien entre les stress biotiques (principalement les virus à cette époque) et abiotiques (carences).

Entre 2006 et mi-2009, je suis parti dans le tout nouveau Australian Research Council Centre of Excellence for Plant Energy Biology (ARC CoE PEB) situé à Perth en Australie pour travailler dans le laboratoire de Ian Small. Ce centre regroupe autour du thème du métabolisme énergétique des équipes aux compétences et thématiques complémentaires et dispose d'un matériel de pointe en transcriptomique, protéomique et métabolomique. J'ai participé à de nombreux projets sur la régulation de l'expression génétique des organites (chloroplaste et mitochondrie) chez *Arabidopsis thaliana* en caractérisant plusieurs acteurs moléculaires dont des protéines à domaines

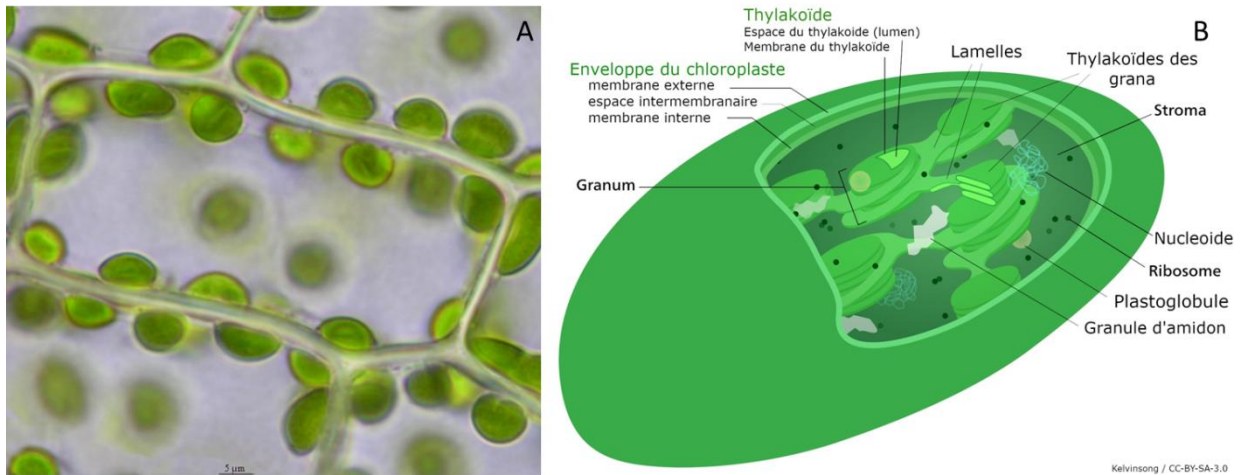
pentatricopeptides répétés (PPR) ou en étudiant leur réponse transcriptomique lors de stress ou du développement. J'ai également séquencé et annoté le génome chloroplastique de l'orchidée souterraine *Rhizanthella gardneri* en utilisant une technique de séquençage de nouvelle génération.

Entre mi-2009 et mi-2010, j'ai travaillé sur la réponse à la carence en phosphate chez *Arabidopsis thaliana* dans l'équipe « Biologie du développement des plantes » dirigée par Laurent Nussaume au CEA de Cadarache. Ce bref séjour m'a permis de participer à des projets de chémo-génomique (criblage de banques de molécules) pour élucider les voies de signalisation de la carence en phosphate et de développer des outils à base de PCR en temps réel.

Depuis 2011, je suis chargé de recherche dans l'équipe « Génomique fonctionnelle d'*Arabidopsis* » de l'URGV qui est devenue « Expression génétique des organites » à l'IPS2. Après avoir été sous la direction de Claire Lurin, j'anime cette équipe depuis février 2021. J'y ai continué la caractérisation de certaines PPR d'*Arabidopsis* mais aussi de pathogènes bactériens. Dans le cadre d'une collaboration avec Marc-André Selosse (MNHN et Université de Gdansk), nous avons développé l'étude de la physiologie d'orchidées européennes dont certaines sont capables de se passer de photosynthèse. Ces travaux se font grâce aux outils développés avec la plateforme transcriptomique dont j'étais le conseiller scientifique avant d'en devenir le responsable scientifique. Ce poste m'a en effet amené à mettre en place les outils tant humides (passage de la technologie puce à ADN vers le RNA-seq) qu'analytiques (bio-informatiques et bio-statistiques) nécessaires à l'évolution de la plateforme. Cela s'est évidemment fait en étroite collaboration avec les membres de la plateforme. De plus, avec Marie-Laure Martin-Magniette, nous menons une méta-analyse des données transcriptomiques générées depuis plus de 15 ans par la plateforme qui nous permet de mettre en évidence le rôle central de la traduction et du chloroplaste dans la réponse d'*Arabidopsis* au stress.

## Introduction

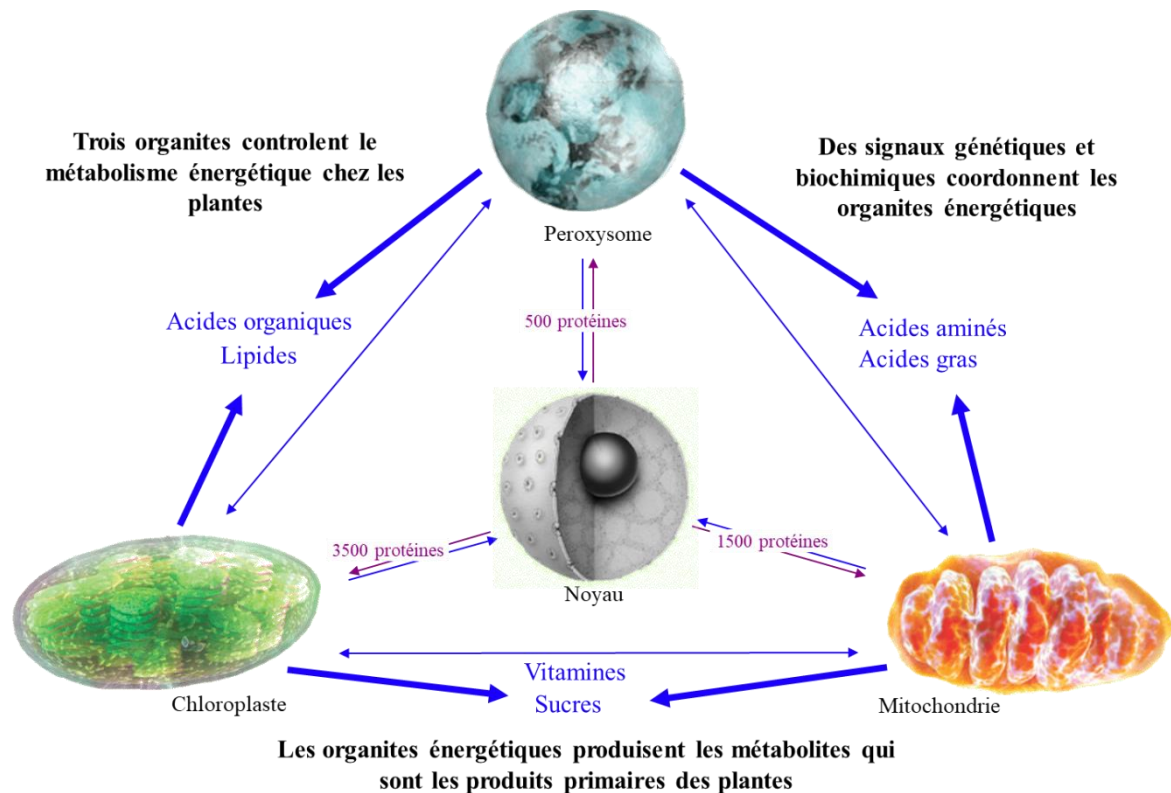
Le chloroplaste est ce qui donne aux plantes leur caractéristique principale aux yeux du grand public : la couleur verte ! Il a été formellement décrit en 1837 (von Mohl, 1837) et pris son nom actuel en 1884 (Strasburger, 1884) qui dérive du grec signifiant «qui est fait de vert ». Le chloroplaste est présent dans les cellules chlorophylliennes des plantes et des algues et leur nombre varie de 1 à plus de 100 par cellule selon les espèces et les tissus (figure 1A). C'est un organite caractérisé par une double membrane et une forte concentration en chlorophylle. A l'intérieur de cette enveloppe, on distingue un système de membranes appelé thylakoïdes qui peuvent former des empilements, les grana (figure 1B).



**Figure 1 : Chloroplaste.** A : chloroplastes dans les cellules de feuille de *Funaria hygrometrica* © Auclair Jean-Jacques. B : Structure du chloroplaste.

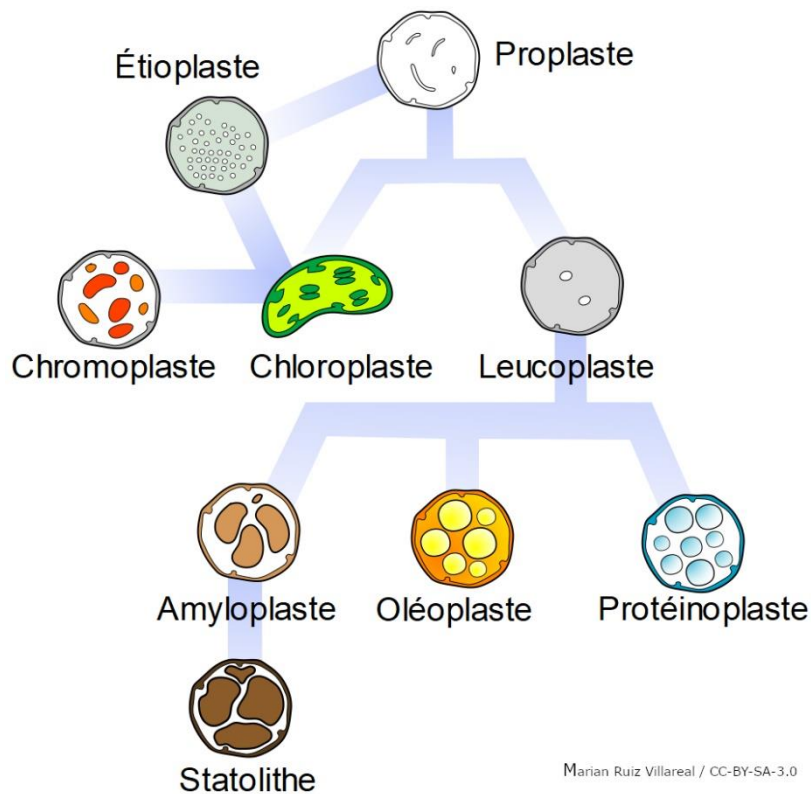
C'est au niveau des thylakoïdes que la chlorophylle est située. Le stroma du chloroplaste inclut entre autres des ribosomes, des grains d'amidon et le nucléoïde qui est un complexe composé d'ADN et de protéines (Majeran et al., 2012). Le chloroplaste contient donc son propre génome (ou plastome) ainsi qu'un système d'expression génétique. Ce génome est une réminiscence de l'origine endosymbiotique du chloroplaste qui dérive d'une cyanobactérie ancestrale qui a été « capturée » par un organisme eucaryote non photo-synthétique il y a plus d'un milliard d'années. Cette hypothèse émise dès 1883 (Schimper, 1883) est maintenant très largement admise par la communauté même si l'identité de cette cyanobactérie ancestrale est toujours débattue (Lawrence et al., 2019; Sánchez-Baracaldo et al., 2017).

Le chloroplaste est le lieu de la photosynthèse ce qui en fait la source principale d'énergie primaire de la vie sur Terre car il produit l'oxygène et le sucre à la base de presque toutes les chaînes « alimentaires » depuis l'intérieur des cellules photosynthétiques jusqu'aux écosystèmes globaux.



**Figure 2: Intégration du chloroplaste dans le métabolisme énergétique de la cellule**

Pour assurer cette fonction, le chloroplaste utilise les gènes de son génome mais suite à l'endosymbiose l'immense majorité des gènes de la cyanobactérie d'origine a été transférée vers le génome nucléaire ou remplacés par des gènes eucaryotes (Nakayama and Archibald, 2012). Ce sont donc plus de 3500 protéines codées par le noyau qui sont envoyées vers le chloroplaste alors que celui-ci ne code qu'environ 100 protéines. De plus, la photosynthèse ne constitue qu'une partie du métabolisme énergétique cellulaire qui implique des échanges de métabolites entre le cytosol, le peroxyosome et la mitochondrie. Ainsi le fonctionnement du chloroplaste implique de nombreuses interactions avec les autres compartiments cellulaires (figure 2). De plus, l'activité du chloroplaste est loin de se limiter à la photosynthèse. Il suffit de réaliser que le chloroplaste n'est qu'une des formes de plastes présentes dans les plantes (figure 3; Solymosi et al., 2018). Les autres formes, qui possèdent toutes le même génome et sont issues du proplaste comme le chloroplaste, sont présentes dans les tissus non photosynthétiques. Ainsi, le chloroplaste est impliqué dans de très nombreuses voies métaboliques et réponses de la plante à son développement et à son environnement.



**Figure 3 : les différentes formes de plastides**

L'étude du chloroplaste implique donc une approche pluri-disciplinaire afin de couvrir toutes ses facettes et obtenir une vue d'ensemble.

## Approches évolutives : étude d'orchidées non autotrophes

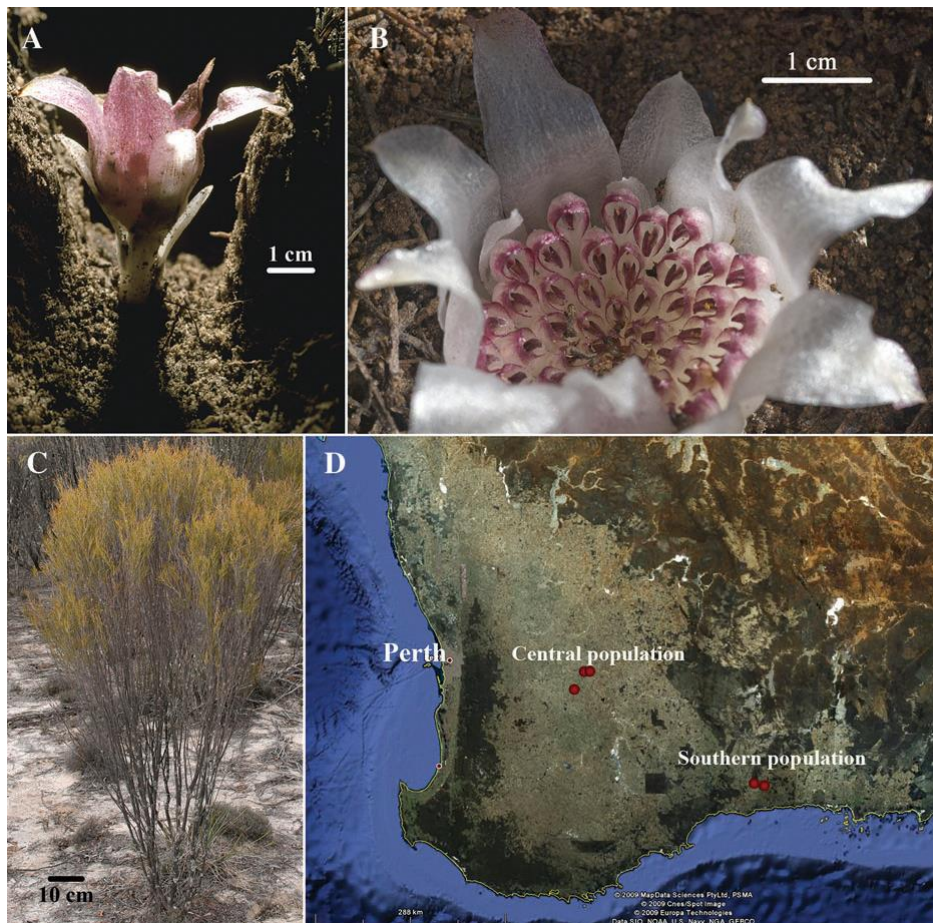
L'étude des fonctions non-photosynthétiques est compliquée dans les organismes photosynthétiques car les perturbations du métabolisme des plastides entraînent des effets pléiotropes qui sont souvent délétères voire létaux. Il existe cependant de nombreuses espèces de plantes capables de se passer de photosynthèse et qui représentent donc des opportunités pour étudier ces fonctions non-photosynthétiques et l'intégration du chloroplaste dans le métabolisme cellulaire. Ce sont des parasites soit d'autres plantes telles que les plantes du genre *Cuscuta*, *Orobanche* ou *Rafflesia* soit des parasites de champignons. Dans ce dernier cas, on parle de plantes mycohétérotrophes. Ces plantes détournent la symbiose mycorhizienne dont bénéficient plus de 85% des plantes vasculaires (Brundrett and Tedersoo, 2018; van der Heijden et al., 2015). Cette relation, généralement mutualiste, donne à la plante une meilleure nutrition minérale et une meilleure résistance aux stress biotiques et abiotiques. Celle-ci fournit à son partenaire fongique des composés organiques issus de la photosynthèse en échange d'eau et de minéraux (N, P, K... ; Rich et al., 2017). Cependant, plus de 500 espèces mycohétérotrophes ont détourné cette relation et dépendent entièrement de leur partenaire fongique pour leur nutrition minérale et organique. Cette stratégie est apparue en

parallèle plus de 50 fois dans 17 familles de plantes mais plus communément chez les *Orchidaceae* (Perez-Lamarque et al., 2020; Merckx and Freudenstein, 2010).

Que ce soit les parasites de plantes ou les mycohétérotrophes, ces espèces non-photosynthétiques ont toutes maintenu un génome chloroplastique ou plastome. La seule exception connue à ce jour est peut-être *Rafflesia lagascae* (Molina et al., 2014). Alors que l'évolution de ce génome depuis l'endosymbiose d'origine est marquée par un transfert massif des gènes vers le noyau, le maintien d'un plastome dans ces plantes indique que ce transfert est limité par la réalisation d'une ou plusieurs fonctions non-photosynthétiques essentielles du plaste. La comparaison du contenu en gènes des plastomes de plante non-photosynthétiques peut permettre l'identification de ces fonctions ainsi que les contraintes limitant le transfert de gènes du plastome vers le génome nucléaire.

## Comparaison de plastomes

En 2008, les plastomes de Dicotylédones parasites étaient déjà disponibles (DePamphilis and Palmer, 1990; Funk et al., 2007) mais aucun de mycohétérotrophe ni de Monocotylédones. Lors de mon séjour en Australie occidentale mon hobby pour les orchidées m'a fait découvrir et observer la rarissime orchidée souterraine *Rhizanthella gardneri* (figure 4). Les 3 espèces (à l'époque, 4 espèces aujourd'hui) présentent la particularité unique avec l'aristolochiacée *Hydnora triceps* de passer l'ensemble de leur cycle de vie sous terre.

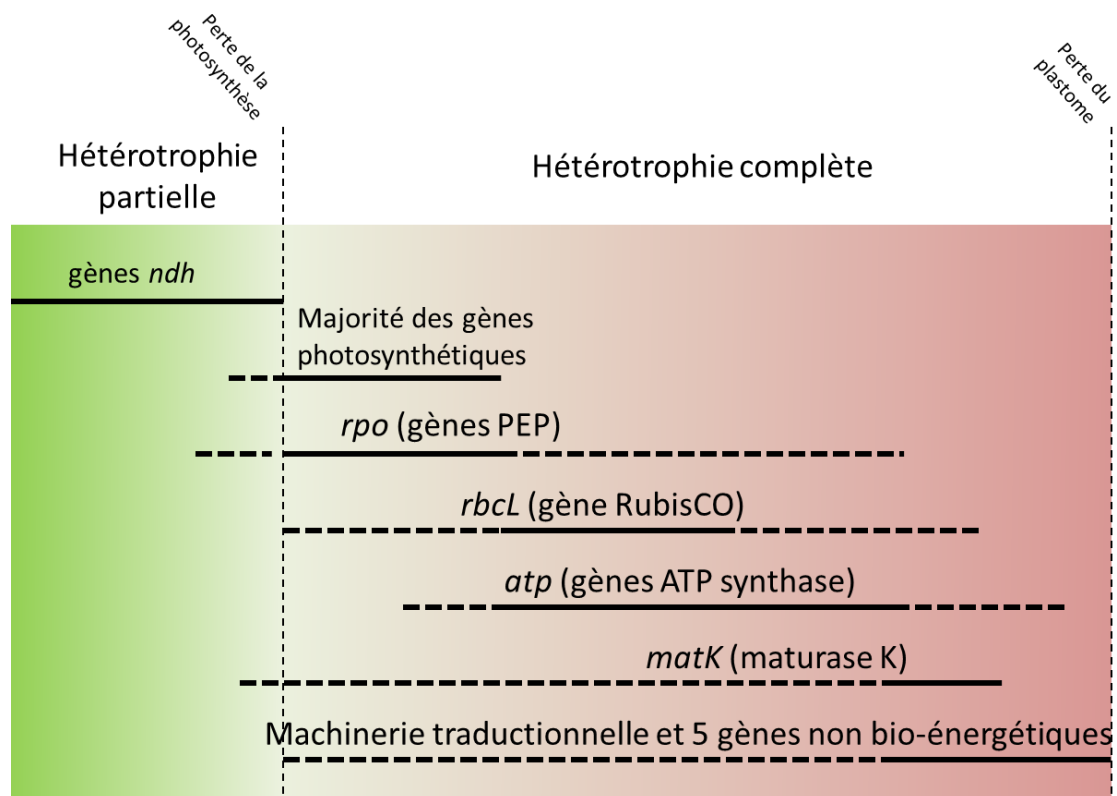


**Figure 4 : Description de *Rhizanthella gardneri*.** A : un capitule détaché de *R. gardneri*. Photo de Susumu Yamaguchi. B : Chaque capitule regroupe environ 50 fleurs simples. C : Habitat de *R.*



ceux associés au métabolisme (transcription et épissage) des ARN (*rpo*, *matK*) alors que des ARN fonctionnels (épissés et édités) ont été détectés. Parmi les gènes conservés, on trouve ceux nécessaires à la traduction (ARN ribosomiaux (ARNr), certains ARN de transfert (ARNt) et la plupart de ceux codant pour des protéines ribosomiales). Cette machinerie traductionnelle, très probablement complétée par des protéines et des ARNt importés du cytosol, semble servir à la production de 4 protéines : ACCD, CLPP, YCF1 et YCF2. ACCD code pour une sous-unité de l'acetyl-CoA carboxylase qui catalyse l'étape initiale et limitante de la biosynthèse d'acides gras dans la plante. CLPP est une sous-unité catalytique de la protéase Clp impliquée dans le contrôle qualité des protéines du plaste. YCF1 et YCF2 font partie de la machinerie d'import de protéines depuis le cytosol vers le plaste. A ces quatre protéines, il faut ajouter l'ARNt glutamate qui est le précurseur de la chlorophylle et des hèmes.

Depuis cette étude, avec l'avènement des techniques de séquençage haut-débit (« NGS »), des dizaines de plastomes de plantes parasites ou mycohétérotrophes ont été séquencées. Des milliers de plastomes sont également disponibles pour des plantes autotrophes mais aussi pour des plantes dites mixotrophes. Ces plantes intermédiaires entre les plantes autotrophes et mycohétérotrophes sont aussi appelées mycohétérotrophes partielles car elles effectuent toujours la photosynthèse mais sont également capables de récupérer des nutriments organiques depuis leur partenaire fongique (Těšitel et al., 2018; Selosse and Roy, 2009). J'ai participé au séquençage de 13 de ces plastomes (Lallemand et al., 2019a) dont les résultats confirment le modèle de dégradation du plastome lors de la perte de la photosynthèse proposé par Graham et al. (2017 ; figure 6).



**Figure 6 : Evolution du plastome après la perte de la photosynthèse.** Repris de Graham et al. (2017). Les lignes pleines indiquent le stade le plus probable de perte des gènes. Les tirets indiquent les extensions possibles.

Ce modèle indique une dégradation progressive mais irréversible des diverses machineries du plaste en fonction de l'ancienneté de la perte de la photosynthèse. La plupart des gènes de la photosynthèse sont très rapidement perdus alors que les gènes les plus conservés parmi les plantes parasites et mycohétérotrophes anciennes sont les 5 gènes mentionnés précédemment : *accD*, *clpP*, *ycf1*, *ycf2* et *trnE* en plus de certains gènes de la machinerie traductionnelle. Ces résultats indiquent que le métabolisme des lipides (avec *accD*) et celui des hèmes (avec *trnE*) sont des fonctions non-photosynthétiques essentielles du chloroplaste.

Cependant l'étude des plastomes n'offre qu'une vue très limitée de l'activité du chloroplaste et de l'impact de la perte de la photosynthèse. En effet, l'évolution du plastome est marquée avant tout par un transfert massif de gènes vers le noyau. L'étude des plastomes permet d'observer la disparition de gènes du plastome qui peut résulter soit d'une perte de ces gènes soit de leur transfert vers le noyau. De plus, à l'image de la vacuole ou du peroxysoxe, le chloroplaste importe la quasi-totalité des protéines dont il a besoin et certaines de ses fonctions sont exclusivement codées par des gènes du noyau et donc impossibles à identifier dans le plastome. L'étude du génome nucléaire des plantes non-photosynthétiques constitue donc un complément indispensable à l'étude de leur plastome.

## Comparaison de génomes

L'avènement des NGS a considérablement facilité l'accès aux génomes nucléaires. De nombreux génomes nucléaires de plantes sont maintenant disponibles et on trouve notamment les génomes de 2 plantes parasites obligatoires, *Striga asiatica* (Yoshida et al., 2019) et *Cuscuta campestris* (Vogel et al., 2018) et d'une orchidée mycohétérotrophe, *Gastrodia elata* (Yuan et al., 2018). Cependant, malgré l'amélioration des techniques, les génomes de plantes sont souvent de très (trop) grande taille pour être facilement séquencés. C'est notamment le cas des orchidées non chlorophylliennes dont les tailles de génome estimées vont de 3 Gbases pour *Corallorhiza trifida* à 16 Gbases pour *Neottia nidus-avis* (Pellicer and Leitch, 2020). Il est cependant possible d'accéder à la séquence des gènes d'un organisme grâce au séquençage de son transcriptome. Cette approche permet de limiter l'effort de séquençage aux séquences transcrites du génome et notamment aux séquences codant pour des protéines. C'est l'approche qui a été choisie par le projet « 1000 plants » (Leebens-Mack et al., 2019) et pour au moins 14 espèces non chlorophylliennes dont 8 espèces de *Cuscuta* (Yang et al., 2019) et 2 espèces d'*Epipogium* (Schelkunov et al., 2018).

Le professeur Marc-André Selosse du Muséum National d'Histoires Naturelles de Paris et de l'université de Gdansk (Pologne) et ses équipes s'intéressent à la symbiose mycorhizienne chez les Pyrolacées et les Orchidacées notamment à la mixotrophie et à la mycohétérotrophie. Les orchidées constituent un cadre phylogénétique intéressant pour l'étude de l'apparition et de la physiologie de ces types trophiques particuliers. De plus, le cycle de vie de toutes les orchidées inclut une phase mycohétérotrophe. En effet, les orchidées produisent des graines de toute petite taille, dépourvues de réserves et dont l'embryon n'est qu'au stade globulaire de son développement (Rasmussen, 1995). Pour germer, les orchidées sont obligées de former une relation mycohétérotrophique avec des champignons mycorhiziens. La plupart d'entre elles se développent pour devenir autotrophes au stade adulte mais d'autres restent mycohétérotrophes toute leur vie durant.

Dans le cadre de plusieurs projets, nous avons appliqué les techniques d'analyse -omiques à cette thématique. Nous avons ainsi analysé le transcriptome de *Neottia nidus-avis* et d'*Epipogium aphyllum*, deux orchidées de la flore eurasiatique (Figure 7, Jąkowski et al., 2021).

La limitation majeure d'une analyse de transcriptome pour étudier les gènes d'un organisme est que cette approche ne permet d'accéder qu'aux gènes exprimés dans le tissu analysé. Comme tous les gènes ne sont pas actifs dans un tissu ou une condition, il est nécessaire d'analyser plusieurs tissus et/ou conditions pour espérer être le plus exhaustif possible. Dans ce projet nous avons prélevé des échantillons de mycorhize, de tige et de fleur de 2 plantes pour chaque espèce. Les ARN messagers ont été séquencés sur la plateforme transcriptomique de l'IP2S avec la technologie Illumina.



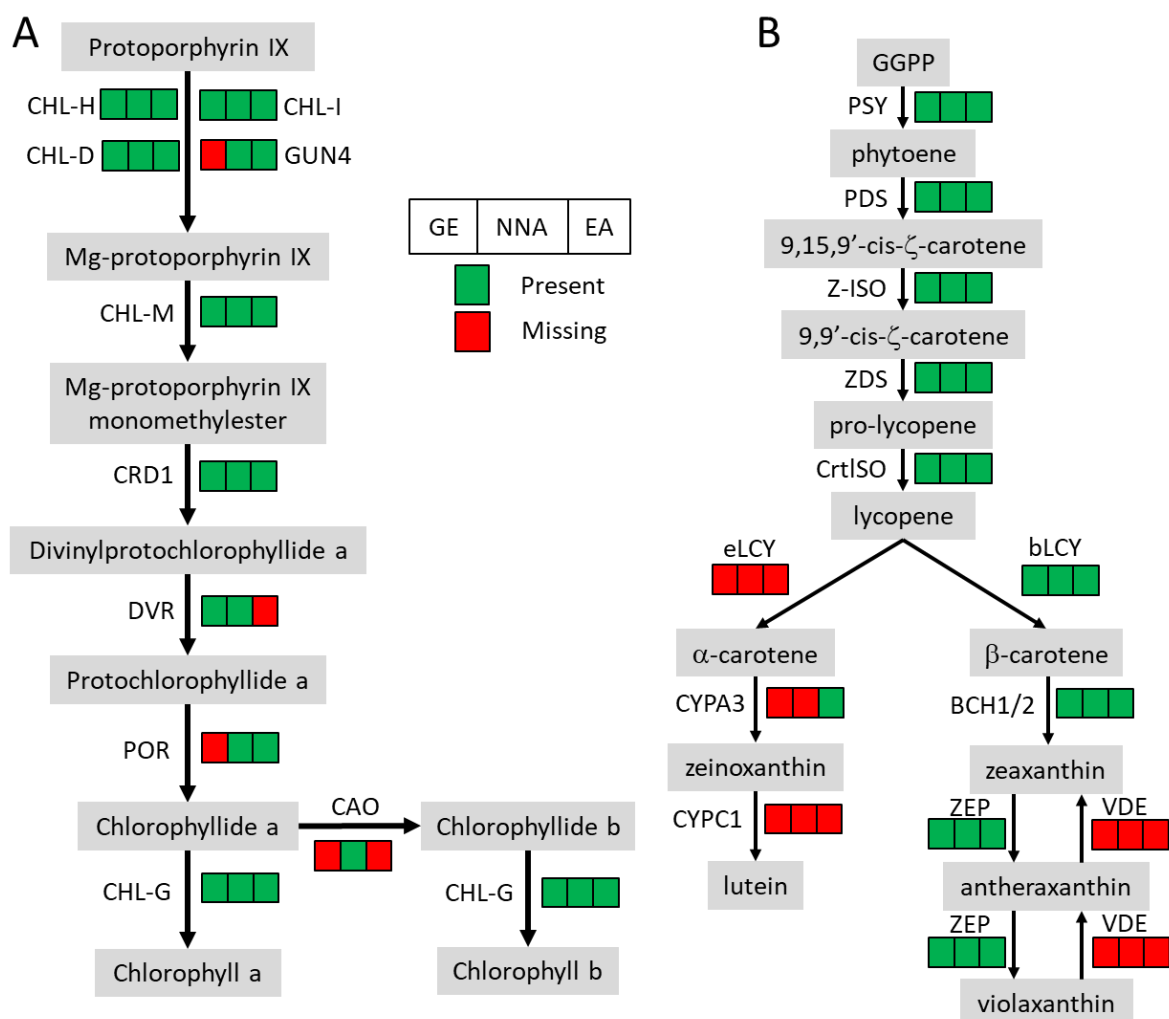
**Figure 7: Inflorescences d'*Epipogium aphyllum* (gauche) et *Neottia nidus-avis* (droite)**

Après un gros travail d'assemblage des séquences de chaque tissu en contig, d'élimination des contigs correspondant à des contaminations (les échantillons ont été prélevés dans la nature et sont donc contaminés par d'autres microorganismes surtout, par définition, les mycorhizes), de fusion des différents assemblages et enfin d'annotation des contigs, nous avons obtenu un jeu de transcrits par espèce que nous avons pu comparer aux contenus géniques d'autres espèces d'orchidées. Pour cela nous avons recherché des gènes correspondant à plus de 8000 orthologues répartis dans plus de 1300 groupes fonctionnels. En comparaison avec le génome de *G. elata*, aucun groupe fonctionnel n'a mis en évidence un manque d'orthologues parmi nos 2 jeux de gènes même en relâchant totalement la stringence des tests (pas de contrôle des faux-positifs) ce qui suggère fortement que nos jeux sont complets et que les 3 espèces d'orchidées mycohétérotrophes ont un contenu en gènes très similaire. Par contre, en comparant ces 3 espèces aux génomes disponibles de 3 espèces d'orchidées autotrophes (*Phalaenopsis equestris*, *Dendrobium catenatum*, *Apostasia shenzhenica*), nous avons mis en évidence des pertes d'orthologues significatives dans plusieurs groupes fonctionnels qui sont tous associés à la photosynthèse.

Même en allant dans le détail des voies métaboliques, les pertes semblent ciblées sur les fonctions strictement associées à la photosynthèse. Par exemple, la voie de biosynthèse des caroténoïdes mène notamment à la lutéine et aux xanthophylles. La lutéine est nécessaire à la photoprotection

des antennes des photosystèmes tout comme les xanthophylles à travers le cycle des xanthophylles. Mais les xanthophylles sont aussi des précurseurs de l'acide abscissique. Les orchidées mycohétérotrophes ont une voie de biosynthèse des caroténoïdes complète jusqu'à la violaxanthine mais qui a perdu la branche menant à la lutéine (figure 8B). De même, la violaxanthine de-epoxydase, nécessaire pour boucler le cycle des xanthophylles, a été perdue.

Comme pour l'analyse des plastomes, on peut déduire de l'analyse du contenu nucléaire en gènes que des fonctions classiquement associées à la photosynthèse mais conservées chez les espèces mycohétérotrophes ont très probablement d'autres fonctions. Ainsi, la voie de synthèse des chlorophylles est quasiment complète chez les orchidées mycohétérotrophes (figure 8A). Cela correspond à ce qui a été observé chez les plantes parasites (Wickett et al., 2011). Cette voie est même complète chez *N. nidus-avis* ce qui est congruent avec des études biochimiques montrant l'accumulation des chlorophylles a et b (ainsi que de la violaxanthine mais pas de lutéine)(Pfeifhofer, 1989). La protoporphyrine IX est le précurseur des hèmes ce qui justifie le maintien de sa synthèse mais les molécules suivantes ne sont actuellement impliquées que dans la synthèse des chlorophylles et sont considérées comme photo-toxiques : leur exposition à la lumière produit des espèces activées de l'oxygène (ROS).



**Figure 8: Altérations des voies de synthèse de pigments chez les orchidées mycohétérotrophes.** GE: *G. elata*. NNA: *N. nidus-avis*. EA: *E. aphyllum*. **A: Synthèse des chlorophylles.** CHL-D, CHL-H, CHL-I, GUN4: *magnesium chelatase*. CHL-M: *Mg-protoporphyrin IX O-méthyltransferase*. CRD1: *Mg-*

*protoporphyrin IX monomethylester cyclase*. *DVR*: divinyl chlorophyllide-*a* 8-vinyl-reductase. *POR*: protochlorophyllide oxidoreductase. *CAO*: chlorophyllide *a* oxygenase. *CHL-G*: chlorophyll synthase. **B: Synthèse des caroténoïdes**. *PSY*: Phytoene synthase. *PDS*: phytoene desaturase. *Z-ISO*:  $\zeta$ -carotene isomerase. *ZDS*:  $\zeta$ -carotene desaturase. *CrtISO*: carotenoid isomerase. *eLCY*: lycopene  $\epsilon$ -cyclase. *bLCY*: lycopene  $\beta$ -cyclase. *BCH1/2*:  $\beta$ -ring carotene hydroxylase. *ZEP*: zeaxanthin epoxidase. *VDE*: violaxanthin de-epoxidase. *CYPA3*:  $\alpha$ -carotene  $\beta$ -ring hydroxylase. *CYPC1*: carotenoid  $\epsilon$ -hydroxylase.

De plus, chez *N. nidus-avis*, le maintien d'une production de chlorophylles a et b s'accompagne du maintien de protéines qui, chez *A. thaliana*, sont impliquées dans la fixation des chlorophylles en réponse à un stress lumineux. Cette fixation doit probablement être destinée à empêcher la chlorophylle de capter la lumière car *N. nidus-avis* n'est pas verte sauf si on chauffe ses tissus. Il est donc étonnant que les activités enzymatiques nécessaires à leur synthèse soient maintenues dans des plantes non-photosynthétiques. La chlorophylle a aussi été détectée chez d'autres orchidées mycohétérotrophes (Barrett et al., 2014) et les auteurs ont proposé que cela pourrait permettre une activité photosynthétique minimale et localisée fournissant un supplément de molécules carbonées pour la production de graines. Cette hypothèse est corroborée par le fait que les produits de la photosynthèse de l'orchidée mixotrophique *Cephalanthera damasomium* sont destinés aux fruits et aux graines (Lallemant et al., 2019b). Elle est aussi renforcée par Menke et Schmid (1976) qui ont observé une photophosphorylation cyclique dans la fleur de *N. nidus-avis*. Cependant cette observation est incompatible avec l'absence de la plupart des gènes nucléaires et plastidiaux codant pour les sous-unités du photosystème I et du cytochrome b6/f.

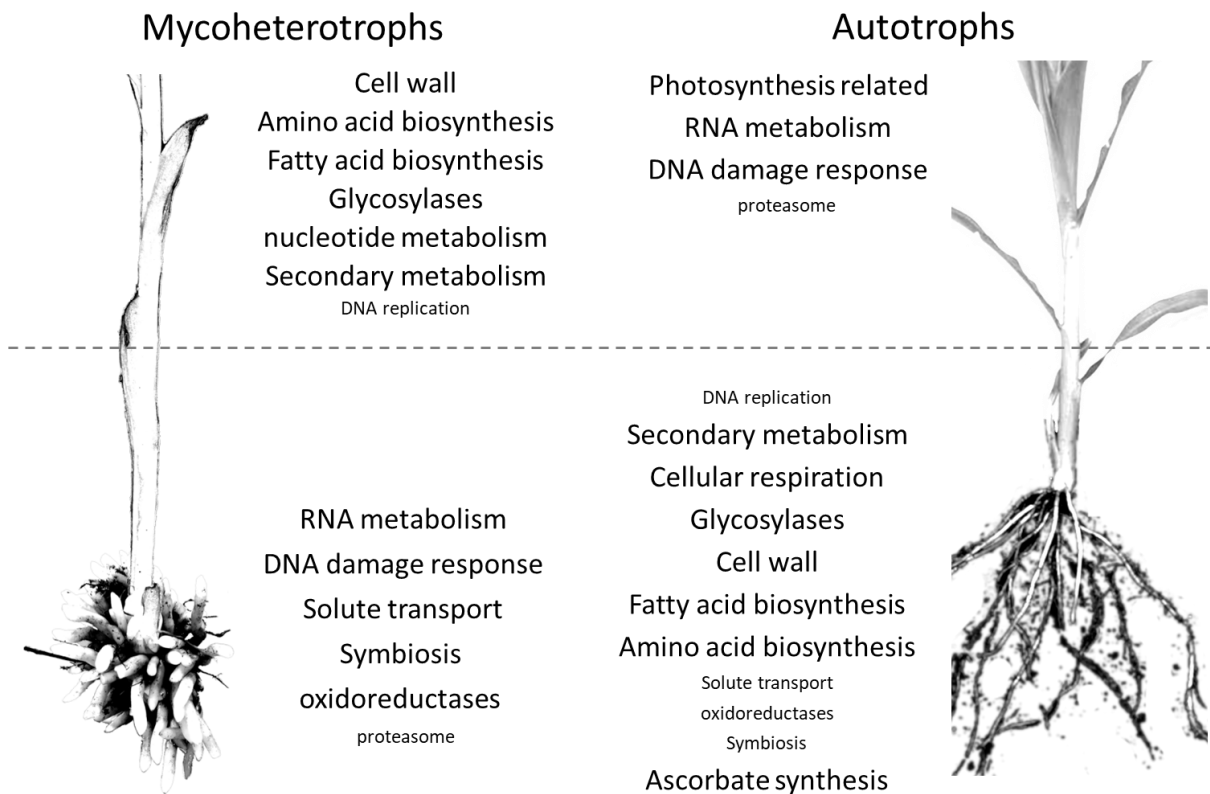
Au contraire, nous n'avons pas détecté de nouveau gène qui pourrait être associé à la mycohétérotrophie.

L'analyse du contenu en gènes des espèces mycohétérotrophes permet d'identifier quelles sont les fonctions non-photosynthétiques potentielles du chloroplaste à travers les pertes et les conservations de gènes. Cependant l'impact de la mycohétérotrophie ne peut pas se limiter à des pertes de gènes et l'analyse du contenu en gènes ne peut pas détecter des néofonctionnalisations ou des changements de profils d'expression contrairement à une analyse quantitative du transcriptome.

## Comparaison de transcriptomes

Avec le même jeu de données que précédemment, l'analyse d'orthologie effectuée mais aussi les outils statistiques développés ces dernières années sur la plateforme transcriptomique, nous avons identifié de nombreuses voies métaboliques différentiellement activées entre les différents organes de nos 2 espèces mycohétérotrophes (Jąkowski et al., 2021). Ces profils sont très similaires entre les 2 espèces mais c'est la comparaison avec des espèces autotrophes qui permet de mettre en évidence les conséquences de la mycohétérotrophie. Faute de jeu de données publiques équivalent au notre pour des orchidées autotrophes (organes similaires et présence de répliques), nous avons utilisé des données de maïs et de *Brachypodium distachyon*. Sur les 8600 orthologues détectés dans les 4 espèces (donc n'incluant pas ceux qui sont directement liés à la photosynthèse), près de 3400 présentaient un ratio racine/tige significativement différent entre les 2 espèces mycohétérotrophes et les 2 espèces autotrophes. Ceci démontre un impact de la mycohétérotrophie dépassant très largement la photosynthèse. Parmi les voies métaboliques impactées (figure 9), la plupart présentent un ratio racine/tige inversé. On peut notamment remarquer que la synthèse des acides aminés et

des lipides présente un maximum d'activité dans les organes puits.



**Figure 9 : Les voies métaboliques impactées par la mycohétérotrophie.** Les voies métaboliques enrichies dans les orthologues présentant un ratio racine/tige significativement différent entre les espèces mycohétérotrophes et autotrophes sont indiquées au niveau de l'organe où leur expression est la plus forte. Les voies indiquées avec une grande police sont également différenciellement exprimées entre racine et tige.

On sait que les mycorhizes sont une zone de transfert d'acides aminés et de lipides de la plante vers le champignon (Fochi et al., 2017; Ghirardo et al., 2020; Rasmussen, 1995). Chez les mycohétérotrophes, ce transfert pourrait se faire dans le sens inverse expliquant un besoin localement moindre pour ces nutriments.

Cependant, avec cette approche, il est difficile de déterminer précisément la cause de ces changements : est-ce la différence de trophie, la différence entre orchidées et graminées et/ou la différence de conditions de culture ?

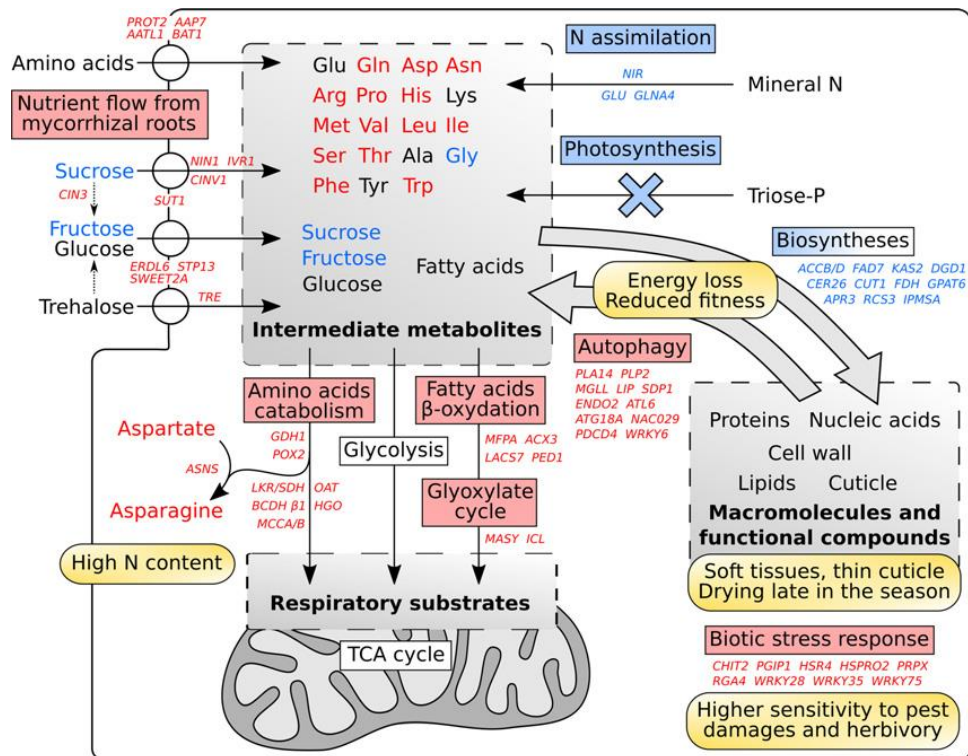
De plus, la perte de photosynthèse doit clairement avoir un impact majeur sur le métabolisme primaire. Même si l'analyse de transcriptome peut donner une idée de ces modifications, une analyse directe du métabolome est nécessaire.

Dans un projet précédent, toujours avec l'équipe de Marc-André Selosse et toujours chez les orchidées, nous avons tiré profit d'un phénomène biologique assez rare : chez certaines orchidées, on observe parfois l'apparition spontanée de plantes totalement albinos, non chlorophylliennes au milieu d'individus normalement verts (figure 10).



**Figure 10 : Plantes de *Cephalanthera damasonium* normale et albinos**

Les individus albinos ont une croissance et une floraison similaires aux individus verts mais ils sont plus sujets aux attaques d'agresseurs et produisent moins de graines (Julou et al., 2005; Roy et al., 2013). Les espèces chez lesquelles on peut observer ce phénomène rare sont donc capables de se passer en très grande partie de la photosynthèse. Des analyses isotopiques ont montré que le carbone et l'azote de ces espèces provient à la fois d'une nutrition minérale et de leur champignon mycorhizien (Selosse and Roy, 2009). Elles sont dites mixotrophes puisqu'à la fois autotrophes et mycohétérotrophes. Les espèces mixotrophes sont considérées comme une étape évolutive vers la mycohétérotrophie. La comparaison des individus chlorophylliens et non chlorophylliens au sein d'une même espèce permet donc d'observer les conséquences de la perte de photosynthèse dans des conditions naturelles et dans un contexte génétique stable. Une autre question découle du fait que les espèces mixotrophes sont capables de maintenir naturellement des individus albinos, contrairement aux espèces autotrophes dont les individus albinos ont besoin de conditions de cultures très particulières. Cette capacité est très certainement liée à leur association à des champignons mycorhiziens différents de ceux des autotrophes mais pourrait aussi nécessiter des modifications de leur physiologie, prélude à la transition vers la mycohétérotrophie. Pour répondre à ces questions, nous avons donc comparé le transcriptome et le métabolome de feuilles d'individus verts et albinos chez 3 espèces d'orchidées mixotrophes (Lallemand et al., 2019c). L'analyse en parallèle de ces 2 types de données -omiques ont montré que les plantes albinos présentent un métabolisme similaire chez les 3 espèces (figure 11).



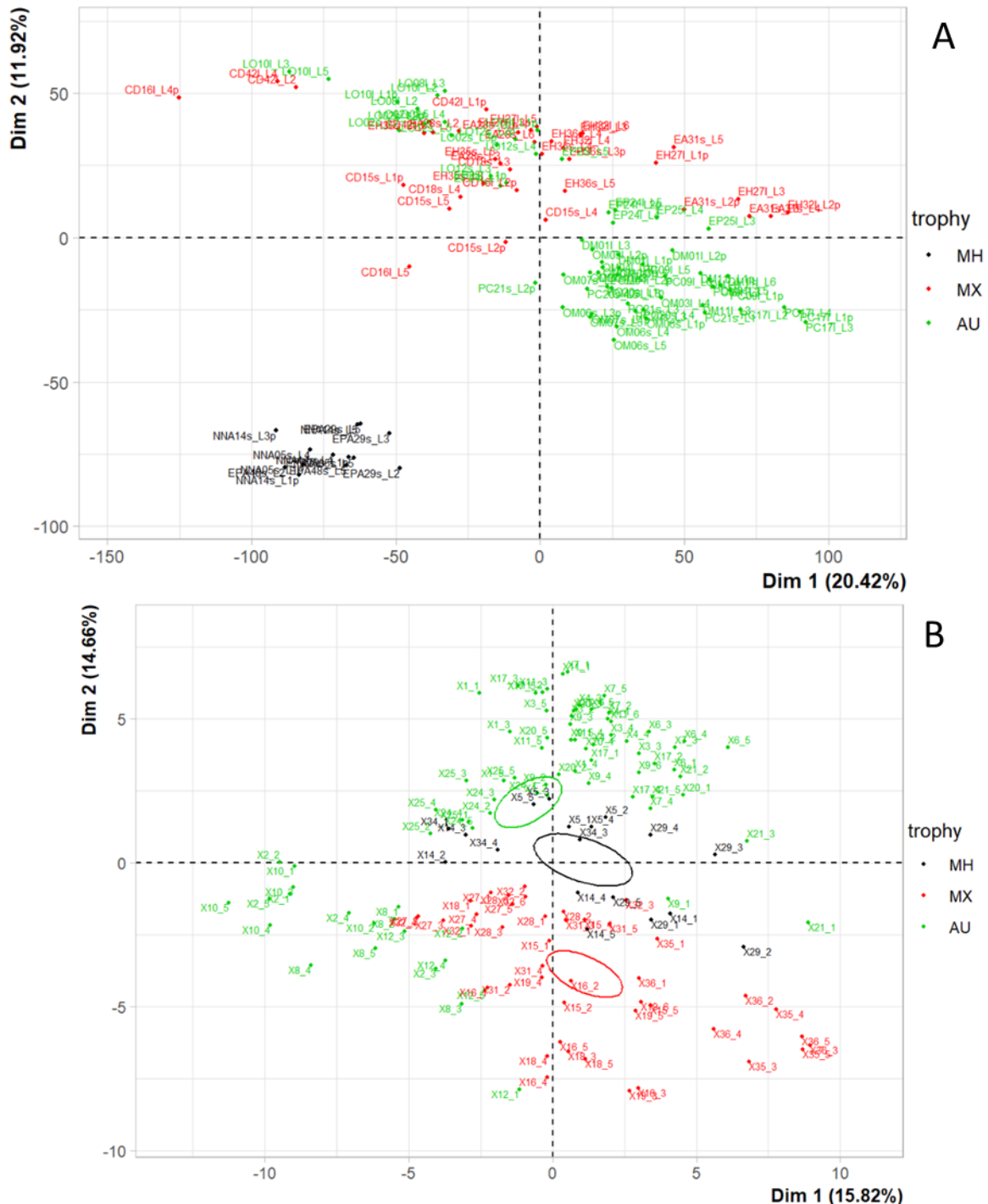
**Figure 11 : Altérations du métabolisme déclenchées par l'abaissement de l'azote.** Le rouge (le bleu) indique une augmentation (une diminution) de l'abondance du métabolite ou du transcrite. Les boîtes de texte indiquent les principales voies métaboliques induites (rouge), réprimées (bleu) ou inchangées (blanc). Les boîtes arrondies jaunes indiquent les phénotypes observés associés à ces altérations.

La perte de photosynthèse entraîne une carence en carbone associée à une suraccumulation d'azote. En l'absence de sucres, le métabolisme énergétique dégrade les acides aminés et les acides gras provenant de la dégradation des macromolécules et probablement des mycorhizes.

Ce profil moléculaire est très similaire à celui observé chez des mutants albinos d'*Arabidopsis* ou dans des zones blanches de tissus panachés de *Pelargonium*. Malgré un contexte expérimental ou phylogénétique très différent, cette similarité suggère que la capacité des orchidées mixotrophes à maintenir spontanément des individus totalement non chlorophylliens n'est pas associée à des innovations métaboliques. Ainsi, plutôt que sur des innovations, l'évolution vers la mycohétérotrophie chez les orchidées se base sur la plasticité du métabolisme végétal et la capacité à exploiter les ressources organiques du champignon mycorhizien pour remplacer les photosynthétats.

Comme l'étude précédente sur les orchidées mycohétérotrophes, cette étude souffre de biais et de manques dans le design expérimental : l'interprétation de la comparaison avec des espèces autotrophes est compliquée par des conditions expérimentales très différentes et il semble que l'organe essentiel pour comprendre les évolutions trophiques soit la mycorhize qui est absente ici. Mais ces projets ont été des préalables incontournables pour une étude de bien plus grande ampleur que je coordonne avec Marc-André Selosse. Le projet Orchidomics vise à étudier les différences de physiologie moléculaires entre des orchidées européennes de types trophiques différents : autotrophes, mixotrophes et mycohétérotrophes. Cela revient à étudier les conséquences d'une dépendance plus ou moins grande à la photosynthèse. Pour ce faire, près de 300 échantillons ont été analysés en transcriptomique et métabolomique. Ces échantillons correspondent aux mycorhizes et

aux feuilles de 10 espèces différentes collectées dans 30 populations naturelles en Pologne. Les données métabolomiques et transcriptomiques ont été générées et l'analyse est en cours. Une première exploration des données se basant sur les ratios « parties aériennes/parties souterraines » montre une distinction entre les espèces autotrophes, mixotrophes et mycohétérotrophes à la fois dans le transcriptome et le métabolome (figure 12). Cela indique que ces données omiques devraient nous permettre d'élucider certaines des différences liées à ces différents types trophiques.



**Figure 12: Analyse en composante principale des ratios « partie aérienne/partie souterraine » du projet Orchidomics. A. Données transcriptomiques B. données métabolomiques. Echantillons de plantes mycohétérotrophes (MH), mixotrophes (MX) et autotrophes (AU).**

On observe cependant plusieurs résultats surprenants. D'abord dans les données métabolomiques, les échantillons mycohétérotrophes se placent comme intermédiaires entre les échantillons mixotrophes et autotrophes alors qu'on aurait pu s'attendre à trouver les échantillons mixotrophes en position intermédiaire. Ensuite, une espèce autotrophe, *Neottia ovata*, se groupe avec les échantillons mixotrophes que ce soit en métabolome ou en transcriptome. Il y a donc de nombreuses pistes à explorer dans ces données mais le plus laborieux commence, à savoir se plonger dans les détails, trouver les bonnes méthodes d'analyses bioinformatiques et statistiques et construire des interprétations biologiques.

Le montage de ce projet et l'interprétation de ce grand jeu de données est impossible sans des compétences en botanique, écologie, mycologie, physiologie végétale, biologie moléculaire, chimie, génomique, bio-informatique et statistiques. Cependant, leur juxtaposition n'est pas suffisante pour que chacune de ces compétences bénéficie au (et du) projet autrement dit pour que le projet soit réellement interdisciplinaire. Il faut être capable de formuler des questions qui soient déclinables dans chacun des domaines de compétences, de formaliser des questions que chaque expert puisse traduire dans son vocabulaire. Pour le projet Orchidomics, les 2 projets précédents avec les équipes de Marc-André Selosse nous ont donné le temps pour apprendre à travailler ensemble, pour apprendre un vocabulaire commun et pour comprendre les attentes et contraintes des uns et des autres.

Les approches évolutives offrent des pistes et des informations très intéressantes pour comprendre le rôle et le fonctionnement du chloroplaste car elles permettent d'observer son fonctionnement dans tous les contextes retenus par l'évolution. C'est particulièrement vrai avec les récentes avancées technologiques dans le séquençage et autres -omiques. Cependant, elles peuvent être très frustrantes car les preuves scientifiques apportées sont quasi-exclusivement circonstancielles : on constate la conservation ou la perte de tel gène ou fonction dans tel groupe d'espèces, tel changement d'expression. A partir de là, on essaie de trouver une explication, un lien de causalité. Or ce type d'étude souffre très souvent de ce que les statisticiens appellent une confusion d'effet. Est-ce que les différences de contenu génique observées entre les 6 espèces d'orchidées sont toutes dues à la perte de photosynthèse ou au caractère épiphyte *versus* terrestre ? Les différences de profils d'expression sont-elles associées au fonctionnement du plaste et/ou à la comparaison d'orchidées et de graminées et/ou à des conditions de croissance différentes ? De plus, l'interprétation des données de génomique ou de transcriptomique requiert la connaissance préalable au moins partielle de la fonction des gènes. Or elle provient presque exclusivement de l'homologie avec des gènes dont la fonction a été élucidée dans d'autres plantes pour lesquelles des outils de génie génétique existent et où les approches de génomique fonctionnelle sont possibles. Ces plantes modèles permettent des expériences interventionnelles dans lesquelles les paramètres génétiques et environnementaux sont contrôlés autant que faire se peut. Il est ainsi possible d'intervenir sur un seul paramètre de l'expérience en maintenant les autres quasiment constants ce qui permet de limiter les confusions d'effet, de se rapprocher du lien de causalité et éventuellement de déterminer la fonction d'un gène. J'ouvre ici une parenthèse : ces expériences en laboratoire sont très puissantes mais elles peuvent être considérées comme très artificielles et éloignées de la réalité du terrain comme le démontre la difficulté à transposer des résultats du laboratoire au champ. Entre les approches évolutives et ces approches « interventionnelles », il y a donc un espace d'interdisciplinarité très intéressant qui est probablement une autre forme de biologie translationnelle.

# Approches de génomique fonctionnelle : étude des plantes modèles

Comprendre le fonctionnement du chloroplaste nécessite de comprendre le fonctionnement de son génome, quelles sont les fonctions de ses gènes et comment leur activité est régulée.

## Régulation de l'expression génétique du chloroplaste

Le plaste possède son propre génome comprenant quelques dizaines de gènes codant pour des protéines de son expression génétique mais aussi de la photosynthèse. Cependant, de nombreuses protéines codées dans le noyau sont importées dans le plaste pour former des complexes avec les protéines du plaste (figure 2). Les expressions génétiques du plaste et du noyau doivent donc être étroitement coordonnées. Du fait de son origine procaryote, le système d'expression du plaste se rapproche de celui des bactéries (figure13).

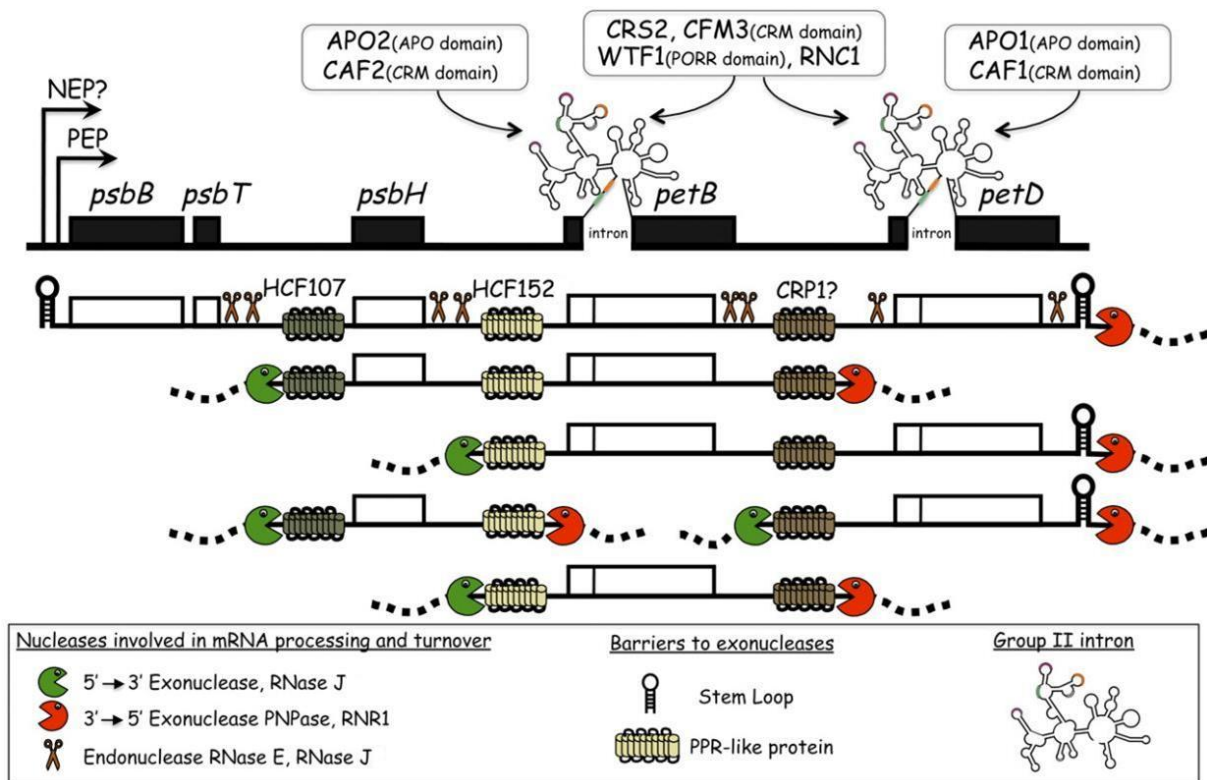


Figure 13 : Expression du polycistron psbB. D'après Barkan, 2011.

En effet, les transcrits du plaste peuvent former des polycistrons, un seul promoteur initiant la transcription de plusieurs gènes sur une seule molécule ce qui est typique des systèmes bactériens et les introns sont de type II comme les introns de bactéries. Ces ARN sont soumis à de nombreux processus moléculaires post-transcriptionnels dont des coupures endonucléolytiques, des digestions exonucléolytiques ainsi que l'édition de certaines bases spécifiques de C vers U qui s'ajoute à la modification des ARNt. Le transcriptome du chloroplaste est donc constitué d'un ensemble particulièrement complexe d'isoformes et d'intermédiaires (figure 13). De plus, les transcrits du plaste ne sont pas poly-adenylés ou plutôt la poly-adenylation marque les molécules destinées à la dégradation alors que dans le noyau la poly-adenylation caractérise les ARN messagers. Or la quasi-totalité des outils d'analyse se focalisent sur ces ARN messagers à travers leur poly-adenylation et ne

sont donc pas adaptés pour l'étude du transcriptome du plaste. En suivant l'évolution des technologies j'ai développé plusieurs protocoles d'analyse d'abord à base de PCR quantitative haut-débit pour l'analyse de l'expression des gènes et leur épissage (Delannoy et al., 2015) puis à base de RNA-seq pour l'analyse exhaustive du transcriptome (figure 14, Malbert et al., 2018). Ce protocole inclut la partie biologie moléculaire mais aussi et surtout les étapes bio-informatiques et statistiques qui ont dû être adaptées aux spécificités du transcriptome des organites. Un pipeline a été mis à disposition de la communauté pour faciliter les analyses (Baudry et al. 2021, <https://forgemia.inra.fr/GNet/pipelineoge>).

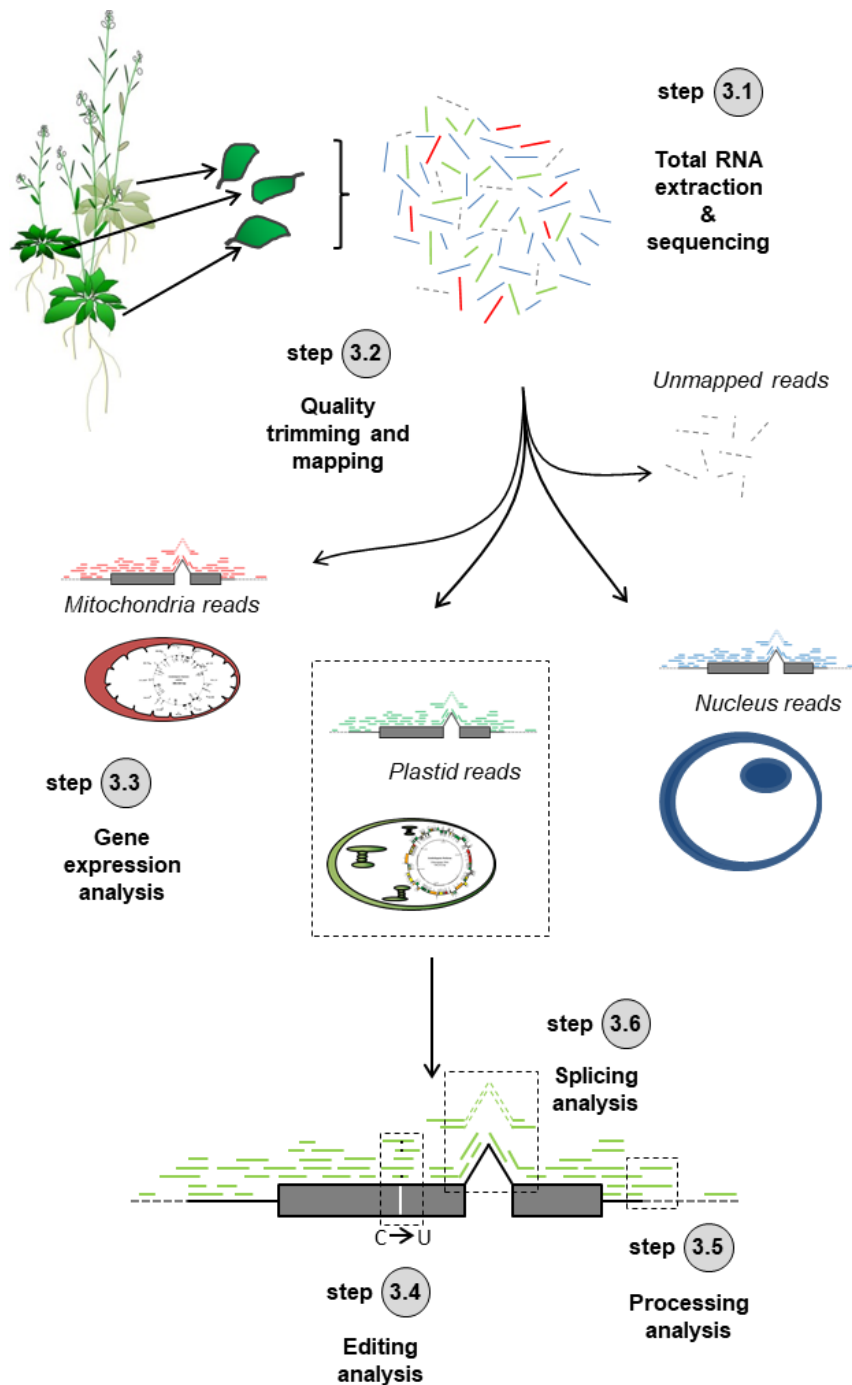
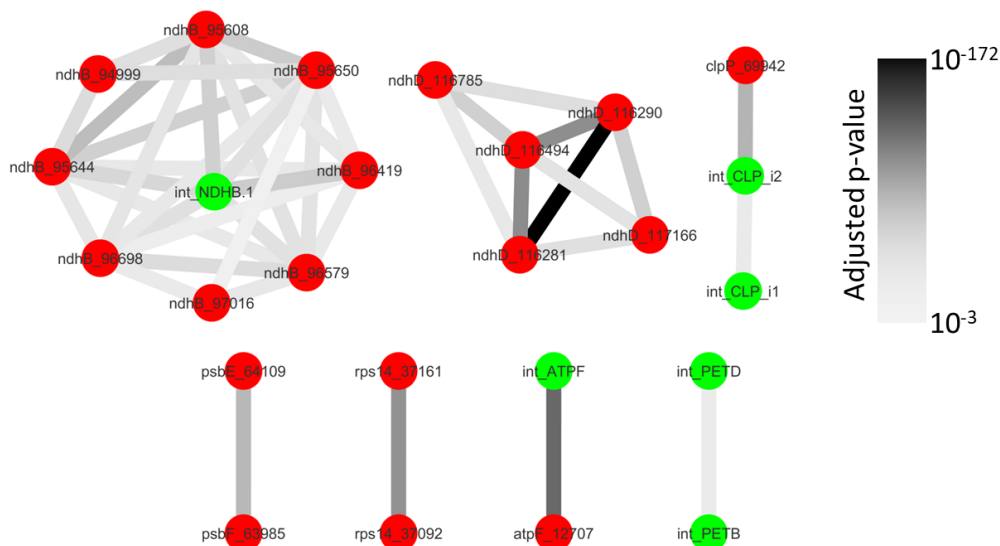


Figure 14 : Etapes du protocole d'analyse du transcriptome des organites par RNA-seq

Il couvre tous les aspects du transcriptome du plaste sauf le processing, c'est-à-dire la digestion des extrémités des transcrits. Ce processus moléculaire modifie les extrémités des transcrits produisant une population dynamique d'isoformes qui ne correspondent souvent pas aux annotations existantes. En conséquence, l'étude de ce processus se fait souvent manuellement et sans support statistique (Castandet et al., 2013). En collaboration avec Guillem Rigail de l'équipe Genomic Networks et Benoit Castandet de mon équipe, je développe ce type analyse grâce à une optimisation des méthodes de segmentation. Ces méthodes identifient les points de rupture dans un signal le long du génome ce qui permet de s'affranchir de l'annotation. Cependant, elles ne prennent pas encore en compte les répliques biologiques ou la comparaison entre conditions. Arnaud Liehrmann, doctorant encadré par Guillem et Benoit, a très récemment proposé un package R, diffSegR, permettant de prendre en compte ces 2 aspects et les premiers résultats sont spectaculaires.

Une autre limitation l'analyse du transcriptome chloroplastique est liée à la technologie de séquençage, actuellement Illumina, qui ne donne accès qu'à des fragments de transcrits et rend très difficile voire impossible la reconstruction des multiples isoformes mentionnées précédemment. Le séquençage dit de troisième génération comme celui développé par la société Oxford Nanopore Technologies permet un séquençage de transcrit pleine longueur et ainsi d'accéder directement au transcriptome complet. Cependant les protocoles existants sont exclusivement dédiés aux ARN poly- adénylés. J'ai modifié ces protocoles pour étudier les ARN des organites et les premières données montrent l'intérêt de cette approche pour l'étude du transcriptome du chloroplaste (Guilcher et al., 2021). Grâce à un transcriptome « pleine longueur » nous avons pu étudier les liens de dépendances entre les différents mécanismes de maturation des transcrits dans le chloroplaste. Les outils bioinformatiques et statistiques que nous utilisons devront être améliorés pour en tirer pleinement avantage mais nous mettons d'ores et déjà en évidence des dépendances entre édition et épissage (figure 15). La première constatation est que nous n'observons pas d'antagonisme mais uniquement des dépendances positives.



**Figure 15: Liens de dépendance entre édition et épissage dans le chloroplaste d'*A. thaliana*.** Les événements d'épissage sont en vert et ceux d'édition en rouge

Dans le détail, on observe que l'édition du site ndhD\_117166, nécessaire pour l'apparition du bon codon start, est fortement activée par l'édition des 4 autres sites d'édition de ce transcrit qui semble initiée d'abord par l'édition du site ndhD\_116785 puis celle du site ndhD\_116494. De cette manière,

il semble que la traduction ne puisse pas démarrer avant que tous les sites ne soient édités. On observe également que l'édition du site *clpP\_69942* favorise l'épissage de l'intron 2 de *clpP* qui lui-même favorise l'épissage de l'intron 1. De même on retrouve la dépendance entre l'édition d'*atpF\_12707* avec l'épissage de l'intron qui a déjà été observée dans le mutant d'édition *aef1* (Yap et al., 2015). Ces liens de dépendance et cette chronologie dans les événements de maturation ne peuvent pas s'expliquer directement par les mécanismes moléculaires connus actuellement.

Avec cette technique nous avons enfin un accès exhaustif au transcriptome des organites, notamment du chloroplaste. Cette vision globale va nous permettre de comprendre les liens complexes entre les différentes étapes de maturation des ARN et aussi de les suivre en réponse aux changements du contexte cellulaire (environnement extérieur et développement). Dans un processus qui est principalement régulé au niveau de la maturation des ARN, c'est indispensable à la compréhension de son fonctionnement.

Avec les outils disponibles, j'ai pu étudier la régulation du transcriptome du plaste et donc sa coordination avec le transcriptome nucléaire. Elle est assurée par un contrôle de l'expression génétique du plaste par le noyau à travers l'envoi de protéines de régulation dans le plaste. La plupart d'entre elles appartiennent à une grande famille de près de 500 membres chez *Arabidopsis thaliana* : les protéines à domaines pentatricopeptides répétés (PPR). Les PPR ont été découvertes il y a plus de 20 ans (Small and Peeters, 2000; Lurin et al., 2004). Elles sont caractérisées par la répétition en tandem de domaines PPR qui peut être accompagnée de domaines supplémentaires E, E+ et DYW à l'extrémité C-terminale (figure 16). Ces motifs définissent plusieurs sous-familles de protéines PPR. Les PPR sont présentes chez tous les eucaryotes mais sont très abondantes chez les plantes vasculaires (Gutmann et al., 2019). Depuis leur découverte, il a été démontré qu'elles sont impliquées dans toutes les étapes de l'expression génétique des organites : transcription, processing, épissage, édition et traduction. Elles agissent en se fixant de manière spécifique à une séquence d'ARN déterminée par certains acides aminés de leurs domaines PPR (figure 16, Delannoy et al., 2007; Barkan et al., 2012). Pour les caractériser fonctionnellement, il faut donc déterminer leur ARN cible et leur action sur cet ARN.

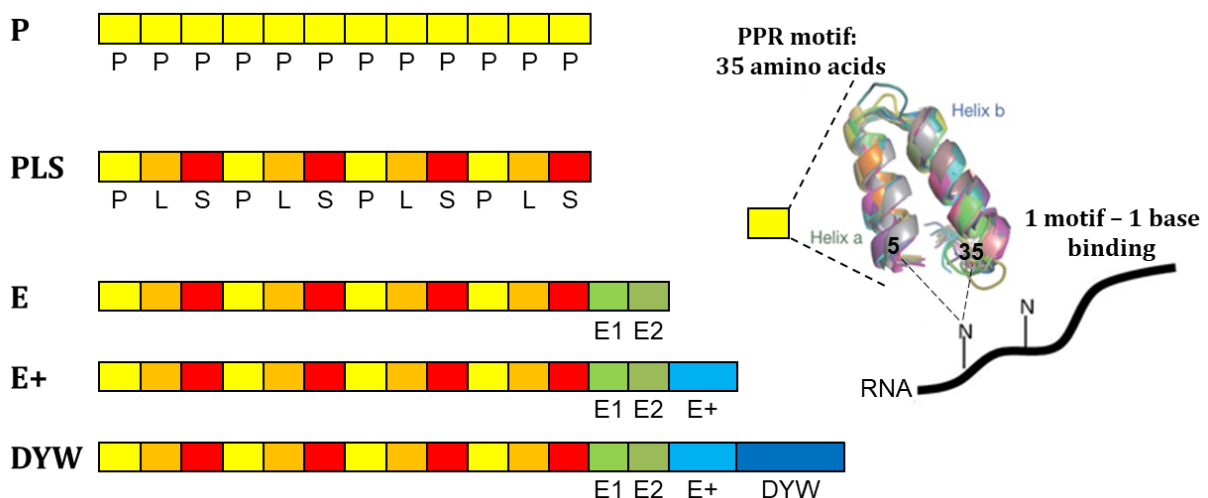
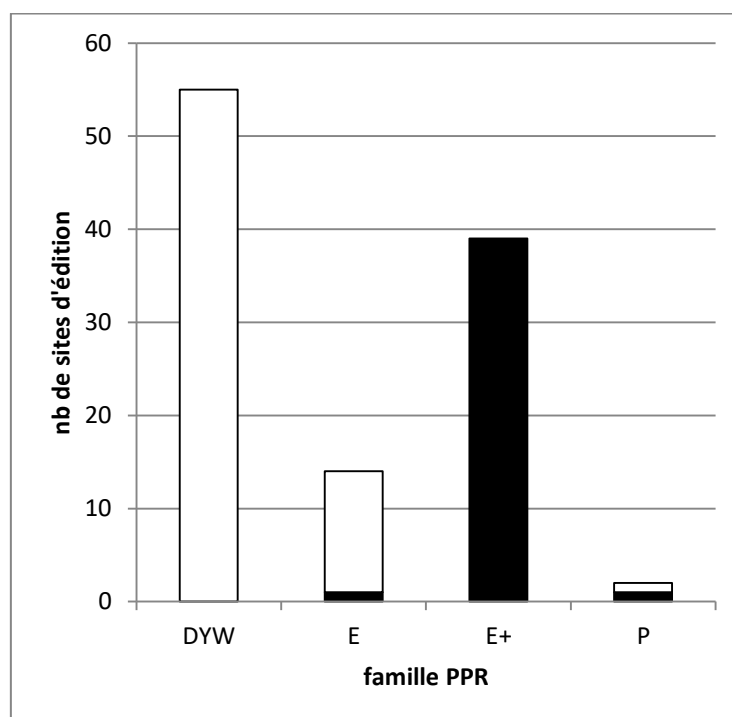


Figure 16 : les protéines à domaines pentatricopeptides répétés

Une partie de mon travail sur les PPR s'est focalisée sur un des domaines qui y sont associés : le domaine DYW. Les PPR à domaine DYW sont impliquées dans l'édition (notamment CLB19 et YS1). Ce domaine présente de fortes similarités avec les cytidines désaminases, sa distribution phylogénétique est identique à celle de l'édition C vers U (Salone et al., 2007), et des mutations ciblées sur les acides aminés correspondant au site actif des cytidines désaminases inhibent l'activité d'édition des PPR (Boussardon et al., 2014). Avec la démonstration récente de l'activité de cytidine désaminase du domaine DYW (Hayes and Santibanez, 2020), il n'y a maintenant plus aucun doute qu'il s'agit bien du domaine responsable de l'activité enzymatique d'édition C vers U. Avec les outils d'analyse du transcriptome, j'ai participé à l'étude de nombreuses PPR. Ainsi OTP51 est nécessaire pour l'épissage de l'intron 2 de *ycf3* (De Longevialle et al., 2008), CLB19 contrôle l'édition de 2 sites dans *rpoA* et *clpP* (Chateigner-Boutin et al., 2008), YS1 est nécessaire à l'édition de *rpoB* (Zhou et al., 2009) et OTP70 à l'épissage de *rpoC1* (Chateigner-Boutin et al., 2011).

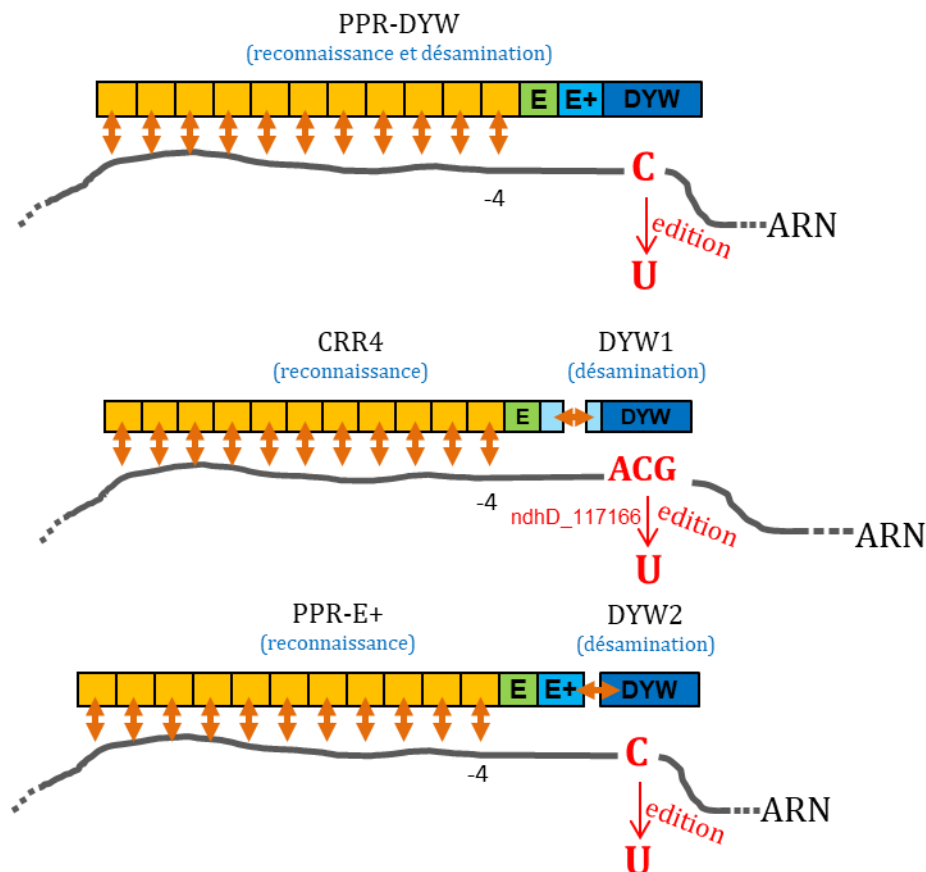


**Figure 17: Les défauts d'édition du mutant *dyw2* définissent 2 classes de sites d'édition.** Le nombre de sites d'édition associés à une PPR d'une famille donnée et dont l'édition est inhibée dans le mutant *dyw2* est indiqué en noir. Le nombre de sites d'édition non

Cependant, de manière intrigante, plusieurs études ont montré que les PPR à domaine DYW peuvent toujours éditer leur cible *in vivo* même si on leur enlève leur domaine DYW (Okuda et al., 2009). Les PPR des sous-familles E et E+ sont également impliquées dans l'édition malgré leur absence de domaine DYW. L'équipe OGE a identifié un petit groupe de protéines PPR atypiques contenant un domaine DYW. L'une d'elle, appelée DYW1, est nécessaire à l'édition du site d'édition *ndhD-117166* mais travaille en complexe avec la PPR CRR4 qui est dépourvue de domaine DYW (Boussardon et al., 2012).

De manière similaire, nous avons identifié la protéine DYW2 en tant qu'interactant physique de la PPR E+ CLB19. Malgré le phénotype embryoléthale du mutant *dyw2*, j'ai analysé son transcriptome par RNA-seq ce qui a permis d'observer que 529 sites d'édition sont affectés que ce soit du côté

mitochondrial ou chloroplastique et que 272 d'entre eux voient leur taux d'édition diminué de 25% ou plus (Guillaumot et al., 2017; Malbert et al., 2020). Plus important, parmi les sites d'édition dont les PPR associées étaient connues, tous ceux associés à des PPR E+ ont un taux d'édition diminué dans le mutant *dyw2* mais aucun de ceux associés aux PPR DYW (figure 17). Ces résultats suggèrent que les PPR de type E+ et DYW2 forment des complexes nécessaires à l'édition de leurs sites. Dans ces complexes, les PPR apportent la reconnaissance du site à éditer par leurs motifs PPR et DYW2 apporte l'activité cytidine désaminase avec son domaine DYW. Ils regroupent ainsi les activités portées par les PPR de la famille DYW (figure 18).



**Figure 18: Trois complexes d'édition regroupant les fonctions de reconnaissance et de désamination.** Les autres membres connus de ces complexes (ORRM1, OZ1, MORFs, RIP, NUWA...) ne sont pas indiqués par soucis de simplicité.

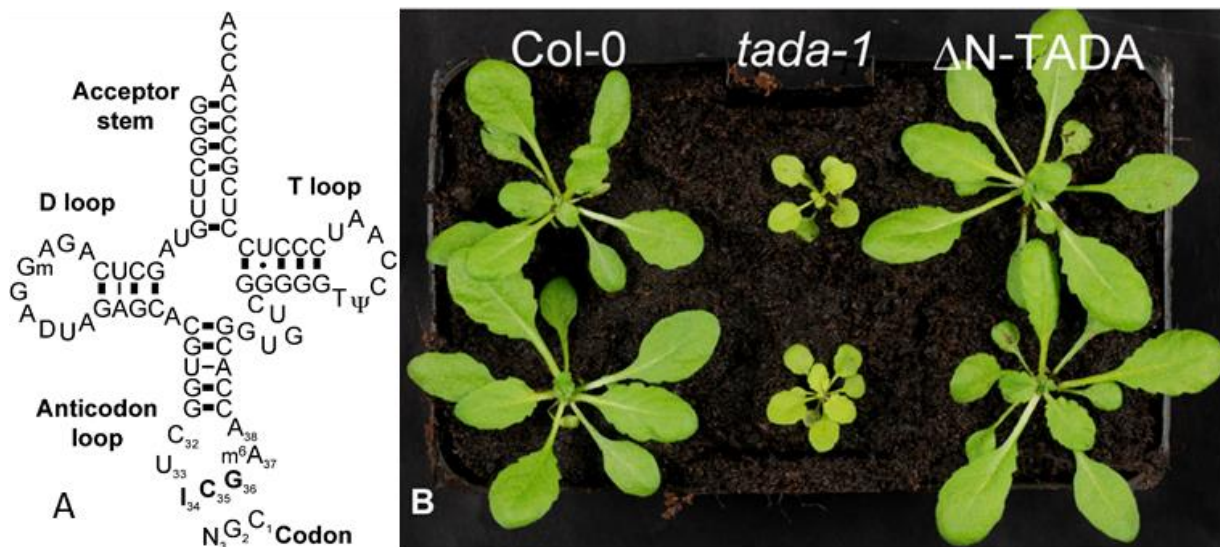
Nous avons renforcé le lien fonctionnel entre les PPR E+ et DYW2 par la caractérisation de 3 PPR d'édition supplémentaires (PGN, OTP90 et MEF37) contrôlant 16 sites d'édition (Malbert et al., 2020) mais ces études ont également largement complexifié notre modèle du fonctionnement de l'édition dans les organites. En effet, le modèle actuel prévoit qu'une protéine PPR vient se fixer sur l'ARN en amont du site à éditer en position -4 et que chaque motif PPR reconnaît une base en amont de cette position -4 (figure 18). Selon ce modèle, cette fixation est nécessaire et suffisante pour éditer le site reconnu. Sur la base de ce modèle et des sites d'édition déjà associés à des PPR, il est possible de déduire un code PPR définissant les préférences de chaque motif PPR pour une base ou une autre en établissant les fréquences d'association (Kobayashi et al., 2019). Selon ce modèle, lorsqu'une PPR d'édition est mutée, l'édition de son site est abolie. Avec l'utilisation du RNA-seq lors de nos études,

nous avons observé chez les mutants de PPR, en plus de quelques sites dont l'édition est abolie, des sites dont l'édition baisse seulement de manière partielle voire augmente.

Les baisses ou hausses partielles d'édition suggèrent que plusieurs PPR peuvent se fixer sur le même site permettant une redondance fonctionnelle (pour les baisses partielles) ou entraînant une compétition (pour les hausses). De plus, la densité des sites d'édition fait que les sites de fixation potentiels des PPR de plus d'un tiers des 795 sites d'édition identifiés contiennent eux-mêmes un site d'édition. Et nous avons identifié des cas où l'édition du site de fixation est nécessaire à l'édition du site correspondant (donc très probablement à la fixation de la PPR associée) ou au contraire l'abolit. Il faudrait donc tenir compte de l'éventuelle édition des sites de fixation dans l'établissement des fréquences d'association entre bases et motifs PPR ; toutefois cela n'est pas possible car on ne connaît pas le statut d'édition de ces sites au moment de la fixation de la PPR. Ensuite pour prédire le site de fixation d'une PPR, il ne faudrait plus se contenter d'estimer son affinité pour une séquence donnée mais également tenir compte des autres protéines se fixant à l'ARN (autres PPR mais également ribosomes, CRM etc...) limitant *in vivo* le nombre de sites de fixation disponibles. Ainsi le modèle actuel de l'édition dans le chloroplaste est simpliste et donc faux. Cependant, il reste utile car il permet toujours de faciliter la caractérisation de PPR d'édition (Kobayashi et al., 2020) et il sert de référence pour évaluer la pertinence des nouvelles connaissances produites dans le domaine.

J'ai également participé à la caractérisation d'autres gènes impliqués dans le contrôle de l'expression du chloroplaste tels que *CP31A* impliqué dans la stabilité de certains transcrits (Tillich et al., 2009), *RPO7mp* qui code une ARN polymérase ciblée au chloroplaste et à la mitochondrie mais dont le rôle semble limité à la transcription dans la mitochondrie (Kühn et al., 2009) ou les *CSP41* impliqués dans la stabilisation de certains ARNs (Qi et al., 2012).

J'ai surtout participé de manière prépondérante à l'identification de la tRNA Arginine Deaminase (TADA, Delannoy et al., 2009). Elle est responsable de l'édition d'un ARNt chloroplastique qui en change l'anti-codon de ACG à ICG (figure 19A). Nous avons confirmé son activité désaminase *in vitro* et montré que la synthèse protéique est fortement diminuée dans le plaste du mutant entraînant un fort retard de croissance (figure 19B). Cependant, en l'absence d'édition de cette base de l'ARNt, la plante aurait dû en théorie être incapable de traduire 2 codons (CGC et CGA) se trouvant notamment dans les gènes *rbcL* ou *atpA*. Par des approches de protéomique nous avons montré que ces codons sont pourtant correctement traduits ce qui implique que ces codons peuvent être reconnus par seulement 2 des trois bases. Une analyse par RT-PCR quantitative a enfin montré que l'inhibition de traduction du plaste dans le mutant entraîne une suraccumulation des transcrits chloroplastiques qui est n'est pas corrélée à l'abondance de leur protéine.



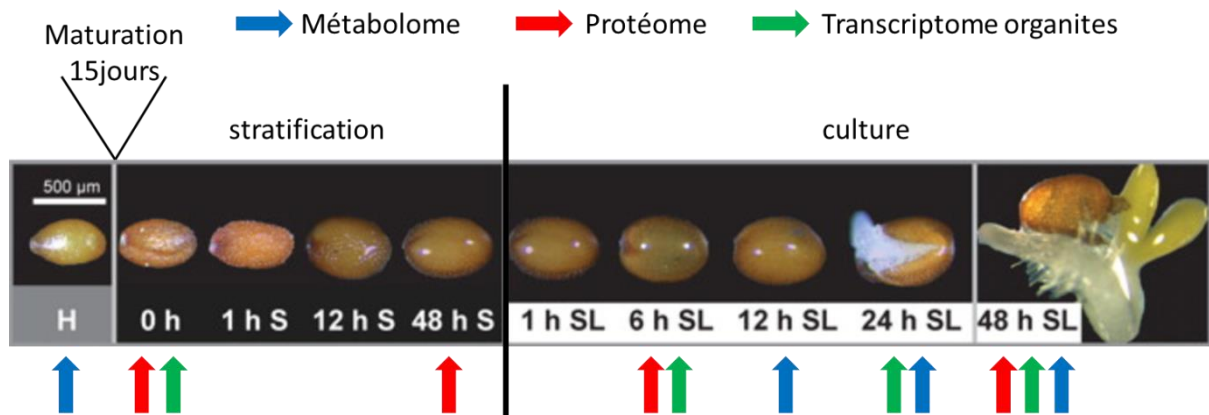
**Figure 19: Rôle de la protéine TADA chez *A. thaliana*.** A : séquence de l'ARNt<sup>Arg</sup>(ICG) avec son anticodon associé à son codon correspondant. I : inosine; N : A, C, U, ou G; m<sup>6</sup>A : N<sup>6</sup>-methyl A; D : dehydro U; Gm : 2-O- methyl G; T : 5-methyl U; ψ : pseudo U. B : phénotype du mutant *tada-1* et de la lignée compléementée ΔN-TADA

J'ai donc participé à la caractérisation de nombreux gènes impliqués dans l'expression génétique du chloroplaste. Cependant, pour être honnête, il est difficile de parler de « régulateurs » en tant que tels. En effet, chacune de ces protéines est nécessaire à un processus moléculaire de l'expression génétique du chloroplaste et leur absence peut entrainer des phénotypes dramatiques ce qui souligne leur importance dans le fonctionnement de la plante. Pour autant, la plupart de ces processus ne semblent pas varier au cours du développement ou en réponse à l'environnement. Par exemple, l'édition de l'immense majorité des sites d'édition ne varie que très peu voire pas du tout. Dans ces conditions, ces protéines semblent plutôt être des piliers, le cadre dans lequel l'expression génétique du chloroplaste se déroule et dans lequel d'autres mécanismes la régulent. Car l'expression génétique du chloroplaste change durant le développement et en réponse au stress.

## Le chloroplaste lors du développement

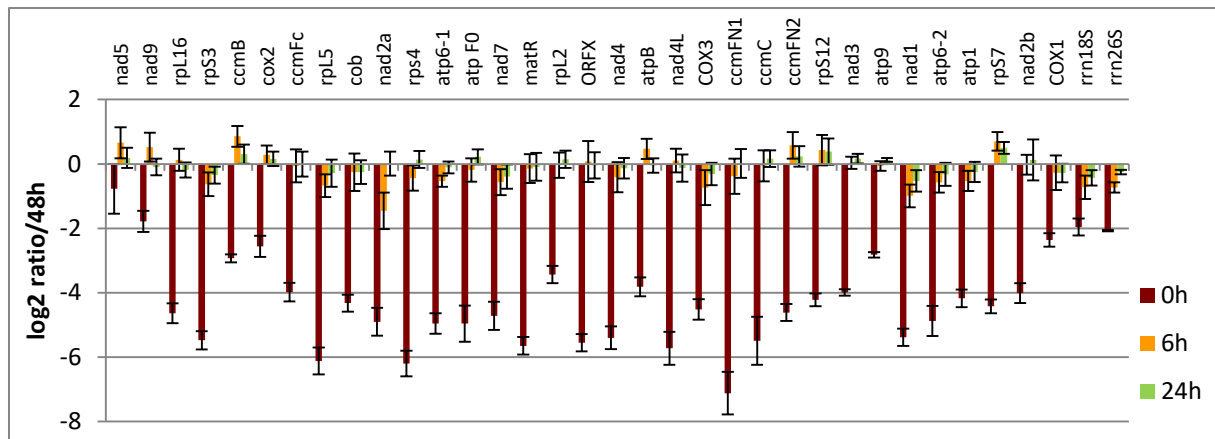
Lors des toutes premières étapes de leur développement, c'est-à-dire lors de la germination, les plantes doivent réussir plusieurs transitions. La première est la mise en route de leur métabolisme qui a été mis en pause lors de la maturation et la dessiccation de la graine. Cette mise en route permet la croissance et l'émergence de la radicule et de l'hypocotyle. La seconde est la transition de l'hétérotrophie vers l'autotrophie. Autrement dit, les plantes doivent passer d'un métabolisme énergétique basé sur les réserves de la graine à un métabolisme énergétique basé sur la photosynthèse. Cette transition majeure doit se faire dès l'émergence du sol par la plantule. Les mitochondries et les plastes font partie intégrante de ces transitions.

Pour étudier leur mise en place et leurs éventuelles transitions lors de la germination, plusieurs équipes du ARC CoE PEB de Perth ont analysé le transcriptome nucléaire (par puces Affymetrix d'expression sur une plateforme dont j'ai été le responsable pendant près de 2 ans), le transcriptome des organites (par RT-qPCR), le métabolome par GC-MS et le protéome par MS/MS depuis la récolte des graines jusqu'à 48h après leur mise en culture (figure 20).



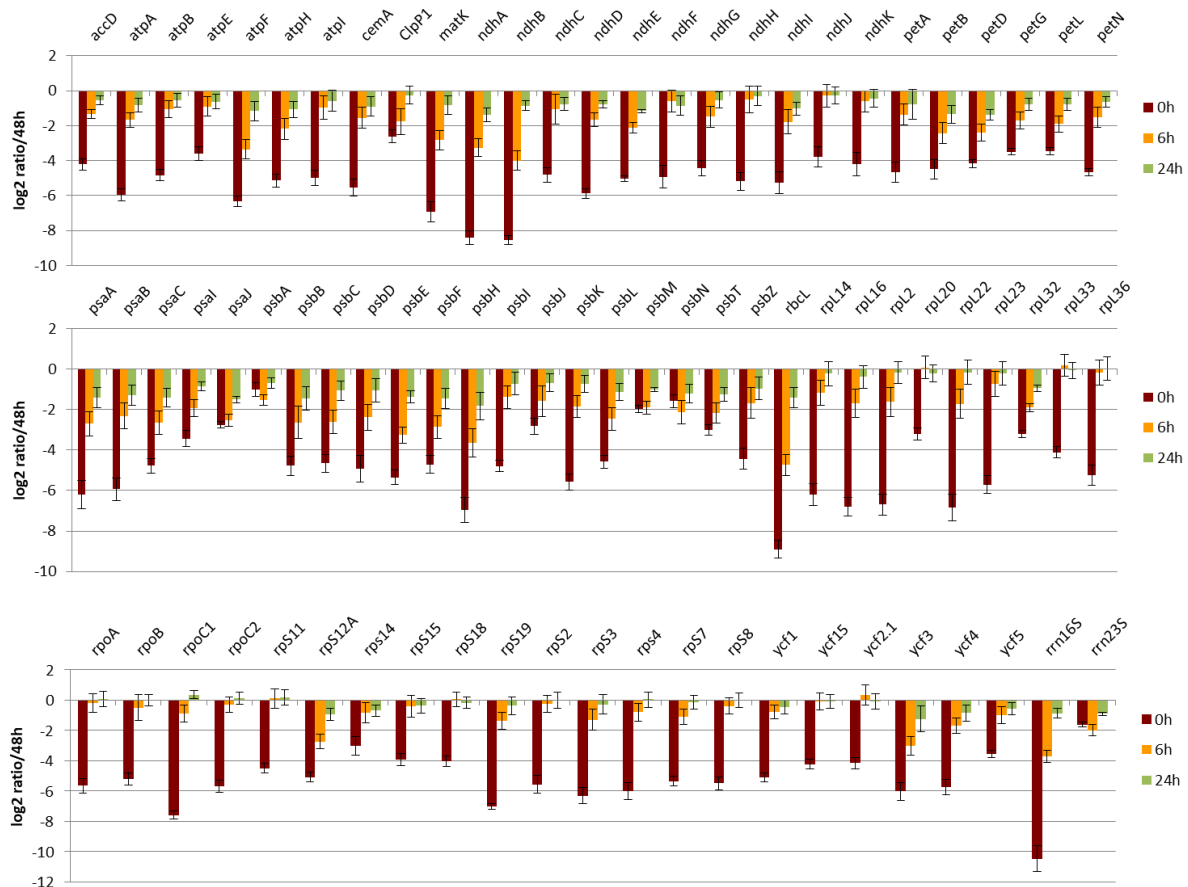
**Figure 20: design expérimental de l'analyse multi-omique de la germination.** Tous les points de la cinétique ont été analysés par puce Affymetrix alors que seuls les points indiqués ont été analysés avec les autres technologies.

Cependant, seule l'analyse du transcriptome nucléaire a inclus la totalité des 10 points de la cinétique. Le transcriptome des organites, le protéome et le métabolome ont été analysés sur 4 points en raison de limitations sur le matériel biologique et le budget disponibles mais également par choix stratégiques : certains points de la cinétique ne sont pas apparus comme pertinents a priori pour toutes les analyses. Les résultats de l'analyse de transcriptome nucléaire ont été publiés à part et ont mis en évidence un pic d'expression de gènes impliqués dans la traduction et le métabolisme des ARN mitochondriaux à la fin de la stratification et au tout début de la mise en culture (Narsai et al., 2011). C'est notamment le cas de 45% des gènes de PPR détectés dans cette cinétique et ce pic d'expression précède celui des composants du métabolisme énergétique. L'activité transcriptionnelle liée au plaste n'apparaît qu'à la fin de la cinétique avec l'émergence des cotylédons. Ces données transcriptionnelles ont été en partie associées aux données protéomiques puis aux données de transcription des organites dans un second temps (Law et al., 2012). En lien avec les données transcriptomiques nucléaires, le transcriptome mitochondrial montre que la quasi-totalité des gènes étudiés atteignent leur niveau d'expression final dès 6h après mise en culture (figure 21). Le choix de ne pas inclure des points de mesure en fin de stratification et en tout début de mise en culture s'est révélé préjudiciable à l'intérêt de ces données. Elles suggèrent cependant qu'il serait intéressant de refaire une étude de ces temps très précoces avec les outils d'analyse actuels pour avoir accès à l'ensemble des processus transcriptionnels (édition, épissage, processing).



**Figure 21: Evolution du transcriptome mitochondrial lors de la germination.** Les données sont normalisées par rapport au temps final de la cinétique (48h) de chaque gène.

Côté transcriptome chloroplastique, les résultats sont bien plus variés mais, comme les auteurs de Law et al. (2012) dont je fais partie ont orienté l'article sur le côté mitochondrial, ils n'ont pas été mis en avant (figure 22). On observe une activation du transcriptome chloroplastique (qui ne se termine pas à 48h) différente selon les gènes. Ceux qui sont impliqués dans la machinerie d'expression du plaste (*rpo*, *rps*, *rpl*, *ycf2*, *ycf1* mais pas les ARN ribosomiaux *rrn*) montrent une cinétique similaire à celle des gènes mitochondriaux avec des niveaux stables dès 6h. A l'inverse, la majorité des gènes impliqués dans la machinerie de photosynthèse (*rbcl*, *psa*, *psb*, *ndh*, *pet*, *ycf3*, *ycf4*, *ycf5*) présentent une accumulation progressive et plus tardive. Cette répartition en 2 groupes de gènes correspond aux gènes contrôlés par chacun des 2 systèmes de transcription existant dans le plaste. Le premier est celui de la Plastid-Encoded RNA Polymerase (PEP), d'origine bactérienne qui est codé par les gènes *rpo* du plaste. Le deuxième est celui de la Nuclear-Encoded RNA Polymerase (NEP), de type phagique qui est codé par le gène *rpoTp* du noyau. La NEP s'accumule de manière précoce lors de la germination et avant la PEP (Demarsy et al., 2006) ce qui est en accord avec les profils d'expression obtenus. Il y a donc d'abord une mise en place des complexes nécessaires à la production des complexes photosynthétiques avant que les gènes codant pour ces complexes ne soient transcrits. Cependant, une observation plus attentive montre qu'il existe des exceptions à ces 2 profils. Il s'agit des gènes *psbA*, *psbM* (qui codent pour des protéines du centre réactionnel du photosystème II), *psbN* (qui code pour une sous-unité du photosystème II) et l'ARN ribosomal 23S. Les transcrits de ces gènes sont déjà quasiment à leur niveau de fin de germination dans les graines. C'est intrigant pour des gènes aussi intimement liés à la photosynthèse et dont les transcrits sont les plus abondants d'une cellule de feuille : *psbA* représente à lui seul en moyenne 20% des ARN messagers des 24000 gènes codant des protéines chez *Arabidopsis* ! C'est deux fois plus que pour le transcrite de la grande sous-unité de la RubisCO, la protéine la plus abondante des plantes.



**Figure 22: Evolution du transcriptome plastidial lors de la germination.** Les données sont normalisées par rapport au temps final de la cinétique (48h) de chaque gène.

Malgré la volonté des chefs d'équipe et du centre de produire un gros jeu de données multi-omiques, ce projet n'a pas permis d'obtenir un jeu de données homogène pour toutes les analyses omiques ce qui rend impossible une analyse intégrée. Il a fallu s'en remettre à l'expertise (et donc aux biais) de chacun pour exploiter laborieusement et valoriser partiellement ces données. La raison principale je pense vient du manque de coordination entre les équipes mais aussi du manque de vision, de question biologique claire. En raison des contraintes techniques (production d'assez de matériel biologique pour toutes les analyses) et financières, il n'était pas possible de tester tous les temps de la cinétique avec toutes les technologies mais une meilleure coordination et une vision claire des objectifs auraient permis d'identifier les temps clés à analyser de manière exhaustive. La participation d'un spécialiste de l'intégration de données aurait aussi permis d'éviter ces erreurs.

## Réponses aux stress chez les plantes : le chloroplaste en filigrane

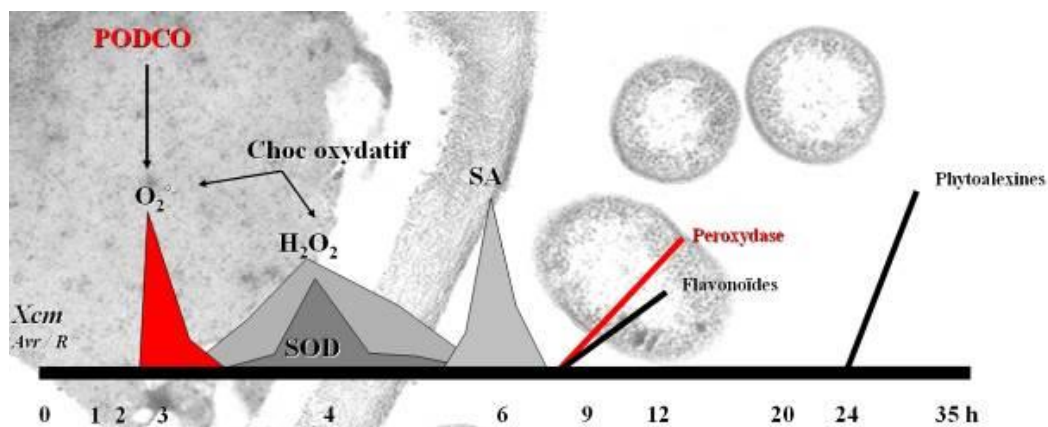
Au cours de mon parcours scientifique j'ai étudié la réponse des plantes à plusieurs stress biotiques et abiotiques en commençant par l'interaction entre le cotonnier et une bactérie pathogène *Xanthomonas campestris* pv. *malvacearum* (Xcm) lors de ma thèse. Leur interaction est de type gène pour gène: en cas de reconnaissance d'un facteur d'avirulence de la bactérie par les produits des gènes de résistance de la plante, l'interaction est incompatible et le déclenchement d'une réponse

hypersensible (RH) contient la maladie. Cette RH entraîne la mort programmée des cellules végétales infectées et de leurs voisines (figure 23). Dans le cas où les facteurs d'avirulence de la bactérie ne sont pas reconnus, la RH n'est pas déclenchée et la maladie se développe. L'interaction est compatible.



**Figure 23: Interaction incompatible entre un cotylédon de cotonnier Réba B50 et Xcm race 18.** Les zones claires correspondent à la mort cellulaire des tissus dans lesquels la bactérie a été infiltrée.

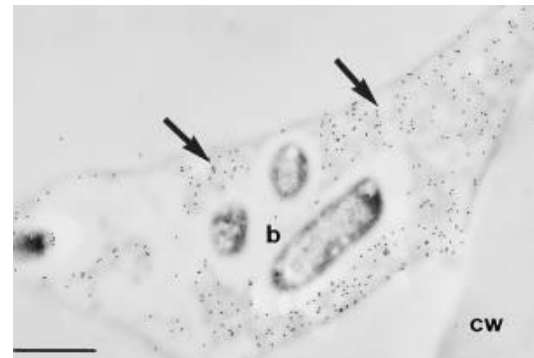
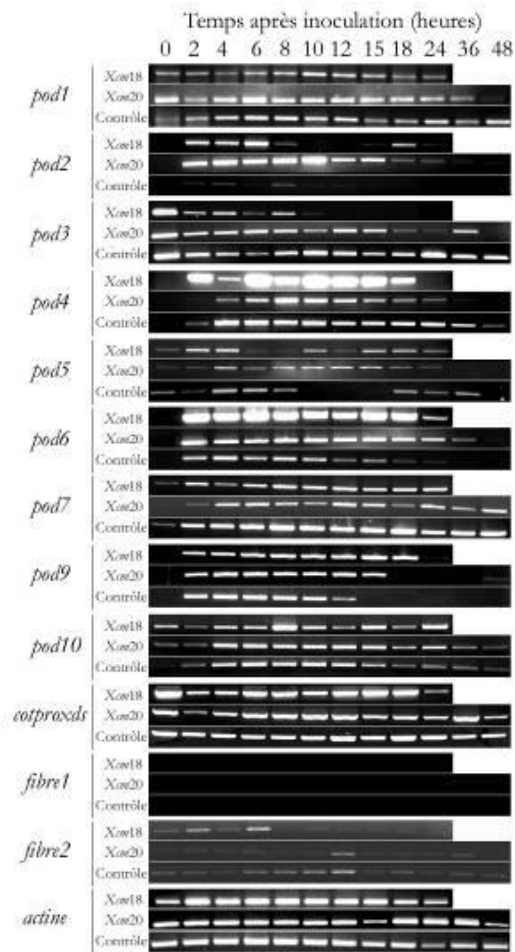
Chez le cotonnier et les autres plantes, la RH est caractérisée par un choc oxydatif en 2 phases (figure 24): une production précoce d'anion superoxyde ( $O_2^{\cdot-}$ ) suivi d'une vague plus tardive de peroxyde d'hydrogène. Grâce à des approches biochimiques, l'équipe avait pu montrer que, contrairement à *A. thaliana* (du moins dans les connaissances de l'époque), la production initiale d'anion superoxyde n'était pas due à une NADPH oxydase mais à une peroxydase cationique de classe III inconnue, nommée PODCO (Martinez et al., 2000).



**Figure 24: Evènements de la réponse hypersensible du cotonnier connus au début du projet.** En heures après infiltration de la bactérie avirulente. SOD: superoxide dismutase. SA: acide salicylique.

Afin de caractériser la voie de perception de la bactérie par le cotonnier, mon projet visait à identifier PODCO par des outils de biologie moléculaire. Le génome du cotonnier était évidemment inconnu à l'époque et les outils d'analyse du transcriptome en étaient alors à leurs balbutiements. Seuls 4 gènes de peroxydases sur les 164 connus actuellement chez le cotonnier étaient disponibles. J'ai donc criblé 2 banques de cDNAs faites à partir d'ARN extrait à 2h et 6h après infiltration de la bactérie grâce à un oligonucléotide dégénéré correspondant au site actif conservé des peroxydases de classe III. J'ai ainsi pu isoler 8 nouveaux gènes. J'ai ensuite analysé le profil d'expression des 12 gènes de peroxydase de cotonnier connus lors de l'interaction avec *Xcm* par PCR semi-quantitative (figure 25A). Cependant, aucune ne présente une induction spécifique de l'interaction incompatible à

3h correspondant à l'augmentation d'activité et à l'accumulation de peroxydases que nous avons mise en évidence par immuno-marquage (figure 25B) et tests biochimiques (Delannoy et al., 2003).



**Figure 25: Accumulation des peroxydases lors de l'interaction avec Xcm.**

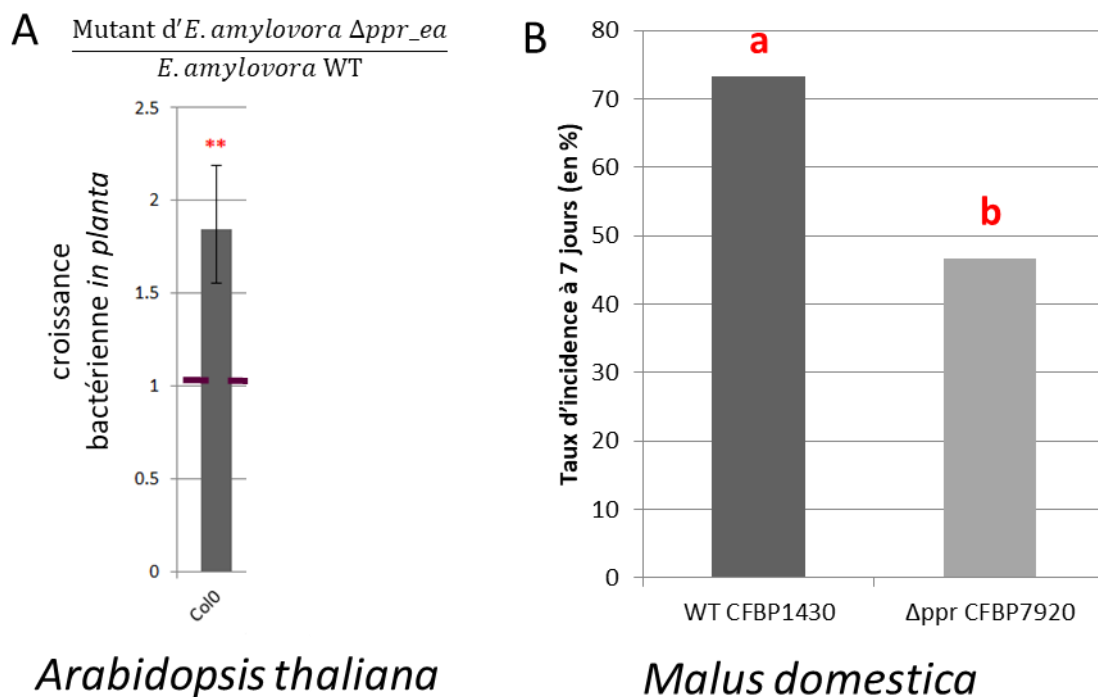
**A:** Expression des gènes de peroxydase de cotonnier connus lors de l'interaction incompatible (*Xcm18*) ou compatible (*Xcm20*) par PCR semi-quantitative.

**B:** immunomarquage à l'or des peroxydases à 3h lors de l'interaction incompatible. b: bactéries; cw: paroi cellulaire. Les flèches indiquent l'accumulation des peroxydases dans une matrice extracellulaire encapsulant les bactéries.

Rétrospectivement, alors même que la biochimie avait donné des indications très précises sur la peroxydase candidate PODCO, la transcriptomique n'a pas permis d'identifier le gène qui la code. Cet échec est très probablement dû à l'absence des outils de transcriptomique adéquats pour analyser sans *a priori* le profil d'expression de toutes les peroxydases de classe III du cotonnier en réponse à l'infection. De fait les peroxydases de classe III forment une large famille multigénique (presque 130 gènes connus actuellement chez le cotonnier). Elles sont facilement identifiables par leur domaine protéique et leur fonction moléculaire est d'oxyder des composés organiques. C'est la diversité de leurs substrats qui fonde la diversité de leurs fonctions. En ce sens leur analyse fonctionnelle pose les mêmes problèmes que celle des PPR.

Bien que cela ne soit pas évident du tout à la lecture des quelques paragraphes précédents, cette thèse a aussi été l'occasion d'un premier contact avec le chloroplaste. En effet, j'ai longtemps été intrigué par le fait que la HR n'avait lieu que si l'infection du cotonnier par *Xcm* se faisait de jour indiquant peut-être un lien avec cet organite. D'autres exemples de cette dépendance à la lumière sont connus (Hamel et al., 2016) et depuis l'importance du chloroplaste dans les interactions plantes pathogènes a largement été documentée (Littlejohn et al., 2020) que ce soit à travers les voies de signalisation qui s'y déroulent ou par le fait que des effecteurs, des molécules ou protéines produites

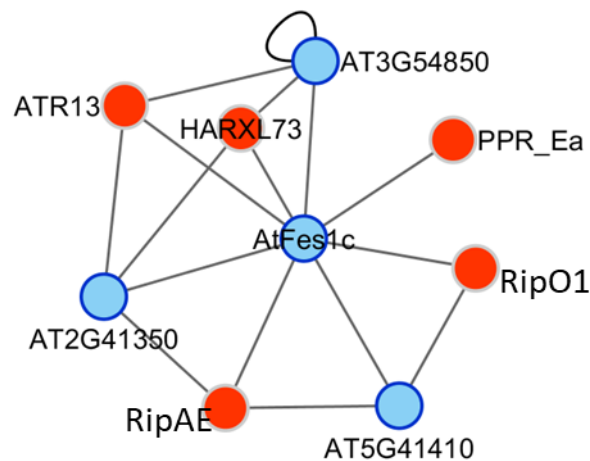
par les pathogènes pour favoriser l'infection, ciblent ce compartiment cellulaire. Dans cette perspective, j'ai étudié avec mon étudiant en thèse Bastien Malbert la fonction de 2 PPR très particulières car codées dans le génome de deux bactéries pathogènes de plantes : *Ralstonia solanacearum*, pathogène de nombreuses espèces végétales et *Erwinia amylovora*, pathogène du pommier. Les PPR sont très abondantes chez les eucaryotes et les plantes en particulier, mais je ne les ai détectées que dans le génome d'une poignée de bactéries pour la plupart des pathogènes de plantes, d'animaux ou d'humains. Nous voulions vérifier que ces PPR étaient des effecteurs ciblant les organites des plantes et identifier leur cible ARN. L'idée était que ces cibles devaient être importantes pour la réponse des plantes aux attaques pathogènes. Ces deux protéines se sont révélées particulièrement récalcitrantes (comme beaucoup de PPR) à nos approches d'analyse fonctionnelle. Nous avons néanmoins pu montrer que la PPR d'*Erwinia amylovora* (PPR\_EA) n'est pas nécessaire à la physiologie de la bactérie : la souche bactérienne mutée pour cette PPR a une croissance normale que ce soit en milieu riche ou minimum mais se comporte différemment de la souche sauvage en interaction avec les plantes (figure 26).



**Figure 26: Caractérisation du mutant bactérien en condition d'infection.** A : ratio de croissance in planta du mutant par rapport au sauvage sur *Arabidopsis thaliana*, plante non hôte d'*Erwinia amylovora*. B : Incidence des lésions bactériennes sur *Malus domestica* (plante hôte). Expérimentation réalisée par l'équipe de Marie-Noëlle Brisset à l'INRAE d'Angers.

La mutation de cette PPR diminue la virulence de la bactérie sur sa plante hôte, le pommier. A l'inverse, elle entraîne une augmentation de la fitness de la bactérie en interaction avec *Arabidopsis thaliana*, plante non-hôte, c'est-à-dire sur une plante sur laquelle la bactérie est incapable de déclencher une maladie. Même si elle est une plante non-hôte, *A. thaliana* est utilisée pour étudier le mécanisme d'infection d'*E. amylovora* (Degraeve et al., 2013). D'ailleurs, des lignées d'*A. thaliana* surexprimant PPR\_EA sont également plus sensibles à *Hyaloperonospora arabidopsidis* mais pas à

*Pseudomonas syringae* (deux micro-organismes pathogènes d'*A. thaliana*, un oomycète et une eubactérie respectivement) ce qui nous permet de conclure qu'elle impacte la réponse de la plante aux pathogènes. Pour comprendre le mécanisme impliqué, nous avons essayé de confirmer l'injection de cette protéine par la bactérie dans les cellules de la plante hôte (une caractéristique des effecteurs) et de déterminer sa localisation subcellulaire une fois dans la cellule végétale mais sans succès. Comme cette protéine est une PPR nous avons examiné le transcriptome des lignées Arabidopsis de surexpression par RNA-seq à la recherche de défauts caractéristiques des PPR (défauts transcriptionnels ou post-transcriptionnels dans les organites) mais sans succès également. Par contre, une analyse double hybride a montré une interaction avec la protéine Fes1c (AT5G02150, figure 27).

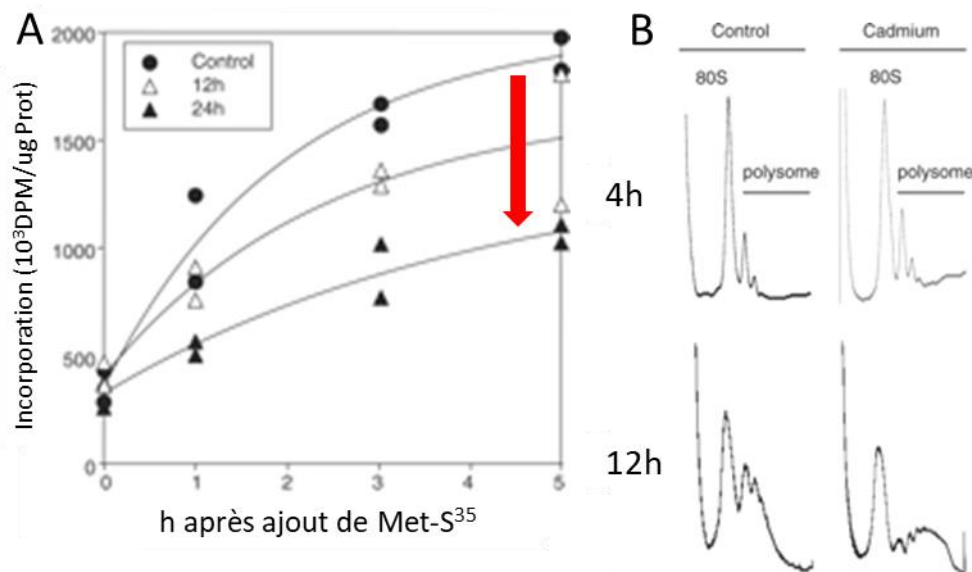


**Figure 27: Réseau d'interactions protéine-protéines de Fes1c.** Les effecteurs de pathogènes sont indiqués en rouge.

Fes1c est une protéine de fonction inconnue annotée comme Nucleotide Exchange Factor par homologie avec *Saccharomyces cerevisiae*. Il existe 3 paralogues chez *A. thaliana* dont Fes1a qui a été impliqué dans la réponse au stress chaud (Zhang et al., 2010). Fes1c est surtout la cible de plusieurs effecteurs en plus de PPR\_EA : 2 effecteurs de *Ralstonia solanacearum* (RipO1 et RipAE) mais surtout 2 effecteurs de *H. arabidopsidis* (ATR13 et HARXL73). Nous avons également montré que le mutant de Fes1c chez *A. thaliana* est plus sensible à l'infection par *H. arabidopsidis* suggérant que notre PPR agit via son interaction avec la protéine Fes1c et que Fes1c semble importante pour la réponse à certains pathogènes. Ainsi, bien que PPR\_EA soit constituée de domaines PPR, il ne semble pas qu'elle agisse comme les autres PPR par une interaction avec un ARN de chloroplaste et/ou de mitochondrie. Son rôle exact dans l'interaction plante pathogène est encore obscur.

A côté de l'étude des stress biotiques, j'ai également participé à plusieurs projets étudiant la réponse des plantes aux stress abiotiques. Le premier est celui induit par le cadmium lors de mon post-doc au CEA. Avec Rodnay Sormani, à l'époque doctorant dans l'équipe, nous avons étudié l'effet du cadmium sur la traduction. Nous avons montré que le cadmium, même à dose trop faible pour entraîner une mort cellulaire, conduit à une réduction globale de la synthèse protéique (Figure 28A, Sormani et al., 2011). Une autre technique pour estimer l'activité de synthèse protéique est la mesure de l'accumulation de polysomes de haut poids moléculaire. Les polysomes sont constitués d'un transcrit associé à un ou plusieurs ribosomes. Plus le nombre de ribosomes est grand, plus le polysome est de haut poids moléculaire. Aussi le modèle couramment admis postule que les

transcrits associés à des polysomes de haut poids moléculaire sont fortement traduits et inversement.



**Figure 28: inhibition de traduction par le stress cadmium.** A : Cinétiques d'incorporation de Met- $^{35}\text{S}$  à 4 et 12h après traitement cadmium. B : profils de polysome à 4 et 12h après traitement cadmium.

De même, une accumulation de polysomes de haut poids moléculaire traduit une forte synthèse protéique. En réponse au stress cadmium, nous avons observé une accumulation de polysomes associée à cette baisse de synthèse protéique (figure 28B) ce qui est en contradiction avec le modèle précédent. Cette contradiction provient du fait qu'une des suppositions de base de ce modèle est fautive dans notre expérience : un transcrit est associé avec de nombreux ribosomes uniquement lorsque sa traduction est élevée car les ribosomes s'enchainent rapidement les uns derrière les autres. Or, ce cas peut également se produire lorsque la course des ribosomes sur un transcrit ralentit voire stoppe, formant ainsi un bouchon comme sur les routes d'Ile de France aux heures de pointe.

Par une analyse transcriptomique utilisant les puces CATMA de l'URGV, nous avons identifié 668 transcrits affectés par le cadmium que ce soit dans leur abondance totale ou dans leur accumulation dans les polysomes. Ces transcrits sont enrichis entre autres en transcrits de protéines ciblées au chloroplaste et en transcrits du ribosome mais nous avons surtout observé que les variations d'accumulation dans les polysomes étaient en accord avec les variations d'accumulation totale pour seulement 150 de ces transcrits (classe 1). 475 transcrits ont montré des changements d'accumulation totale sans changement dans les polysomes (classe 2) et les 43 autres montrent au contraire des changements dans les polysomes sans variation de leur accumulation totale (classe 3). Ainsi, le cadmium inhibe la synthèse protéique mais entraîne également une très large régulation traductionnelle (Sormani et al., 2011).

Lors d'un autre projet avec l'équipe de Françoise Budard de l'IJPB (Versailles), nous avons étudié la réponse à la carence en azote d'accessions d'*Arabidopsis thaliana* dont les génomes cytoplasmiques (mitochondrie et chloroplaste) avaient été échangés créant ainsi de nouvelles combinaisons nucléocytoplasmiques. Comme l'activité des organites dépend de la coordination entre les 3 génomes de la

cellule qui ont eu le temps de co-évoluer, l'équipe avait montré dans un autre article que les accessions regroupant des génomes moins adaptés les uns aux autres montraient des différences dans de nombreux traits adaptatifs (nombre de graines, tailles des siliques, floraison etc... (Roux et al., 2016)). Afin de comprendre les mécanismes moléculaires en jeu, nous avons donc analysé le transcriptome, métabolome et protéome de certaines de ces nouvelles combinaisons nucléo-cytoplasmiques (Chardon et al., 2020). Plus de 12250 gènes, 77 protéines et 68 métabolites ont été affectés par la carence en azote. Parmi les gènes différentiellement exprimés on trouve un fort enrichissement en gènes de protéines ciblées au chloroplaste ainsi que 76% des gènes du chloroplaste démontrant l'impact majeur de cette carence sur le chloroplaste. Les gènes nucléaires codant pour des sous-unités des photosystèmes étaient largement inhibés alors que les gènes chloroplastiques des mêmes complexes étaient induits. Par contre aucune variation des protéines de ces gènes n'a été détectée montrant une forte régulation post-transcriptionnelle. De même, plus de 7000 gènes répondaient (presque exclusivement en inhibition) à la rupture de co-adaptation entre génomes nucléaire et cytoplasmiques contre seulement une protéine et neuf métabolites suggérant que des contrôles post-transcriptionnels maintiennent l'homéostasie des protéines.

Je dois admettre que je suis assez déçu par ce que nous avons réussi à tirer de ce gros projet et de toutes ces données mais il est difficile de voir ce que je ferais différemment si je pouvais le refaire. Le design expérimental a été établi soigneusement avec les statisticiens afin de favoriser une bonne intégration de cette étude multi-omique. Mixomics (Rohart et al., 2017), un outil dédié à l'intégration multi-omique, a été utilisé pour l'analyse intégrée des données produites. Cependant, ce sont les sorties des analyses statistiques individuelles de chaque jeu de données omiques qui ont été exploitées pour l'interprétation des données et la publication. De plus, il nous a fallu presque 5 ans d'analyses et d'interprétations pour écrire cet article. Cela illustre parfaitement que les sorties d'outils statistiques adaptés aux données et qui sont donc sophistiqués sont parfois très difficiles à interpréter. Cette difficulté peut venir soit du manque de connaissances des biologistes pour mettre en contexte les liens statistiques identifiés soit de l'interprétation même de ces liens (comment interpréter des corrélations entre des combinaisons linéaires de transcrits, de protéines et de métabolites ?). De plus, fondamentalement, Mixomics permet de trouver des marqueurs moléculaires distinguant des catégories d'échantillons (typiquement séparer les patients malades des non malades). Pour une analyse visant plutôt à identifier des perturbations à l'échelle du métabolisme cellulaire, c'est probablement un désavantage. Même pour les analyses sur chaque jeu de données omiques qui étaient basées sur des modèles linéaires (certes à 3 facteurs biologiques et avec des interactions d'ordre 3), le simple volume des résultats a rendu extrêmement difficile l'obtention d'une vision à la fois globale et détaillée. Au final, des exemples de gènes, protéines et métabolites ont été choisis pour illustrer les résultats ce qui est assez frustrant je trouve. Nous avons peut-être fait l'erreur de combiner un effet fort (la carence en azote) avec l'effet qui nous intéressait le plus (la rupture de co-adaptation cytoplasme-noyau) ce qui a généré cette masse de variations. Une autre piste est d'augmenter le nombre de répliques. Cela peut apparaître comme une fuite en avant vers plus de données mais les relations recherchées dans les analyses multi-omiques sont complexes et deux ou trois mesures sont notablement insuffisantes pour les estimer correctement.

L'autre point illustré par ce projet est l'importance d'outils statistiques « performants » dans les projets omiques. Les analystes se basent sur les résultats de ces outils pour identifier les entités (gènes, transcrits, protéines, métabolites etc...) potentiellement intéressantes pour leur question, bref pour faire le tri dans la masse des possibles et indiquer où focaliser leur attention. Autrement

dit, cela revient éliminer l'immense majorité des entités étudiées. Il faut donc s'assurer que les outils utilisés répondent effectivement à la question posée par le biologiste et qu'ils fonctionnent correctement sur les données analysées. Avec l'évolution rapide des technologies générant les données analysées, cette question se pose sans cesse.

Au cours des 10 dernières années sur la plateforme transcriptome, j'ai piloté une transition technologique majeure depuis les outils d'hybridation sur puce vers les outils de séquençage à la fois sur le plan de la génération des données et sur celui de leur analyse bio-informatique et statistique. Comme à chaque transition technologique, le fantasme que la nouvelle technologie allait nous épargner les analyses a resurgi. Plusieurs scientifiques éminents et enthousiastes m'ont dit que puisque le RNA-seq permet de compter les transcrits un par un et donc d'accéder directement à l'expression des gènes, il n'y avait plus besoin de bio-informatique ni de statistiques... Cela ne pouvait pas être plus éloigné de la vérité. Autant les protocoles générant des données par RNA-seq se sont stabilisés assez rapidement, autant ces données nécessitent énormément de bio-informatique avec d'innombrables paramètres à toutes les étapes (contrôle qualité, assignation aux gènes (« mapping »), comptage) pour arriver aux tables de comptage. Et ce d'autant plus que le RNA-seq ouvrait la possibilité d'étudier l'expression des gènes dans n'importe quelle espèce. Pour les statistiques, les outils ont été proposés assez rapidement mais nous avons été très surpris de constater que notre bonne « vieille » puce à ADN CATMA faisait mieux que le tout nouveau RNA-seq en ce qui concerne la détection de gènes différentiellement exprimés. Comme les données générées par le RNA-seq étaient de meilleure qualité que celles de la puce à ADN (notamment une gamme dynamique incomparable et la quasi absence de bruit de fond), je me suis plongé dans les analyses statistiques du RNA-seq avec les statisticiens de l'unité. Pour comprendre l'importance de chaque paramètre de cette analyse et sans chercher à développer un nouvel outil statistique, nous avons créé un jeu de données synthétique permettant de faire une comparaison neutre des outils disponibles (Rigaill et al., 2018).

Cette comparaison est neutre car nous n'avons pas cherché à promouvoir une éventuelle nouvelle méthode par rapport aux autres mais juste à identifier les paramètres clés de l'analyse différentielle par RNA-seq. Plusieurs questions se posent lorsqu'on compare des méthodes d'analyse différentielle.

- Comment les modèles statistiques s'ajustent aux données ?

En raison du très faible nombre de réplicas (petite taille du jeu de données dans la terminologie statistique ce qui peut prêter à confusion pour du RNA-seq), les outils statistiques ne peuvent pas estimer correctement la variabilité de l'expression d'un gène dans une condition donnée (sa variance). Or, pour une différence d'expression donnée, selon que la variance est grande ou pas, un gène pourra être déclaré différentiellement exprimé (DE) ou pas. C'est donc un paramètre clé et son estimation est l'objet de nombreuses recherches en statistiques et constitue ce qui distingue les principaux outils d'analyse différentielle. En effet, afin d'estimer au mieux la variance, les outils les plus populaires (et performants) ont recours à des suppositions sur les données analysées : leur aléa suit telle ou telle loi statistique dont les paramètres mathématiques seront estimés à partir de quelques observations disponibles. Les résultats de ce modèle ajusté d'aléa sont ensuite comparés aux observations pour voir si elles sont aléatoires (gène non DE) ou non (gène DE). Il est donc fondamental de s'assurer que le modèle mathématique utilisé correspond bien aux données analysées.

- Les p-valeurs calculées discriminent-elles bien les gènes DE et non DE ?

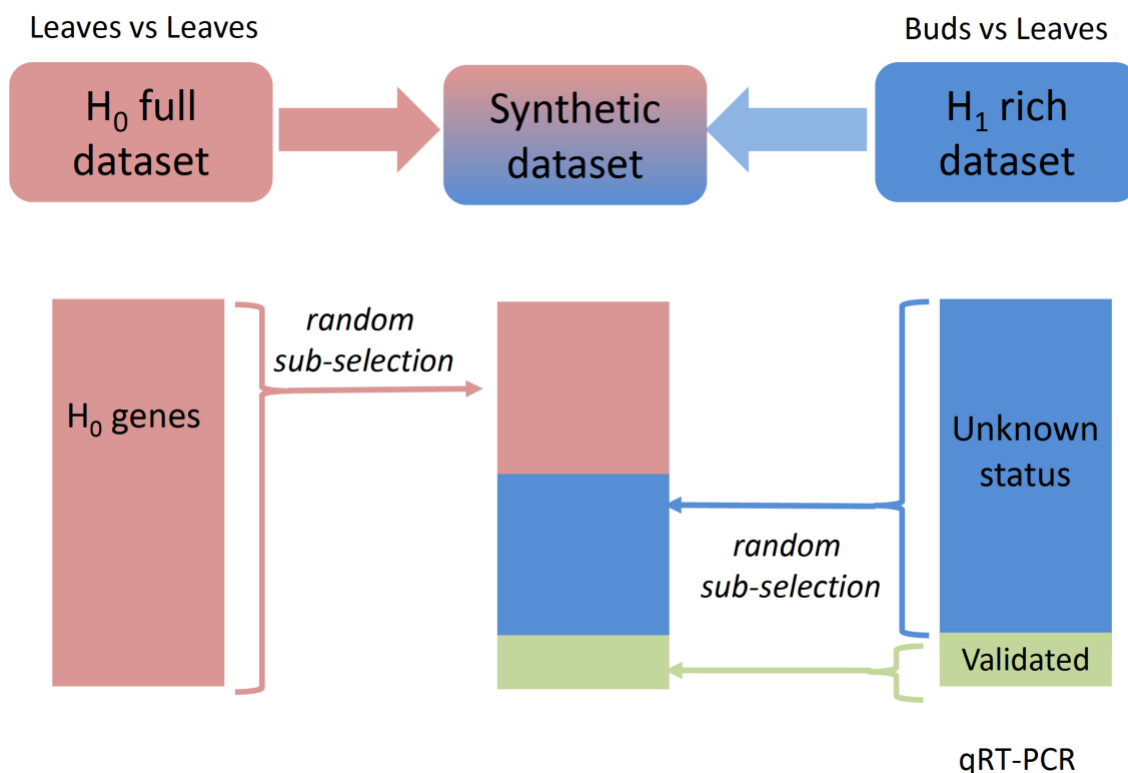
Une p-valeur est la probabilité d'observer les résultats d'expression mesurés par RNA-seq selon le modèle d'aléa ajusté choisi. En d'autres termes, plus la p-valeur est petite, moins on se trompe en affirmant qu'un gène est DE. C'est une valeur dont l'utilisation est largement discutée (Wasserstein and Lazar, 2016) mais qui reste le juge de paix pour identifier les gènes DE : ceux dont la p-valeur est sous un certain seuil (arbitraire..) sont déclarés DE. Il faut donc que la p-valeur classe bien les gènes selon qu'ils soient DE ou non.

- Les faux-positifs sont-ils bien contrôlés ?

En fixant un seuil sur les p-valeurs, l'analyste accepte un risque déterminé d'obtenir des faux positifs c'est-à-dire des gènes déclarés comme DE alors qu'ils ne le sont pas. Une méthode qui contrôle bien les faux positifs sera donc celle qui donne un nombre de faux positifs attendu (pas plus, mais pas moins non plus...).

- Les méthodes sont-elles puissantes ?

La puissance est la capacité à identifier les gènes réellement DE. C'est tout aussi important (voire plus) que la capacité à contrôler les faux-positifs car tout gène ne faisant pas partie de la liste des gènes déclarés DE sera presque définitivement écarté du reste du projet.



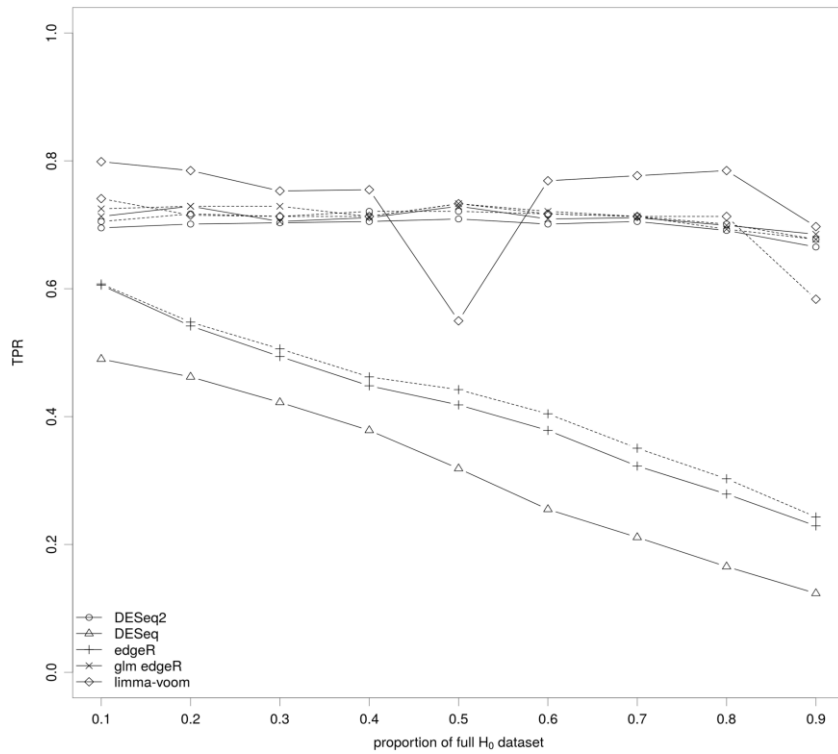
**Figure 29: Construction de jeux de données synthétiques à partir de données réelles.**

Pour répondre à ces questions, il est nécessaire d'avoir un jeu de données permettant de mesurer tous ces paramètres. C'est possible en utilisant des données simulées pour lesquelles on connaît la

« vérité » mais dans ce cas il faut choisir un modèle d'aléa ce qui favorisera forcément les méthodes statistiques utilisant le même modèle. De plus, ces données simulées ne sont pas réalistes. Une autre option est d'utiliser des données réelles mais dans ce cas la vérité n'est pas connue et il est impossible de tester différents scénarios (plus ou moins de gènes DE notamment). Nous avons choisi une troisième option, celle des données synthétiques (figure 29).

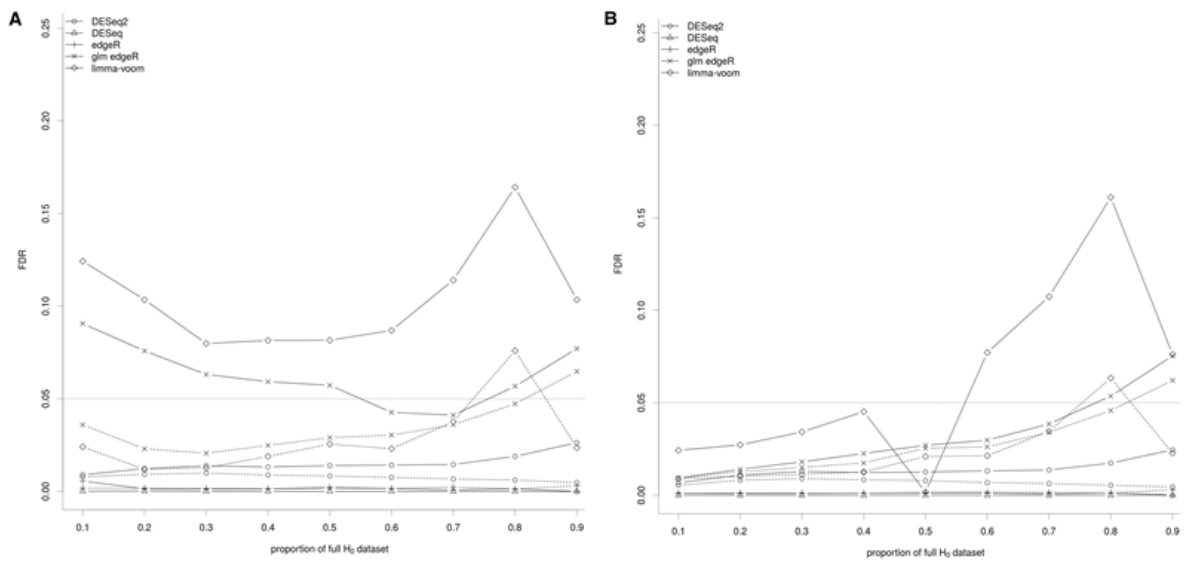
Il s'agit de partir de 2 comparaisons, l'une comparant des échantillons de la même condition donc dans laquelle il ne doit y avoir aucun gène DE (le «  $H_0$  full ») et une autre comparant des échantillons très différents donc avec de nombreux gènes DE (le «  $H_1$  rich »). Nous avons choisi de comparer des feuilles à des boutons floraux. Pour cette 2<sup>ème</sup> comparaison, nous avons testé l'expression de 333 gènes au hasard par RT-qPCR pour valider le caractère DE de 251 d'entre eux. Il va sans dire que l'analyse des données de RT-qPCR est sujet à caution mais cette technique est considérée comme le standard pour identifier les gènes DE. En combinant une fraction des gènes (incluant systématiquement les 333 gènes testés en RT-qPCR) du jeu «  $H_1$  rich » avec les autres gènes du jeu «  $H_0$  full », on obtient des jeux de données réalistes avec des proportions de gènes DE variables avec lesquels il est possible d'estimer la puissance des méthodes (avec les 251 gènes DE validés). L'estimation du taux de faux positifs est délicate (mais possible) car il faut identifier les gènes non DE. Ils correspondent aux gènes venant du jeu «  $H_0$  full » mais aussi aux gènes non DE venant du jeu «  $H_1$  rich ». Avec quelques astuces, nous avons pu proposer un encadrement de ce taux de faux positifs. La principale limitation de cette étude est qu'elle n'inclut que 2 répliques biologiques ce qui ne correspond plus aux standards actuels.

Avec ces jeux de données, nous avons testé l'impact de la modélisation de la moyenne, de l'estimation de la variance et du filtrage des gènes faiblement exprimés sur les performances des analyses lorsque la proportion de gènes DE varie. Pour cela nous avons comparé plusieurs méthodes illustrant ces différents paramètres. DESeq et edgeR utilisent une analyse basée sur la binomiale négative mais sans modélisation de la moyenne. Ces méthodes diffèrent par leur manière d'estimer la variance. DESeq2 et glm edgeR utilisent des modèles linéaires généralisés. Pour faire simple, ils font des modèles linéaires mais en utilisant la binomiale négative au lieu de la loi normale. Enfin, limma voom utilise une transformation des données de comptage (voom) pour revenir au cadre analytique connu pour les puces à ADN (limma) qui se base sur un modèle linéaire.



**Figure 30: Impact des paramètres de l'analyse différentielle sur la puissance.**

TPR : True positive rate, proportion de gènes réellement DE déclarés DE. Proportion of full H<sub>0</sub> dataset : plus elle est élevée, moins il y a de gènes DE dans le jeu de données synthétique. Les courbes représentent la médiane de 10 jeux de données synthétiques. Les courbes pleines correspondent aux analyses sans filtrage des gènes peu exprimés et celles en pointillés aux analyses avec filtrage.

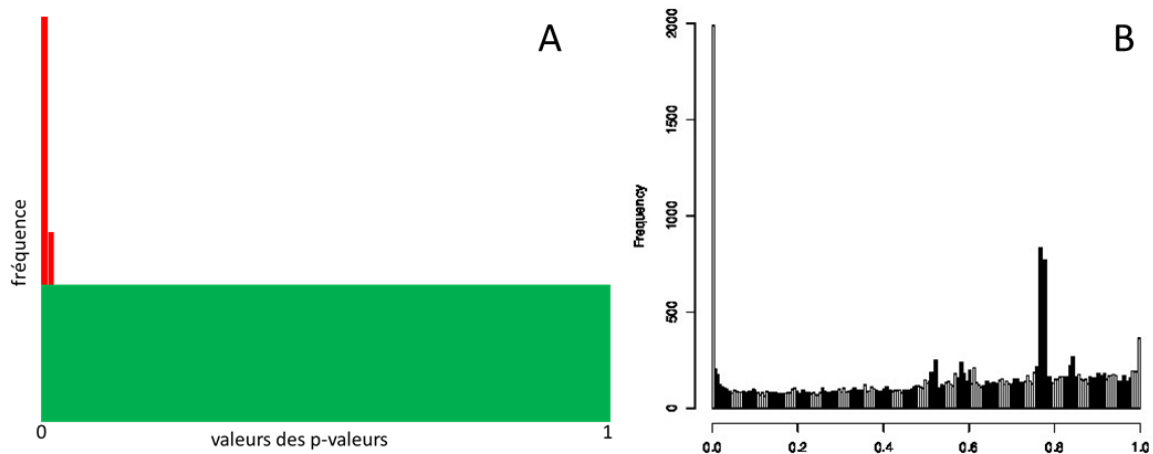


**Figure 31: Impact des paramètres de l'analyse différentielle sur le taux de faux positifs.**

A : borne supérieure du taux de faux positifs. B : borne inférieure du taux de faux positifs. Les courbes représentent la médiane de 10 jeux de données synthétiques pour des analyses où les gènes sont déclarés DE si leur p-valeur ajustée par FDR est inférieure à 5%. Les courbes pleines

correspondent aux analyses sans filtrage des gènes peu exprimés et celles en pointillés aux analyses avec filtrage.

Nous avons observé que le paramètre le plus important de l'analyse différentielle des données RNA-seq est la modélisation de la moyenne. Les méthodes qui ne le font pas (edgeR et DESeq) sont moins puissantes (figure 30) car trop stringentes (figure 31). C'était un résultat inattendu car le design expérimental est très simple avec seulement un facteur biologique avec 2 modalités (feuille et bouton floral) et un facteur technique correspondant aux réplicas biologiques. Il est donc étonnant qu'une modélisation de la moyenne ait un tel impact dans ce contexte. De plus, les efforts des équipes de statisticiens travaillant sur ces outils se focalisent principalement sur l'estimation de la variance. C'est la principale différence (avec la normalisation) entre DESeq(2) et (glm) edgeR. Il est donc crucial de proposer une modélisation de la moyenne qui corresponde aux données pour obtenir une analyse statistique performante. Le corollaire est que, si la modélisation de la moyenne au sein du modèle linéaire n'est pas bonne, les résultats de l'analyse différentielle n'ont pas de pertinence. Cependant l'évaluation de la correspondance entre les données et la modélisation de la moyenne n'est pas facile à évaluer en dehors de jeux de données synthétiques ou simulés. C'est d'autant plus vrai que les modèles linéaires permettent de prendre en compte des designs expérimentaux complexes avec plusieurs facteurs biologiques et des interactions. Par exemple, une expérience de plus en plus courante chez les collaborateurs de la plateforme transcriptomique consiste à évaluer l'impact d'une mutation sur la réponse à un changement de l'environnement. Classiquement, les biologistes recherchent les gènes DE en réponse à l'environnement dans le génotype 1 puis font de même pour le génotype 2. Ils font ensuite une intersection des 2 listes pour voir ce qui diffère entre les 2 génotypes. Cette approche souffre de nombreuses limitations (effet seuil sur la p-valeur, non détection des gènes DE et de même signe dans les 2 génotypes mais avec un logratio différent, détection partielle des gènes DE, absence de soutien statistique...). Dans un modèle linéaire, il suffit de tester l'interaction entre les 2 facteurs. Afin d'évaluer la bonne concordance entre le modèle linéaire et les données pour tous les projets, nous avons proposé un critère qui se base sur l'histogramme de distribution des p-valeurs brutes (Rigaill et al., 2018). Par définition, dans toute analyse différentielle, les gènes peuvent être divisés en 2 populations, ceux qui sont DE et ceux qui ne le sont pas. Ceux qui sont DE ont des p-valeurs faibles ce qui crée un pic vers 0. Pour les gènes non DE, une petite analyse mathématique montre que leurs p-valeurs brutes sont distribuées uniformément entre 0 et 1, l'histogramme est plat. Ainsi tout histogramme de p-valeur peut montrer un pic vers 0 (s'il y a des gènes DE) mais doit autrement être plat (figure 32). Si ça n'est pas le cas avec les p-valeurs calculées par l'analyse statistique c'est que le modèle utilisé ne correspond pas bien aux données et dans ce cas, ce modèle (ou les données) doit être modifié.



**Figure 32: utilisation de l'histogramme des p-valeurs brutes pour l'évaluation de l'analyse statistique.**

A : représentation théorique d'un histogramme de p-valeurs brutes. Les p-valeurs des gènes DE sont en rouge et celles des gènes non DE en vert. B : exemple d'histogramme de p-valeurs montrant un problème de modélisation.

L'utilisation des modèles linéaires dans l'analyse des données omiques présente des avantages qui sont de plus en plus nombreux à mesure que les designs expérimentaux deviennent plus complexes. Ces designs correspondent en général à des questions complexes posées par le biologiste. Or le recours aux contrastes permet de répondre précisément à ces questions. En fait, le modèle linéaire est un remarquable outil de discussion avec le collaborateur avant même la production des données car il l'oblige à formuler précisément ses questions et nous permet d'en déduire le design expérimental nécessaire. Il en résulte une inversion, surprenante pour les biologistes, des priorités où c'est le modèle/outil statistique qui contraint et guide le design expérimental. C'est pourquoi il est au cœur de l'école chercheurs « de l'expression des gènes aux réseaux » que j'ai construite et anime avec Marie-Laure Martin depuis 2015. Lors de cette formation de 4 jours, nous abordons l'analyse statistique du transcriptome depuis l'analyse différentielle à l'inférence et l'analyse de réseau de gènes en passant par la co-expression.

Avec ces outils et dans le cadre de mon activité de plateforme transcriptomique sur de nombreux projets, il est rapidement apparu que les gènes associés aux activités du chloroplaste sont très souvent impactés par les conditions de stress. Une méta-analyse des données transcriptome de la plateforme depuis sa création en 2003 ont permis de confirmer cette impression.

## **Approches de méta-analyse : la place centrale du chloroplaste dans la réponse au stress**

Afin de mieux comprendre la réponse aux stress chez *Arabidopsis thaliana*, l'équipe Genomic Network a initié la méta-analyse des données transcriptomiques générées pour les projets en lien avec le stress, soit à l'époque, 387 comparaisons faites avec la puce à ADN CATMA. Cette méta-

analyse a débuté par l'identification des gènes co-exprimés dans ces conditions. L'analyse de co-expression recherche les gènes dont les profils d'expression sont similaires et est généralement basée sur les outils de corrélation (Usadel et al., 2009). Bien que la corrélation soit un mauvais estimateur de la similarité quand il y a beaucoup de variables (c'est le cas pour les gènes lors d'une analyse de co-expression), ces outils sont utilisés pour une analyse rapide de la similarité des profils d'expression des paires de gènes qui est résumée par cette valeur de corrélation variant de -1 (anti-corrélation parfaite) à 1 (corrélation parfaite). Une fois ces valeurs obtenues, l'application d'un seuil (arbitraire...) permet d'identifier les paires de gènes co-exprimés. Pour éviter l'application de ce seuil, il est possible d'utiliser les valeurs de corrélation comme une mesure de distance dans des outils de classification non-supervisée tels que le clustering hiérarchique ou l'analyse de K-means pour identifier des groupes (clusters) de gènes co-exprimés. Dans ces analyses, il faut cependant souvent fournir le nombre de clusters attendus ce qui influence très fortement l'analyse et les résultats obtenus. En partant de l'analyse des paires de gènes, il est ainsi possible d'essayer d'identifier la structure de co-expression globale.

A l'inverse de l'approche par corrélation, l'analyse de co-expression effectuée sur notre jeu de données part de la population globale de profils de transcription pour chercher sa structure sous-jacente. Elle est basée sur les modèles de mélange. Par analogie avec les spectres d'activité de la photosynthèse, les modèles de mélange font de la dé-convolution : ils supposent que le signal observé vient de la superposition de plusieurs groupes « homogènes » de profils d'expression, c'est-à-dire de plusieurs groupes de gènes partageant un profil d'expression similaire donc de gènes co-exprimés. Comme les suppositions mathématiques de cette approche (« il existe des groupes homogènes de profils ») correspondent directement à des suppositions biologiques (« il existe des gènes co-exprimés »), ses résultats (les groupes de gènes) sont directement interprétables (groupes de co-expression). De plus, ces modèles permettent d'identifier les groupes et de calculer les probabilités pour chaque gène d'appartenir à chacun d'entre eux. Il est donc possible de ne pas affecter un gène à un groupe si son appartenance n'est pas claire (ce qui se justifie biologiquement). Enfin, le nombre de groupes est déterminé par une approche de sélection de modèles sur la base des données et non de manière arbitraire. Cette sélection de modèles permet aussi de savoir si l'analyse a réussi à trouver la structure sous-jacente ou pas. Si le jeu de données est basé sur plusieurs structures sous-jacentes alors l'analyse n'aboutit pas. Ainsi il a été nécessaire de subdiviser les 387 comparaisons en plusieurs groupes jusqu'à ce que l'analyse aboutisse dans chaque groupe. Nous sommes finalement arrivés à 18 groupes associant des comparaisons correspondant à une même condition de stress. D'après moi, cela signifie qu'il n'existe pas une structure sous-jacente de co-expression unique et identique pour toutes les conditions ce qui s'interprète facilement comme le fait que des gènes co-exprimés ne le sont pas dans toutes les conditions.

Les quelques 700 clusters identifiés dans les 18 analyses de co-expression pour chacune des conditions de stress ont été annotés et regroupés au sein de la base données Gem2Net (Zaag et al. 2015). Pour pouvoir intégrer ces 18 analyses de co-expression en une seule analyse de co-régulation, nous avons fait l'hypothèse que des gènes co-exprimés dans plusieurs stress sont co-régulés. Nous avons ainsi calculé le nombre de fois où chaque paire de gènes était co-exprimée (entre 0 et 18 fois) ce qui permet d'obtenir un réseau de co-expression dont les arêtes ont un score égal au nombre de co-expressions. En comparant ce réseau à des réseaux aléatoires de mêmes caractéristiques (même nombre de nœuds et d'arêtes et même distribution du degré des nœuds) nous avons vu que les arêtes avec un score de 3 ou plus (donc les paires de gènes co-exprimés dans 3 conditions de stress

ou plus) avaient moins de 1% de chance d'être dues au hasard. Nous avons ainsi obtenu un réseau de co-régulation composé de 4476 gènes et 56489 arêtes. On observe que l'immense majorité des arêtes sont soutenues à la fois par des stress biotiques et abiotiques et les gènes sont fortement enrichis en gènes annotés « réponse au stress », « photosynthèse » et « traduction ». Après ce travail de plusieurs mois/années nous étions très fiers de ce réseau. Cependant il est très vite apparu que sa taille nous empêchait d'en avoir une interprétation globale et notamment de voir comment ces différentes fonctions s'articulent dans le réseau. Pour en estimer la pertinence biologique, nous devons nous contenter d'exemples ciblés sur certains gènes ce qui est frustrant. Des indicateurs nous suggéraient que ce réseau n'est pas homogène mais est constitué de clusters de gènes. Grâce à un outil d'analyse de la topologie du réseau, basé sur les modèles de mélange, qui regroupe les nœuds du réseau (les gènes dans notre cas) ayant le même voisinage, nous avons détecté 52 clusters. Pour estimer la stabilité topologique de ce réseau de co-régulation, nous avons répété toute la construction du réseau en enlevant à son tour chacune des 18 conditions de stress. Seuls les gènes qui restaient dans le même cluster dans au moins 12 des 18 analyses ont été conservés pour obtenir le réseau de réponse au stress d'*A. thaliana*. Il est composé de 3407 gènes et 45788 arêtes. Le fait de conserver une majorité de gènes suggère fortement que, même si la réponse à chaque stress est spécifique, il existe un réseau de co-régulation commun aux différents stress chez *A. thaliana*.

Dans ce réseau, nous avons identifié 47 clusters de gènes (figure 33) dont 42 présentent des enrichissements associés notamment à la réponse au stress (12 clusters), à des fonctions photosynthétiques (8 clusters), à la respiration cellulaire (4 clusters) et à la traduction (5 clusters). Ces résultats suggèrent que ces clusters correspondent à des modules fonctionnels c'est-à-dire des groupes de gènes impliqués dans la même fonction. A une échelle supérieure, les clusters de fonction similaires se regroupent en 10 super-clusters qui peuvent être considérés comme les fonctions biologiques qui sous-tendent la réponse d'*A. thaliana*. On trouve notamment 2 fonctions biologiques associées à la photosynthèse à côté d'autres associées aux réponses au stress (3), à la respiration cellulaire (1) et au métabolisme des protéines (3) notamment la traduction.



condition de stress. Cela correspond parfaitement aux prédictions du réseau. Nous avons également étudié les mutants de 9 gènes de fonction inconnue (dont 5 de modules associés à des fonctions photosynthétiques) en réponse à la chaleur et au sel, deux stress choisis uniquement sur la base de la simplicité de leur mise en œuvre. Huit d'entre eux sont impliqués dans la réponse à au moins un des 2 stress dont 6 aux 2. Ces résultats confirment le très fort pouvoir prédictif de notre méta-analyse (que nous avons appelé EMERGENT) ainsi que la validité biologique des regroupements de gènes dans notre réseau.

On peut en déduire que la réponse au stress chez *Arabidopsis* s'articule autour de fonctions de réponse qui sont communes à plusieurs stress mais aussi et surtout autour de 3 grandes fonctions cellulaires basales : photosynthèse, respiration et traduction. Ces fonctions ont pour point commun d'être les acteurs majeurs du métabolisme énergétique : la photosynthèse et la respiration pour en produire et la traduction pour en consommer une grande partie. Il semble donc qu'en cas de stress la plante agisse pour limiter la source de ce stress (par les modules « réponse au stress ») tout en ajustant son métabolisme énergétique. Certains pourraient dire (et nous ont dit) qu'on ne fait que retrouver le squelette du métabolisme de la cellule. Cela semble plus complexe mais quand bien même ce serait le cas, ce qui a pris des dizaines d'années de recherche par des approches multiples pour comprendre le métabolisme d'une cellule, nous l'avons fait par une méta-analyse sans a priori sur des données transcriptomiques.

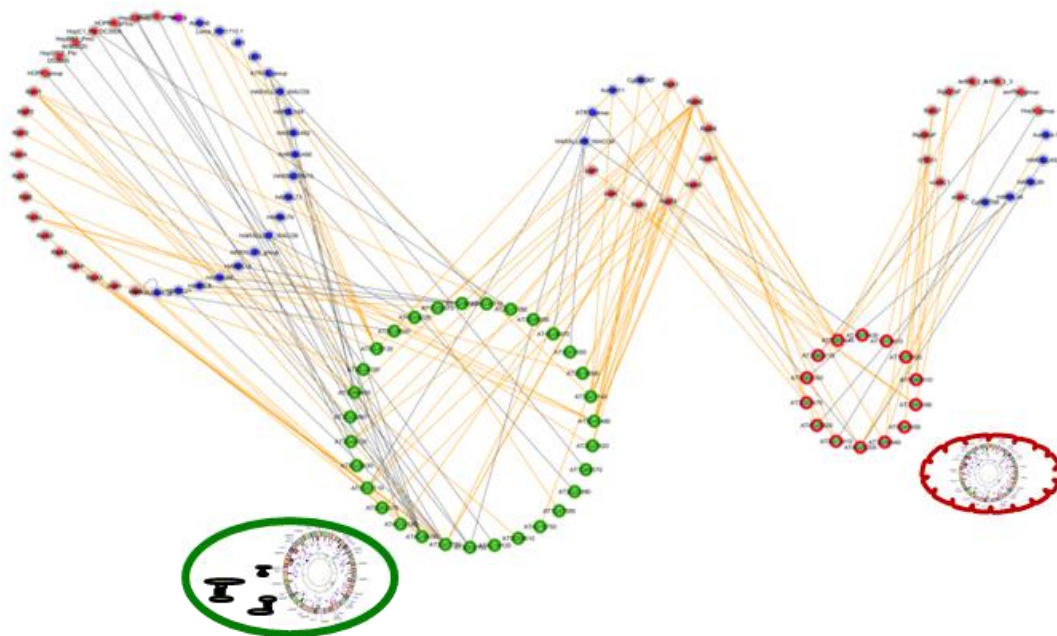
Nous avons fait la même analyse sur des données transcriptomiques de *Levure* (en quelques mois cette fois) et nous retrouvons des fonctions biologiques identiques hormis la photosynthèse évidemment et des gènes conservés entre ces deux organismes. Il semble donc que nous ayons identifié l'architecture de la réponse au stress chez les eucaryotes et qu'elle se définisse autour du métabolisme énergétique. Cela suggère que le stress est perçu par la cellule comme une perturbation de son homéostasie énergétique cellulaire.

Ce travail est une très belle illustration du potentiel de la modélisation en biologie et donc du travail interdisciplinaire. Cependant, nous avons passé énormément de temps avec Marie-Laure Martin à formaliser chacune des étapes de cette analyse complexe en vérifiant que les suppositions sous-jacentes aux outils mathématiques employés avaient une correspondance en biologie. Je crois que c'est la raison principale (voire unique ?) de la très forte pertinence des résultats obtenus, de la facilité à les interpréter aux différentes échelles du réseau (gène, cluster, super-cluster) et de la possibilité d'appliquer cette approche à d'autres espèces.

## **Projet : Vers le réseau de régulation du chloroplaste en réponse au stress**

Cette méta-analyse a mis en évidence le rôle central du chloroplaste dans la réponse au stress chez *A. thaliana*. De fait le nombre de publications liant le stress au chloroplaste a fortement augmenté ces 20 dernières années, atteignant un quart de la littérature associée au chloroplaste en 2020. Cependant, le rôle de l'expression chloroplastique est quasiment ignoré à part dans quelques études relativement récentes s'intéressant notamment au signal rétrograde qui permet de coordonner le fonctionnement du chloroplaste et du noyau (Robles et al., 2015; Robles and Quesada, 2019; Chan et al., 2016; De Souza et al., 2017; Hernández-Verdeja and Strand, 2018). Pourtant 3 de nos 47 modules

fonctionnels sont associés plus précisément à l'expression génétique du plaste ce qui fournit de nombreuses pistes et candidats pour relier la régulation de l'expression du chloroplaste et la réponse au stress. D'autres résultats de notre équipe confirment ce rôle lors des interactions avec les pathogènes. Durant les interactions compatibles, les pathogènes déploient des effecteurs, en général des protéines transférées vers la cellule hôte, dont les fonctions sont de supprimer les réponses immunitaires de la plante, d'interférer avec la physiologie de l'hôte pour permettre l'acquisition de nutriments. Or certains sont ciblés au chloroplaste pour en modifier son fonctionnement (Medina-Puche et al., 2020; Kretschmer et al., 2020) et son expression. Les cibles de ces effecteurs sont donc très probablement des facteurs clés de la réponse des plantes aux pathogènes. En utilisant la plateforme de double hybride de levure haut-débit de l'équipe, nous avons recherché les interactants protéiques de 200 effecteurs bactériens et fongiques (Mukhtar et al., 2011; González-Fuente et al., 2020 et non publié). Nous avons identifié 31 protéines ciblées au chloroplaste (figure 34) dont certaines impliquées dans la régulation de l'expression du chloroplaste. Ce sont autant d'autres candidats potentiels.



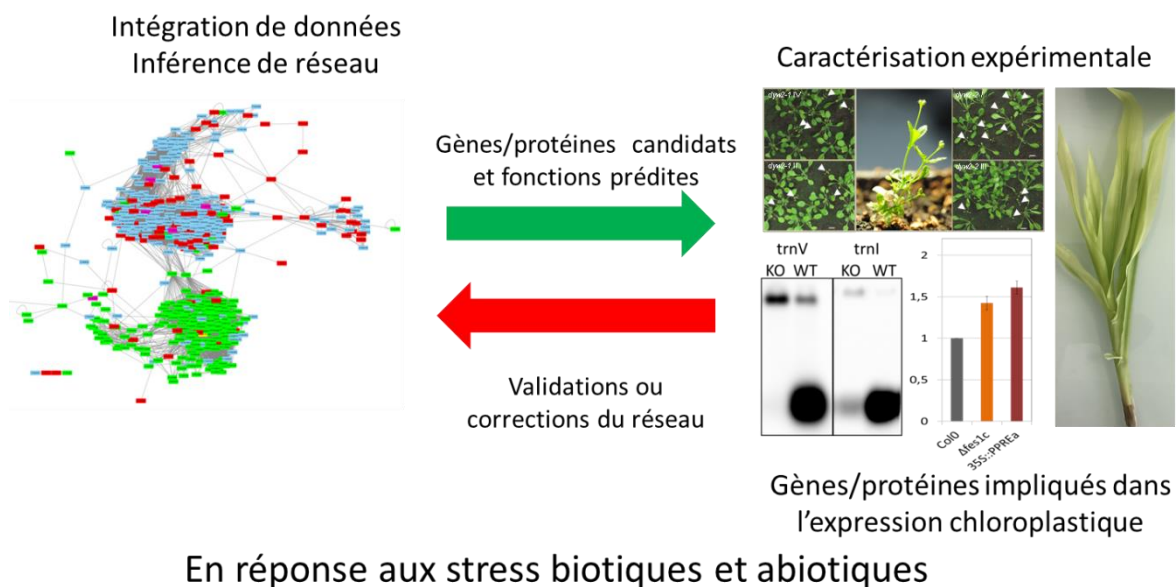
**Figure 34: Réseau des interactions protéines-protéines connues entre les protéines du plaste (en vert), de la mitochondrie (cerclé de rouge) et les effecteurs de différents pathogènes de plantes (en haut).**

Afin de comprendre les mécanismes moléculaires de la réponse adaptative du chloroplaste au stress, il est nécessaire de connaître les réseaux moléculaires qui organisent son activité.

Le travail de méta-analyse transcriptomique et les cribles de double hybride donnent une idée globale de ces réseaux mais ils ont plusieurs limitations. La première est que ces deux approches explorent des niveaux de régulation moléculaire distincts. On peut considérer que leur étude conjointe est complémentaire, mais leur intégration n'est pas triviale. La deuxième est qu'ils n'incluent pas les gènes et protéines codés directement par le chloroplaste. Comme indiqué précédemment, les transcrits chloroplastiques ne sont pas polyadénylés et sont donc exclus de la quasi-totalité des jeux de données transcriptomiques, notamment de ceux ayant servis à la construction de notre réseau de co-régulation. De la même manière, les protéines codées par le

génomique du plaste ne font pas partie de la banque de protéines disponibles pour nos cribles de double-hybride. Or ces gènes jouent un rôle central dans l'activité du chloroplaste. Il est donc nécessaire de les inclure dans toute étude prétendant modéliser son fonctionnement. La troisième est que ces approches réseaux permettent d'associer des gènes-protéines et donc de prédire qu'ils fonctionnent ensemble mais elles ne donnent pas la fonction exacte. Pour interpréter un réseau en termes fonctionnels il est nécessaire d'avoir une connaissance déjà solide de la fonction d'une fraction significative (c'est une formulation très floue mais il est difficile d'être plus précis) des gènes d'un réseau. Cette connaissance se base sur la caractérisation fonctionnelle expérimentale de chacun des gènes, caractérisation qui peut être guidée par les prédictions issues du réseau.

Ainsi mon projet fera l'aller-retour entre l'intégration de données et la caractérisation expérimentale des gènes candidats pour tendre vers la compréhension de la régulation du chloroplaste en réponse au stress (figure 35). J'appliquerai cette stratégie sur 2 plantes, *A. thaliana* et le maïs car ce sont deux modèles très largement utilisés dans l'étude du chloroplaste, pour l'étude des photosynthèses en C3 et en C4 et pour l'étude des Dicotylédones et des Monocotylédones. Pour ces raisons, il existe déjà de nombreuses ressources génétiques et de nombreuses données notamment transcriptomiques pour ces espèces.



**Figure 35: Stratégie scientifique du projet**

L'inférence de réseau sera faite en intégrant :

- les données de transcriptomiques. Pour *A. thaliana*, j'utiliserai le réseau de co-régulation déjà identifié. Il sera complété par des données transcriptomiques générées par l'équipe de recherche ou par des collaborateurs de la plateforme transcriptomique de l'IPS2 permettant l'analyse du transcriptome chloroplastique. Il devrait ainsi être possible de placer les gènes chloroplastiques dans le réseau de co-régulation. Le réseau de co-régulation du maïs est en cours de construction grâce aux données publiques déjà disponibles
- les données d'interactions protéine-protéine. La carte des interactions protéine-protéine d'*A. thaliana* est déjà partiellement disponible grâce aux cribles de double hybride de levure haut-débit

déjà faits (Dreze et al., 2011) et à venir sur la plateforme interactome de l'IPS2. L'intégration des protéines codées par le chloroplaste dans cette carte est en cours grâce à un crible double hybride. Elle pourra être complétée par les prédictions d'amarrage moléculaire (« molecular docking ») développées par Marie-Hélène Muchielli-Giorgi de l'IPS2. Cette approche devrait bénéficier grandement de la percée récente dans le domaine des prédictions de structure (Tunyasuvunakool et al., 2021). Pour le maïs, l'approche double hybride de Levure n'est pas actuellement envisageable mais si les prédictions d'interactions sont suffisamment fiables, elles pourraient être employées pour inférer le réseau des interactions protéine-protéine.

- les compositions protéique et nucléique du nucléoïde. C'est a priori dans ce complexe du chloroplaste que se font les régulations d'expression du plaste. Par des approches de biochimie disponibles dans l'équipe, le contenu protéique de ce complexe commence à être connu chez le maïs et *A. thaliana*. En les combinant à de la transcriptomique et de la protéomique, il sera possible de connaître la composition en protéines et en transcrits de ce complexe voire d'étudier sa dynamique en réponse à différents stress.

Il existe plusieurs manières d'intégrer des données omiques de nature différente. Dans mon cas, nous fusionnerons les réseaux obtenus avec chacun des types de données ce qui permet de s'assurer que chaque type de données est analysé avec les outils optimisés. Ensuite la topologie de ce réseau sera analysée pour identifier les modules fonctionnels et les fonctions biologiques. Par contre, comme les arêtes de ce réseau seront de natures différentes (co-régulation, interaction protéique, appartenance au nucléoïde), il est probable que l'outil utilisé précédemment pour le réseau de co-régulation ne soit plus approprié. Ce sera une nouvelle occasion d'interagir avec mes collègues statisticiens.

Les réseaux *A. thaliana* et maïs ainsi obtenus seront comparés sur la base de l'orthologie des gènes/protéines mais aussi par une comparaison de leur topologie. L'objectif sera d'identifier les gènes/protéines qui partagent une position commune dans leur réseau respectif et ainsi mettre en évidence l'architecture de la réponse au stress commune aux plantes. J'aurai là aussi besoin de mes collègues statisticiens. Il sera ensuite possible d'identifier les gènes de fonction mal ou pas connues, appartenant aux modules fonctionnels impliqués dans la régulation de l'expression génétique du chloroplaste et conservés chez *A. thaliana* et le maïs.

A partir de là, ces candidats seront soumis à une caractérisation expérimentale en testant notamment la réponse de leur mutant à différents stress. Le choix des stress à appliquer est toujours complexe mais il sera dans ce cas guidé par le réseau de co-régulation : ce gène a été intégré au réseau car il a été identifié comme co-exprimé avec d'autres dans une série de stress connus. Il sera donc opportun de tester d'abord l'implication de ce gène spécifiquement dans ces stress. Par ailleurs, le rôle de ce gène/protéine dans l'expression du chloroplaste sera mesuré en analysant le transcriptome chloroplastique de son mutant avec tous les outils de transcriptomique que j'ai développés. Si cette protéine n'a pas encore été testée en double hybride, elle le sera. Les données omiques ainsi obtenues seront intégrées dans l'inférence des réseaux et ainsi de suite.

Ce projet, comme une bonne part de mon activité de recherche passée, se situe donc à plusieurs interfaces : une interface thématique entre la réponse au stress chez les plantes et l'étude de l'expression du chloroplaste et une interface disciplinaire entre plusieurs disciplines de la biologie, la

bioinformatique et les statistiques. Ce positionnement, solidement ancré dans mon expertise (et ma communauté) de biologie moléculaire végétale mais en interaction étroite avec d'autres disciplines, ouvre de multiples possibilités toutes aussi passionnantes les unes que les autres !

## Bibliographie

- Barkan, A.** (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* **155**: 1520–1532.
- Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I.** (2012). A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genet.* **8**: e1002910.
- Barrett, C.F., Freudenstein, J. V., Li, J., Mayfield-Jones, D.R., Perez, L., Pires, J.C., and Santos, C.** (2014). Investigating the Path of Plastid Genome Degradation in an Early-Transitional Clade of Heterotrophic Orchids, and Implications for Heterotrophic Angiosperms. *Mol. Biol. Evol.* **31**: 3095–3112.
- Bougoure, J.J., Brundrett, M.C., and Grierson, P.F.** (2010). Carbon and nitrogen supply to the underground orchid, *Rhizanthella gardneri*. *New Phytol.* **186**: 947–956.
- Boussardon, C., Avon, A., Kindgren, P., Bond, C.S., Challenor, M., Lurin, C., and 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., Berthomé, R., Hammani, K., Okuda, K., Shikanai, T., Small, I., and Lurina, C.** (2012). Two interacting proteins are necessary for the editing of the *ndhD-1* site in *Arabidopsis* plastids. *Plant Cell* **24**: 3684–3694.
- Brundrett, M.C. and Tedersoo, L.** (2018). Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol.* **220**: 1108–1115.
- Castandet, B., Hotto, A.M., Fei, Z., and Stern, D.B.** (2013). Strand-specific RNA sequencing uncovers chloroplast ribonuclease functions. *FEBS Lett.* **587**: 3096–3101.
- Chan, K.X., Phua, S.Y., Crisp, P., McQuinn, R., and Pogson, B.J.** (2016). Learning the Languages of the Chloroplast: Retrograde Signaling and beyond. *Annu. Rev. Plant Biol.* **67**: 25–53.
- Chardon, F. et al.** (2020). The consequences of a disruption in cyto-nuclear coadaptation on the molecular response to a nitrate starvation in *arabidopsis*. *Plants* **9**.
- 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., and Leo, P.** (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J.* **56**.
- Chateigner-Boutin, A.-L.A.-L., des Francs-Small, C.C.C.C., Delannoy, E., Kahlau, S., Tanz, S.K.S.K., de Longevialle, A.F.A.F., Fujii, S., and Small, I.** (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J.* **65**: 532–42.
- Degrave, A., Moreau, M., Launay, A., Barny, M.A., Brisset, M.N., Patrit, O., Taconnat, L., Vedel, R., and Fagard, M.** (2013). The bacterial effector *DspA/E* is toxic in *Arabidopsis*

- thaliana and is required for multiplication and survival of fire blight pathogen. *Mol. Plant Pathol.* **14**: 506–517.
- Delannoy, E.** (2012). L'orchidée souterraine, *Rhizanthella gardneri*. *La Garance Voyag.*: 22–27.
- Delannoy, E., Falcon de Longevialle, A., and Francs-Small, C.C. des** (2015). Mitochondrial RNA Transcript Analysis Assay of Arabidopsis Leaf Tissues. *Bio-protocol* **5**: e1620.
- Delannoy, E., Fujii, S., Colas Des Francs-Small, C., Brundrett, M., and Small, I.** (2011). Rampant Gene loss in the underground orchid *Rhizanthella Gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* **28**.
- Delannoy, E., Jalloul, A., Assigbetsé, K., Marmey, P., Geiger, J.P., Lherminier, J., Daniel, J.F., Martinez, C., and Nicole, M.** (2003). Activity of Class III Peroxidases in the Defense of Cotton to Bacterial Blight. *Mol. Plant-Microbe Interact.* **16**.
- Delannoy, E., Ret, M.L., Faivre-Nitschke, E., Estavillo, G.M., Bergdoll, M., Taylor, N.L., Pogson, B.J., Small, I., Lmbault, P., and Gualberto, J.M.** (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNA<sup>Arg</sup>(ACG) and is essential for efficient chloroplast translation. *Plant Cell* **21**.
- Delannoy, E., Stanley, W.A., Bond, C.S., and Small, I.D.** (2007). Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem. Soc. Trans.* **35**.
- Demarsy, E., Courtois, F., Azevedo, J., Buhot, L., and Lerbs-Mache, S.** (2006). Building up of the plastid transcriptional machinery during germination and early plant development. *Plant Physiol.* **142**: 993–1003.
- DePamphilis, C.W. and Palmer, J.D.** (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **348**: 337–339.
- Dreze, M. et al.** (2011). Evidence for Network Evolution in an Arabidopsis Interactome Map. *Science* (80-. ). **333**: 601–607.
- Fochi, V., Chitarra, W., Kohler, A., Voyron, S., Singan, V.R., Lindquist, E.A., Barry, K.W., Girlanda, M., Grigoriev, I. V., Martin, F., Balestrini, R., and Perotto, S.** (2017). Fungal and plant gene expression in the *Tulasnella calospora*–*Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytol.* **213**: 365–379.
- Funk, H.T., Berg, S., Krupinska, K., Maier, U.G., and Krause, K.** (2007). Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* **7**: 45.
- Ghirardo, A., Fochi, V., Lange, B., Witting, M., Schnitzler, J.-P., Perotto, S., and Balestrini, R.** (2020). Metabolomic adjustments in the orchid mycorrhizal fungus *Tulasnella calospora* during symbiosis with *Serapias vomeracea*. *New Phytol.* **228**: 1939–1952.
- González-Fuente, M. et al.** (2020). EffectorK, a comprehensive resource to mine for *Ralstonia*, *Xanthomonas*, and other published effector interactors in the Arabidopsis proteome. *Mol. Plant Pathol.* **21**: 1257–1270.
- Graham, S.W., Lam, V.K.Y., and Merckx, V.S.F.T.** (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* **214**: 48–55.

- Guilcher, M., Liehrmann, A., Seyman, C., Blein, T., Rigail, G., Castandet, B., and Delannoy, E.** (2021). Full Length Transcriptome Highlights the Coordination of Plastid Transcript Processing. *Int. J. Mol. Sci.* **22**.
- Guillaumot, D., Lopez-Obando, M., Baudry, K., Avon, A., Rigail, G., Falcon De Longevialle, A., Broche, B., Takenaka, M., Berthomé, R., De Jaeger, G., Delannoy, E., and Lurin, C.** (2017). Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **114**: 8877–8882.
- Gutmann, B., Royan, S., Schallenberg-Rüdinger, M., Lenz, H., Castleden, I.R., McDowell, R., Vacher, M.A., Tonti-Filippini, J., Bond, C.S., Knoop, V., and Small, I.D.** (2019). The expansion and diversification of pentatricopeptide repeat RNA editing factors in plants. *Mol. Plant.*
- Hamel, L.P., Sekine, K.T., Wallon, T., Sugiwaka, Y., Kobayashi, K., and Moffett, P.** (2016). The chloroplastic protein THF1 interacts with the Coiled-Coil domain of the disease resistance protein N<sup>1</sup> and regulates light-dependent cell death. *Plant Physiol.* **171**: 658–674.
- Hayes, M. and Santibanez, P.** (2020). A plant pentatricopeptide repeat protein with a DYW-deaminase domain is sufficient for catalyzing C-to-U RNA editing in vitro. *J. Biol. Chem.* **295**: 3497–3505.
- van der Heijden, M.G.A., Martin, F.M., Selosse, M.A., and Sanders, I.R.** (2015). Mycorrhizal ecology and evolution: The past, the present, and the future. *New Phytol.* **205**: 1406–1423.
- Hernández-Verdeja, T. and Strand, Å.** (2018). Retrograde signals navigate the path to chloroplast development. *Plant Physiol.* **176**: 967–976.
- Jałalski, M., Minasiewicz, J., Caius, J., May, M., Selosse, M.-A., and Delannoy, E.** (2021). The Genomic Impact of Mycoheterotrophy in Orchids. *Front. Plant Sci.* **0**: 1084.
- Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., and Selosse, M.A.** (2005). Mixotrophy in orchids: Insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytol.* **166**: 639–653.
- Kobayashi, T., Yagi, Y., and Nakamura, T.** (2019). Comprehensive Prediction of Target RNA Editing Sites for PLS-Class PPR Proteins in *Arabidopsis thaliana*. *Plant Cell Physiol.* **60**: 862–874.
- Kobayashi, T., Yagi, Y., Tamura, T., and Nakamura, T.** (2020). Identification of new pentatricopeptide repeat proteins, MREF1 and 2, involved in mitochondrial RNA editing, using computational target RNA prediction. *Plant Biotechnol.* **37**: 111–116.
- Kretschmer, M., Damoo, D., Djamei, A., and Kronstad, J.** (2020). Chloroplasts and plant immunity: Where are the fungal effectors? *Pathogens* **9**.
- Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., De Longevialle, A.F., Otoole, N., Börner, T., Millar, A.H., Small, I.D., and Whelan, J.** (2009). Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of *Arabidopsis thaliana*. *Plant Cell* **21**.
- Lallemant, F. et al.** (2019a). Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids. *Genome Biol Evol* **11**: 2457–2467.

- Lallemand, F., Figura, T., Damesin, C., Fresneau, C., Griveau, C., Fontaine, N., Zeller, B., and Selosse, M.A.** (2019b). Mixotrophic orchids do not use photosynthates for perennial underground organs. *New Phytol.* **221**: 12–17.
- Lallemand, F., Martin-Magniette, M.-L., Gilard, F., Gakière, B., Launay-Avon, A., Delannoy, É., and Selosse, M.-A.** (2019c). In situ transcriptomic and metabolomic study of the loss of photosynthesis in the leaves of mixotrophic plants exploiting fungi. *Plant J.* **98**.
- Law, S.R., Narsai, R., Taylor, N.L., Delannoy, E., Carrie, C., Giraud, E., Harvey Millar, A., Small, I., and Whelan, J.** (2012). Nucleotide and RNA metabolism prime translational initiation in the earliest events of mitochondrial biogenesis during arabidopsis germination. *Plant Physiol.* **158**.
- Lawrence, T.J., Amrine, K.C.H., Swingley, W.D., and Ardell, D.H.** (2019). TRNA functional signatures classify plastids as late-branching cyanobacteria. *BMC Evol. Biol.* **19**.
- Leebens-Mack, J.H. et al.** (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.
- Littlejohn, G.R., Breen, S., Smirnov, N., and Grant, M.** (2020). Chloroplast immunity illuminated. *New Phytol.*
- De Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H., and Small, I.** (2008). The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana*. *Plant J.* **56**.
- Lurin, C. et al.** (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**: 2089–103.
- Majeran, W., Friso, G., Asakura, Y., Qu, X., Huang, M., Ponnala, L., Watkins, K.P., Barkan, A., and van Wijk, K.J.** (2012). Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: A new conceptual framework for nucleoid functions. *Plant Physiol.* **158**: 156–189.
- Malbert, B., Burger, M., Lopez-Obando, M., Baudry, K., Launay-Avon, A., Härtel, B., Verbitskiy, D., Jörg, A., Berthomé, R., Lurin, C., Takenaka, M., and Delannoy, E.** (2020). The analysis of the editing defects in the *dyw2* mutant provides new clues for the prediction of rna targets of arabidopsis e+-class ppr proteins. *Plants* **9**.
- Malbert, B., Rigail, G., Brunaud, V., Lurin, C., and Delannoy, E.** (2018). Bioinformatic analysis of chloroplast gene expression and rna posttranscriptional maturations using RNA sequencing.
- Martinez, C., Baccou, J.-C., Bresson, E., Baissac, Y., Daniel, J.-F., Jalloul, A., Montillet, J.-L., Geiger, J.-P., Assigbetsé, K., and Nicole, M.** (2000). Salicylic Acid Mediated by the Oxidative Burst Is a Key Molecule in Local and Systemic Responses of Cotton Challenged by an Avirulent Race of *Xanthomonas campestris pv malvacearum*. *Plant Physiol.* **122**: 757.
- Medina-Puche, L., Tan, H., Dogra, V., Wu, M., Rosas-Diaz, T., Wang, L., Ding, X., Zhang, D., Fu, X., Kim, C., and Lozano-Duran, R.** (2020). A Defense Pathway Linking Plasma Membrane and Chloroplasts and Co-opted by Pathogens. *Cell* **182**: 1109-1124.e25.
- Menke, W. and Schmid, G.H.** (1976). Cyclic Photophosphorylation in the Mykotrophic Orhid *Neottia nidus-avis*. *Plant Physiol.* **57**: 716–719.

- Merckx, V. and Freudenstein, J. V.** (2010). Evolution of mycoheterotrophy in plants: A phylogenetic perspective. *New Phytol.* **185**: 605–609.
- von Mohl, H.** (1837). Ueber die Vermehrung der Pflanzen-Zellen durch Teilung. *Flora*.
- Molina, J. et al.** (2014). Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol. Biol. Evol.* **31**: 793–803.
- Mukhtar, M.S. et al.** (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* (80-. ). **333**: 596–601.
- Nakayama, T. and Archibald, J.M.** (2012). Evolving a photosynthetic organelle. *BMC Biol.* **10**.
- Narsai, R., Law, S.R., Carrie, C., Xu, L., and Whelan, J.** (2011). In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in *Arabidopsis*. *Plant Physiol.* **157**: 1342–1362.
- Okuda, K., Chateigner-Boutin, A.-L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T.** (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* **21**.
- Pellicer, J. and Leitch, I.J.** (2020). The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytol.* **226**: 301–305.
- Perez-Lamarque, B., Selosse, M.A., Öpik, M., Morlon, H., and Martos, F.** (2020). Cheating in arbuscular mycorrhizal mutualism: a network and phylogenetic analysis of mycoheterotrophy. *New Phytol.*
- Pfeifhofer, H.W.** (1989). Evidence for Chlorophyll b and Lack of Lutein in *Neottia nidus-avis* Plastids. *Biochem. und Physiol. der Pflanz.* **184**: 55–61.
- Pulido, P., Zagari, N., Manavski, N., Gawronski, P., Matthes, A., Scharff, L.B., Meurer, J., and Leister, D.** (2018). CHLOROPLAST RIBOSOME ASSOCIATED Supports Translation under Stress and Interacts with the Ribosomal 30S Subunit. *Plant Physiol.* **177**: 1539–1554.
- Qi, Y., Armbruster, U., Schmitz-Linneweber, C., Delannoy, E., De Longevialle, A.F., Rühle, T., Small, I., Jahns, P., and Leister, D.** (2012). *Arabidopsis* CSP41 proteins form multimeric complexes that bind and stabilize distinct plastid transcripts. *J. Exp. Bot.* **63**.
- Rasmussen, H.N.** (1995). *Terrestrial orchids : from seed to mycotrophic plant* (Cambridge University Press).
- Rich, M.K., Nouri, E., Courty, P.E., and Reinhardt, D.** (2017). Diet of Arbuscular Mycorrhizal Fungi: Bread and Butter? *Trends Plant Sci.* **22**: 652–660.
- Rigall, G. et al.** (2018). Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief. Bioinform.* **19**.
- Robles, P., Micol, J.L., and Quesada, V.** (2015). Mutations in the plant-conserved MTERF9 alter chloroplast gene expression, development and tolerance to abiotic stress in *Arabidopsis thaliana*. *Physiol. Plant.* **154**: 297–313.

- Robles, P. and Quesada, V.** (2019). Transcriptional and post-transcriptional regulation of organellar gene expression (Oge) and its roles in plant salt tolerance. *Int. J. Mol. Sci.* **20**.
- Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A.** (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Comput. Biol.* **13**: e1005752.
- Roux, F., Mary-Huard, T., Barillot, E., Wenes, E., Botran, L., Durand, S., Villoutreix, R., Martin-Magniette, M.L., Camillerif, C., and Budar, F.** (2016). Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 3687–3692.
- Roy, M., Gonneau, C., Rocheteau, A., Berveiller, D., Thomas, J.-C., Damesin, C., and Selosse, M.-A.** (2013). Why do mixotrophic plants stay green? A comparison between green and achlorophyllous orchid individuals in situ. *Ecol. Monogr.* **83**: 95–117.
- Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C.** (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* **581**: 4132–4138.
- Sánchez-Baracaldo, P., Raven, J.A., Pisani, D., and Knoll, A.H.** (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E7737–E7745.
- Schelkunov, M.I., Penin, A.A., and Logacheva, M.D.** (2018). RNA-seq highlights parallel and contrasting patterns in the evolution of the nuclear genome of fully mycoheterotrophic plants. *BMC Genomics* **19**.
- Schimper, A.F.W.** (1883). Ueber die Entwicklung der Chlorophyllkörner und Farbkörper. *Bot. Zeitung* **41**: 105–160.
- Sellosse, M.-A. and Roy, M.** (2009). Green plants that feed on fungi: facts and questions about mixotrophy. *Trends Plant Sci.* **14**: 64–70.
- Small, I.D. and Peeters, N.** (2000). The PPR motif - A TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**: 45–47.
- Solymosi, K., Lethin, J., and Aronsson, H.** (2018). Diversity and plasticity of plastids in land plants. In *Methods in Molecular Biology* (Humana Press Inc.), pp. 55–72.
- Sormani, R., Delannoy, E., Lageix, S., Bitton, F., Lanet, E., Saez-Vasquez, J., Deragon, J.M., Renou, J.P., and Robaglia, C.** (2011). Sublethal cadmium intoxication in *Arabidopsis thaliana* impacts translation at multiple levels. *Plant Cell Physiol.* **52**.
- De Souza, A., Wang, J.Z., and Dehesh, K.** (2017). Retrograde Signals: Integrators of Interorganellar Communication and Orchestrators of Plant Development. *Annu. Rev. Plant Biol.* **68**: 85–108.
- Strasburger, E.** (1884). *Das botanische Praktikum* G. Fischer, ed (Jena).
- Těšitel, J., Těšitelová, T., Minasiewicz, J., and Selosse, M.A.** (2018). Mixotrophy in Land Plants: Why To Stay Green? *Trends Plant Sci.* **23**: 656–659.
- Tillich, M., Hardel, S.L.S.L.S.L., Kupsch, C., Armbruster, U., Delannoy, E., Gualberto, J.M.J.M., Lehwark, P., Leister, D., Small, I.D.I.D., and Schmitz-Linneweber, C.** (2009). Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 6002–6007.

- Tunyasuvunakool, K. et al.** (2021). Highly accurate protein structure prediction for the human proteome. *Nat.* 2021: 1–9.
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., and Provart, N.J.** (2009). Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant, Cell Environ.* **32**: 1633–1651.
- Vogel, A. et al.** (2018). Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat. Commun.* **9**: 2515.
- Wasserstein, R.L. and Lazar, N.A.** (2016). The ASA’s Statement on p-Values: Context, Process, and Purpose. *Am. Stat.* **70**: 129–133.
- Wickett, N.J., Honaas, L.A., Wafula, E.K., Das, M., Huang, K., Wu, B., Landherr, L., Timko, M.P., Yoder, J., Westwood, J.H., and Depamphilis, C.W.** (2011). Transcriptomes of the parasitic plant family orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr. Biol.* **21**: 2098–2104.
- Yang, Z. et al.** (2019). Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nat. Plants* **5**: 991–1001.
- Yap, A., Kindgren, P., Colas des Francs-Small, C., Kazama, T., Tanz, S.K., Toriyama, K., and 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.
- Yoshida, S. et al.** (2019). Genome Sequence of *Striga asiatica* Provides Insight into the Evolution of Plant Parasitism. *Curr. Biol.* **29**: 3041-3052.e4.
- Yuan, Y. et al.** (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* **9**.
- Zhang, J.X., Wang, C., Yang, C.Y., Wang, J.Y., Chen, L., Bao, X.M., Zhao, Y.X., Zhang, H., and Liu, J.** (2010). The role of *Arabidopsis* AtFes1A in cytosolic Hsp70 stability and abiotic stress tolerance. *Plant J.* **62**: 539–548.
- Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.-L., Delannoy, E., Hammani, K., Small, I., and Huang, J.** (2009). The *Arabidopsis* gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant J.* **58**.

# Curriculum vitae

## Etienne Delannoy

Né le 6 juillet 1976 à Lille

Adresse : Equipe Expression génétique des organites  
Institut de Sciences des Plantes Paris-Saclay  
Batiment 630, rue Noetzelin  
91190 Gif-sur-Yvette

Mél : etienne.delannoy@inrae.fr

## Situation actuelle :

Chargé de recherche INRAE

Coordinateur de l'équipe [Expression génétique des organites](#)

Responsable scientifique de la plateforme [SPOmics transcriptome](#) de l'IPS2

Coordinateur de la plateforme [SPOmics](#)

Coordinateur du groupe de travail Bio-informatique de l'EUR Saclay Plant Sciences

Co-animateur du Réseau [NetBio](#)

Responsable de la commission scientifique de la [Société Française d'Orchidophilie](#)

## Travaux de recherche

**Depuis fin 2015** : chercheur à l'IPS2 dans l'équipe Expression génétique des organites. Responsable scientifique de la plateforme transcriptomique de l'IPS2

**Décembre 2010-2015** : recrutement en tant que chargé de recherche INRA (CR1) à l'Unité de Recherche en Génomique Végétale, Evry dans l'équipe Génomique Fonctionnelle d'Arabidopsis. Conseiller scientifique de la plateforme transcriptomique de l'URGV.

**2009-2010** : CEA Cadarache (France): Post-doc. Laboratoire de Biologie de Développement des Plantes/ IBEB. Depuis 12 octobre 2009. Directeur de laboratoire: Laurent Nussaume  
Analyse des voies de signalisation du phosphate chez *A. thaliana* par la recherche et l'analyse de molécules chimiques perturbant les réponses à la carence en phosphate. Méta-analyse des paramètres contrôlant ces réponses.

**2006-2009** : ARC Centre of Excellence for Plant Energy Biology (Australie): Assistant professeur de recherche. 28 Mars 2008 - 15 Aout 2009. Directeur de laboratoire: Ian Small  
Expression génétique dans les organites d'*Arabidopsis thaliana*.

**2004-2006** : CEA-Université Aix-Marseille 2 (France): Post-doc, Laboratoire de Génétique et Biophysique des Plantes, 12 Janvier 2004 - 28 février 2006. Directeur de laboratoire: Christophe Robaglia  
Régulation traductionnelle en réponse au stress et rôle des miRNA chez *Arabidopsis thaliana*.

**1999-2003** : Université Montpellier 2 (France): Etudiant en DEA puis thèse, IRD Montpellier, GeneTrop, UR Résistance des Plantes. Directeur de thèse : Michel Nicole

Etude des peroxydases de classe III du cotonnier dans ses interactions avec *Xanthomonas campestris* pv. *malvacearum*.

**Eté 1998:** Kansas State University (USA): Department of Plant Pathology. Isolation of trans-acting factors that regulate expression of Rice Pox 22.3 during race specific resistant and susceptible interactions with *Xanthomonas oryzae* pv. *oryzae*. Encadrant: Jan E. LEACH.

**Eté 1997 :** ENS Lyon (France): Reproduction et Développement des Plantes. Analyse du rôle des aquaporines dans la germination des grains de pollen. Encadrant: Thierry Gaude

## **Etudes supérieures**

**1999-2003** Doctorat de Ressources Phytogénétiques et Interactions Biologiques de l'école doctorale de Biologie Intégrative de l'université Montpellier 2. Directeur de thèse: M. NICOLE

**1996-1999** Master de Biologie cellulaire et biochimie, ENS-Universités Paris 6, 7 and 11,

**1998-1999** DEA de Biologie, Diversité et Adaptation des Plantes Cultivées option Phytopathologie, Université Paris 6

**1997-1998** Maitrise de Biologie Cellulaire – Physiologie ENS-Universités Paris 6, 7 and 11 (modules de Génétique *in vitro* (Paris 6) et de Génétique de Développement des Plantes (Paris 11))

**1996-1997** Licence de Biologie Cellulaire – Physiologie ENS-Universités Paris 6, 7 and 11 (modules de Physiologie végétale et d'Ecologie- Evolution)

**1996** Admission à l'École Normale Supérieure de Paris

## **Enseignement:**

**Depuis 2016 :** Co-animateur avec Marie-Laure Martin-Magniette de l'École Chercheur et de l'École d'été "De l'expression des gènes aux réseaux pour l'identification de gènes candidats" (Paris 2016, Paris 2017, Grenoble 2016, Toulouse 2017, Québec 2018, Saint Quentin 2019, Avignon 2020, Montpellier 2021)

**Depuis 2015 :** enseignements autour de l'analyse du transcriptome et des réseaux de gènes

M1 Systems and Synthetic Biology (Université Paris-Saclay)

M1 Biologie Intégrative et Physiologie Végétale (Université Paris-Saclay)

M2 Sciences du Végétal (Université Paris-Saclay)

M2 Ingénierie de Plate-Forme en Biologie (Université de Paris)

**2001-2003 :** Moniteur normalien, Université Montpellier 2. Enseignement en TP et TD de Physiologie Végétale en DEUG.

## **Encadrement de thèses:**

- Bastien MALBERT: Subversion of plant organelle gene expression by bacterial PPR proteins. (2015-2018).

- Rim ZAAG : Enrichissement de profils transcriptomiques par intégration de données hétérogènes : annotation fonctionnelle de gènes d'*Arabidopsis thaliana* impliqués dans la réponse aux stress (2014-2016). Co-encadrement avec Marie-Laure Martin-Magniette.

# Liste des publications

## 1) Publications dans des journaux à comité de lecture.

Le nom des étudiants encadrés sont soulignés. \* correspondance.

- Guilcher M, Liehrmann A, Seyman C, Blein T, Rigail G, Castandet B, **Delannoy E**. Full Length Transcriptome Highlights the Coordination of Plastid Transcript Processing. *Int J Mol Sci*. 2021 Oct 19;22(20):11297. doi: 10.3390/ijms222011297.
- Baudry K, **Delannoy E**, Colas des Francs-Small C. Analysis of the Plant Mitochondrial Transcriptome. *Methods Mol Biol*. 2022; 2363:235-262. doi: 10.1007/978-1-0716-1653-6\_17.
- Jąkowski M, Minasiwicz J, Caius J, May M, Selosse M-A, **Delannoy E\***. The Genomic Impact of Mycoheterotrophy in Orchids. *Front Plant Sci*. 2021 Jun 9; 12:1084. doi:10.3389/fpls.2021.632033
- Deveaux Y, Conde E Silva N, Manicacci D, Le Guilloux M, Brunaud V, Belcram H, Joets J, Soubigou-Taconnat L, **Delannoy E**, Corti H, Balzergue S, Caius J, Nadot S, Damerval C. Transcriptome Analysis Reveals Putative Target Genes of APETALA3-3 During Early Floral Development in *Nigella damascena* L. *Front Plant Sci*. 2021 Jun 4;12:660803. doi: 10.3389/fpls.2021.660803.
- Sakai, K., Citerne, S., Antelme S., Le Bris P., Daniel S., Boudier A., D'Orlando A., Cartwright A., Tellier F., Pateyron S., **Delannoy E.**, Laudencia-Chingcuanco D., Mouille G., Palauqui J.C., Vogel J., Sibout R. BdERECTA controls vasculature patterning and phloem-xylem organization in *Brachypodium distachyon*. *BMC Plant Biol* 21, 196 (2021)
- Blein T, Balzergue C, Roulé T, Gabriel M, Scalisi L, François T, Sorin C, Christ A, Godon C, **Delannoy E**, Martin-Magniette ML, Nussaume L, Hartmann C, Gautheret D, Desnos T, Crespi M. Landscape of the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to Phosphate Starvation. *Plant Physiol*. 2020 Jul;183(3):1058-1072. doi: 10.1104/pp.20.00446.
- Chardon, F.; Cuff, G.; **Delannoy, E.**; Aubé, F.; Lornac, A.; Bedu, M.; Gilard, F.; Pateyron, S.; Rogniaux, H.; Gargaros, A.; Mireau, H.; Rajjou, L.; Martin-Magniette, M.-L.; Budar, F. The Consequences of a Disruption in Cyto-Nuclear Coadaptation on the Molecular Response to a Nitrate Starvation in Arabidopsis. *Plants* 2020, 9(5), 573. doi : 10.3390/plants9050573
- Malbert B, Burger M, Lopez-Obando M, Baudry K, Launay-Avon A, Hartel B, Verbitskiy D, Jorg A, Berthome R, Lurin L, Takenaka M, **Delannoy E\***. The Analysis of the Editing Defects in the *dyw2* Mutant Provides New Clues for the Prediction of RNA Targets of Arabidopsis E+-Class PPR Proteins. *Plants*. 2020;9(2) :280. doi : 10.3390/plants9020280
- Cuello C, Baldy A, Brunaud V, Joets J, **Delannoy E**, Jacquemot MP, Botran L, Griveau Y, Guichard C, Soubigou-Taconnat L, Martin-Magniette ML, Leroy P, Méchin V, Reymond M, Coursol S. A systems biology approach uncovers a gene co-expression network associated with cell wall degradability in maize. *PLoS One*. 2019 Dec 31;14(12):e0227011. doi: 10.1371/journal.pone.0227011. eCollection 2019.
- Lallemand F, Logacheva M, Le Clainche I, Bérard A, Zheleznaia E, May M, Jakalski M, **Delannoy E**, Le Paslier MC, Selosse MA. Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids. *Genome Biol Evol*. 2019 Sep 1;11(9):2457-2467. doi: 10.1093/gbe/evz170.
- Niazi AK, **Delannoy E**, Iqbal RK, Mileschina D, Val R, Gabryelska M, Wyszko E, Soubigou-Taconnat L, Szymanski M, Barciszewski J, Weber-Lotfi F, Gualberto JM, Dietrich A. Mitochondrial Transcriptome Control and Intercompartment Cross-Talk During Plant Development. *Cells*. 2019 Jun 13;8(6). pii: E583. doi: 10.3390/cells8060583

- Lallemand F, Martin-Magniette ML, Gilard F, Gakière B, Launay-Avon A, **Delannoy E\***, Selosse MA. In situ transcriptomic and metabolomic study of the loss of photosynthesis in the leaves of mixotrophic plants exploiting fungi. *Plant J.* 2019 Feb 8. doi: 10.1111/tbj.14276
- Damerval C, Citerne H, Conde E Silva N, Deveaux Y, **Delannoy E**, Joets J, Simonnet F, Staedler Y, Schönenberger J, Yansouni J, Le Guilloux M, Sauquet H, Nadot S. Unraveling the Developmental and Genetic Mechanisms Underpinning Floral Architecture in Proteaceae. *Front Plant Sci.* 2019 Jan 25;10:18.
- Sakai K, Taconnat L, Borrega N, Yansouni J, Brunaud V, Paysant-Le Roux C, **Delannoy E**, Martin Magniette ML, Lepiniec L, Faure JD, Balzergue S, Dubreucq B. Combining laser-assisted microdissection (LAM) and RNA-seq allows to perform a comprehensive transcriptomic analysis of epidermal cells of Arabidopsis embryo. *Plant Methods.* 2018 Feb 3;14:10. doi: 10.1186/s13007-018-0275-x
- Lloyd A, Blary A, Charif D, Charpentier C, Tran J, Balzergue S, **Delannoy E**, Rigail G, Jenczewski E. Homoeologous exchanges cause extensive dosage dependent gene expression changes in an allopolyploid crop. *New phytol.* 2018 Jan;217(1):367-377. doi: 10.1111/nph.14836
- Guillaumot D, Lopez-Obando M, Baudry, K, Avon A, Rigail G, Falcon de Longevialle A, Broche B, Takenaka M, Berthomé R, De Jaeger G, **Delannoy E**, Lurin C. Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc Natl Acad Sci U S A.* 2017 Aug 15;114(33):8877-8882.
- Balzergue C, Dartevelle T, Godon C, Laugier E, Meisrimler C, Teulon JM, Creff A, Bissler M, Brouchoud C, Hagège A, Müller J, Chiarenza S, Javot H, Becuwe-Linka N, David P, Péret B, **Delannoy E**, Thibaud MC, Armengaud J, Abel S, Pellequer JL, Nussaume L, Desnos T. Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. *Nat Commun.* 2017 May 15;8:15300. doi: 10.1038/ncomms15300
- Citerne HL, Reyes E, Le Guilloux M, **Delannoy E**, Simonnet F, Sauquet H, Weston PH, Nadot S, Damerval C. Characterization of CYCLOIDEA-like genes in Proteaceae, a basal eudicot family with multiple shifts in floral symmetry. *Ann Bot.* 2017 Feb;119(3):367-378.
- Pedroza-García JA, Mazubert C, Del Olmo I, Bourge M, Domenichini S, Bounon R, Tariq Z, **Delannoy E**, Piñeiro M, Jarillo JA, Bergounioux C, Benhamed M, Raynaud C. Function of the Plant DNA Polymerase Epsilon in Replicative Stress Sensing, a Genetic Analysis. *Plant Physiol.* 2017 Mar;173(3):1735-1749.
- Rigail G, Balzergue S, Brunaud V, Blondet E, Rau A, Rogier O, Caius J, Maugis-Rabusseau C, Soubigou-Taconnat L, Aubourg S, Lurin C, Martin-Magniette ML, **Delannoy E\***. Synthetic datasets for the identification of key ingredients for RNA-seq differential analysis. *Brief Bioinform.* 2016 bbw092
- Pedroza-Garcia JA, Domenichini S, Mazubert C, Bourge M, White C, Hudik E, Bounon R, Tariq Z, **Delannoy E**, Del Olmo I, Piñeiro M, Jarillo JA, Bergounioux C, Benhamed M, Raynaud C. Role of the Polymerase  $\epsilon$  sub-unit DPB2 in DNA replication, cell cycle regulation and DNA damage response in Arabidopsis. *Nucleic Acids Res.* 2016 May 18. pii: gkw449
- Kanno S, Arrighi JF, Chiarenza S, Bayle V, Berthomé R, Péret B, Javot H, **Delannoy E**, Marin E, Nakanishi TM, Thibaud MC, Nussaume L. A novel role for the root cap in phosphate uptake and homeostasis. *Elife.* 2016 Apr 6;5:e14577.
- Bonnot C, Pinson B, Clément M, Bernillon S, Chiarenza S, Kanno S, Kobayashi N, **Delannoy E**, Nakanishi TM, Nussaume L, Desnos T. A chemical genetic strategy identifies the PHOSTIN, a synthetic molecule that triggers phosphate starvation responses in Arabidopsis thaliana. *New Phytol.* 2016 Jan;209(1):161-76.

- [Zaag R](#), Tamby JP, Guichard C, Tariq Z, Rigail G, **Delannoy E**, Renou JP, Balzergue S, Mary-Huard T, Aubourg S, Martin-Magniette ML, Brunaud V. GEM2Net: from gene expression modeling to -omics networks, a new CATdb module to investigate Arabidopsis thaliana genes involved in stress response. *Nucleic Acids Res.* 2015 Jan 28;43 (Database issue):D1010-7.
- Arnaud C, Clément M, Thibaud MC, Javot H, Chiarenza S, **Delannoy E**, Revol J, Soreau P, Balzergue S, Block MA, Maréchal E, Desnos T, Nussaume L. Identification of phosphatin, a drug alleviating phosphate starvation responses in Arabidopsis. *Plant Physiol.* 2014 Nov;166(3):1479-91.
- Law SR, Narsai R, Taylor NL, **Delannoy E**, Carrie C, Giraud E, Millar AH, Small ID, Whelan J. (2012) Nucleotide and RNA metabolism prime translational initiation in the earliest events of mitochondrial biogenesis during Arabidopsis thaliana germination. *Plant Physiol.* 2012 Apr;158(4):1610-27
- Qi Y, Armbruster U, Schmitz-Linneweber C, **Delannoy E**, de Longevialle AF, Rühle T, Small I, Jahns P, Leister D. (2012) Arabidopsis CSP41 proteins form multimeric complexes that bind and stabilize distinct plastid transcripts. *J Exp Bot* 63(3):1251-70
- **Delannoy E**, Fujii S, Colas des Francs C, Brundrett M, Small ID (2011) Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol* 28(7):2077-2086
- Sormani R, **Delannoy E**, Lageix S, Bitton F, Lanet E, Saez-Vasquez J, Deragon JM, Renou JP, Robaglia C. (2011) Sub-Lethal Cadmium Intoxication In Arabidopsis thaliana Impacts Translation At Multiple Levels. *Plant Cell Physiol* 52(2):436-447
- Chateigner-Boutin AL, Colas Des Francs-Small C, **Delannoy E**, Kahlau S, Tanz SK, de Longevialle AF, Fujii S, Small ID (2011) OTP70 is a pentatricopeptide protein of the E subgroup involved in splicing of the plastid transcript rpoC1 . *Plant Journal* 65(4):532-542
- Albrecht V, Simková K, Carrie C, **Delannoy E**, Giraud E, Whelan J, Small ID, Apel K, Badger MR, Pogson BJ. (2010) The cytoskeleton and the peroxisomal-targeted snowy cotyledon3 protein are required for chloroplast development in Arabidopsis. *The Plant Cell* 22(10):3423-38
- Kühn K, Richter U, Meyer EH, **Delannoy E**, de Longevialle AF, O'Toole N, Börner T, Millar AH, Small ID, Whelan J. (2009) Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of Arabidopsis thaliana. *The Plant Cell* 21(9):2762-79
- Meyer EH, Tomaz T, Carroll AJ, Estavillo G, **Delannoy E**, Tanz SK, Small ID, Pogson BJ, Millar AH. (2009) Remodeled respiration in ndusf4 with low phosphorylation efficiency suppresses Arabidopsis germination and growth and alters control of metabolism at night. *Plant Physiology* 151(2):603-19
- **Delannoy E**, Le Ret M, Faivre-Nitschke SE, Estavillo G, Bergdoll M, Taylor NL, Pogson B, Small ID, Imbault P, Gualberto JM. (2009) Arabidopsis tRNA adenosine deaminase (TADA) is responsible for editing the wobble nucleotide of chloroplast tRNA<sup>Arg</sup>(ACG), a modification essential for efficient chloroplast translation. *The Plant Cell* 21(7):2058-71
- [Lanet E](#), **Delannoy E**, Sormani R, Floris M, Brodersen P, Crete P, Voinnet O, Robaglia C. (2009) Biochemical evidence for translational repression by plant microRNAs. *The Plant Cell* 21(6):1762-8
- Tillich M, Hardel SL, Kupsch C, Armbruster U, **Delannoy E**, Gualberto JM, Lehwark P, Leister D, Small ID, Schmitz-Linneweber C. (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proceedings of the National Academy of Science* 106(14):6002-6007
- Okuda K, Chateigner-Boutin AL, Nakamura T, **Delannoy E**, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small ID, Shikanai T. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *The Plant Cell* 21(1):146-56.

- Zhou W, Cheng Y, Yap A, Chateigner-Boutin AL, **Delannoy E**, Hammani K, Small ID, Huang J. (2008) The Arabidopsis gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant Journal* 58(1):82-96
- Garmier M, Carroll AJ, **Delannoy E**, Vallet C, Day DA, Small ID, Millar AH.(2008) Complex I dysfunction redirects cellular and mitochondrial metabolism in Arabidopsis. *Plant Physiology* 148(3):1324-41.
- Chateigner-Boutin AL, Ramos-Vega M, Guevara-García A, Andrés C, de la Luz Gutiérrez-Nava M, Cantero A, **Delannoy E**, Jiménez LF, Lurin C, Small ID, León P. (2008) CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. *Plant Journal* 56(4):590-602.
- de Longevialle AF, Hendrickson L, Taylor NL, **Delannoy E**, Lurin C, Badger M, Millar AH, Small ID. (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. *Plant Journal* 56(1):157-168.
- **Delannoy E**, Stanley WA, Bond CS, Small ID. (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochemical Society Transactions* 35(Pt 6):1643-1647.
- **Delannoy E.**, Marmey P., Jalloul A., Etienne H., Nicole M. (2006) Molecular analysis of some cotton peroxidases. *Journal of Cotton Science* 9:229-236
- Voloudakis A. E., Marmey P., **Delannoy E.**, Jalloul A, Martinez C and Nicole M. (2006) Molecular cloning and characterization of *Gossypium hirsutum* superoxide dismutase genes during cotton-*Xanthomonas campestris* pv. *malvacearum* interaction. *Physiological and Molecular Plant Pathology* 68 (4-6): 119-127
- **Delannoy E.**, Jalloul A., Marmey P., Daniel J.F. , Montillet J.L. , Lyon B. , Essenberg M. and Nicole M. (2005) The hypersensitive response of cotton to *Xanthomonas campestris* pathovar *malvacearum*. *Annual Review of Phytopathology*. 43: 2.1-2.20
- **Delannoy E**, Marmey P, Penel C, Nicole M. (2004) Les peroxydases vegetales de classe III *Acta Botanica Gallica*, 151 (4), 353-380.
- **Delannoy E.**, Daniel J.F., Assigbetsé K., Marmey P., Geiger J.P., Lherminier J., Martinez C., Nicole, M. (2003) Class III peroxidase activity during resistance of cotton to bacterial blight. *Mol Plant-Microbe Interact* 16 (11) : 1030-1038.
- Jalloul A, Montillet JL, Assigbetse K, Agnel JP, **Delannoy E**, Triantaphylides C, Daniel JF, Marmey P, Geiger JP, Nicole M. (2002) Lipid peroxidation in cotton: *Xanthomonas* interactions and the role of lipoxygenases during the hypersensitive reaction. *Plant J*. 32 : 1-12

## 2) Chapitre d'ouvrage

- Malbert B, Rigai G, Brunaud V, Lurin C, **Delannoy E\***. Bioinformatic Analysis of Chloroplast Gene Expression and RNA Posttranscriptional Maturations Using RNA Sequencing. *Methods Mol Biol*. 2018;1829:279-294. doi:10.1007/978-1-4939-8654-5\_19.

-

## 3) Articles dans des journaux grand public

- Delannoy E. (2021) Le feu et les orchidées d'Australie occidentale : amis ou ennemis ? *L'Orchidophile* 52 (229), 131-136
- Nicole M. & Delannoy E. (2020) L'immunité des plantes. *L'Orchidophile* 51 (225), 123-131
- Delannoy E. (2015). *Rhizanthella gardneri*, l'orchidée souterraine. *L'Orchidophile*, 46 (205), 161-168.

- Delannoy E. (2016). Focus sur *Rhizanthella gardneri* R.S. Rodgers : l'orchidée souterraine australienne. Journal de Botanique (73), 85-90.
- Delannoy E. (2012). L'orchidée souterraine, *Rhizanthella gardneri*. La Garance Voyageuse (100).

#### **4) Intervention dans les médias**

- Interview pour l'émission télévisée « Silence ça pousse » diffusée en mai 2014.