



HAL
open science

Maturation des ARN des organites de plante

Benoît Castandet

► **To cite this version:**

Benoît Castandet. Maturation des ARN des organites de plante. Sciences du Vivant [q-bio]. Université Paris, 2022. tel-04716928v2

HAL Id: tel-04716928

<https://hal.inrae.fr/tel-04716928v2>

Submitted on 25 Oct 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.

Université de Paris

Document de synthèse en vue de l'obtention d'une

Habilitation à Diriger des Recherches

Mention : Sciences du Vivant

Par Benoît Castandet

Maître de conférences de l'Université de Paris

Institut des Sciences des Plantes-Paris Saclay (IPS2)
UMR CNRS-INRAE-Université Paris Saclay, Université
Evry-Val d'Essonne, Université de Paris

Sommaire

SOMMAIRE	2
SYNTHESE DES ACTIVITES DE RECHERCHE	3
Les années Bordelaises	4
Premiers pas	4
L'édition des ARN	5
Mais pourquoi éditer ?	7
L'édition dans les introns est nécessaire à l'épissage	10
Les différents évènements de maturation des ARN sont connectés	13
Que faire ?	16
Heureux qui comme Ulysse	18
La découverte des approches transcriptomiques	18
RNA-Seq et édition	20
Stress et métabolisme des ARN chloroplastiques	22
Les ribonucléases au cœur de l'expression du génome chloroplastique	26
RNA-Seq et ribonucléases	27
Il est temps de rentrer	32
Retour en France	33
Le Terminome-Seq	33
Terminome-Seq avec des mutants de ribonucléases	35
Bilan	39
PROJET DE RECHERCHE	41
Etude du transcriptome chloroplastique	42
La RNase J et la production du transcriptome mature	44
Etude structure-fonction de la RNase J	46
Identification des partenaires de la RNase J	47
Conclusion	49
BIBLIOGRAPHIE	50

Synthèse des activités de recherche

Les années Bordelaises

Premiers pas

Puisqu'il faut bien commencer, trouver un point de départ à cette introspection scientifique qu'est l'HDR, j'ai décidé de débiter en 2006. Non pas que les années précédentes aient été sans importance, mais 2006 marque un tournant. Avec du recul, peut-être que la découverte des introns codants (un oxymore jusque-là pour moi) dans le cours de génétique moléculaire sur les mutants *petites* de levure a eu un impact non négligeable sur le devenir du futur enseignant chercheur que j'étais. Plus sûrement, c'est lors du stage de Master 1 dans l'équipe de Gérard Barroso (UMR 5097, Université de Bordeaux) que mon inclination pour les organites à génomes s'est développée. En plus de la thématique propre du stage, le site d'intégration d'un plasmide linéaire dans le génome mitochondrial du champignon basidiomycète *Agrocybe aegerita*, j'ai pu découvrir l'incroyable richesse des banques de données bibliographiques. Au milieu des heures passées à naviguer de mots clés en mots clés, un peu comme un enfant s'amusant à feuilleter au hasard les pages d'un dictionnaire, je me suis finalement forgé une certitude sur le genre de thématiques qui me séduisait. La lecture des articles de 1999 de Gray, Lang et Burger sur l'évolution du génome mitochondrial et l'origine des eucaryotes reste pour moi un moment fort (Gray *et al.*, 1999; Lang *et al.*, 1999), tout comme les dissertations de Forterre sur l'origine des virus, des domaines du vivant ou LUCA (Forterre, 2006; Forterre, 2005; Forterre *et al.*, 2005). Il fallait que je travaille sur quelque chose de singulier et les organites à génomes sont des champions de l'excentricité.

Ce stage a bien sûr aussi été le moment de ma découverte d'une pratique de la biochimie et de la biologie moléculaire qui pourrait paraître archaïque à côté des standards actuels mais qui était incroyablement formatrice. Surtout, j'ai eu la chance d'identifier un fragment d'ADN de 137 nucléotides qui a permis de montrer que l'intégration du plasmide linéaire dans le génome mitochondrial s'accompagnait de la duplication du gène *nad4*. Bien que les deux gènes *nad4* demeurent intacts (identité stricte de séquence), on observe une érosion des séquences plasmidiques intégrées. Cette érosion résulterait d'événements de mutations ponctuelles associées à des recombinaisons conduisant ainsi à la perte de fonction du gène *polB* (qui code pour l'ADN polymérase permettant la réplication du plasmide) intégré. Enfin, j'ai également pu découvrir le processus de publication scientifique de l'intérieur puisque j'ai été cosignataire du manuscrit décrivant les résultats (Ferandon *et al.*, 2008).

Gérard Barroso a ensuite été le tuteur de mon monitorat d'initiation à l'enseignement supérieur et l'essentiel de mon activité d'enseignement lors de ma thèse s'est déroulé au sein du master professionnel Biotechnologie des Champignons dont il est le responsable. J'ai encadré des travaux personnels au cours desquels les étudiants découvraient les techniques de base de microbiologie et biologie moléculaire en manipulant des échantillons fongiques d'intérêt pour le laboratoire. C'est lors de ce monitorat qu'une deuxième certitude est apparue, l'enseignement j'aimais ça et il est indissociable de l'activité de recherche. Par la recherche, nous créons des connaissances qui ne sont rien si nous ne les partageons pas. Parce qu'il nous force à sortir de l'hyperspécialisation, l'enseignement, lui, nous permet de rendre la recherche plus intelligible. Ils sont les deux faces d'une même activité plutôt que la simple addition de métiers différents.

L'édition des ARN

Lors de la période 2007-2010 mon travail de recherche s'est déroulé dans le cadre du stage de master 2 puis de celui de la thèse, au sein de l'équipe d'Alejandro Araya dans l'UMR CNRS 5234 à l'université de Bordeaux. Je voulais travailler sur quelque chose de singulier ou d'excentrique, l'édition des ARN semblait avoir toutes les qualités requises. Elle désigne tous les événements qui changent la séquence d'un ARN comparativement à celle du gène correspondant, à l'exception des événements de polyadénylation, capping et d'épissage (Gray and Covello, 1993). L'édition permet donc la modification sur le transcrit de l'information codée par le gène. Suite à sa découverte dans les mitochondries de plantes (Gualberto *et al.*, 1989; Covello and Gray, 1989; Hiesel *et al.*, 1989) il a même été suggéré dans un éditorial du *Journal of Molecular Evolution* qu'elle remettait en cause le « dogme central de la biologie moléculaire » (Jukes, 1990). Il n'est en effet plus possible de déduire directement la séquence d'une protéine à partir de celle de l'ADN, il faut maintenant analyser le transcrit correspondant. Mais peut-on pour autant vraiment conclure que l'édition rend le dogme caduc ? En fait, l'édition pose plus de difficultés aux versions popularisées du dogme (notamment par Watson (Watson, 1965)) qu'à sa formulation originale par Crick (Crick, 1958; Crick, 1970). Pour lui, comme pour Jacob et Monod (Jacob and Monod, 1961), le fondement principal du dogme est le sens de l'information. Elle va du gène vers la protéine, jamais dans

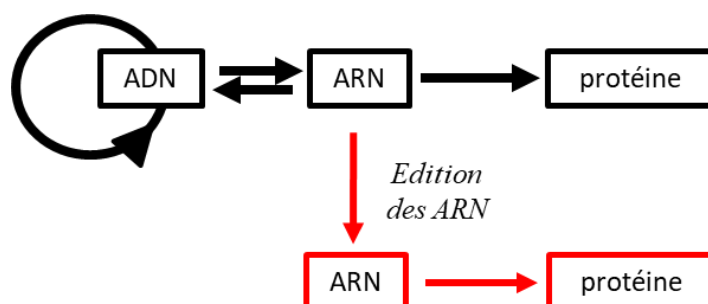


Figure 1: Le dogme central de la biologie moléculaire amendé avec l'édition

Les flèches représentent le sens des flux d'information entre les polymères. Les flèches noires représentent les transferts correspondant à la version traditionnelle du dogme et les flèches rouges les transferts provoqués par l'édition des ARN

le sens inverse. L'édition rend seulement le trajet un peu plus sinueux (Figure 1).

Dans les organites de plante, l'édition procède par la désamination de certaines cytosines en uraciles sur les ARN messagers, modifiant plus de 400 sites dans la mitochondrie et une quarantaine dans le chloroplaste. Ces modifications permettent de restaurer des résidus conservés dans les organites d'espèces ne faisant pas d'édition. Elle a donc un rôle de correction de l'information génétique. Une combinaison d'approches *in vivo*, *in vitro* et *in organello* a permis de montrer que la spécificité de reconnaissance des sites d'édition se trouvait dans une courte région d'une trentaine de nucléotides (éléments *cis*) principalement en amont du site d'édition. Cette région est reconnue par des facteurs *trans* qui sont des protéines à répétition de motifs pentatricopeptide (PPR). Ces facteurs permettent le ciblage de la désaminase sur la cytosine à éditer. Pratiquement toutes les PPR d'édition identifiées appartiennent à une sous-famille appelée PLS qui est quasi exclusivement présente dans les plantes (Benoît Castandet and Araya, 2011a; Small *et al.*, 2020). Au sein de cette sous famille il existe un groupe de protéines PPR (PPR DYW) qui ont un domaine cytidine déaminase en C-terminal. L'existence d'une corrélation stricte entre l'édition et la présence de PPR DYW chez les embryophytes suggérait que cette classe de protéines correspondait au facteur catalytique permettant l'édition (Salone *et al.*, 2007; Rudinger *et al.*, 2008). Il aura cependant fallu attendre plus de 10 ans pour voir ces prédictions vérifiées expérimentalement (Hayes and Santibanez, 2020; Oldenkott *et al.*, 2019). Entre temps, le modèle simple (simpliste !) d'édition présenté lors de ma thèse s'était déjà largement complexifié (Benoît Castandet and Araya, 2011a; Baudry, 2019) (Figure 2).

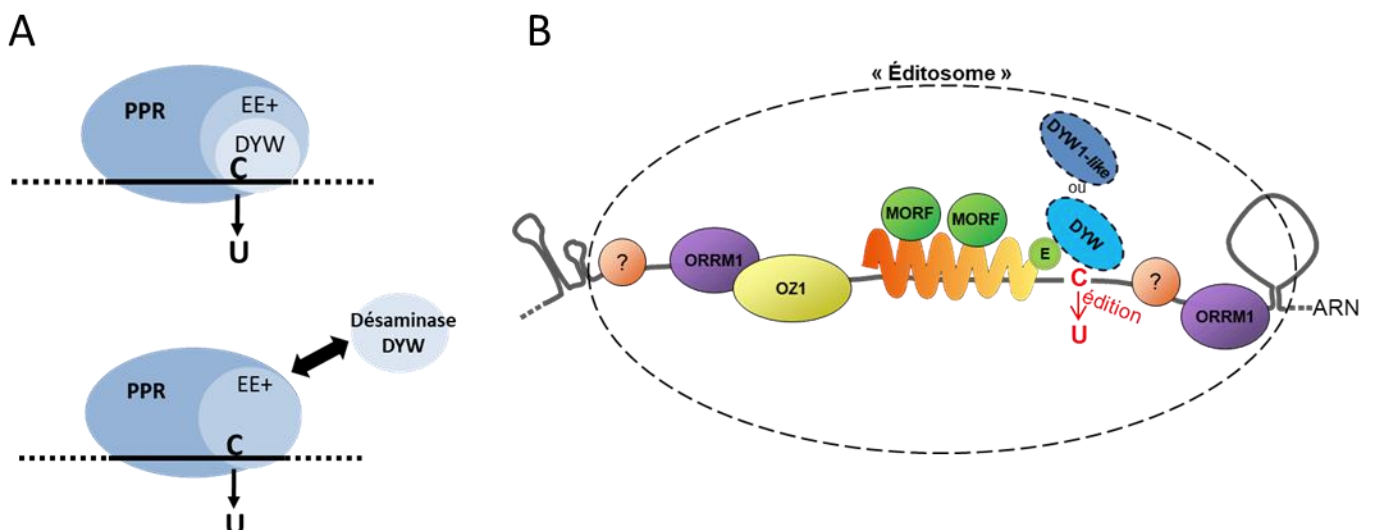


Figure 2: Modèles d'éditosome dans les organites de plantes en A) 2010 et B) 2020

Le schéma en A est issu de ma thèse (Castandet and Araya, 2011) et celui en B de la thèse de Kévin Baudry (Baudry, 2019). Dans les deux cas, une protéine PPR se lie à l'élément *cis* de reconnaissance sur l'ARN et effectue la désamination ou recrute en *trans* une PPR permettant la désamination. En dix ans le nombre de protéines impliquées a augmenté, notamment grâce à la découverte des facteurs MORFS/RIP, OZ et ORRM dont le rôle n'est pas encore totalement élucidé.

Mais pourquoi éditer ?

Lorsque l'on parle de l'édition des ARN, la question du « pourquoi » revient quasi systématiquement. Pourquoi cette étape existe-t-elle, ne serait-il pas plus simple d'avoir directement des gènes « corrects » dans les organites ? Je n'ai pas fait exception et cela a bien sûr été l'une des premières questions que j'ai posées lors de mon entretien en vue du stage de master 2. Le modèle communément accepté pour expliquer l'émergence de l'édition chez les plantes fait de celle-ci un « accident historique », un événement neutre devenu indispensable par le simple fait de son existence. Dans cette hypothèse, l'apparition par dérive génétique d'une activité d'édition a rendu neutre des transitions T>C au niveau de l'ADN puisqu'elles étaient corrigées au niveau de l'ARN (Covello and Gray, 1993). Ce modèle est un exemple de « Constructive Neutral Evolution » (CNE), un processus évolutif qui permet d'expliquer les mécanismes d'interdépendance moléculaire et une bonne partie de la complexité des organismes, sans faire appel à la sélection (Lukeš *et al.*, 2011; Gray *et al.*, 2010; Stoltzfus, 1999; Stoltzfus, 2012). Le grand mérite de ce modèle est qu'il permet d'éviter l'écueil d'une vision pan-adaptationiste trop souvent adoptée par les biologistes moléculaires pour qui la plupart des traits sont vus comme des adaptations ayant une fonction créée par un processus de sélection positive ou directionnelle (Linguist *et al.*, 2020). Dans le cas de l'édition, il n'est par exemple pas rare de lire qu'elle est apparue pour corriger les mutations apparues dans les génomes des organites (Maier *et al.*, 2008). Une telle vision peut cependant difficilement expliquer comment les intermédiaires contenant les mutations sans mécanisme de correction n'ont pas été éliminés et constitue un exemple de ce que Gould et Lewontin dénoncent comme « just so stories » (Gould and Lewontin, 1979).

Alors que j'avais ces questions en tête, la lecture d'un petit livre qui traitait, entre autre, de la sexualité du thym, m'a initié à la théorie du conflit nucléo-cytoplasmique (Gouyon, 2001). Chez les organismes où l'hérédité des génomes nucléaires et cytoplasmiques est différente (bi parentale et maternelle, respectivement), un gène cytoplasmique qui favorise sa propre transmission va être sélectionné, même si cela se fait au dépend du génome nucléaire (Havird *et al.*, 2019). Une conséquence bien connue de ce conflit chez les plantes est la stérilité mâle cytoplasmique (CMS) qui est contrebalancée par l'émergence de gènes restaurateurs de fertilité (gènes *Rf*) appartenant à la famille des protéines PPR (Budar *et al.*, 2003; Touzet and Meyer, 2014). J'ai immédiatement été séduit par l'esthétique du modèle et fait le lien avec le phénomène d'édition. Après deux ans de maturation je suis tombé sur l'ébauche de la même idée dans l'ouvrage « Genes in conflict » de Burt et Trivers (Burt and Trivers, 2008) ce qui eut pour effet de me décider à tenter de la publier. Mon hypothèse est que la similarité entre les mécanismes de restauration de fertilité et d'édition (les deux passent par des protéines PPR) peut s'expliquer s'ils sont tous les deux

la conséquence de la même force motrice, le conflit nucléo-cytoplasmique. Je propose ainsi que les mutations dans le génome des organites sont apparues avant le mécanisme de correction et qu'elles ont pu se répandre, à la condition qu'elles augmentent la transmission de leur propre génome. L'équilibre dans la transmission des deux types de génomes a été rétabli par la correction au niveau de l'ARN des mutations organellaires, c'est-à-dire par l'édition (Figure 3). Je propose également que ce modèle élaboré pour les plantes pourrait expliquer la prévalence des événements d'éditations dans les organites d'autres organismes (Castandet and Araya, 2012).

Presque dix ans plus tard, quel regard porter sur cette hypothèse ? Une de mes craintes était qu'elle soit considérée comme une élucubration inconséquente par les chercheurs du domaine. A Ithaca, lors de mon postdoctorat; j'ai eu la chance de pouvoir inviter et discuter longuement avec Mike Gray. Bien qu'il ne soit pas convaincu, je l'ai au moins entendu m'assurer qu'elle était « loin d'être stupide » et il l'a cité dans une de ses revues comme contre-proposition au CNE (Gray, 2012). Un des points contestés du modèle était l'équivalence que je faisais entre les PPR *Rf* et les PPR d'édition. A l'époque de la publication il était en effet connu que les PPR *Rf* se distinguaient des autres PPR par leur évolution. A la différence des PPR d'édition, elles montrent des signes de diversification et un regroupement en cluster sur le génome, fruit de nombreuses recombinaisons. Cela s'expliquerait par la position des gènes *Rf* dans la « course à l'armement » que se livrent les génomes mitochondriaux et nucléaires (Fujii *et al.*, 2011). Dans le cadre de mon hypothèse il était ainsi légitime de s'attendre à voir un comportement similaire pour au moins une partie des PPR d'édition, ce qui n'était pas le cas. La critique était donc légitime mais une analyse récente du jeu de donnée OneKP semble maintenant montrer un tel comportement chez les plantes non vasculaires (Gutmann *et al.*, 2019). Ce qui n'était qu'une intuition n'est donc peut-être pas si loin de la réalité et je considère que mon hypothèse n'a en rien perdu de sa légitimité. Un paradoxe plus personnel est que, bien que soutenant mon modèle comme une alternative plausible, je suis moi-même devenu un promoteur zélé du CNE en biologie des ARN.

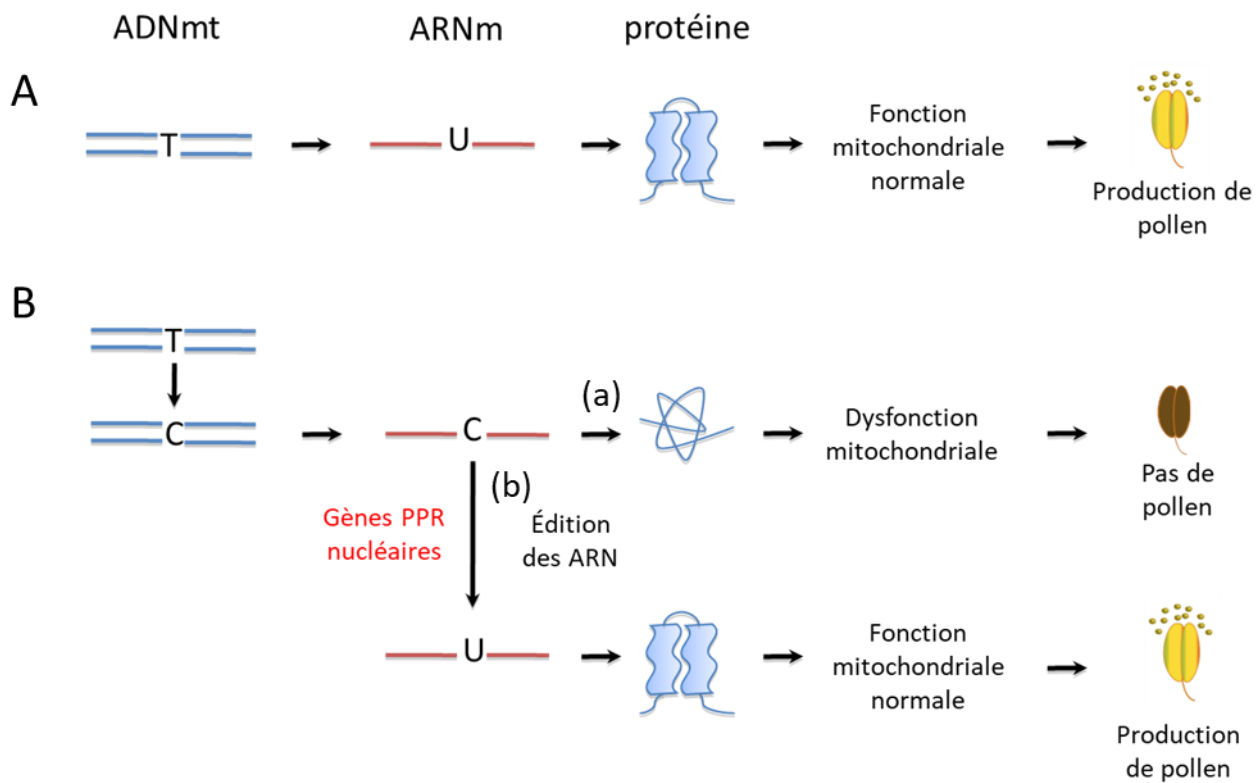


Figure 3: Modèle pour l'origine de l'édition dans les organites de plante.

(A) Avant l'apparition de l'édition, les gènes mitochondriaux codent pour des protéines fonctionnelles. La physiologie mitochondriale est normale et les plantes produisent du pollen, assurant une transmission équilibrée des gènes mitochondriaux et nucléaires. (B) Lorsque des mutations T>C apparaissent dans le génome mitochondrial, des protéines partiellement fonctionnelles sont produites (a) ce qui entraîne une dysfonction mitochondriale et l'arrêt de la production de pollen (plantes mâle-stériles). Par conséquent, la transmission du génome mitochondrial augmente. La dissémination de ces mutations dans la descendance crée une forte pression de sélection sur le génome nucléaire pour faire rétablir l'équilibre dans la transmission des génomes. Cela se fait par la mobilisation de gènes PPR (b) qui vont corriger les mutations au niveau de l'ARN. La fonction mitochondriale est rétablie ainsi que la production de pollen et la transmission équilibrée des génomes mitochondriaux et nucléaires

L'édition dans les introns est nécessaire à l'épissage

La partie expérimentale de ma thèse a d'abord consisté en l'étude du lien entre édition et épissage des ARN dans les mitochondries de plantes. Bien que la grande majorité des sites d'édition soient dans les séquences codantes des gènes et permettent le rétablissement d'un acide aminé conservé, certains sites se trouvent dans les introns. La quasi-totalité des introns mitochondriaux de plantes sont des introns de groupe II connus pour avoir une structure secondaire canonique nécessaire à l'épissage et il avait été proposé que l'édition permette de rétablir ou d'améliorer des appariements de base nécessaire pour l'épissage (Bonen, 2008).

Pour compenser l'absence de système de transformation du génome des mitochondries de plante j'ai utilisé une méthode d'électroporation mise au point au laboratoire. Des mitochondries isolées d'embryons de blé ou de tubercules de pomme de terre sont incubées en présence d'un plasmide contenant un gène que l'on veut exprimer. Petit clin d'œil de l'histoire, la méthode utilisée pour purifier les embryons de blé avait été mise au point par Herbert Stern, le père de David Stern que je n'allais pas tarder à rejoindre (Johnston and Stern, 1957). Un choc électrique permet de faire rentrer le plasmide dans la mitochondrie et l'incubation dure toute la nuit. Afin que l'ADN introduit soit pris en charge par la machinerie de transcription, il est nécessaire de placer le gène dans un contexte contenant des éléments d'expression mitochondriale. Dans ce but, le promoteur du gène *cox2*, dont on sait qu'il permet d'initier efficacement la transcription, est utilisé avec le site terminateur du gène *cob* en aval de l'ORF. Cela permet en plus de disposer d'une séquence spécifique du transgène pour les analyses des produits d'expression par RT-PCR (Farré and Araya, 2001).

J'ai d'abord utilisé le transcrit *rps10* de pomme de terre comme modèle biologique. Deux sites d'édition de *rps10* (C2 et C3) sont situés dans des régions impliquées dans la réaction d'épissage (Figure 4A). C2 se trouve dans la seconde zone de liaison à l'intron (IBS2) et est impliqué dans une interaction tertiaire entre l'IBS2 et la zone de liaison à l'exon (EBS2) mais les mutations introduites *in organello* montrent que son édition n'est pas connectée à l'épissage. Contrairement au résidu C2, l'édition du C3 est indispensable à la réaction d'épissage (Figure 4B). La conversion de C3 en U permet de rétablir un appariement de bases canonique A:U dans la tige du domaine 1 de l'intron. Cet événement d'édition est essentiel puisque des mutations sur le site d'édition ou le nucléotide interagissant avec le résidu édité abolissent l'épissage (C3A et A3T dans la Figure 4). Nous avons pu démontrer que c'est la nécessité d'appariement qui rend la réaction d'édition indispensable et non l'édition en tant que telle. En effet, des mutations changeant les

base (SDxB dans la Figure 5C) ne permet pas de retrouver l'épissage. Il y aurait donc un mécanisme en jeu autre que le simple appariement de base. Il est aussi à noter que l'édition observée dans l'exon *nad1e* introduit par électroporation est bien plus faible que dans le sauvage pour les sites situés dans la partie 3' de l'exon (Figure 5D), un résultat sur lequel je reviendrai plus tard.

La mise au point du *trans*-épissage *in organello* s'est faite juste à la fin de ma thèse et j'ai par la suite, au début de mon postdoctorat, co-établi avec Alejandro le plan de travail de la technicienne du laboratoire, Cindy Aknin, pour réaliser les dernières expériences. Cela explique ma position comme dernier auteur de la publication (Farré *et al.*, 2012).

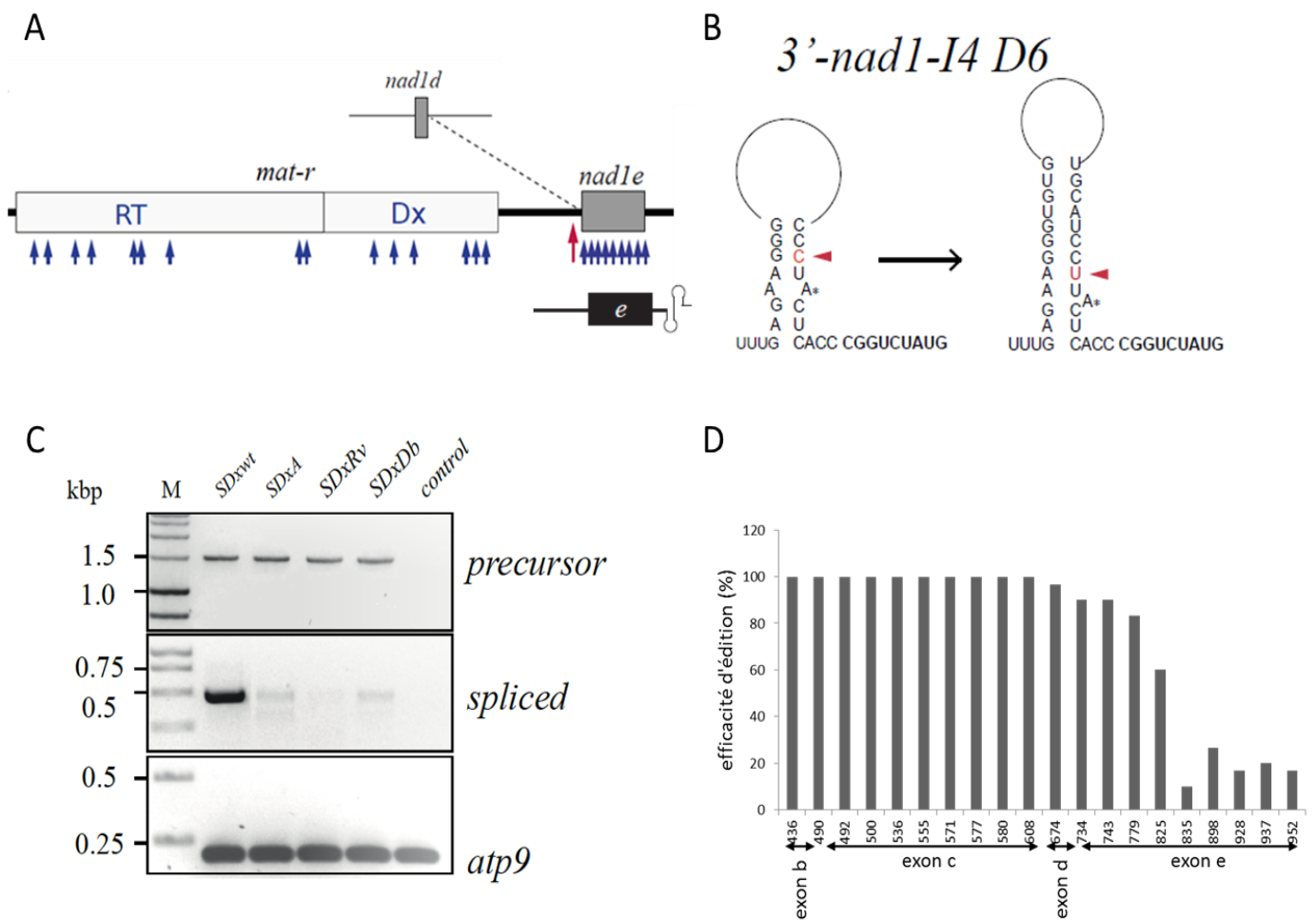


Figure 5: Epissage et édition *in organello* de transcrits *nad1* de blé.

(A) schéma représentant l'unité transcriptionnelle contenant l'intron 4 de *nad1* avec *mat-r* et l'exon e. C'est cette unité qui est ensuite modifiée et introduite par électroporation dans les mitochondries isolées. L'exon e est *trans*-épissé avec l'exon d. Les sites d'édition sont indiqués par des flèches bleues. La flèche rouge indique le site d'édition dans le domaine 6 de l'intron. (B) Changement de structure du domaine 6 de l'intron induit par l'édition. (C) produits de RT-PCR issus des différentes constructions *nad1-1-e* sauvages (SDxwt) et mutants introduites par électroporation. Dans le mutant SDxA le C éditible est muté en A, le mutant SDxRv contient le nucléotide interagissant avec le C édité dans le domaine 6 muté en T et le mutant SDxDb associe les deux mutations. La ligne contrôle représente des mitochondries électroporées avec un plasmide vide. La qualité des ARN a été vérifiée en amplifiant le transcrit *atp9* endogène. (D) Niveau d'édition au sein du transcrit épissé chimérique *nad1*. Le transcrit contient des exons endogènes (b, c et d) et l'exon e issu de l'électroporation.

Avec Dominique Bégu nous avons enfin étendu l'étude de l'interaction entre édition et épissage aux introns de groupe I que l'on trouve principalement dans les génomes mitochondriaux des fougères. L'analyse des séquences de transcrits *cox1* de différentes espèces nous a permis de montrer que l'édition existait également dans ces introns. De plus, les prédictions de structure montrent que ces sites d'édition se trouvent dans des domaines hautement conservés et qu'ils permettent de restaurer des appariements de bases ou la structure essentielle du cœur catalytique. Nous faisons donc l'hypothèse que, de la même manière que pour les introns de groupe II, l'édition pourrait également être indispensable à l'épissage des introns de groupe I. Nous n'avons cependant pas de preuve expérimentale puisque nous n'avons pas pu mettre au point de test *in vitro* ou *in organello* d'épissage pour les introns de fougères (Bégu *et al.*, 2011)

Ce travail de biochimie et biologie moléculaire, bien qu'extrêmement formateur d'un point de vue expérimental, n'a somme toute qu'une portée limitée dans la compréhension des mécanismes d'expression génétique des mitochondries de plante. Nous n'avons fait qu'étendre aux introns un résultat qui était déjà connu depuis longtemps pour les séquences codantes : au lieu de rétablir le codage pour le bon acide aminé, l'édition dans les introns rétablit la structure nécessaire à l'épissage. Je vais maintenant présenter des résultats plus inattendus qui soulèvent des questions fondamentales quant au fonctionnement de la maturation des ARN et qui ont eu un rôle déterminant dans la poursuite de mon parcours scientifique.

Les différents évènements de maturation des ARN sont connectés

Le système *in organello* avait permis de montrer que les résidus directement adjacents en -1 et +1 de la cytosine éditée jouaient un rôle critique pour l'édition de deux sites sur le transcrit *cox2* de blé (Choury *et al.*, 2004; Farré *et al.*, 2001). Lors de ma thèse, j'ai essayé de confirmer et d'étendre ce résultat aux 18 sites d'édition de ce transcrit. Pour cela j'ai muté systématiquement les nucléotides -1 et +1 de tous les sites d'édition sur le gène *cox2* cloné dans un plasmide et j'ai électroporé ces constructions dans des mitochondries isolées de blé. Sur cette base j'ai pu montrer que les sites d'édition pouvaient être classés en quatre familles : (a) dépendance du résidu +1, (b) dépendance du résidu -1, (c) dépendance des deux résidus et (d) indépendance. Tous les cas possibles existaient donc ce qui, en termes d'intérêt, n'avait rien d'exceptionnel. Nous interprétons ce résultat comme la preuve qu'il existe au moins quatre modes différents de reconnaissance des cytosines à éditer dans les mitochondries de plantes, ce qui donnait une indication de la diversité des protéines impliquées dans la réaction (Benoît Castandet and Araya, 2011b). Le

mode de reconnaissance des sites était en effet une question importante à l'époque. L'absence de séquence consensus évidente dans les éléments *cis* de reconnaissance suggérait que l'hypothèse « un site-un facteur d'édition » était plausible, même si elle impliquait l'existence de centaines de protéines dédiées à l'édition. Au début de ma thèse en Octobre 2007, seules deux protéines PPR, CRR4 et CRR21 (Kotera *et al.*, 2005; Okuda *et al.*, 2007; Okuda *et al.*, 2006) étaient connues comme facteurs *trans* d'édition et chacune de ces protéines est responsable de l'édition d'un résidu différent sur le même ARNm. Bien que la famille PPR soit très vaste, elle ne contient pas assez de membres pour cibler de manière individuelle tous les sites (Lurin *et al.*, 2004) et il était alors légitime de proposer que des sites d'édition différents puissent être reconnus par un même facteur *trans*. Dans ce cas, il serait possible de distinguer des familles de sites ayant un même mode de reconnaissance, ce qui avait justifié mon travail.

Un résultat marquant de ce travail fut cependant l'observation d'effets à distance. Certaines mutations introduites influencent l'édition d'un site lointain. Cela montre que certains sites ne sont pas autonomes pour la réaction et que d'autres éléments que la séquence *cis* définie (une vingtaine de nucléotides en amont et une dizaine en aval du C à éditer) ont aussi un rôle à jouer. Ces résultats ont depuis été confirmés par une autre équipe en exprimant des versions mutées du *cox2* de maïs dans des mitochondries isolées de chou-fleur (Grüttner *et al.*, 2013) et font écho à des observations plus anciennes faites lors de test d'édition *in vitro* (Takenaka *et al.*, 2004). Afin d'expliquer ces résultats il était couramment avancé que certains sites pouvaient avoir des éléments de reconnaissance différents et plus long, ou que les mutations pouvaient perturber certaines structures secondaires importantes (Takenaka *et al.*, 2004; Staudinger *et al.*, 2005). De notre point de vue, une explication possible était que les mutations introduites perturbent l'engagement du transcrit dans la maturation, et ce de manière globale. Cela aurait pour résultat une édition atypique sur certains sites, expliquant ainsi les profils observés. Il faudrait qu'un transcrit soit reconnu comme « légitime » pour être pris en charge.

Afin de vérifier l'hypothèse selon laquelle le devenir des transcrits pouvait avoir une influence sur l'efficacité de l'édition j'ai décidé de perturber artificiellement la maturation des ARN en construisant un gène *cox2* de blé dépourvu d'introns (Figure 6A). La construction s'exprime comme celle contenant l'intron mais nous observons une très forte inhibition de l'édition par rapport aux transcrits sauvages, ce qui renforce l'hypothèse que les mécanismes de maturation doivent être interconnectés dans les mitochondries de plante (Figure 6C et 6D). Ce résultat obtenu chez le blé semble généralisable puisque la même observation a été faite avec des mitochondries de *Solanum tuberosum* pour les transcrits *rps10* sans introns (Figure 6B, 6C et 6E) (Castandet *et al.*, 2010). Sur la base de ces observations, on peut penser qu'il existe dans les mitochondries de plantes un complexe de

maturation capable de prendre en charge les transcrits après leur synthèse jusqu'à ce qu'ils soient prêts pour la traduction.

Dans notre cas, des transcrits contenant naturellement un intron ne seraient plus reconnus par la machinerie de maturation lorsqu'ils sont dépourvus de cet intron. L'engagement dans la maturation se ferait par la reconnaissance de l'ARN par des facteurs protéiques jouant en quelque sorte le rôle de chaperonnes, les protéines PPR semblant à l'époque être les candidats idéaux pour jouer ce rôle. Cette idée d'un complexe de maturation qui nécessite qu'un ARN soit reconnu comme légitime aurait le mérite d'expliquer de nombreuses bizarreries observées dans les différents modèles expérimentaux. Par exemple, alors que le site C259 du transcrit *cox2* de blé est toujours très fortement édité lorsqu'il est placé à différents endroits du transcrit (avec sa séquence *cis*) ou en contexte hétérologue (mitochondrie de pomme de terre), il n'est plus édité dans les mitochondries de blé lorsqu'il est greffé sur le transcrit *rps10* de pomme de terre. Là encore, il faut noter que l'intron de *rps10* de pomme de terre n'est pas épissé dans les mitochondries de blé (Choury *et al.*, 2005; Choury and Araya, 2006). Une explication de ce comportement serait que le transcrit *rps10* n'est pas reconnu par la machinerie de maturation de la mitochondrie de blé, l'excluant par ricochet de la machinerie d'édition, même pour un site de blé normalement reconnu (ici le C259). De la même manière, les résultats de la Figure 5D qui montrent une édition plus faible dans la partie 3' de l'exon e introduit par électroporation (Farré *et al.*, 2012), ceux montrant un rôle déterminant de la région 5'UTR pour l'édition de transcrits *atp6* (Staudinger *et al.*, 2005) ou la très faible efficacité des systèmes d'édition *in vitro* (Blanc *et al.*, 1995) pourraient tous s'expliquer par une perturbation, une non reconnaissance du transcrit dans sa globalité. La séquence *cis* de reconnaissance ne peut jouer son rôle que si le transcrit est correctement engagé dans la maturation.

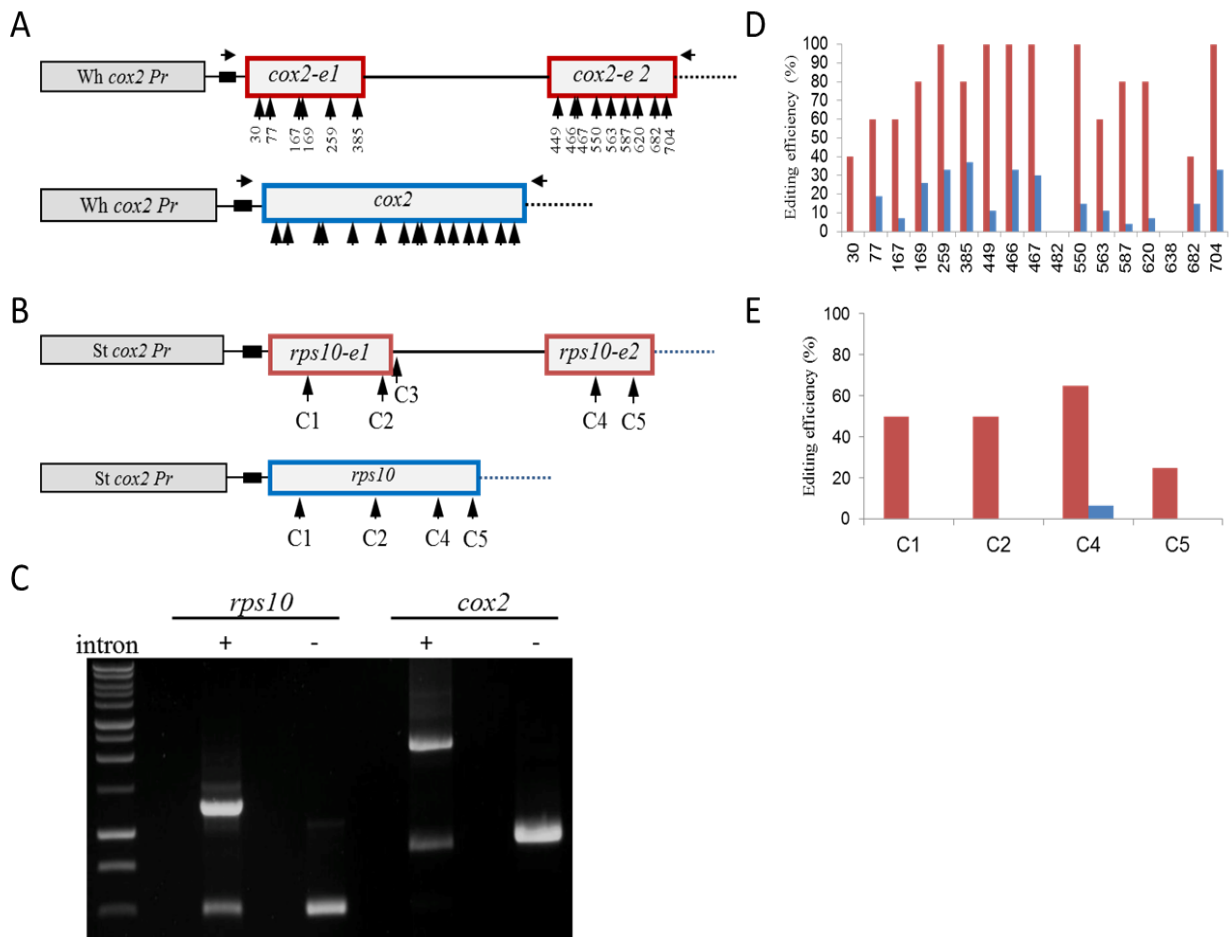


Figure 6: Enlever l'intron perturbe l'édition des transcrits *cox2* de blé et *rps10* de pomme de terre.

Schéma des gènes *cox2* de blé (A) et *rps10* de pomme de terre (B) qui sont électroporés, avec ou sans intron, et localisation des sites d'édition. Les gènes sont sous la dépendance d'un promoteur *cox2* de blé ou pomme de terre. (C) expression des constructions dans les mitochondries isolées de pomme de terre (*rps10*) et blé (*cox2*). D et E, comparaison de l'efficacité d'édition dans les constructions avec (rouge) et sans (bleu) introns.

Que faire ?

Ces résultats, bien qu'excitants, me faisaient toucher du doigt les limites de notre approche expérimentale. Avec l'électroporation de mitochondries isolées, il est possible d'introduire des constructions mimant des gènes qui portent des mutations et de suivre la maturation des transcrits qui en découle mais on se limite forcément à une approche gène par gène. Il est impossible d'étudier l'influence de notre construction sur la maturation du reste des transcrits ou même de regarder de manière plus globale les tendances. Ce que l'on observe est vrai pour un type de transcrit, celui que l'on étudie, pas forcément pour le transcriptome. On se confronte en somme aux limites de l'approche réductionniste. Malgré sa puissance et ses succès, celle-ci ne peut pas aboutir en biologie où, pour paraphraser Richard Levins, « le tout est la vérité ». Il m'apparaissait donc nécessaire de développer une approche plus globale d'étude du transcriptome des organites et mon postdoctorat a été l'occasion de s'atteler à cette tâche.

Avec le recul, en plus des aspects pratiques et méthodologiques, cette expérience de thèse m'a peut-être surtout permis de renforcer mes intuitions quant à ma vision du métier et de l'encadrement. Avec le monitorat j'ai appris à conjuguer enseignement et recherche (ce qu'Alejandro appelait la dialectique du scientifique !) et acquis la conviction qu'il fallait un maximum intégrer des aspects recherche ou vie de labo, dans nos enseignements. Autre conviction renforcée, il est hors de question que l'état d'esprit (supposément) valorisé dans le laboratoire et l'université ne s'arrête aux portes de ces lieux ou ne se limite à l'étroit contour d'une thématique de recherche. Les interactions avec mon directeur de thèse étaient loin de se limiter à mon travail expérimental. Je me souviens encore de vendredis après-midi entiers à parler de littérature, d'histoire des sciences ou de politique, des livres échangés et de la confrontation de nos regards critiques. Bien qu'incroyablement enrichissant, j'ai pu constater que cet aspect de l'encadrement est l'exception plutôt que la règle. Ma conviction s'est faite certitude, je ne conçois pas l'encadrement d'étudiants, et a fortiori celui du chercheur en formation qu'est le doctorant, dans une logique de rentabilité scientifique personnelle ou immédiate. Mon rôle est de l'accompagner dans la construction de son propre parcours, de l'aider à réaliser des connexions entre les différents domaines, pas de choisir une direction à sa place.

Heureux qui comme Ulysse

La découverte des approches transcriptomiques

J'ai commencé en Juin 2011 mon postdoctorat au sein de l'équipe de David Stern au Boyce Thompson Institute sur le campus de l'université de Cornell à Ithaca, NY. Le laboratoire étudie la maturation des ARN dans le chloroplaste et plus particulièrement le rôle des endo- et exoribonucléases dans ce processus, un des aspects que je n'avais pas étudié lors de ma thèse. Les discussions avec David avaient initialement tourné autour de deux thématiques de recherche, mettre en place l'électroporation de chloroplastes isolés d'*Arabidopsis* et essayer de voir si la ribonucléase RNase J est responsable de la maturation en 5' des ARN chloroplastiques chez l'algue *Chlamydomonas reinhardtii*. Autre clin d'œil de l'histoire puisque la RNase J tient une grande place dans l'ANR jeune chercheur que j'ai pu obtenir des années plus tard.

Un des grands avantages de travailler avec David est la liberté totale qu'il m'a permis d'avoir dans les recherches que j'ai poursuivies. Devant le peu d'avancées et les difficultés rencontrées avec *Chlamydomonas* et mes électroporations de chloroplastes il ne s'est pas offusqué de me voir petit à petit explorer de nouveaux horizons. David est en fait ouvert à tout, à condition de pouvoir le convaincre qu'on ne perd pas notre temps ou trop d'argent, ce qui dans sa tête est un quasi synonyme. J'ai donc pu explorer l'hypothèse faite pendant ma thèse de la connexion entre les différents événements de maturation et j'ai pour cela utilisé des approches transcriptomiques (RNA-Seq). Grâce aux formations offertes par l'Université de Cornell et le Boyce Thompson Institute, j'ai acquis les bases de la bioinformatique dédiée au RNA-Seq et je me suis formé aux langages de programmation bash et perl. Je maîtrisais donc toutes les étapes d'une analyse transcriptomique depuis l'extraction des ARN jusqu'à l'analyse des données de séquençage.

Depuis la première expérience de RNA-Seq publiée pour *A.thaliana* en 2007 (Weber *et al.*, 2007), les banques de données telles que la SRA (www.ncbi.nlm.nih.gov/sra) se sont enrichies de dizaines de milliers de jeu de données, et ce pour de multiples groupes de plantes. Cela représente sans conteste une opportunité incroyable pour l'étude de l'expression des génomes des organites, mais force est de constater que quasiment personne ne regarde ce qu'il s'y passe (Smith, 2013). Une des raisons de la quasi invisibilité des transcrits d'organites dans les approches RNA-Seq est d'ordre purement technique. Il est en effet connu depuis leur découverte que les ARNm ne représentent qu'une très faible proportion des ARN totaux (principalement ARNr et ARNt) (Ycas and Vincent, 1960; Brenner *et al.*, 1961; Gros *et al.*, 1961; Astrachan and Volkin, 1958). La plupart des protocoles de

préparation de bibliothèques de RNA-Seq cherchent donc à éliminer ou réduire les ARNr en sélectionnant les ARN contenant une queue polyA. L'ennui, c'est que la polyadénylation est un signal de dégradation dans les chloroplastes et les mitochondries (Lange *et al.*, 2009), ce qui entraîne une très faible représentation de ces ARN dans les bibliothèques préparées de cette manière. L'alternative consiste à éliminer spécifiquement les ARNr par hybridation soustractive et cette approche est de plus en plus utilisée car elle donne aussi accès aux ARN non codants. L'argument technique éliminé, un autre obstacle apparaît : la plupart des outils développés se focalisent sur le transcriptome nucléaire et ne prennent malheureusement pas en compte les spécificités de l'expression génétique des organites comme l'édition par exemple.

Cela a été la motivation pour le développement d'un outil dédié à l'analyse du transcriptome chloroplastique à partir de données RNA-Seq. Le but était de pouvoir étudier simultanément tous les événements de maturation en une seule expérience. Étant donné l'importance des mécanismes de signalisation rétrograde (Chan *et al.*, 2016), la possibilité de regarder en même temps les transcriptomes nucléaires et d'organites était aussi une perspective alléchante. Il s'est écoulé à peu près quatre ans entre mes premiers gribouillages de lignes de commandes et l'écriture de ChloroSeq (Castandet *et al.*, 2016) (accessible à <https://github.com/BenoitCastandet/chloroseq>). Même s'il n'a pas été facile de publier le manuscrit qui était sous un format hybride, à cheval entre la biologie et la bioinformatique, l'outil a été plutôt bien accueilli et a fait l'objet d'un éditorial de David Roy Smith dans le journal *Briefing in Bioinformatics* (Smith and Sanità Lima, 2016). Suite au développement de ChloroSeq, j'ai eu l'opportunité d'encadrer Elena Michel, une étudiante en thèse du département « Plant Biology » de l'Université de Cornell lors de sa rotation dans notre équipe. Nous avons essayé de rendre l'outil plus accessible en produisant un protocole qui explique, étape par étape, comment étudier le transcriptome chloroplastique. Elena a également produit des exemples de code qui permettent de produire des figures dans R basées sur les résultats de ChloroSeq (Michel *et al.*, 2018).

Un des avantages de travailler sur l'expression du génome chloroplastique, c'est sa petite taille, aux alentours de 150 kb. Il est donc possible d'aller au-delà de l'analyse d'expression différentielle des gènes et de regarder le transcriptome avec une résolution simple nucléotide. Cela donne une précision d'analyse comparable aux techniques traditionnelles de biologie moléculaire mais à l'échelle du transcriptome. En d'autres termes, il devient possible de regarder une approche -omique avec les yeux et le mode de pensée d'un biologiste moléculaire. La contrepartie de ce choix est que je n'ai absolument pas développé les approches statistiques adéquates pour ce genre de travail, je me suis uniquement contenté de mes yeux. En gros, je regardais le RNA-Seq comme un RNA-blot.

Cela a d'incomparables avantages mais l'absence de statistiques constitue tout de même une lacune que je m'attèle à combler depuis mon recrutement à l'Université de Paris.

RNA-Seq et édition

Avant mon arrivée à Ithaca, il avait été effectué un séquençage massif brin spécifique des ARN de l'écotype col0 d'*Arabidopsis thaliana* et d'un mutant pour la polynucléotide phosphorylase (PNPase, mutant *pnp1-1*), une exoribonucléase 3'-5' impliquée dans la maturation 3' et la dégradation polyA dépendante des ARN. L'étude avait permis de mettre en évidence une grande quantité d'ARN non codants antisens aux transcrits connus dans le chloroplaste, ou pncRNAs (Hotto *et al.*, 2011). Ces données ont constitué mon premier terrain de jeu et mes premiers scripts ont été écrits pour les faire parler. J'ai bien entendu commencé par regarder ce qu'elles pouvaient nous apprendre sur l'édition des ARN.

Le premier objectif était de vérifier qu'il était bien possible d'identifier les sites d'édition connus du transcriptome chloroplastique. Au début de ce travail il en était connu 34 chez *Arabidopsis*, la quasi-totalité dans les régions codantes (Chateigner-Boutin and Small, 2007). Toutes les méthodes d'études utilisées jusque-là nécessitaient une connaissance *a priori* de la position du site d'édition puisqu'il fallait d'abord amplifier une région d'intérêt et comparer la séquence de l'ADNc à celle de l'ADN. Avec le RNA-Seq tout le transcriptome est représenté sans *a priori* et un site d'édition doit apparaître comme un « single nucleotide polymorphism (SNP)»: Nous avons non seulement pu identifier tous les sites d'édition déjà connus mais nous avons également pu en découvrir neuf de plus, portant le total à 43 (Table 1). Ces sites sont faiblement édités en comparaison des sites préalablement décrits.

Trois de ces neuf sites sont uniquement présents dans le mutant *pnp1-1*. La PNPase ayant pour rôle de dégrader les ARN, son absence permet de préserver certaines parties des transcrits qui sont normalement éliminées. Nous avons en effet pu montrer que ces sites étaient sur des portions d'ARN qui ne sont normalement pas ou très peu abondantes. D'autre part, les séquences *cis* probables de reconnaissance de ces nouveaux sites sont similaires à certaines séquences *cis* de sites déjà connus dont le taux d'édition est bien plus élevé. Nous pensons donc qu'il pourrait s'agir de sites secondaires, édités car reconnus par la machinerie d'édition de manière aspécifique.

Table 1: Nouveaux sites d'édition identifiés par RNA-Seq. Neuf nouveaux sites ont été identifiés dont trois sont uniquement présents dans le mutant *pnp1-1*. L'édition est donnée en pourcentage de transition C en U. Les éléments *cis* potentiels des nouveaux sites sont indiqués.

	Position sur le génome	WT		<i>pnp1-1</i>		Elément <i>cis</i>
		couverture	édition	couverture	édition	
atpH 3'UTR	13210	2702	4%	4254	4%	GTAGTTTTTTTTAATTCTATC
ycf3 Intron 2	43350	390	12%	1685	1%	GACTAGATATGCCTAAATAC
rps4 3'UTR	45095	69	10%	205	1%	ATTTTTCCTATTTCATGTATC
ndhK-ndhJ	49209	1342	6%	864	0%	CTTCATAAATTAGAATTAAC
rps18 3'UTR	68453	25	0%	721	26%	ATTTCTACTCTACCTTCCCC
ycf2 as	91535	2	0%	47	11%	TCATCAATATCGATATCATC
ndhB 3'UTR	94622	2	0%	324	6%	CTACTTTTTTACATATCTCTC
ndhB	96439	419	6%	597	2%	TCACTGTAGGAATTGGGTTC
<u>ndhB</u>	96457	518	5%	820	2%	CAATTGCGCTTATATTCATC

Ce travail a été publié en collaboration avec Christian Schmitz-Linneweber et Hannes Ruwe son étudiant en thèse (Ruwe *et al.*, 2013). Alors que je présentais ces premiers résultats en 2012 à la « Gordon Research Conference on Mitochondria and Chloroplasts », Christian est venu me voir pour me montrer, qu'avec Hannes, ils avaient fait exactement la même chose sur le même jeu de données. Je dois être honnête et reconnaître qu'ils étaient même plus avancés que ce que je présentais dans mon poster. A l'heure où les chercheurs sont de plus en plus secrets sur leurs avancées, où tout le monde a peur que le collègue de l'équipe voisine ne s'approprie des résultats ou des idées, je tiens à souligner la classe de Christian. A ce moment-là j'avais plutôt en tête l'idée de faire un papier plus global sur le RNA-Seq, un travail encore loin d'être finalisé. Il aurait pu ne rien dire et publier quelques mois plus tard. Il nous a proposé de lui-même de publier ensemble nos résultats et je le remercie chaleureusement d'avoir démontré que le partage désintéressé des résultats avait toujours un sens.

En lien avec l'hypothèse avancée pendant ma thèse, nous avons également pu montrer que le mutant *pnp1-1* présentait des altérations spécifiques de l'édition de certains sites alors qu'il n'est pas un mutant d'édition à proprement parler. Cela appuie l'idée que les différents événements de maturation des ARN dans les organelles de plante sont étroitement reliés, et surtout que le RNA-Seq est une approche prometteuse pour l'étudier.

L'autre intérêt d'un tel outil est qu'il permet d'utiliser des données de séquençage déposées dans des banques (comme la SRA du NCBI) et d'extraire les informations sur les organites qui avaient été laissées de côté. Il est possible de réinterroger des données en leur posant des questions qui n'avaient pas été la motivation première de leur génération. Cette

possibilité a été saluée en 2016 par un éditorial du *New England Journal of Medicine* qui se félicitait du développement de ce qu'il nommait avec malice des « research parasites » (Longo and Drazen, 2016). Il existe même maintenant un « Parasite award » qui cherche à promouvoir la réutilisation rigoureuse des données (<https://researchparasite.com/>)!

Stress et métabolisme des ARN chloroplastiques

ChloroSeq a été l'occasion d'expérimenter ce travail de sporozoïte ou parasite junior, pour rester dans la même métaphore. J'ai analysé les données de Di *et al.* (2014) qui ont étudié la régulation de longs ARN non codants en réponse à 6 et 12 heures de différents stress abiotiques (froid, chaud, salinité, sécheresse). Les banques de séquençage ont été réalisées en sélectionnant des ARN polyA(-) et en enlevant les ARNr, ce qui permettait donc une analyse des ARN chloroplastiques. Il apparaît clairement que l'épissage est inhibé après 12 heures de stress thermique (Figure 7A). L'efficacité d'épissage médiane passe d'environ 50% dans la condition témoin à 10% après 12 h de stress. Cette inhibition semble spécifique à ce stress puisqu'elle n'est pas visible dans les autres conditions. De la même manière, on observe une forte inhibition de l'efficacité d'édition, là aussi spécifique au stress chaud (Figure 7B). Alors que la valeur médiane d'édition dans l'expérience témoin est de l'ordre de 90%, celle-ci est réduite à 70% après 3h de stress et même 55% après 12 h de stress chaud. Lorsque l'on regarde les sites d'édition de manière individuelle on observe de grandes variations, certains subissant même une légère amélioration de leur édition alors que d'autres la voient pratiquement abolie.

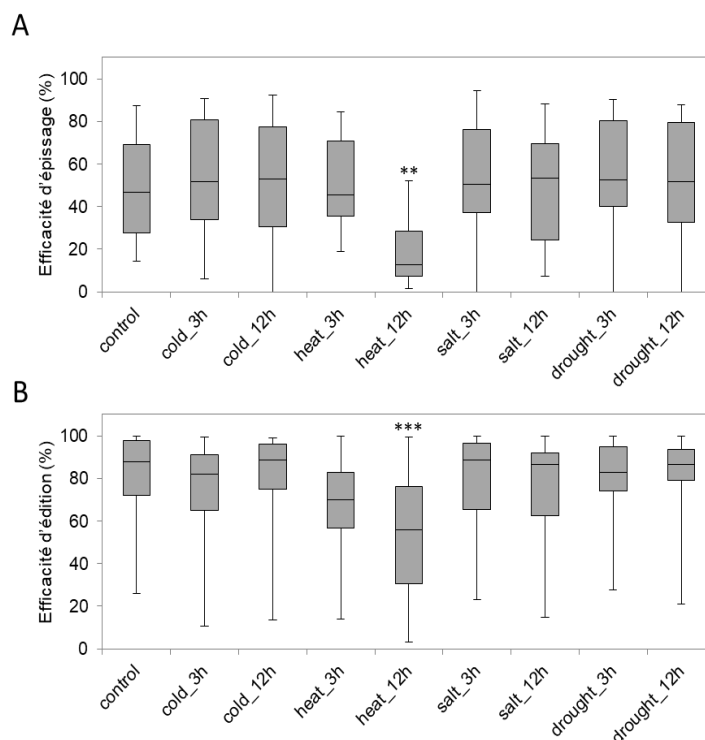


Figure 7: Le stress chaud inhibe l'épissage et l'édition des transcrits chloroplastiques.

Les efficacités globales d'épissage (A) et édition (B) sont représentées par des box plots. La barre du milieu correspond à l'efficacité médiane et les extrémités des boîtes représentent 25% et 75% de la distribution des efficacités. Les moustaches indiquent les valeurs minimales et maximales. ** $p < 0,01$ et *** $p < 0,001$.

Encore une fois, il semble bien qu'édition et épissage soient deux réactions liées puisqu'elles varient de la même manière en condition de stress. Ces résultats confirment ce qui avait déjà été observé dans des tests *in vitro* d'édition et d'épissage dans les années 1990 et 2000 (Karcher and Bock, 1998; Karcher and Bock, 2002) mais l'avantage du RNA-Seq est qu'il donne une vision sur tous les évènements à la fois. Cela peut permettre d'éviter des généralisations hasardeuses basées sur le comportement particulier d'un site. Pour expliquer cette baisse généralisée j'ai quantifié l'expression des facteurs protéiques qui permettent l'édition et l'épissage mais je n'ai pas pu trouver de corrélation convaincante.

Cette absence de lien direct semble invalider l'idée parfois émise que ces facteurs pourraient être des éléments régulateurs de l'expression génétique des organites. Bien que de plus en plus de résultats montrent que cette expression varie en fonction du développement ou des conditions de stress, ces effets globaux n'ont pour l'instant pas pu être reliés à des éléments causaux simples (García-Andrade *et al.*, 2013; Zhang *et al.*, 2014; Kakizaki *et al.*, 2012; Zhao *et al.*, 2020; Leister *et al.*, 2017). Avec Elena, à titre d'exemple, nous avons quantifié les niveaux d'édition chloroplastique lors du développement de la feuille et montré que l'édition des transcrits du complexe NDH varie considérablement entre 16 et 22 jours après émergence, un possible reflet d'une transition entre croissance et sénescence (Figure 8). Là encore, nous n'avons pas pu identifier de lien direct entre les taux d'édition et les facteurs d'édition (Michel *et al.*, 2018).

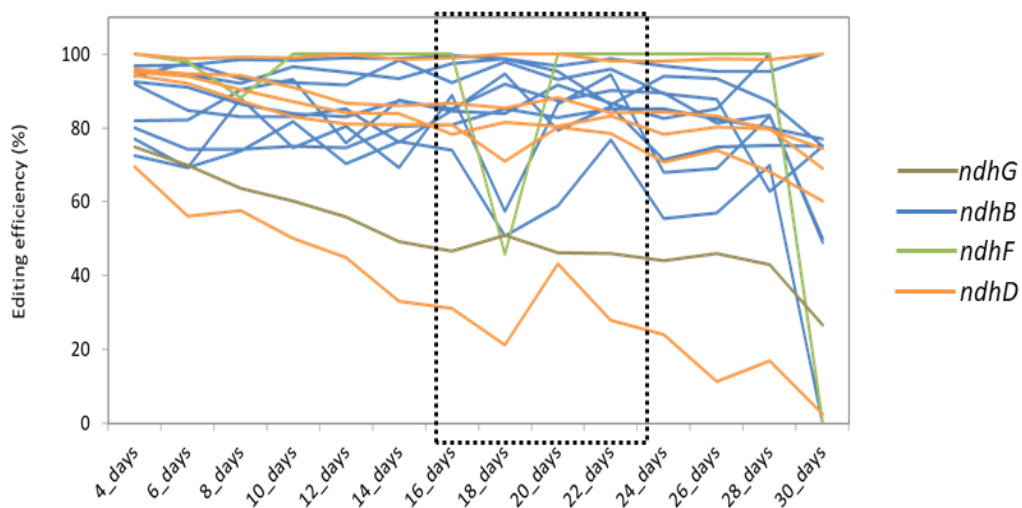


Figure 8 : Le développement de la feuille influence le niveau d'édition.

Quantification de l'édition des sites des transcrits *ndh* durant le développement de la feuille (de 4 à 30 jours après émergence). Les sites d'un même transcrit sont représentés avec une même couleur, le code couleur est indiqué sur la droite de la figure. Les pointillés soulignent la période entre 16 et 22 jours après l'émergence de la feuille où les variations sont très importantes.

Avec le RNA-Seq, il est bien sûr possible d'aller au-delà de l'analyse des événements particuliers que sont l'édition et l'épissage et d'avoir une vision plus globale des effets des différents stress sur le transcriptome chloroplastique. Cela est même indispensable lorsque l'on considère les résultats que je viens de décrire. On peut par exemple représenter l'intégralité du transcriptome sous forme d'une « heat map » créée à partir des résultats donnant la couverture pour des fenêtres de 100 nt chevauchant de 50 nt (Figure 9A). Il apparaît évident qu'à la différence des autres stress, le stress « chaud » (i.e. 37°C pendant 3 et 12h) a un effet majeur sur le transcriptome. L'abondance des transcrits sur les deux brins augmente quasi systématiquement en comparaison avec l'expérience témoin. La plupart des gènes chloroplastiques sont surexprimés, mais pas seulement. Il y a aussi des transcrits antisens aux régions annotées ce qui laisse penser que le stress chaud augmente la quantité de d'ARN non codants (pncRNA) dans le chloroplaste. Nous avons ensuite validé cette augmentation grâce à des northern blots. Le transcrit *rbcl* qui code pour la large sous-unité de la rubisco est un exemple où un pncRNA antisens s'accumule (Figure 9B). Ce pncRNAs d'environ 3,7 kb avait été préalablement identifié par l'équipe dans un mutant déficient pour une ribonucléase, la RNaseJ, et il interfère avec la traduction car il s'apparie avec le brin sens de *rbcl*, réduisant ainsi l'accès au polysome (Sharwood *et al.*, 2011).

Cette analyse faite sur des données que je n'avais pas générées a donc permis de montrer l'intérêt de l'approche RNA-Seq. Les résultats obtenus font en plus écho à des expériences publiées plusieurs décades auparavant. L'augmentation de l'abondance de la plupart des transcrits en réponse à un stress chaud avait par exemple était décrite dans les cotylédons de citrouille en 1993 (Kusnetsov *et al.*, 1993). Au milieu des années 1980, il avait également été montré que la chaleur provoquait une inhibition générale de la traduction chloroplastique chez le haricot (Süss and Yordanov, 1986). Avec notre pipeline, nous avons donc pu mettre en évidence trois mécanismes possibles pour expliquer une telle inhibition : apparition de duplex ARN sens et antisens, inhibition de l'épissage et inhibition de l'édition. Comme une confirmation, *psbA* était le seul gène de la publication de 1986 dont la traduction n'était pas inhibée et il ne contient ni sites d'édition, ni intron.

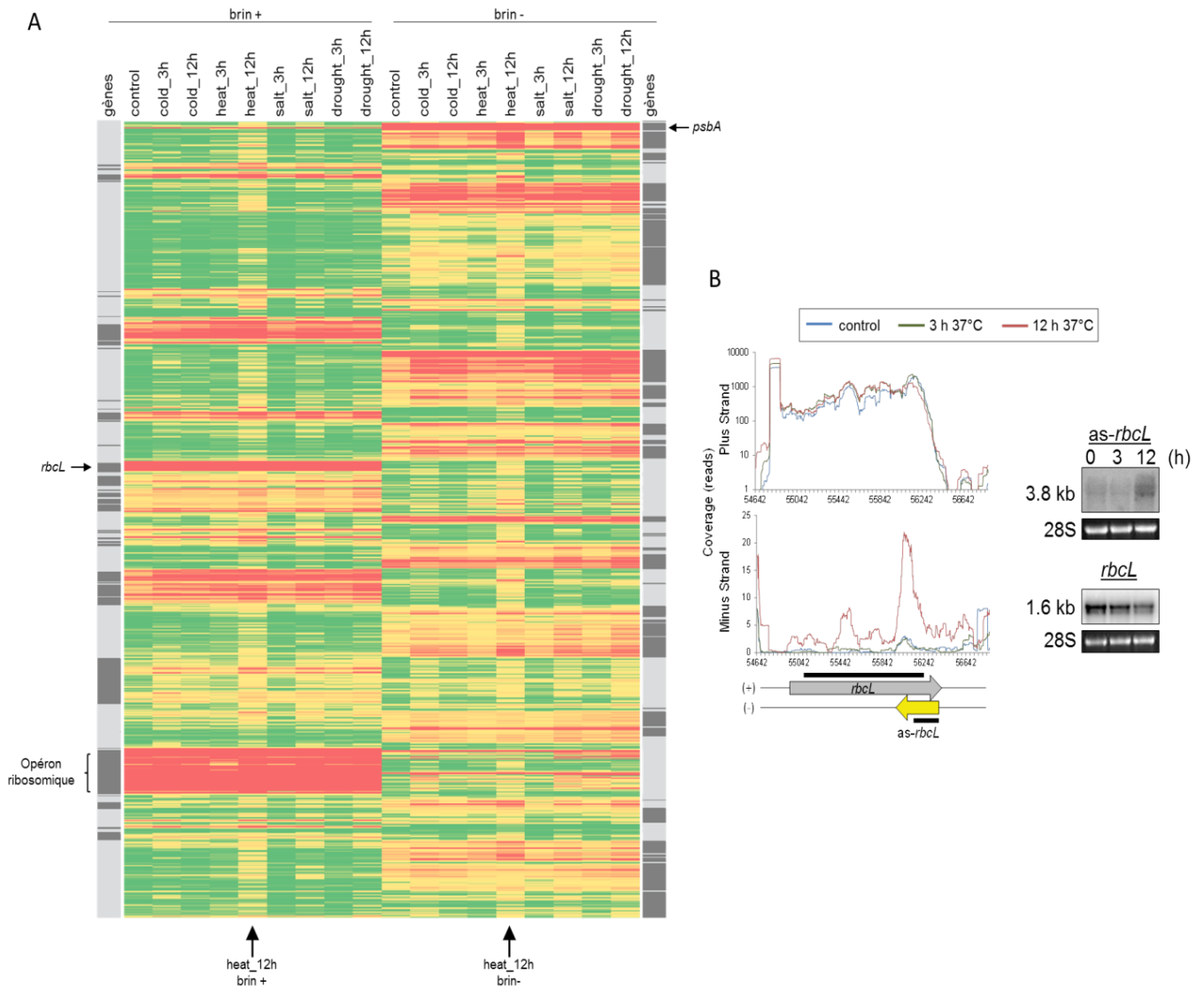


Figure 9: Le stress chaud a un effet majeur sur le transcriptome chloroplastique.

(A) Heat map créée pour les différentes conditions de stress et expériences témoins en utilisant la couverture sur des fenêtres de 100 nt se chevauchant par 50 nt. Le vert correspond aux régions les moins exprimées et le rouge aux plus exprimées. Trois régions connues pour être hautement transcrites sont indiquées par des flèches horizontales : les gènes *psbA* et *rbcL* et l'opéron ribosomique. Des flèches verticales en bas de la heat map indiquent pour les deux brins la colonne correspondant au stress chaud de 12 heures. (B) Accumulation du pncRNA antisens à *rbcL* après 12 heures de stress chaud. La partie gauche montre la couverture simple nucléotide pour les deux brins et la partie droite comporte deux northern blot brin spécifique réalisés avec les sondes indiquées par un trait noir.

Les ribonucléases au cœur de l'expression du génome chloroplastique

La mise en évidence en condition de stress chaud d'un pncRNA qui s'accumule également dans un mutant de ribonucléase montre tout l'intérêt qu'il pourrait y avoir à utiliser les approches globales pour étudier ce type d'enzyme. Elles permettent les maturations en 5' et 3', le clivage des polycistrons, et la dégradation des ARN et ont donc un rôle majeur dans la production des transcrits matures et fonctionnels (modèle dans la Figure 10). La maturation des extrémités 5' des ARN est réalisée par une exoribonucléase 5'-3', la RNaseJ alors que la Polynucléotide phosphorylase (PNPase) et RNR1, deux exoribonucléases 3'-5', travaillent de concert pour former les extrémités 3'. Les unités polycistroniques sont aussi clivées par des endoribonucléases mais le partage des rôles entre la RNase J, RNaseE ou CSP41 n'est pas clair. Une redondance fonctionnelle est aussi très probable. Les ARN sont protégés d'une dégradation excessive par des structures secondaires ou des protéines spécifiques de liaison à l'ARN comme les protéines à répétition de motif pentatricopeptide (PPR). Le transcriptome du chloroplaste est donc un mélange d'un ensemble d'intermédiaires de maturation ou de transcrits avortés qui doivent être éliminés et des transcrits correctement maturés qui doivent être préservés. Les ribonucléases qui permettent la dégradation des ARN sont les mêmes que celles utilisées pour la maturation. La dégradation d'un transcrit est initiée par un clivage endoribonucléolytique (RNaseJ, RNaseE, CSP41) suivi de l'addition d'une courte queue riche en polyA aux extrémités 3' des fragments ainsi créés et enfin la digestion 3'-5' par la PNPase et RNR1 (MacIntosh and Castandet, 2020; Germain *et al.*, 2013).

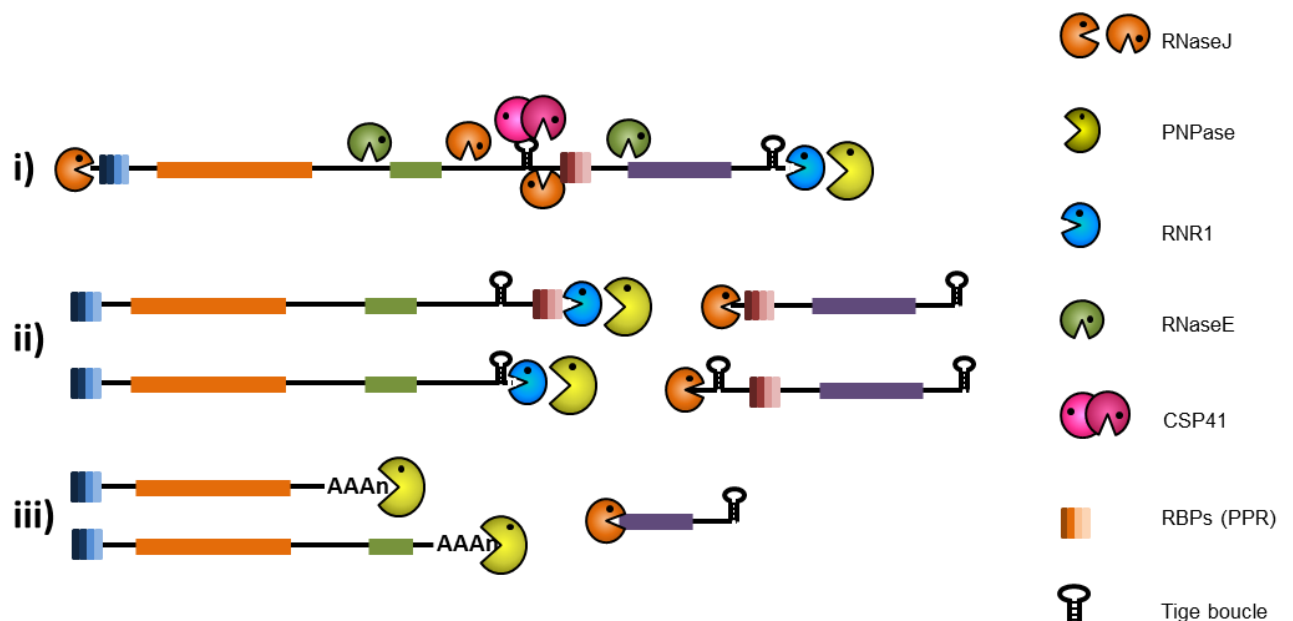


Figure 10: Rôle des ribonucléases dans la maturation des ARN chloroplastiques.

i) La Rnase J mature l'extrémité 5' d'un ARN polycistronique alors que la PNPase et RNR1 travaillent de concert pour former l'extrémité 3'. Le transcrit est aussi clivé par des endo-ribonucléases mais le partage des rôles entre la RNase J, RNaseE ou CSP41 n'est pas clair. **ii)** Suite au clivage, les extrémités 3' et 5' nouvellement créées sont également maturées par les mêmes exo-ribonucléases. Les extrémités des transcrits matures sont protégées des protéines de liaison à l'ARN (PPR par exemple) ou des structures secondaires. **iii)** Les transcrits mal maturés ou illégitimes sont dégradés par la PNPase suite à l'addition d'une courte queue polyA. La RNaseJ joue aussi un rôle dans la dégradation et le contrôle qualité des ARN, notamment pour prévenir la formation de duplex ARN-ARN.

L'étude des ribonucléases chloroplastiques a toujours été un des intérêts, si ce n'est l'intérêt principal, de David. Son équipe a eu un rôle prépondérant dans la caractérisation des cinq ribonucléases majeures dont je viens brièvement de décrire le rôle. Jusque-là, la méthode de choix était de comparer les profils d'accumulation des transcrits entre une plante sauvage et un mutant déficient pour la ribonucléase à étudier. Bien que couronnée de succès, cette approche est extrêmement biaisée dans le choix des transcrits analysés. A cause de l'impossibilité matérielle de réaliser des RNA blots sur l'ensemble du transcriptome, la plupart des études s'étaient focalisées sur les ARNr et quelques unités de transcription célèbres comme *psbB*, *rbcL* ou *psbA*. L'idée que j'avais en tête était qu'il était possible d'utiliser l'approche RNA-Seq pour étudier les ribonucléases. En théorie cela devait permettre de s'affranchir du biais d'échantillonnage des transcrits et d'avoir une vue sur l'intégralité du transcriptome. Encore fallait-il le montrer pour pouvoir éventuellement convaincre David de se lancer dans cette voie.

RNA-Seq et ribonucléases

J'ai donc utilisé le jeu de données déjà présent au laboratoire qui comparait les transcriptomes de plantes *col0* et mutantes pour la PNPase. Lors de sa thèse qu'il était en train de terminer, Arnaud Germain avait montré en utilisant abondamment les RNA blots que la PNPase permettait la maturation 3' d'une multitude d'ARN (Germain *et al.*, 2011). J'ai cherché à voir s'il était possible d'observer dans le RNA-Seq une signature correspondant aux cibles identifiées par Arnaud. L'intérêt de la petite taille du génome chloroplastique est que l'on peut se permettre d'inspecter visuellement tous les résultats de la bioinformatique. Il suffit de dessiner la couverture simple nucléotide pour toutes les positions du génome et de comparer l'expression entre le sauvage et le mutant. Avec le recul, je peux affirmer que si cela a fonctionné c'est parce que j'ai eu l'idée de représenter la couverture avec une échelle logarithmique ce qui permet d'identifier rapidement à l'œil des différences d'un facteur 10 entre les expériences (Figure 11).

J'ai ainsi pu observer un effet sur toutes les cibles précédemment identifiées, telles que l'ARNm de *rbcL*. Comme illustré par la Figure 11A, l'extension 3' de l'ARNm de *rbcL* visible en northern blot est également visible en utilisant la couverture simple nucléotide fournie par ChloroSeq. La couverture en 3' du gène est nettement supérieure dans le mutant en comparaison avec le WT. Les deux courbes commencent à diverger une dizaine de nucléotides en aval de la fin d'une structure tige boucle (Figure 11B) qui est connue pour arrêter *in vitro* l'activité de la PNPase (Yehudai-Resheff *et al.*, 2001). En comparant les profils transcriptionnels WT et *pnp1-1* sur tout le génome nous avons ainsi pu mettre en évidence et confirmer par northern blot de nouveaux transcrits ciblés par la PNPase comme

ndhJ par exemple (Figure 11C). Les transcrits sont plus longs dans le mutant car ils contiennent une extension en 3' qui n'est pas dégradée. Lorsque l'on fait un RNA blot en utilisant une sonde spécifique pour cette extension, le signal est uniquement visible dans le mutant (Figure 11A et 11C). Un autre résultat marquant aura été l'identification d'extensions 5' des ARNt dans le mutant que nous interprétons comme des ARN s'accumulant car non dégradés suite au clivage de l'extrémité 5' de l'ARNt par la RNaseP. (Castandet *et al.*, 2013)

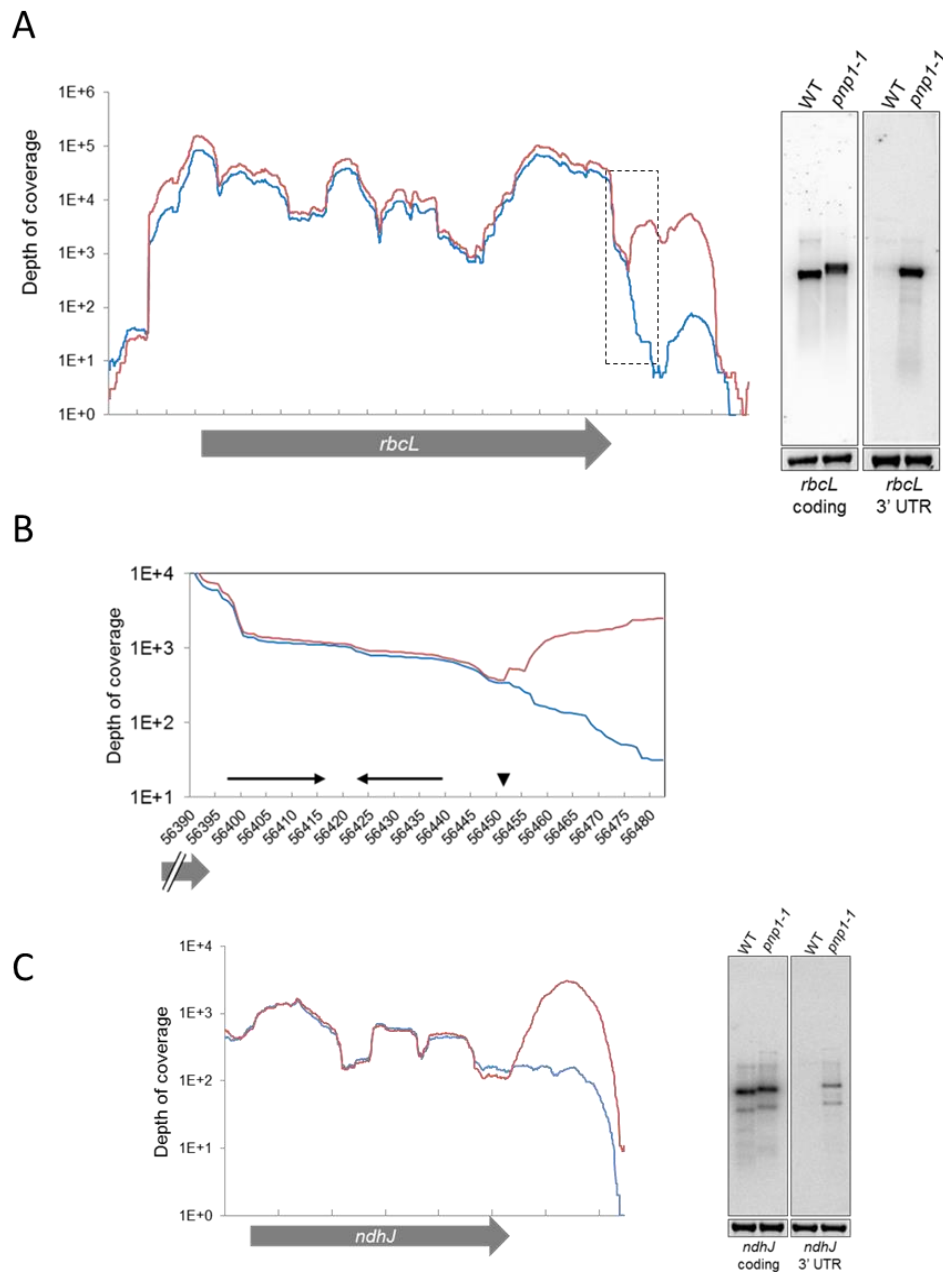


Figure 11: Le RNA-Seq prédit le rôle de la PNPase dans la maturation 3' des transcrits.

La profondeur de séquençage simple nucléotide pour *rbcL* (A et B) et *ndhJ* (C) est indiquée. Les northern blots correspondants sont sur la droite. Deux types de sondes ont été utilisés, une dans la région codante des gènes et une dans la partie 3' non codante. Les 2 flèches horizontales en B indiquent une tige boucle et la position où les deux courbes divergent est indiquée par un triangle noir.

Ces résultats étaient la confirmation que notre stratégie pouvait être utilisée pour découvrir le rôle de protéines de maturation de l'ARN jusqu'à présent inconnues. David était convaincu, il fallait donc maintenant passer à l'étape d'après, générer nos propres données sur une enzyme dont le rôle n'avait pas été décrit. Amber Hotto travaillait au même moment sur la caractérisation de la Mini ribonucléase III, Mini-III. Il avait été montré par l'équipe de Ciarán Condon que l'enzyme de *Bacillus subtilis* permettait la maturation 5' et 3' de l'ARNr 23S. Les deux extrémités s'apparient et forment une structure double brin qui est clivée endonucléolytiquement par la mini III (Redko *et al.*, 2008). Il existe deux paralogues de cette enzyme chez les plantes dont les protéines sont localisées dans le chloroplaste, RNC3 et RNC4. Alors que les simples mutants n'ont pas de phénotype, Amber avait montré qu'un double mutant *mnc3/4* possède un léger défaut de croissance. Comme chez *Bacillus*, les premières analyses moléculaires montraient également des défauts de maturation des ARNr chez le chloroplaste.

Mini III était donc le candidat idéal pour tester la puissance du RNA-Seq et nous nous sommes attelés à la tâche. L'analyse des résultats de RNA-Seq montre une plus grande expression de l'ARNr 4.5S dans le double mutant que dans le sauvage (Figure 12B). La validation de ce résultat par northern blot révèle que c'est en fait le précurseur de l'ARN 4.5S qui est plus abondant, la région intergénique 4.5S-5S étant toujours présente dans le mutant alors qu'elle est dégradée en condition contrôle (Figure 12C). D'autre part, l'utilisation de la couverture simple nucléotide nous a permis de mettre en évidence l'existence d'un ARN antisens à la région intergénique 4.5S-5S. La présence de cet ARN a été validée par northern blot et nous pensons qu'il pourrait être impliqué dans la maturation de l'ARNr sens 4.5S (Figure 12D). Enfin le RNA-Seq a également permis de montrer que Mini-III jouait un rôle dans la maturation 5' de l'ARN 23S (Figure 12E). Il est à noter que toutes les extensions prédites par RNA-Seq ont également été validées par des techniques moléculaires (RACE et cRT-PCR) et les résultats ont toujours été concordants. Nous avons développé un modèle de maturation de l'ensemble ARNr 23S-ARNr 4.5S qui est discuté en détail dans la publication correspondante (Hotto *et al.*, 2015). L'appariement de base entre la partie 5' du 23S et la région 3' du 4.5S crée un site de clivage pour Mini III qui mature donc les deux extrémités en même temps. Alternativement, l'ARN antisens à la région intergénique 4.5S-5S pourrait permettre la maturation de la partie 3' du 4.5S en créant également un appariement de base reconnu par Mini III (Figure 13).

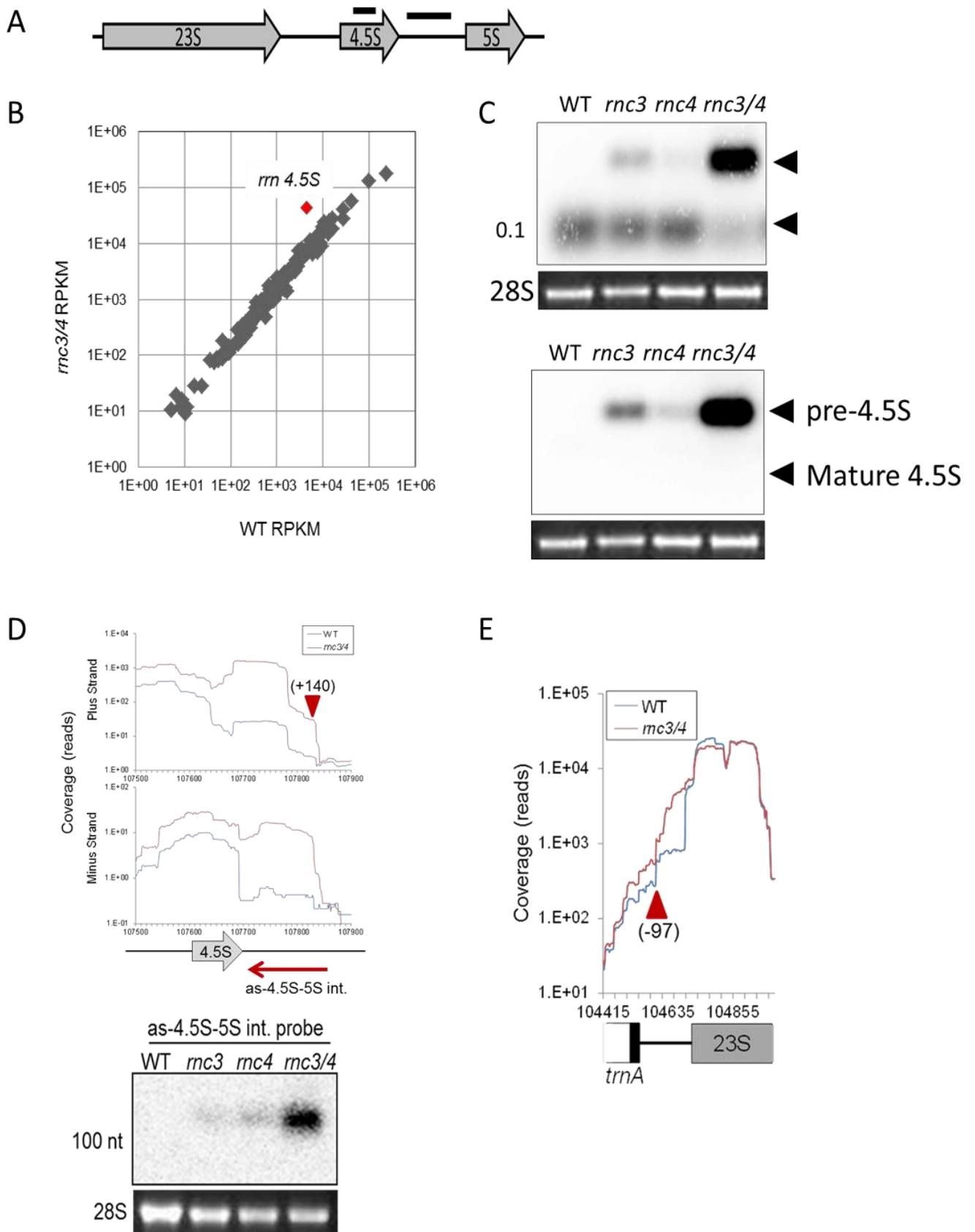


Figure 12: Le RNA-Seq révèle le rôle de Mini-III dans la maturation des ARNr du chloroplaste.

A) Schéma de l'enchainement des ARNr 23S, 4.5S et 5S. La localisation des sondes utilisées pour les northern blots est indiquée. (B) Expression différentielle des transcrits entre WT et *rnc3/4* révélée par RNA-Seq. L'ARNr 4.5S est indiqué en rouge. (C) Northern blots pour sur l'ARN 4.5S (haut) ou la région intergénique entre le 4.5S et le 5S (bas). (D) Couverture simple nucléotide de la région de l'ARN 4.5S sur les deux brins. Le northern blot est réalisé avec une sonde spécifique de l'ARN antisens identifié. E Couverture simple nucléotide de la région 5' de l'ARN 23S.

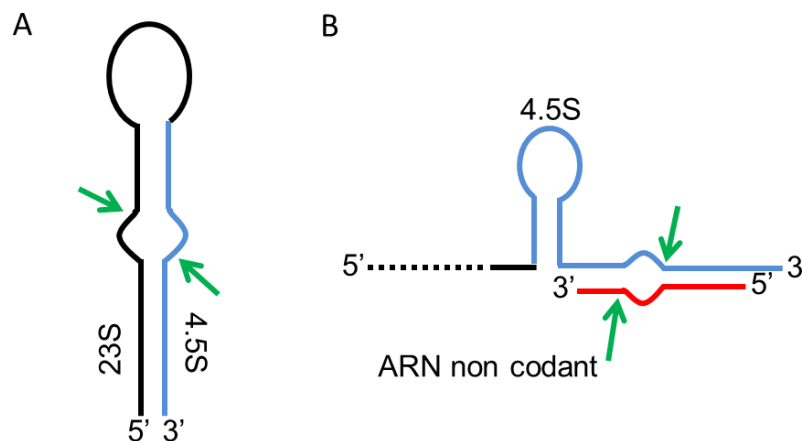


Figure 13: Modèles pour les clivages Mini III dans les ARNr 23S et 4.5S.

(A) Le modèle dit classique où les régions 5' de l'ARNr 23S et 3' de l'ARNr 4.5S s'apparient pour former une structure clivée par Mini III, formant ainsi les extrémités matures des deux ARN. (B) Modèle alternatif où un ARN antisens à la région 4.5S-5S vient s'apparier avec l'ARN issu du brin sens pour former une structure reconnue par Mini III qui peut former l'extrémité mature du 4.5S. Dans ce modèle, la formation de l'extrémité 5' de l'ARNr 23S est découplée de la maturation du 3' de l'ARNr 4.5S.

Enfin, l'aspect qui à posteriori est peut-être le plus intéressant du travail est une observation faite sur les introns. Le RNA-Seq semblait montrer un signal plus important pour les régions introniques dans le mutant *rnc3/4* et les investigations moléculaires nous ont permis de montrer que la Mini III était probablement impliquée dans la dégradation des introns suite à l'épissage. Nous nous sommes donc naturellement demandé si cela était aussi le cas chez la bactérie. L'équipe de Ciarán a exploré l'idée et a pu confirmer que la dégradation des deux introns de *Bacillus subtilis* était également ralentie dans une souche dépourvue de Mini III. Ce résultat, absolument loin d'être extraordinaire ou susceptible de susciter un immense intérêt, permet pourtant d'illustrer le gros avantage de notre approche : la connaissance apportée par l'analyse RNA-Seq dans le chloroplaste a permis de prédire et identifier dans le modèle bactérien un rôle de l'enzyme qui était jusque-là passé inaperçu. L'idée que les génomes des organites permettent de tester à peu de frais et de manière systématique des stratégies qui pourront ensuite s'appliquer à des échelles supérieures venait d'entrer dans ma tête et elle n'en est toujours pas sortie.

La démonstration de l'intérêt des approches RNA-Seq pour étudier le transcriptome chloroplastique et ses différents événements de maturation était donc faite. Une des limites de l'approche est cependant qu'elle ne donne accès qu'au pool d'ARN mature, celui que l'on peut traditionnellement étudier en réalisant des northern blots. Il ne permet pas d'étudier directement la transcription ou d'identifier avec précision les extrémités des ARN. Cette question est pourtant fondamentale lorsque l'on veut avoir une compréhension fine du rôle d'une ribonucléase par exemple. J'ai donc participé (avec Arnaud Germain et Amber Hotto) à la mise au point d'une technique de séquençage systématique des extrémités 5' et 3' des ARN que nous avons appelé le Terminome Sequencing (Terminome-Seq).

Il est temps de rentrer

Mon postdoc aura donc été marqué par l'appropriation du RNA-Seq, de la construction des banques de séquençage jusqu'à l'analyse bioinformatique des résultats. Je serais ingrat si je ne remerciais pas David de m'avoir ainsi laissé prendre le temps de me former, probablement au détriment d'une « productivité bibliométrique » (expression que j'abhorre) plus importante. Cette expérience États-Unienne aura également été l'occasion de me confronter à un mode de fonctionnement de l'Université et de la recherche totalement différent de ce que j'avais pu observer jusque-là. David étant directeur d'institut et très ouvert à la discussion, j'ai pu avoir accès, sans filtre ou intermédiaire, à toute la vision managériale de la recherche qui prévaut là-bas et se fait plus qu'insistante maintenant ici aussi : l'obsession court-termiste du « grant », la cristallisation de toute l'entreprise scientifique autour de la seule figure du « PI », le rôle des fondations privées de mécénat dans le financement ou la disparition annoncée des postes « tenured » au profit de contrats renouvelables. J'ai pu constater les dissonances entre les pratiques et convictions, les tiraillements que cela pouvait provoquer ainsi que ce sentiment de malaise joliment diagnostiqué par d'autres (Lazebnik, 2015).

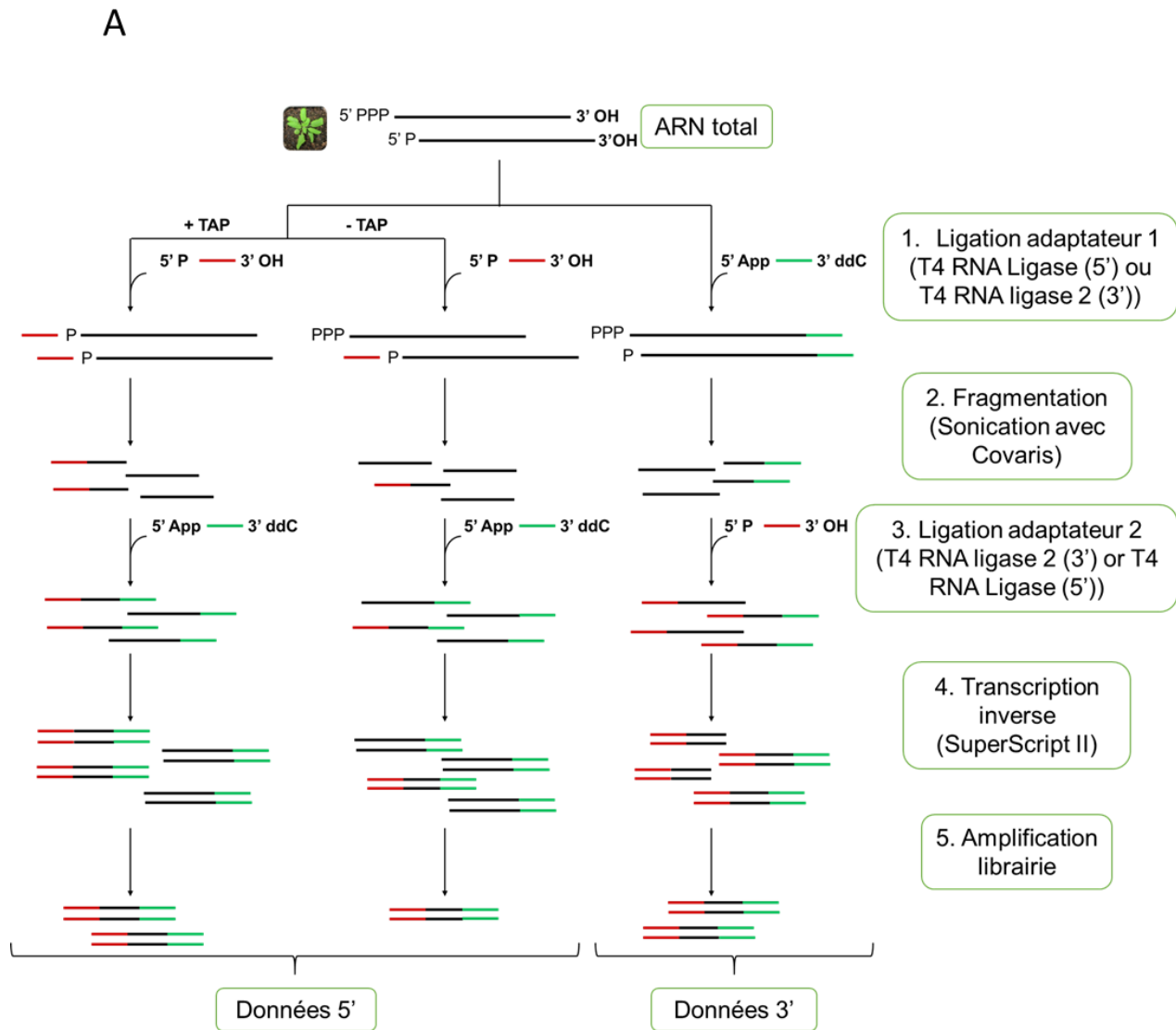
Une des conséquences du financement sur projet est qu'il peut être soumis aux aléas des changements de direction stratégique de l'organisme financeur. Le département de l'énergie américain (DOE) a subitement perdu tout intérêt pour les thématiques de recherche portant sur l'expression génétique des chloroplastes. Cette toute nouvelle précarité financière a coïncidé avec l'ouverture d'un poste de maître de conférences à l'Université de Paris qui correspondait à mon profil. J'ai donc traversé l'Atlantique en emportant avec moi les données de Terminome-Seq afin de pouvoir les analyser sur le plateau de Saclay.

Retour en France

Le Terminome-Seq

Depuis Octobre 2017, je suis maître de conférences en physiologie végétale à l'Université de Paris et j'ai rejoint l'équipe « OGE » animée par Claire Lurin à l'institut des plantes de Paris-Saclay (IPS2). Que l'on laisse trainer ses oreilles dans les couloirs de l'Université ou que l'on discute directement avec des collègues, on est immédiatement frappé par la quantité d'histoires relatant des intégrations problématiques (j'emploie un euphémisme) dans les équipes de recherche. A rebours de ces expériences, je dois m'estimer chanceux d'avoir atterri dans le collectif OGE. Je n'ai jamais été considéré comme une paire de main au service du projet d'un chef d'équipe omnipotent. Claire aime à décrire son rôle comme devant s'assurer que « la mayonnaise prene » (ce sont ses mots !) et j'ai bénéficié d'une totale liberté dans la mise en place de mes thématiques. J'ai donc pu me consacrer pleinement à l'analyse des extrémités des ARN chloroplastiques en utilisant les données du Terminome-Seq.

L'idée originale revient à Arnaud Germain. Pour identifier une extrémité 5' ou 3', il faut passer par un protocole de fabrication de librairie de séquençage qui utilise une ligation d'adaptateurs sur l'ARN. Après transcription inverse, ces adaptateurs sont utilisés pour les étapes d'amplification. Lors de l'analyse, le nucléotide situé immédiatement après l'adaptateur est donc celui sur lequel la ligation s'est faite et correspond à une extrémité ARN. Le problème est que les protocoles de fabrication de bibliothèques se font sur des ARN préalablement fragmentés, noyant les extrémités « réelles » dans un océan d'extrémités « artificielles », nouvellement créées. L'alternative a été de lier le premier adaptateur avant la fragmentation des ARN afin de sauvegarder l'information sur les extrémités natives. Il faut donc faire 2 types de bibliothèques, une pour les extrémités 5' et une pour les 3' (Figure 14A). Un autre souci est la distinction entre les extrémités 5' directement issues de la transcription (dites primaires ou TSS) et celles créées après maturation par une ribonucléase (dites processées ou secondaires). L'adaptateur 5' se lie uniquement aux extrémités 5' monophosphates, cela entraîne une sous-représentation des extrémités primaires qui contiennent 3 phosphates. Une digestion des extrémités primaires par l'enzyme TAP permet de libérer 2 phosphates et donc de les rendre compatibles pour la ligation (Figure 14B).



B

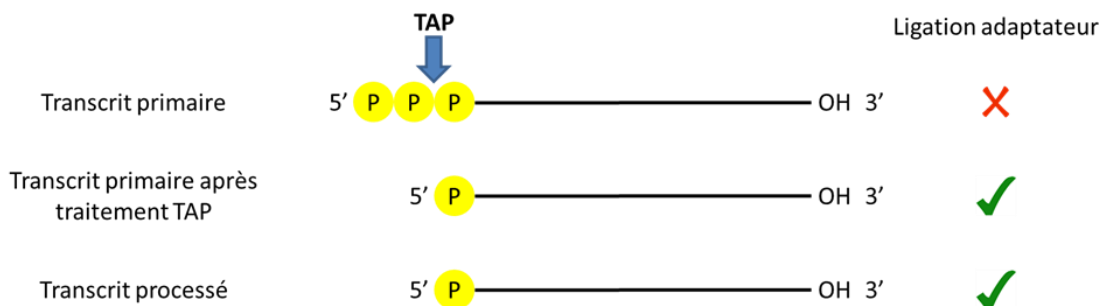


Figure 14: Protocole du Terminome-Seq

(A) Trois bibliothèques de séquençage différentes sont créées à partir d'ARN totaux de feuilles d'Arabidopsis. Pour les bibliothèques 5', l'adaptateur 5' est ligé avant la fragmentation des ARN et la ligation de l'adaptateur 3'. Pour les bibliothèques 3', c'est l'inverse : l'adaptateur 3' est ligé avant la fragmentation et la ligation de l'adaptateur 5'. Cela permet de s'assurer que les ADNc amplifiés représentent bien une extrémité qui était présente au départ dans le transcriptome. (B) Pour pouvoir séquencer les extrémité 5' primaires qui contiennent 3 phosphates il est nécessaire d'utiliser un traitement enzymatique à la TAP avant la première ligation.

L'avantage d'utiliser un traitement TAP est que l'on peut identifier les sites d'initiation de la transcription (TSS) en comparant l'abondance des extrémités obtenues avec ou sans traitement. Avec cette simple approche et en utilisant des filtres conservateurs nous avons pu identifier 215 TSS dans le génome chloroplastique. Ces résultats sont en accord avec ceux déjà obtenus chez l'orge et démontrent que les unités de transcription multi géniques classiquement décrites contiennent toutes des sites internes d'initiation de la transcription (Zhelyazkova *et al.*, 2012; Castandet *et al.*, 2019). Nous identifions également des évènements d'initiation massifs dans des régions inattendues comme *ndhF* (sur les deux brins) et l'intron 1 de *clpP* (Castandet *et al.*, 2019).

En plus des TSS, nous trouvons environ 1600 et 1300 extrémités 5' et 3', souvent organisées en clusters. Beaucoup se trouvent immédiatement en amont ou aval des régions correspondants aux empreintes ARN laissées par les protéines de liaison à l'ARN (PPR par exemple) ou les structures secondaires (Ruwe *et al.*, 2016; Castandet *et al.*, 2019). De notre point de vue, cela confirme, à l'échelle du transcriptome chloroplastique, le modèle de maturation des ARN où ces protéines et structures permettent de désigner les lieux de clivage de l'ARN et de protéger les extrémités contre la dégradation par les exo-ribonucléases.

Terminome-Seq avec des mutants de ribonucléases

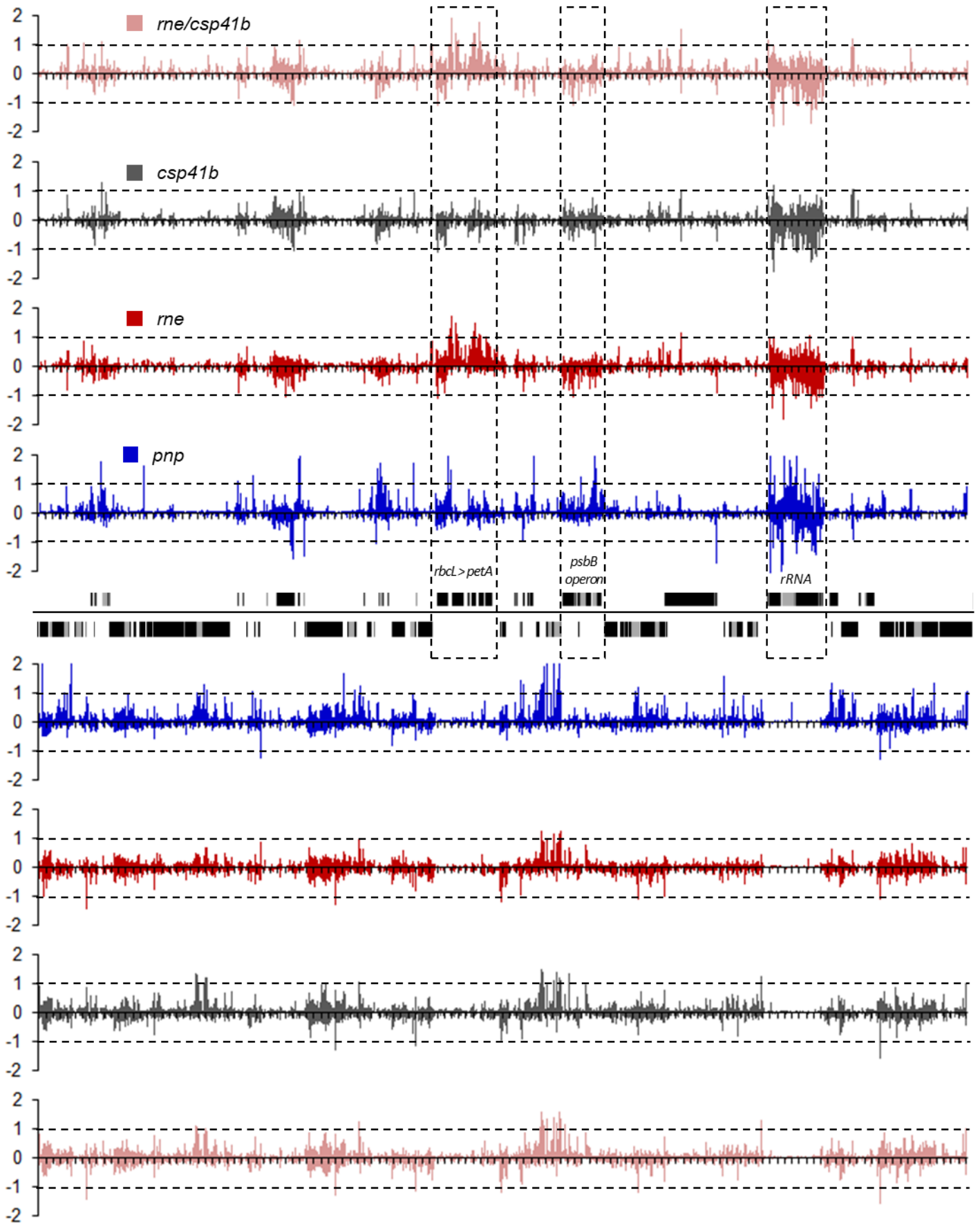
Cette analyse est renforcée par les résultats obtenus dans le mutant pour la PNPase. Comme attendu, nous trouvons un effet prédominant pour les extrémités 3' et nous confirmons que les extensions 3' précédemment identifiées (voir par exemple la Figure 11) sont corrélées avec l'apparition de nouvelles extrémités 3'. Cela est particulièrement visible pour certaines régions génomiques, comme par exemple l'unité de transcription *psbB* ou *rbcL* (Figure 15). La même stratégie a été appliquée à des mutants pour la RNase E, CSP41 et un double mutant *rne/csp41*. En supprimant ainsi des endo-ribonucléases majeures nous nous attendions à observer des profils d'extrémités complètement perturbés. L'analyse montre au contraire des profils bien moins différents du sauvage que ce que nous pouvons observer avec le mutant de la PNPase. La RNase E semble avoir un rôle majeur dans la maturation 3' de *rbcL* et au-delà dans toute l'unité de transcription qui s'étend de *accD* jusqu'à *petA*. Son interaction décrite avec RHON1 dans cette région-là suggère qu'elle pourrait être impliquée dans la terminaison de la transcription (He *et al.*, 2015; Stoppel *et al.*, 2012). CSP41 a un rôle plus confidentiel et beaucoup d'extrémités internes aux ARNt semblent touchées. Il est à noter que les effets des mutations sur les profils d'extrémités sont

plutôt additifs. Le Terminome-Seq du double mutant *rne/csp41* ressemble en effet fortement à une simple addition des profils des motifs individuels (Figure 15).

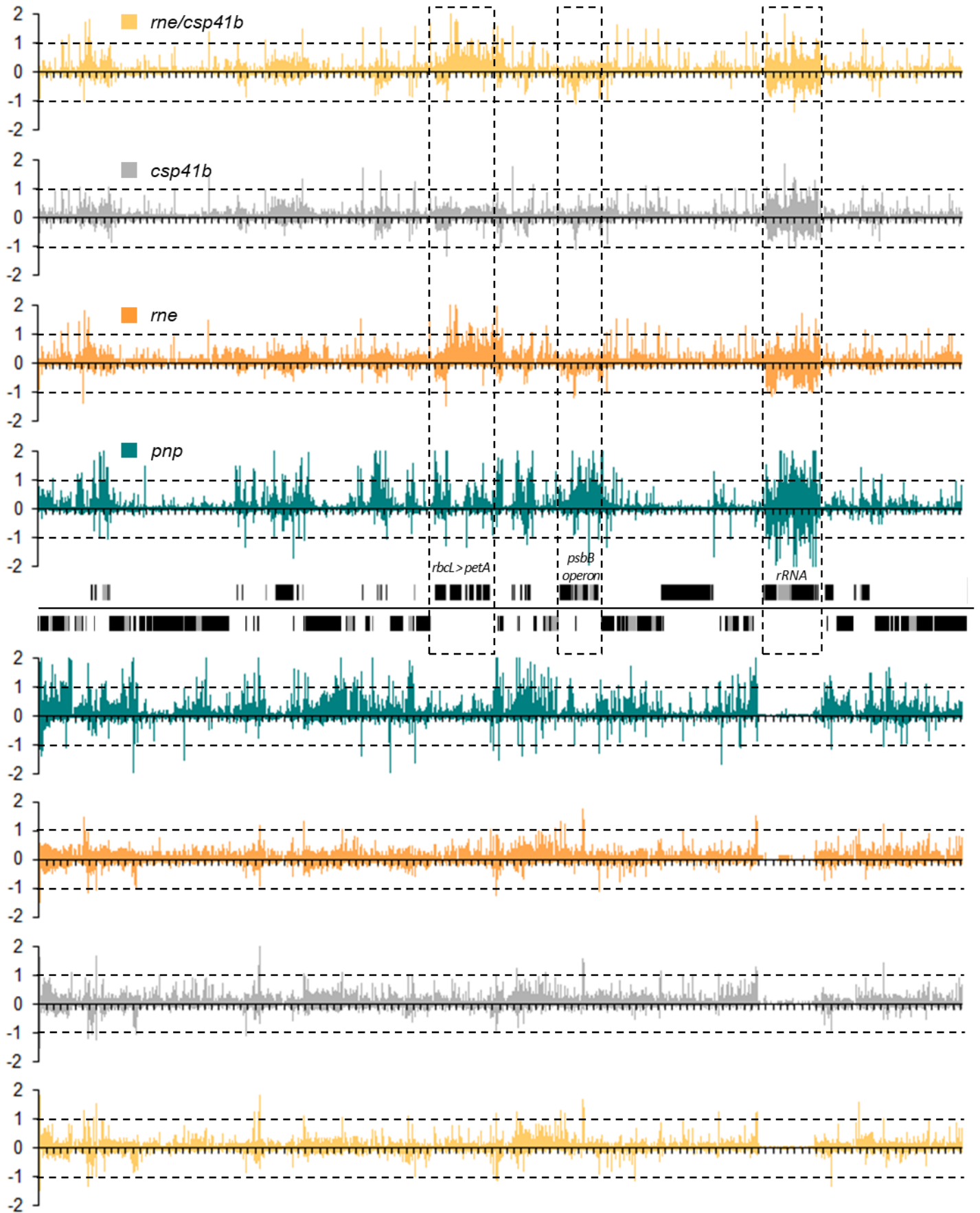
Figure 15: Représentation à l'échelle du génome chloroplastique de l'influence des génotypes *pnp*, *rne*, *csp41b* et *rne/csp41b* sur l'accumulation des extrémités A) 5' et B) 3' en comparaison avec le sauvage (*col-0*).

L'axe des Y représente le $\log_{10} [(mutant\ RPM + 1)/(WT\ RPM + 1)]$. Les gènes sont indiqués entre les pistes qui correspondent aux brins + et - du plastome. Une copie de la région inversée répétée a été omise. Les graduations horizontales sont tous les 1000 nucléotides.

A 5' ends



B 3' ends



Bilan

Il serait difficile de nier que l'expression génétique des organites a constitué le cœur de mon activité de recherche. Si la thématique est restée constante, les compétences, méthodes et modèles utilisés tout au long de mon parcours ont constamment évolué. Lors de ma thèse, j'ai utilisé une approche de biochimie traditionnelle basée sur le fractionnement cellulaire (purification de mitochondries) couplée à de la biologie moléculaire pour amplifier des transcrits spécifiques. Mes modèles d'étude, blé et pomme de terre, étaient justifiés par l'approche biochimique mais ils souffraient du manque de ressources dédiées (mutants, génomes...). Au contraire, les approches RNA-Seq mises au point pendant le post-doctorat ont toutes été appliquées à *Arabidopsis thaliana*, ce qui permet de profiter de l'incroyable richesse des ressources disponibles pour cette plante. Je suis maintenant dans la position où je maîtrise l'intégralité de la boîte à outils génomiques permettant d'étudier toutes les étapes de l'expression génétique des organites, de la transcription à la traduction.

Je suis entré dans la transcriptomique par une voie dérobée, celle d'un biologiste moléculaire. Pour moi, la faible taille du génome chloroplastique était un avantage car elle permettait de ne pas se départir d'une approche empirique et de regarder « à l'œil » l'intégralité des événements identifiés. J'ai ainsi pu passer des heures à regarder et comparer des profils de couverture, nucléotide par nucléotide. Il est ensuite facile d'aller valider les résultats prédits ou identifiés par la transcriptomique (et l'œil) avec des techniques de biologie moléculaire. Nous avons par exemple pu exclure la présence de faux positifs parmi les nouveaux sites d'édition du transcriptome chloroplastique découverts par RNA-Seq, tâche impossible pour un transcriptome nucléaire et à l'origine de controverses sans fin (Li *et al.*, 2012; Pickrell *et al.*, 2012; Kleinman and Majewski, 2012; Lin *et al.*, 2012; Li *et al.*, 2011). De la même manière, les événements de maturation identifiés dans les mutants de la PNPase et de Mini III l'ont tous été grâce à une inspection visuelle des profils de couverture et cela nous a même permis de prédire un événement encore inconnu chez les bactéries. La contrepartie, c'est qu'en plus d'être longues et fastidieuses, ces inspections visuelles nécessitent une bonne connaissance du comportement des profils de couverture pour pouvoir être analysées. En somme, alors qu'à la fin de ma thèse j'avais touché du doigt les limites de l'approche moléculaire classique je suis maintenant confronté aux limites de l'approche haut débit lorsqu'elle est réalisée par un expérimentateur qui réfléchit toujours en bas débit.

Mon arrivée à l'IPS2 a été l'occasion de dépasser ces limitations en bénéficiant de la longue histoire de collaboration entre les équipes OGE et GNet (Réseaux Génomiques, animée par Marie-Laure Martin Magniette). L'équipe est spécialisée dans le développement

de méthodes statistiques dédiées au traitement des données haut débit et cogère la plateforme de transcriptomique de l'institut avec OGE, un atout non négligeable pour mon activité de recherche. Mon idée première est bien sûr de continuer sur la même voie que celle entreprise jusqu'à présent mais d'y associer la puissance des approches de modélisation. Je me place résolument à l'interface entre une vision moléculaireiste de la biologie, que d'aucuns qualifieraient de réductionniste, et une vision plus holistique, de système, qui utilise les outils omiques sans pour autant les fétichiser.

Mon but sera d'éviter deux écueils. Le premier est ce que Carl Woese a décrit comme étant le fondamentalisme réductionniste qui transforme la biologie en une science de l'ingénieur. Arrivée au bout de sa vision d'ensemble, la biologie moléculaire se complait maintenant dans la résolution d'étroites questions pratiques. Bien que ce chemin continue de faire la démonstration de sa productivité, surtout à l'aune des indicateurs de performance qui ont le vent en poupe actuellement, il tourne sciemment le dos au rôle fondamental de la biologie qui est de comprendre le vivant (Woese, 2004). Le deuxième est l'avènement de la science sans hypothèses, cette dernière n'étant pas autre chose qu'un handicap pour reprendre un point de vue exprimé dans *Genome Biology* (Yanai and Lercher, 2020). Avec le déluge des données omiques, l'hypothèse serait en effet devenue superflue, la recherche de corrélation aurait rendu la méthode scientifique obsolète (Anderson, 2008; Mazzocchi, 2015).

Le chemin tracé serpente sans cesse entre ces deux difficultés. Il n'y a plus qu'à se mettre en route.

Projet de Recherche

L'écriture d'un projet de recherche s'accorde difficilement avec ce qui m'a attiré dans la carrière de maître de conférences. L'exercice a peut-être des vertus mais noircir des pages d'AAP en adaptant l'écriture aux désirs du financeur est totalement antinomique de la manière dont j'envisage le métier. Mon projet, c'est de suivre mes envies, d'aller explorer les choses qui m'attirent et ce pour la seule raison qu'elles piquent ma curiosité. On l'aura compris, si la thématique, les objectifs et les méthodes que je vais présenter correspondent en tout point à ce que je vais faire (ou essayer de...) ces prochaines années, l'écriture de la partie projet doit être vue comme un exercice de style plutôt que comme l'expression de ma vision quant à ma pratique future.

La thématique dominante est toujours la même. Je veux comprendre le contrôle qualité des ARN des organites. Mon but est d'arriver à la fois à une description la plus fine possible de ce qu'il se passe suite à la transcription (à quoi ressemble vraiment le transcriptome ? les événements de maturation sont-ils connectés ? comment ?) et de comprendre les mécanismes moléculaires (quelles enzymes sont impliquées ? comment agissent-elles ?) qui permettent de produire ce transcriptome mature. Ma stratégie sera de tirer profit au maximum des capacités des outils de séquençage et des méthodes statistiques de modélisation.

Etude du transcriptome chloroplastique

Grâce aux données récentes de transcriptomique, il est maintenant clair que la quasi-totalité du génome chloroplastique est transcrit (Sanità Lima and Smith, 2017). Avec le Terminome-Seq nous avons pu montrer qu'il y avait des centaines de sites d'initiation de la transcription répartis sur tout le génome ce qui, avec le concours de l'inefficacité des mécanismes de terminaison, donne un transcriptome primaire extrêmement complexe (Castandet *et al.*, 2019; Stern and Gruijsem, 1987). Il suffit de penser à l'unité de transcription de *psbB* qui produit plus de 20 transcrits différents pour en être convaincu. Le chloroplaste contient aussi des centaines d'ARN non codants dont certains peuvent former des duplex ARN-ARN avec les transcrits géniques (Hotto *et al.*, 2011; Sharwood *et al.*, 2011). La question de savoir s'ils ne sont qu'un bruit de fond de l'activité transcriptionnelle ou si au contraire ils ont une véritable fonction biologique est toujours ouverte. Enfin, comme expliqué plus haut, certains transcrits doivent en plus être édités ou épissés.

En dépit du fort développement des approches RNA-Seq, nous ne connaissons toujours pas l'identité des isoformes des transcrits chloroplastiques. Le RNA blot est encore et toujours la méthode de choix pour résoudre la complexité d'une unité de transcription et l'outil ne donne malheureusement accès qu'à la taille du transcrit, pas son identité. Il n'est donc pas adapté pour une étude à l'échelle du génome entier. La technologie Illumina

séquence uniquement des petits fragments qui doivent ensuite être réassemblés bioinformatiquement, une limite difficilement surmontable pour un génome condensé comme celui du chloroplaste. Le Terminome-Seq a permis d'identifier les différentes extrémités des ARN mais nous avons perdu l'information sur les associations d'extrémités : Quelle extrémité 5' avec quelle extrémité 3' ? Si nous n'avons jamais été aussi loin dans la connaissance du transcriptome chloroplastique, le problème de l'identité des isoformes reste entier.

La solution est de pouvoir séquencer à haut débit des ARN pleine longueur et c'est justement la promesse de la technologie nanopore (Parker *et al.*, 2020). La préparation des bibliothèques de séquençage nécessite cependant une queue polyA en 3' des ARN ce qui la rend non utilisable pour les transcrits des organites de plante. Avec Etienne Delannoy, nous avons donc décidé d'utiliser une méthode alternative basée sur la ligation d'un oligonucléotide en 3' des ARN, comme dans le protocole du Terminome-Seq. Cette séquence est ensuite utilisée pour initier la transcription inverse et obtenir des ADNc pleine longueur. La mise au point des détails du protocole aura été un effet bénéfique du confinement du printemps 2020. Etienne est ensuite passé de la théorie à la pratique et a réussi à séquencer des bibliothèques de plantes col0 et mutantes pour la PNPase. Ce mutant doit nous permettre de valider le fait que l'on peut bien identifier des différences de maturation dans les ARN puisque nous connaissons déjà son comportement. Notre but à terme est de répondre à la question que je me pose depuis ma thèse et de vérifier les éventuelles interdépendances entre les événements de maturation.

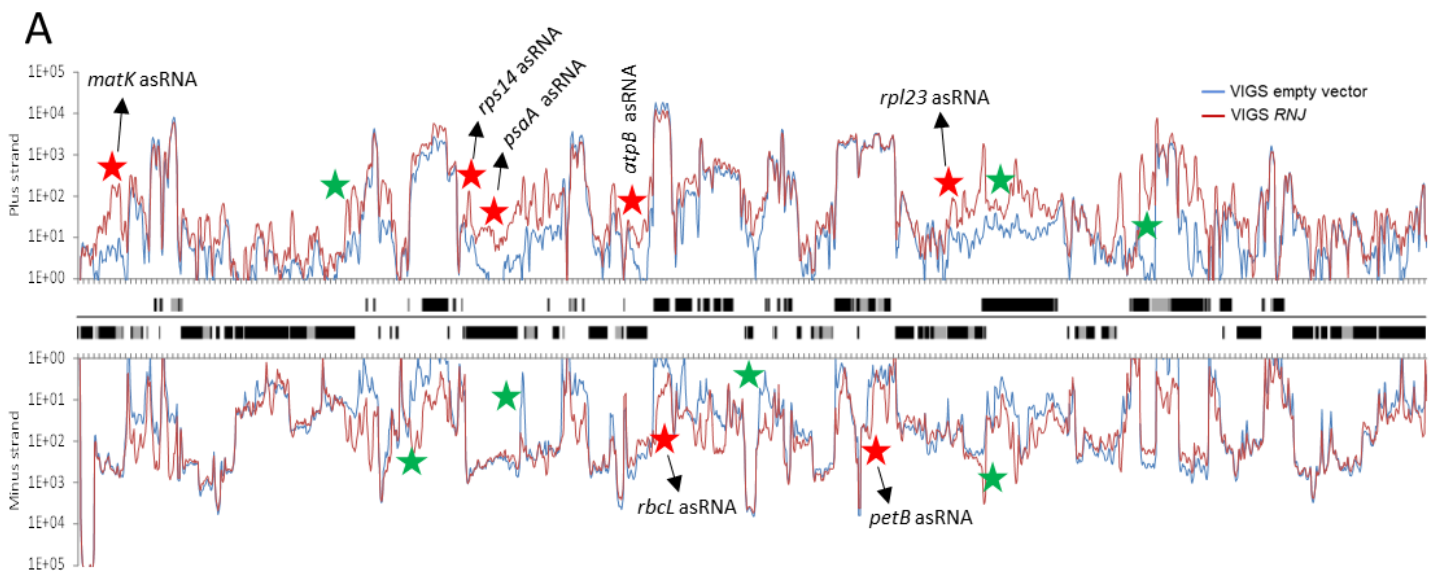
Ce travail, à l'image de ce que j'envisage pour le reste de mon projet, est une œuvre collective qui regroupe les membres des équipes OGE et GNet. Etienne a produit les données, Marine Guilcher (doctorante dans l'équipe) est en charge de l'analyse et les aspects bioinformatiques et statistiques sont principalement traités par Guillem Rigail, Arnaud Lierhmann et Chloé Seyman (stage de L3) membres de GNet. Symbole de cette collaboration, je codirige avec Guillem la thèse d'Arnaud. Il développe DiffSegR, une méthode qui a pour but d'automatiser la détection des événements de maturation différentielle dans des profils de RNA-Seq. En gros, il fait automatiquement et en quelques secondes ce qui m'a pris des mois d'observation pour arriver aux Figures 11 et 12. Ce travail va être essentiel à l'exploitation rapide et rigoureuse des données de RNA-Seq qui sera menée lors du projet.

La RNase J et la production du transcriptome mature

Ce transcriptome chloroplastique, que la première partie de mon projet de recherche devrait permettre de connaître avec une grande précision, est le résultat de l'action des ribonucléases indiquées dans la Figure 10. Elles ont pour rôle essentiel de dégrader les ARN non fonctionnels tout en préservant les futurs transcrits matures. Les mécanismes de ce contrôle qualité capable de cibler ou d'exclure certains transcrits pour la dégradation ne sont pas connus, d'autant plus que les ribonucléases n'ont *a priori* pas de spécificité de séquence. Il a été montré que certaines étaient présentes au sein de complexes de haut poids moléculaire et une hypothèse communément avancée est que la spécificité de reconnaissance pourrait être assurée par des partenaires protéiques membres de ces complexes (Hayes *et al.*, 1996; Stoppel *et al.*, 2012). Alors que les rôles de la PNPase et de RNR1 dans la dégradation des transcrits assistée par la polyadénylation sont plutôt bien compris, l'identité de la (les) ribonucléase(s) assurant le premier clivage endonucléolytique est toujours discutée (MacIntosh and Castandet, 2020). Les résultats de Terminome-Seq présentés dans la Figure 15 semblent par contre contredire un rôle majeur pour la RNase E et CSP41. Reste à explorer le rôle de la RNase J.

La plupart des Rnase J décrites possèdent à la fois une activité endo- et 5'-3' exo-ribonucléase. Elle est de plus en plus considérée comme l'acteur clé du contrôle qualité des ARN chloroplastiques, avec la PNPase (Hotto *et al.*, 2020). Chez les embryophytes, l'enzyme est plus longue que les homologues d'algues, de bactéries ou d'archaea grâce à une extension C-terminale qui est similaire au domaine de liaison à l'ADN GT-1 généralement trouvé sur des facteurs de transcription. Ce domaine reconnaît une séquence dégénérée qui pourrait expliquer une spécificité de reconnaissance pour certains transcrits (Hotto *et al.*, 2020). Les mutants homozygotes de l'enzyme sont létaux chez *Arabidopsis* et les seules expériences fonctionnelles ont dû être réalisées par Virus-Induced Gene Silencing (VIGS). Il a pu être montré que la RNase J permet la maturation des extrémités 5' des transcrits chloroplastiques et que les plantes déficientes accumulent de nombreux ARN antisens (Sharwood *et al.*, 2011; Luro *et al.*, 2013).

Rob Sharwood, auteur des premières expériences sur la RNase J chez David et maintenant en poste à la Western Sydney University, m'avait fait parvenir en 2016 des données de RNA-Seq issues de plantes contrôles et VIGS pour l'enzyme. L'analyse du transcriptome montre clairement une augmentation de l'accumulation des ARN en amont des gènes connus ainsi qu'une grande augmentation des ARN antisens (Figure 16). Alors que ces derniers sont normalement bien moins abondants que les ARN sens correspondants en condition sauvage, les plantes dépourvues de RNase J voient leur quantité augmenter très fortement (Figure 16B). En théorie, cette augmentation pourrait conduire à la production de duplex d'ARN qui pourraient inhiber la traduction de transcrits essentiels. L'ensemble des données préliminaires semble bien confirmer que la RNase J a un rôle fondamental dans les mécanismes de contrôle qualité des ARN du chloroplaste.



B

gene	RNJ effect on sense/antisense
<i>psaA</i>	16
<i>rpl23</i>	16
<i>atpB</i>	16
<i>matK</i>	15
<i>rpl32</i>	14
<i>rps14</i>	12
<i>rpl16_e1</i>	10
<i>psaI</i>	10
<i>rbcL</i>	10

Figure 16: Accumulation des transcrits chloroplastiques dans les plantes contrôles et VIGS pour la RNase J.

A) Couverture en RNA-Seq obtenue sur des fenêtres chevauchantes de 100 nt dans des plantes contrôles (bleu, infectées avec un vecteur vide) et VIGS pour la RNase J (rouge). Les gènes (exons en noir et introns en gris) sont indiqués entre les pistes qui correspondent aux brins + et - du plastome. Une copie de la région inversée répétée a été omise. Les étoiles rouges signalent les régions où une suraccumulation d'ARN antisens modifie fortement le ratio sens/antisens. Les étoiles vertes signalent des régions, souvent en 5' de transcrits décrits, avec une grosse différence de couverture entre les plantes témoins et VIGS.

B) Liste des gènes dont le ratio sens/antisens est le plus fortement perturbé dans les plantes VIGS en comparaison des plantes témoins. Par exemple, le ratio sens/antisens des transcrits *psaA* et *rbcL* est respectivement 16 et 10 fois plus faibles dans les plantes VIGS que les témoins.

Etude structure-fonction de la RNase J

Une analyse fine du rôle de la RNase J dans la maturation des ARN chloroplastiques nécessite de faire des complémentations avec des versions modifiées de l'enzyme. Nous allons pour cela transformer un mutant hétérozygote de l'enzyme (*rnj-2*, line CS 815990) avec différentes versions de la RNase J contenant un épitope FLAG. Les plantes transformées T0 contenant à la fois la mutation et le transgène d'intérêt seront autofécondées afin d'obtenir des mutants homozygotes *rnj-2* complémentés par la version testée de la RNase J. La condition nécessaire est que la version de l'enzyme utilisée pour la complémentation restaure la viabilité. Afin de s'affranchir de la létalité du mutant homozygote, une des constructions exprimera la version sauvage de l'enzyme sous la dépendance du promoteur ABI3 qui est actif uniquement lors de l'embryogenèse (Despres *et al.*, 2001). Cela devrait permettre d'obtenir un mutant total à partir de la fin du développement embryonnaire qui pourra ensuite être utilisé pour des croisements avec des plantes portant des versions de la RNase J qui ne restaurent pas la viabilité. Les différentes lignées envisagées sont présentées dans la Figure 17.

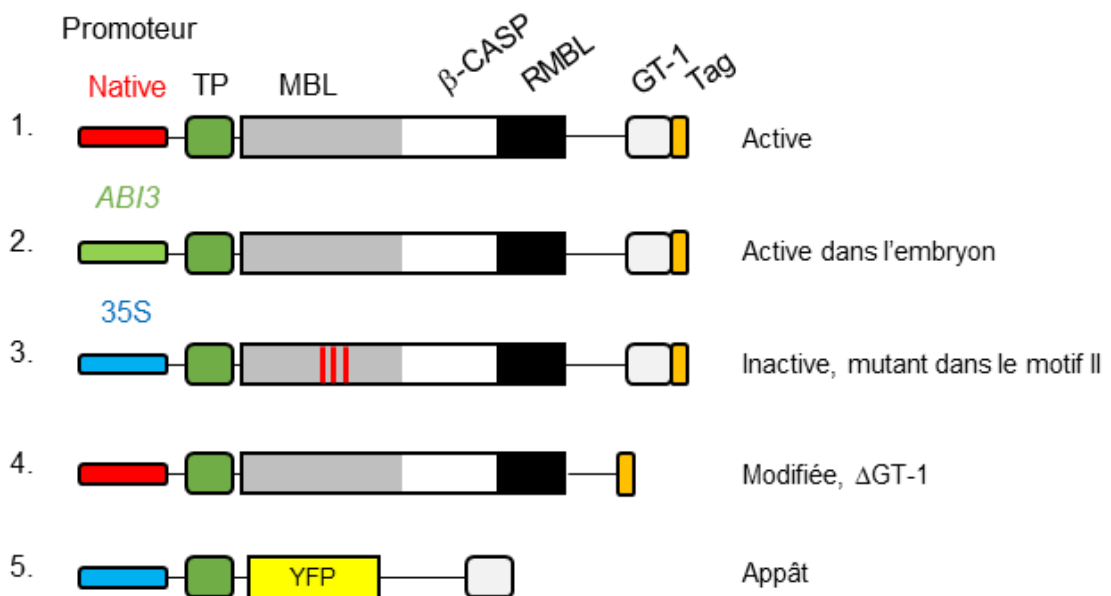


Figure 17: Les différentes constructions utilisées pour transformer le mutant hétérozygote *rnj-2*

Trois types de promoteur vont être utilisés, natif (rouge), spécifique du développement embryonnaire (ABI3 en vert) et constitutif fort (35S en bleu). Le transit peptide (TP) permet l'adressage au chloroplaste. La construction 3 est inactive grâce à 3 mutations dans le site actif (H177A, D179K, H180A) représentées par des barres rouges. La construction 5 sera utilisée pour élucider le rôle du domaine GT-1.

L'ensemble des lignées créées sera soumis à une analyse poussée des phénotypes ARN en utilisant le RNA-Seq et le Terminome-Seq. Nous pourrions ainsi vérifier l'accumulation des ARN à l'échelle du plastome et identifier les apparitions et disparitions d'extrémités entre les différentes lignées. Nous devrions être capables de documenter l'intégralité des événements de maturation produits par la RNase J et finalement découvrir si elle est impliquée dans la coupure endonucléolytique des ARN. Enfin, même si je ne développe pas cet aspect ici, il est évident qu'en fonction des résultats obtenus certaines lignées seront utilisées pour les approches de séquençage nanopore et ARN double brins décrites plus haut. Les informations obtenues avec les différentes lignées illustrées dans la Figure 17 devraient être les suivantes :

La lignée 1 va servir de contrôle WT pour les comparaisons avec les lignées portant des versions modifiées de la RNase J.

La lignée 2 produira la RNase J uniquement durant l'embryogenèse et les feuilles devraient donc être dépourvues de l'enzyme. C'est la lignée qui servira de référence mutante et la comparaison des phénotypes ARN avec la lignée 1 sera essentielle pour la compréhension du rôle de l'enzyme. Je m'attends à voir de multiples extensions en 5' des transcrits et une augmentation des ARN non codants.

Dans la lignée 3 l'enzyme sera inactive mais retiendra ses capacités de liaison à l'ARN (ou ADN). La comparaison avec les profils des lignées 1 et 2 devrait donc nous permettre de vérifier si la liaison seule au substrat ou la présence du domaine GT-1 en tant que tel permet une réversion partielle du phénotype mutant.

Le rôle du domaine GT-1 sera étudié en comparant les profils de la lignée 4 (qui en est dépourvu) avec ceux obtenus pour les lignées 1 et 2. Il a été montré *in vitro* que le domaine n'était pas essentiel à la catalyse (Halpert *et al.*, 2019) mais il pourrait jouer un rôle important dans la spécificité de clivage.

Identification des partenaires de la RNase J

Il a déjà été montré en modèle bactérien que la RNase J interagissait avec de multiples partenaires protéiques (Płociński *et al.*, 2019; Mu *et al.*, 2019). L'identification de ces partenaires est alors essentielle pour arriver à une compréhension globale du fonctionnement de l'enzyme. Des résultats préliminaires obtenus par l'équipe de Gadi Schuster (Tecnion) suggèrent par exemple que certaines protéines de liaison à l'ARN pourraient guider l'enzyme et améliorer ses capacités catalytiques. Les partenaires protéiques seront identifiés en utilisant deux approches complémentaires, la co-immunoprécipitation *in vivo* et le crible double hybride de levure.

Ce crible sera réalisé sur la plateforme InterATOME de l'IPS2 que nous gérons au sein de l'équipe OGE et qui permet de tester les interactions avec plus de 12000 ORF d'Arabidopsis. Au moins trois constructions différentes seront utilisées comme appâts. Une sera bien sûr la version complète de l'enzyme, sans le peptide d'adressage. Les deux autres viseront plus spécifiquement à étudier le rôle du domaine GT-1. Nous préparerons donc une construction correspondant à la RNase J sans le domaine GT-1 et une construction correspondant au domaine seul. L'approche par immunoprécipitation sera faite *in vivo* en utilisant un anticorps FLAG qui reconnaîtra l'étiquette clonée en C-terminal des constructions de la Figure 17. La lignée 1 permettra d'identifier les protéines partenaires et la lignée 4 permettra de vérifier le rôle du domaine GT-1.

L'immunoprécipitation avec l'anticorps FLAG sera aussi utilisée pour identifier les acides nucléiques qui interagissent avec l'enzyme. Pour cette approche, la lignée 3 avec l'enzyme inactive sera particulièrement intéressante puisqu'elle devrait permettre d'éviter les clivages sur les acides nucléiques co-purifiés. Une telle stratégie a déjà été employée avec succès en modèle bactérien (Lioliou *et al.*, 2012; Gatewood *et al.*, 2012). La lignée 1 ne sera pas non plus négligée car elle reste la situation la plus proche des conditions physiologiques sauvage. La lignée 4 permettra encore une fois de vérifier le rôle du domaine GT-1 dans l'interaction avec les acides nucléiques. Cette étude sera enfin renforcée par l'utilisation de la lignée 5. Nous utiliserons les mêmes stratégies que celles précédemment décrites mais avec un anticorps anti YFP pour l'immunoprécipitation. Le but est de vérifier si le domaine GT-1 peut agir indépendamment du reste de l'enzyme ou non.

Conclusion

En guise de conclusion je tiens à réitérer ce que j'ai déjà pu écrire : ce projet est une collaboration poussée entre les équipes OGE et GNet, une œuvre collective. Les plateformes POPS et InterATOME qui vont produire une part importante des données sont par exemple gérées par les deux équipes. Fruit de cette collaboration, le travail de thèse d'Arnaud Liehrmann sur la détection des différences de maturation dans les données de RNA-Seq aura un rôle essentiel au succès du projet. L'analyse structure fonction de la RNase J constituera le gros du travail de thèse de Sébastien Skiada qui rejoindra l'équipe à l'automne 2021. Une partie du travail moléculaire de validation sera réalisé en collaboration avec l'équipe de David Stern au BTI et sera, je l'espère, le début d'une longue et fructueuse collaboration.

Plus égoïstement j'espère que mes convictions profondes quant au rôle de l'encadrant vont résister à la confrontation avec la pratique quotidienne. La place centrale que j'accorde aux principes de coopérations, de participation et d'esprit critique est en dissonance totale avec la mise en compétition permanente des structures et personnes par l'entremise des appels à projet ou la future mise en place des « tenure track » par exemple. J'espère pouvoir transmettre à mes doctorants cette nécessité d'être radical, au sens premier du terme, *radix*. Toujours questionner, toujours revenir à la racine. Peu importe ce qu'ils décideront quant à leur future carrière, j'ai espoir que le temps passé avec moi les aura au moins aidés à choisir en toute conscience.

Bibliographie

- Anderson, C.** (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. Available at: <https://www.wired.com/2008/06/pb-theory/> [Accessed June 11, 2021].
- Astrachan, L. and Volkin, E.** (1958) Properties of ribonucleic acid turnover in T2-infected Escherichia coli. *BBA - Biochim. Biophys. Acta*, **29**, 536–544.
- Baudry, K.** (2019) *L'éditosome du chloroplaste : questions, éléments de réponses et digressions*. Université Paris-Saclay.
- Bégu, D., Castandet, B. and Araya, A.** (2011) RNA editing restores critical domains of a group I intron in fern mitochondria. *Curr. Genet.*, **57**, 317–325.
- Blanc, V., Litvak, S. and Araya, A.** (1995) RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Lett.*, **373**, 56–60.
- Bonen, L.** (2008) Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion*, **8**, 26–34.
- Brenner, S., Jacob, F. and Meselson, M.** (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, **190**, 576–581.
- Budar, F., Touzet, P. and Paepe, R. De** (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica*, **117**, 3–16.
- Burt, A. and Trivers, R.** (2008) *Genes in Conflict The Biology of Selfish Genetic Elements*, Belknap Press.
- Castandet, Benoît and Araya, A.** (2011a) RNA editing in plant organelles. Why make it easy? *Biochem.*, **76**, 924–931.
- Castandet, B. and Araya, A.** (2012) The nucleocytoplasmic conflict, a driving force for the emergence of plant organellar RNA editing. *IUBMB Life*, **64**, 120–125.
- Castandet, Benoît and Araya, A.** (2011b) The RNA editing pattern of cox2 mRNA is affected by point mutations in plant mitochondria. *PLoS One*, **6**.
- Castandet, B., Choury, D., Bégu, D., Jordana, X. and Araya, A.** (2010) Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res.*, **38**, 7112–7121.
- Castandet, B., Germain, A., Hotto, A.M. and Stern, D.B.** (2019) Systematic sequencing of chloroplast transcript termini from *Arabidopsis thaliana* reveals >200 transcription initiation sites and the extensive imprints of RNA-binding proteins and secondary structures. *Nucleic Acids Res.*, **47**, 11889–11905.
- 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.
- Castandet, B., Hotto, A.M., Strickler, S.R. and Stern, D.B.** (2016) ChloroSeq, an Optimized Chloroplast RNA-Seq Bioinformatic Pipeline, Reveals Remodeling of the Organellar Transcriptome Under Heat Stress. *G3 Genes, Genomes, Genetics*, **6**, 2817–2827.
- 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.
- Chateigner-Boutin, A.L. and Small, I.** (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.*, **35**.
- Choury, D. and Araya, A.** (2006) RNA editing site recognition in heterologous plant mitochondria. *Curr. Genet.*, **50**, 405–416.
- Choury, D., Farré, J.C., Jordana, X. and Araya, A.** (2004) Different patterns in the recognition of editing sites in plant mitochondria. *Nucleic Acids Res.*, **32**, 6397–6406.
- Choury, D., Farré, J.C., Jordana, X. and Araya, A.** (2005) Gene expression studies in isolated mitochondria: *Solanum tuberosum* rps10 is recognized by cognate potato but not by the transcription, splicing and editing machinery of wheat mitochondria. *Nucleic Acids Res.*, **33**, 7058–7065.
- Covello, P.S. and Gray, M.W.** (1993) On the evolution of RNA editing. *Trends Genet.*, **9**, 265–268.
- Covello, P.S. and Gray, M.W.** (1989) RNA editing in plant mitochondria. *Nature*, **341**, 662–666.
- Crick, F.** (1970) Central dogma of molecular biology. *Nature*, **227**, 561–563.
- Crick, F.H.C.** (1958) ON PROTEIN SYNTHESIS. *Symp Soc Exp Biol*, **12**, 138–63.
- Despres, B., Delseny, M. and Devic, M.** (2001) Partial complementation of embryo defective mutations: a general strategy to elucidate gene function. *Plant J*, **27**, 149–159.
- Di, C., Yuan, J., Wu, Y., et al.** (2014) Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J.*, **80**, 848–861.
- Farré, J.-C., Leon, G., Jordana, X. and Araya, A.** (2001) cis Recognition Elements in Plant Mitochondrion RNA Editing. *Mol. Cell. Biol.*, **21**, 6731–6737.

- Farré, J.-C.J.-C.C., Aknin, C., Araya, A. and Castandet, B.** (2012) RNA Editing in Mitochondrial Trans-Introns Is Required for Splicing. S. Maas, ed. *PLoS One*, **7**, e52644.
- Farré, J.C. and Araya, A.** (2001) Gene expression in isolated plant mitochondria: High fidelity of transcription, splicing and editing of a transgene product in electroporated organelles. *Nucleic Acids Res.*, **29**, 2484–2491.
- Ferandon, C., Chatel, S.E.K., Castandet, B., Castroviejo, M. and Barroso, G.** (2008) The *Agrocybe aegerita* mitochondrial genome contains two inverted repeats of the *nad4* gene arisen by duplication on both sides of a linear plasmid integration site. *Fungal Genet. Biol.*, **45**, 292–301.
- Forterre, P.** (2006) The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.*, **117**, 5–16.
- Forterre, P.** (2005) The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. In *Biochimie*. Biochimie, pp. 793–803.
- Forterre, P., Gribaldo, S. and Brochier, C.** (2005) Luca: The last universal common ancestor. *Medecine/Sciences*, **21**, 860–865.
- Fujii, S., Bond, C.S. and Small, I.D.** (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 1723–1728.
- García-Andrade, J., Ramírez, V., López, A. and Vera, P.** (2013) Mediated Plastid RNA Editing in Plant Immunity. *PLOS Pathog.*, **9**, e1003713.
- Gatewood, M.L., Bralley, P., Weil, M.R. and Jones, G.H.** (2012) RNA-Seq and RNA Immunoprecipitation Analyses of the Transcriptome of *Streptomyces coelicolor* Identify Substrates for rnaIII. *J. Bacteriol.*, **194**, 2228–2237.
- Germain, A., Herlich, S., Larom, S., Kim, S.H., Schuster, G. and Stern, D.B.** (2011) Mutational analysis of Arabidopsis chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation. *Plant J.*, **67**, 381–94.
- Germain, A., Hotto, A.M., Barkan, A. and Stern, D.B.** (2013) RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA*, **4**, 295–316.
- Gould, S.J. and Lewontin, R.C.** (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. London - Biol. Sci.*, **205**, 581–598.
- Gouyon, P.-H.** (2001) *Les harmonies de la nature à l'épreuve de la biologie*, Versailles: Editions Quæ.
- Gray, M.W.** (2012) Evolutionary origin of RNA editing. *Biochemistry*, **51**, 5235–5242.
- Gray, M.W., Burger, G. and Lang, B.F.** (1999) Mitochondrial evolution. *Science (80-)*, **283**, 1476–1481.
- Gray, M.W. and Covello, P.S.** (1993) RNA editing in plant mitochondria and chloroplasts. *FASEB J.*, **7**, 64–71.
- Gray, M.W., Lukeš, J., Archibald, J.M., Keeling, P.J. and Doolittle, W.F.** (2010) Irremediable complexity? *Science (80-)*, **330**, 920–921.
- Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W. and Watson, J.D.** (1961) Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*. *Nature*, **190**, 581–585.
- Grüttner, S., Hopf, C., Kumar, A. and Kempken, F.** (2013) Deletions in *cox2* mRNA Result in Loss of Splicing and RNA Editing and Gain of Novel RNA Editing Sites. S. Maas, ed. *PLoS One*, **8**, e82067.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.-H. and Grienenberger, J.-M.** (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature*, **341**, 660–662.
- Gutmann, B., Royan, S., Schallenberg-Rüdinger, M., et al.** (2019) The expansion and diversification of pentatricopeptide repeat RNA editing factors in plants. *Mol. Plant*.
- Halpert, M., Liveanu, V., Glaser, F. and Schuster, G.** (2019) The Arabidopsis chloroplast RNase J displays both exo- and robust endonucleolytic activities. *Plant Mol. Biol.*, **99**, 17–29.
- Havird, J.C., Forsythe, E.S., Williams, A.M., Werren, J.H., Dowling, D.K. and Sloan, D.B.** (2019) Selfish Mitonuclear Conflict. *Curr. Biol.*, **29**, R496–R511.
- Hayes, M.L. and Santibanez, P.I.** (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.*, jbc.RA119.011790.
- Hayes, R., Kudla, J., Schuster, G., Gabay, L., Maliga, P. and Grissem, W.** (1996) Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins. *EMBO J.*, **15**, 1132–41.
- He, B., Mu, Y. and Chi, W.** (2015) Effects of inefficient transcription termination of *rbcl* on the expression of *accD* in plastids of *Arabidopsis thaliana*. *Photosynth. Res.*, **126**, 323–330.
- Hiesel, R., Wissinger, B., Schuster, W. and Brennicke, A.** (1989) RNA editing in plant mitochondria. *Science*, **246**, 1632–4.
- Hotto, A.M., Castandet, B., Gilet, L., Higdon, A., Condon, C. and Stern, D.B.** (2015) Arabidopsis Chloroplast Mini-Ribonuclease III Participates in rRNA Maturation and Intron Recycling. *Plant Cell*, **27**, 724–740.
- Hotto, A.M., Schmitz, R.J., Fei, Z., Ecker, J.R. and Stern, D.B.** (2011) Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the Arabidopsis transcriptome. J. K. Kim, ed. *G3 Genes, Genomes, Genetics*, **1**, 559–570.

- Hotto, A.M., Stern, D.B. and Schuster, G.** (2020) Plant Ribonuclease J: An Essential Player in Maintaining Chloroplast RNA Quality Control for Gene Expression. *Plants*, **9**, 334.
- Jacob, F. and Monod, J.** (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
- Johnston, F.B. and Stern, H.** (1957) Mass isolation of viable wheat embryos [24]. *Nature*, **179**, 160–161.
- Jukes, T.H.** (1990) Random walking. *J. Mol. Evol.*, **30**, 1–2.
- Kakizaki, T., Yazu, F., Nakayama, K., Ito-Inaba, Y. and Inaba, T.** (2012) Plastid signalling under multiple conditions is accompanied by a common defect in RNA editing in plastids. *J. Exp. Bot.*, **63**, 251–260.
- Karcher, D. and Bock, R.** (1998) Site-selective inhibition of plastid RNA editing by heat shock and antibiotics: A role for plastid translation in RNA editing. *Nucleic Acids Res.*, **26**, 1185–1190.
- Karcher, D. and Bock, R.** (2002) Temperature sensitivity of RNA editing and intron splicing reactions in the plastid *ndhB* transcript. *Curr. Genet.*, **41**, 48–52.
- Kleinman, C.L. and Majewski, J.** (2012) Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science (80-.)*, **335**.
- Kotera, E., Tasaka, M. and Shikanai, T.** (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, **433**, 326–330.
- Kusnetsov, V. V., Mikulovich, T.P., Kukina, I.M., Cherepneva, G.N., Herrmann, R.G. and Kulaeva, O.N.** (1993) Changes in the level of chloroplast transcripts in pumpkin cotyledons during heat shock. *FEBS Lett.*, **321**, 189–193.
- Lang, B.F., Gray, M.W. and Burger, G.** (1999) Mitochondrial Genome Evolution and the Origin of Eukaryotes. *Annu. Rev. Genet.*, **33**, 351–397.
- Lange, H., Sement, F.M., Canaday, J. and Gagliardi, D.** (2009) Polyadenylation-assisted RNA degradation processes in plants. *Trends Plant Sci.*, **14**, 497–504.
- Lazebnik, Y.** (2015) Are scientists a workforce? - Or, how Dr. Frankenstein made biomedical research sick: A proposed plan to rescue US biomedical research from its current “malaise” will not be effective as it misdiagnoses the root cause of the disease. *EMBO Rep.*, **16**.
- Leister, D., Wang, L. and Kleine, T.** (2017) Organellar Gene Expression and Acclimation of Plants to Environmental Stress. *Front. Plant Sci.*, **08**, 387.
- Li, M., Wang, I.X. and Cheung, V.G.** (2012) Response to comments on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science (80-.)*, **335**.
- Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M. and Cheung, V.G.** (2011) Widespread RNA and DNA sequence differences in the human transcriptome. *Science (80-.)*, **333**.
- Lin, W., Piskol, R., Tan, M.H. and Li, J.B.** (2012) Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science (80-.)*, **335**.
- Linguist, S., Doolittle, W.F. and Palazzo, A.F.** (2020) Getting clear about the F-word in genomics. *PLoS Genet.*, **16**.
- Lioliou, E., Sharma, C.M., Caldelari, I., Helfer, A.C., Fechter, P., Vandenesch, F., Vogel, J. and Romby, P.** (2012) Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genet.*, **8**.
- Longo, D.L. and Drazen, J.M.** (2016) Data Sharing. *N. Engl. J. Med.*, **374**, 276–277.
- Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F. and Gray, M.W.** (2011) How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life*, **63**, 528–537.
- Lurin, C., Andrés, C., Aubourg, S., et al.** (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089–103.
- Luro, S., Germain, A., Sharwood, R.E. and Stern, D.B.** (2013) RNase J participates in a pentatricopeptide repeat protein-mediated 5' end maturation of chloroplast mRNAs. *Nucleic Acids Res.*, **41**, 9141–51.
- MacIntosh, G.C. and Castandet, B.** (2020) Plastidial and secretory ribonucleases, important players in plant RNA homeostasis Title. *Plant Physiol.*, **accepted**.
- Maier, U.G., Bozarth, A., Funk, H.T., Zauner, S., Rensing, S.A., Schmitz-Linneweber, C., Börner, T. and Tillich, M.** (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol.*, **6**, 36.
- Mazzocchi, F.** (2015) Could Big Data be the end of theory in science? *EMBO Rep.*, **16**.
- Michel, E.J.S.S., Hotto, A.M., Strickler, S.R., Stern, D.B. and Castandet, B.** (2018) A Guide to the Chloroplast Transcriptome Analysis Using RNA-Seq. *Methods Mol. Biol.*, **1829**, 295–313.
- Mu, R., Shinde, P., Zou, Z., Kreth, J. and Merritt, J.** (2019) Examining the Protein Interactome and Subcellular Localization of RNase J2 Complexes in *Streptococcus* mutans. *Front. Microbiol.*, **10**.
- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., Shikanai, T. and Haselkorn, R.** (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A.*, **104**, 8178–83.
- Okuda, K., Nakamura, T., Sugita, M., Shimizu, T. and Shikanai, T.** (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.*, **281**, 37661–37667.

- Oldenkott, B., Yang, Y., Lesch, E., Knoop, V. and Schallenberg-Rüdinger, M.** (2019) Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. *Commun. Biol.*, **2**, 85.
- Parker, M.T., Knop, K., Sherwood, A. V., Schurch, N.J., Mackinnon, K., Gould, P.D., Hall, A.J.W., Barton, G.J. and Simpson, G.G.** (2020) Nanopore direct RNA sequencing maps the complexity of arabidopsis mRNA processing and m6A modification. *Elife*, **9**.
- Pickrell, J.K., Gilad, Y. and Pritchard, J.K.** (2012) Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science (80-.)*, **335**.
- Płociński, P., Maclos, M., Houghton, J., et al.** (2019) Proteomic and transcriptomic experiments reveal an essential role of RNA degradosome complexes in shaping the transcriptome of *Mycobacterium tuberculosis*. *Nucleic Acids Res.*, **47**, 5892–5905.
- Redko, Y., Bechhofer, D.H. and Condon, C.** (2008) Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. *Mol. Microbiol.*, **68**, 1096–1106.
- Rüdinger, M., Polsakiewicz, M. and Knoop, V.** (2008) Organellar RNA Editing and Plant-Specific Extensions of Pentatricopeptide Repeat Proteins in Jungermanniid but not in Marchantiid Liverworts. *Mol. Biol. Evol.*, **25**, 1405–1414.
- Ruwe, H., Castandet, B., Schmitz-Linneweber, C. and Stern, D.B.** (2013) Arabidopsis chloroplast quantitative editotype. *FEBS Lett.*, **587**, 1429–1433.
- Ruwe, H., Wang, G., Gusewski, S. and Schmitz-Linneweber, C.** (2016) Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.*, **44**, 7406–7417.
- 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.
- Sanitá Lima, M. and Smith, D.R.** (2017) Pervasive, Genome-Wide Transcription in the Organelle Genomes of Diverse Plastid-Bearing Protists. *G3 Genes, Genomes, Genetics*, **7**, g3.300290.2017.
- Sharwood, R.E., Halpert, M., Luro, S., Schuster, G. and Stern, D.B.** (2011) Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA. *RNA*, **17**, 2165–2176.
- Small, I., Schallenberg-Rüdinger, M., Takenaka, M., Mireau, H. and Ostersetzter-Biran, O.** (2020) Plant organellar RNA editing: what 30 years of research has revealed. *Plant J.*, **101**.
- Smith, D.R.** (2013) RNA-Seq data: A goldmine for organelle research. *Brief. Funct. Genomics*, **12**, 454–456.
- Smith, D.R. and Sanitá Lima, M.** (2016) Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline. *Brief. Bioinform.*, **18**, bbw088.
- Staudinger, M., Bolle, N. and Kempken, F.** (2005) Mitochondrial electroporation and in organello RNA editing of chimeric atp6 transcripts. *Mol. Genet. Genomics*, **273**, 130–136.
- Stern, D.B. and Grussem, W.** (1987) Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell*, **51**, 1145–57.
- Stoltzfus, A.** (2012) Constructive neutral evolution: Exploring evolutionary theory’s curious disconnect. *Biol. Direct*, **7**, 35.
- Stoltzfus, A.** (1999) On the possibility of constructive neutral evolution. *J. Mol. Evol.*, **49**, 169–181.
- Stoppel, R., Manavski, N., Schein, A., Schuster, G., Teubner, M., Schmitz-Linneweber, C. and Meurer, J.** (2012) RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the Arabidopsis chloroplast. *Nucleic Acids Res.*, **40**, 8593–8606.
- Süss, K.-H. and Yordanov, I.T.** (1986) Biosynthetic Cause of in Vivo Acquired Thermotolerance of Photosynthetic Light Reactions and Metabolic Responses of Chloroplasts to Heat Stress . *Plant Physiol.*, **81**, 192–199.
- Takenaka, M., Neuwirt, J. and Brennicke, A.** (2004) Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.*, **32**, 4137–4144.
- Touzet, P. and Meyer, E.H.** (2014) Cytoplasmic male sterility and mitochondrial metabolism in plants. *Mitochondrion*, **19**, 166–171.
- Watson, J.** (1965) *Molecular Biology of the Gene* New York., New York: W.A. Benjamin, Inc.
- Weber, A.P.M., Weber, K.L., Carr, K., Wilkerson, C. and Ohlrogge, J.B.** (2007) Sampling the arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol.*, **144**, 32–42.
- Woese, C.R.** (2004) A New Biology for a New Century. *Microbiol. Mol. Biol. Rev.*, **68**.
- Yanai, I. and Lercher, M.** (2020) A hypothesis is a liability. *Genome Biol.*, **21**.
- Ycas, M. and Vincent, W.S.** (1960) A RIBONUCLEIC ACID FRACTION FROM YEAST RELATED IN COMPOSITION TO DESOXYRIBONUCLEIC ACID. *Proc. Natl. Acad. Sci.*, **46**, 804–811.
- Yehudai-Resheff, S., Hirsh, M. and Schuster, G.** (2001) Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts. *Mol. Cell. Biol.*, **21**, 5408–16.
- Zhang, F., Tang, W., Hedtke, B., Zhong, L., Liu, L., Peng, L., Lu, C., Grimm, B. and Lin, R.** (2014) Tetrapyrrole biosynthetic

enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 2023–2028.

Zhao, X., Huang, J. and Chory, J. (2020) Unraveling the Linkage between Retrograde Signaling and RNA Metabolism in Plants. *Trends Plant Sci.*, **25**.

Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J. and Borner, T. (2012) The Primary Transcriptome of Barley Chloroplasts: Numerous Noncoding RNAs and the Dominating Role of the Plastid-Encoded RNA Polymerase. *Plant Cell*, **24**, 123–136.