



**HAL**  
open science

**Biogénèse des siRNAs endogènes chez *Arabidopsis thaliana*: étude fonctionnelle de DRB7.2, une nouvelle protéine de fixation à l'ARN double brin et développement d'outils moléculaires pour la caractérisation du mode d'action de DCL4**

Thomas Montavon

► **To cite this version:**

Thomas Montavon. Biogénèse des siRNAs endogènes chez *Arabidopsis thaliana*: étude fonctionnelle de DRB7.2, une nouvelle protéine de fixation à l'ARN double brin et développement d'outils moléculaires pour la caractérisation du mode d'action de DCL4. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Strasbourg, 2017. Français. NNT: 2017STRAJ002. tel-01681635v1

**HAL Id: tel-01681635**

**<https://theses.hal.science/tel-01681635v1>**

Submitted on 11 Jan 2018 (v1), last revised 11 Jan 2018 (v2)

**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.

ECOLE DOCTORALE DE LA VIE ET DE LA SANTE

# THÈSE

Présentée par

**Thomas MONTAVON**

à

**L'Institut de biologie moléculaire des plantes, CNRS UP2357**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ DE STRASBOURG**

Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

## **Biogénèse des siRNAs endogènes chez *Arabidopsis thaliana***

**Etude fonctionnelle de DRB7.2, une nouvelle protéine de  
fixation à l'ARN double brin et développement d'outils  
moléculaires pour la caractérisation du mode d'action de  
DCL4**

Soutenue le 6 Janvier 2017 devant la Commission d'Examen :

Directeur de Thèse  
Rapporteur Externe  
Rapporteur Externe  
Examineur Interne

Dr Patrice DUNOYER  
Dr Étienne BUCHER  
Dr Thierry LAGRANGE  
Dr Sébastien PFEFFER





## *Remerciements*

Je tiens tout d'abord à remercier messieurs les docteurs. Étienne Bucher, Thierry Lagrange et Sébastien Pfeffer, d'avoir accepté d'évaluer ce manuscrit.

Je tiens tout particulièrement à remercier mon directeur de thèse Patrice Dunoyer d'avoir accepté que je devienne son "Padawan". Je le remercie pour son soutien et sa patience, tout au long de cette passionnante thèse. Je le remercie également de m'avoir transmis son expérience, sa passion et ses connaissances (scientifiques et autres).

Je tiens bien entendu à remercier également les membres (passés et présents) de mon équipe. "The happy team" (Shah et Damien Garcia, Clara Chicois) pour leurs conseils, leur bonne humeur (bien entendu) et leur soutien sans faille, Fabrice Michel pour ses précieuses connaissances et aides dans de nombreux domaines ainsi que pour son aide pour le papier DRB7.2. Je remercie également Yerim Kwon et Aude Zimmerman pour leur aide sur ce projet. Je remercie aussi Jérôme Zervudacki pour son aide lors des premiers mois au laboratoire. Je remercie Jean Molinier et Christophe Himber pour leurs conseils techniques et pratiques. Je voudrais aussi remercier Valérie Cognat, Timothée Vincent et toute la plateforme de protéomique de l'IBMC pour leur aide sur les différents projets auxquels j'ai pu participer.

Je voudrais également remercier tout le personnel technique (Jardiniers, laverie, séquençage, Bio informaticien, Accueil, magasin) et toutes les personnes de l'IBMP (Alyssa, Clément, Nina, Natka, Todd, Kamal, Khalid, Ahmed, Vianney....) que j'ai pu côtoyer pour leur aide et soutien tout au long de cette thèse. Je voudrais également remercier le ministère de l'enseignement supérieur et de la recherche d'avoir financé une partie de ma thèse ainsi que le CNRS et le Labex Net RNA.

Je remercie plus qu'il n'est possible mon "Coloc de thèse" Marco Incarbone alias "IL CALVO" pour tout les excellents moments que j'ai passé avec lui pendant cette thèse. Je le remercie également pour toutes les discussions (scientifiques ou autre) que j'ai eu avec lui. Je le remercie aussi pour son soutien durant ces années et pour son amitié tout simplement. Je remercie également Marion Clavel sa copine pour ses conseils et son aide sur la partie DRB7.2

Je tiens également à remercier Léon et Lidoux mes potes pour leur soutien.

Je remercie chaleureusement bien évidemment ma famille, Simone et René mes parents, et Julia ma sœur, pour leurs aides, leur soutien et pour m'avoir supporté durant une bonne partie de cette thèse (et de mon existence aussi..).

Enfin je remercie énormément, et je pense que je la remercierai indéfiniment, celle qui partage ma vie et mon cassoulet, Anne-Laure. Je la remercie notamment (la liste est trop longue), pour son amour, sa patience, sa compréhension et pour son soutien indéfectible tout au long de ces années et durant cette thèse. Je la remercie de m'avoir poussé à faire ce que je voulais, et pour tout les moments que j'ai passé avec elle.

# TABLE DES MATIERES

<b>TABLE DES FIGURES</b> .....	- 5 -
<b>ABREVIATIONS</b> .....	- 7 -
<b>INTRODUCTION</b> .....	- 10 -
Du pétunia au prix Nobel: Les premiers pas du RNA silencing .....	- 11 -
Le RNA silencing chez les plantes .....	- 15 -
1 Les enzymes clés de la biogenèse des sRNA chez les plantes .....	- 16 -
2 Les différentes voies endogènes du RNA silencing chez <i>Arabidopsis thaliana</i> .....	- 21 -
Objectifs de la thèse.....	- 29 -
<b>RÉSULTATS</b> .....	<b>31</b>
<b>Chapitre 1: Etude fonctionnelle de DRB7.2, une nouvelle protéine de fixation à l'ARN double brin</b> .....	<b>32</b>
1.1 Histoire évolutive des protéines DRBs chez les plantes .....	33
1.2 Identification d'un complexe spécifique de DRBs séquestrant les précurseurs d'endoIR-siRNA.....	52
1.3 Résultats complémentaires.....	81
1.4 Discussion, perspectives .....	85
<b>Chapitre 2: Etude du mode d'action de DCL4 dans la biogenèse des sRNA endogène</b> .....	<b>91</b>
2.1 Introduction .....	92
2.2 Présentation et caractérisation des allèles utilisés.....	93
2.3 Développement des outils moléculaires.....	96
2.4 Utilisation des outils moléculaires.....	98
2.5 Discussion, perspectives .....	106

<b>CONCLUSION GENERALE.....</b>	<b>111</b>
<b>MATERIEL.....</b>	<b>- 115 -</b>
Souches bactériennes.....	- 115 -
<i>Escherichia Coli</i> .....	- 115 -
<i>Agrobacterium tumefaciens</i> .....	- 115 -
Plantes .....	- 116 -
<i>Arabidopsis thaliana</i> .....	- 116 -
<i>Nicotiana benthamiana</i> .....	- 116 -
Vecteurs plasmidiques.....	- 116 -
pGEM®-T Easy (Promega®).....	- 116 -
pDONR™/Zeo (Invitrogen®) .....	- 117 -
pCambia1300.....	- 117 -
Anticorps, amorces oligonucléotidique et sondes .....	- 118 -
<b>METHODES.....</b>	<b>- 119 -</b>
PCR.....	- 119 -
PCR pour l'amplification de gènes d'intérêt.....	- 119 -
PCR pour la vérification de la présence d'insert dans les plasmides ou pour les génotypages.....	- 120 -
Clonage.....	- 120 -
Clonage Gateway® .....	- 121 -
Clonage par enzymes de restriction .....	- 121 -
Transformations.....	- 122 -
Transformation d' <i>Escherechia coli</i> par choc thermique.....	- 122 -
Transformation <i>d'Agrobacteruim tumefaciens</i> par la méthode d'électroporation.....	- 123 -
Transformation <i>d'Arabidopsis thaliana</i> .....	- 123 -
Extraction d'ADN plasmidique et purification de produit de PCR.....	- 124 -

Extraction d'ADN <i>d'Arabidopsis thaliana</i> .....	- 124 -
Extraction à partir de bourgeons floraux .....	- 124 -
Extraction à partir de feuilles .....	- 125 -
Préparation des ADNc et RT-qPCR.....	- 126 -
Rétrotranscription des ARN.....	- 126 -
qRT-PCR .....	- 127 -
Infection virale.....	- 127 -
Agroinfiltration de feuilles <i>d'Arabidopsis thaliana</i> .....	- 127 -
Bi-molecular fluorescence complementation (BiFC) .....	- 128 -
Constructions plasmidiques .....	- 128 -
Agroinfiltration de feuilles de <i>Nicotiana benthamiana</i> .....	- 129 -
Observation microscopique.....	- 130 -
Extraction d'ARN et des protéines totales .....	- 131 -
Extraction d'ARN totaux.....	- 131 -
Extraction des protéines.....	- 131 -
Northern blot .....	- 132 -
Electrophorèse sur gel d'agarose/formaldehyde.....	- 132 -
Electrophorèse sur gel de poly-acrylamide/urée .....	- 133 -
Synthèse et purification des sondes radioactives .....	- 133 -
Western blot .....	- 134 -
Electrophorèse sur gel de poly-acrylamide .....	- 134 -
Immunoprécipitation des ARN .....	- 135 -
Immunoprécipitation avec billes d'agarose.....	- 135 -
Immunoprécipiataion avec billes magnétiques.....	- 136 -
Immunoprécipitation des protéines .....	- 137 -
Immunoprécipitation avec billes d'agarose.....	- 137 -
Immunoprécipiataion avec billes magnétiques.....	- 138 -

Analyses des protéines par spectrométrie de masse.....	139 -
Digestion des protéines.....	139 -
Analyses par Nano-liquid Chromatography-electrospray Ionization TripleTOF MS/MS .....	140 -
Analyses informatiques.....	140 -
Séquençage à haut débit et analyses bio-informatiques des ARN associés à DCL4-6 et DCL4-8.....	142 -
Séquençage à haut débis .....	142 -
Alignement et analyse différentielle .....	142 -
Annotations .....	143 -
Alignement multiple de séquences .....	143 -
<b><i>ANNEXES</i></b> .....	<b>145</b>
<b><i>BIBLIOGRAPHIE</i></b> .....	<b>151</b>



# ***TABLE DES FIGURES***

## **Introduction:**

**Figure I:** Premier modèle du PTGS (**page 13**)

**Figure II:** Modèle général et domaines fonctionnels des enzymes clés du RNA silencing (**page 15**)

**Figure III:** Voie des miRNAs et des tasiRNAs chez *Arabidopsis thaliana* (**page 22**)

**Figure IV:** Voie des siRNAs hétérochromatiques chez *Arabidopsis thaliana* (**page 26**)

**Figure V:** Voie des endoIR-siRNAs chez *Arabidopsis thaliana* (**page 28**)

## **Chapitre 1:**

**Figure 1.1:** Principe du système SUC:SUL (**page 81**)

**Figure 1.2:** Effet des mutations *drb7.2* et *drb4* sur l'accumulation de différents exoIR-siRNAs (**page 82**)

**Figure 1.3:** Analyse de l'interactome de DRB7.2 (**page 84**)

**Figure 1.4:** Effet de la mutation *d7ip1* sur l'accumulation de divers sRNAs endogènes (**page 85**)

## **Chapitre 2:**

**Figure 2.1:** Caractérisation des allèles découplants de DCL4 (**page 93**)

**Figure 2.2:** Stratégie expérimentale pour l'étude du mode d'action de DCL4 (**page 96**)

**Figure 2.3:** Caractérisation des outils moléculaires (**page 97**)

**Figure 2.4:** Localisation subcellulaire des différents allèles de DCL4 (**page 99**)

**Figure 2.5:** Expériences de Co-IP et d'IP-MS à partir des différentes protéines DCL4 (**page 100**)

**Figure 2.6:** Analyse de l'interactome des différentes protéines DCL4 (**page 101**)

**Figure 2.7:** Analyse des efficacités d'immunoprécipitation des expériences de RIP (**page 103**)

**Figure 2.8:** Etude de la fixation des différents précurseurs de sRNAs DCL4-dépendant par DCL4-8 (**page 104**)

**Figure 2.9:** Stabilisation des précurseurs des sRNAs DCL4-dépendant (**page 105**)

**Figure 2.10:** Identification des précurseurs de sRNAs associés à DCL4-6 ou DCL4-8 (**page 106**)

**Figure 2.11:** Conservation des acides aminés mutés dans *DCL4-6* et *DCL4-8* (**page 107**)

# ***ABREVIATIONS***

ADN: Acide déoxyribonucléique

AGO: Argonaute

ARN: Acide ribonucléique

ARNm: ARN messenger

ATP: Adénosine triphosphate

CLSY: Classy

CMT3: Chromomethyltransferase 3

CP: Capside protein

db: Double brin

DCL: Dicer-like protéine

DDM1: Decreased DNA Methylation 1

DRB: Double-stranded RNA binding protéine

DRM2: Domains Rearranged Methyltransferase 2

dsRBD: Double-stranded RNA binding domain

DUF: Domain of unknown function

easiRNA: epigenetically-activated siRNA

FDR: False Discovery Rate

HEN1: HUA enhancer 1

HST: Hasty

HYL1: Hyponastic leaves 1

IR: Inverted- repeat

IRsiRNA: Inverted- repeat -derived siRNA

KTF1: Kow-Domain Containing Transcription Factor 1

KYP: Kryptonite

MASS SEPC: Spectrométrie de Masse

MET1: Methyltransferase 1

sb: Simple brin

mgf: Mascot Generic File

MID: Middle

miRNA: micro-RNA

MS: Murashige-Skoog

MS: Spectrométrie de masse

MT: Méthyltransférase

nt: nucleotide

PAZ: Piwi/Argonaute/Zwille

PIWI: P-element-induced wimpy testes

Pol: ADN polymérase ARN-dépendante

PTGS: Post-transcriptional gene silencing

PVX: *Potato X virus*

RdDM: RNA directed DNA Methylation

RDR: RNA polymerase RNA-dependant

RNAi: interference par l'ARN

RNase: Ribonucléase

RISC: RNA-induced silencing complex

RTL: RNaseIII-like

SA: Acide salicylique

SE: Serrate

SHH1: Sawadee Homeodomain Homologue 1

siRNA: small-interfering RNA

sRNA: small RNA

TAIR: The Arabidopsis Information Resource

tasiRNA: trans-acting siRNA

T-DNA: ADN de transfert

TEV: *Tobacco etch virus*

TGS: Transcriptional gene silencing

TMV: *Tobacco mosaic virus*

TRV: *Tobacco rattle virus*

vasiRNA: virus-activated siRNA

vsRNA: viral-derived siRNA

# ***INTRODUCTION***

# DU PETUNIA AU PRIX NOBEL: LES PREMIERS PAS DU RNA SILENCING

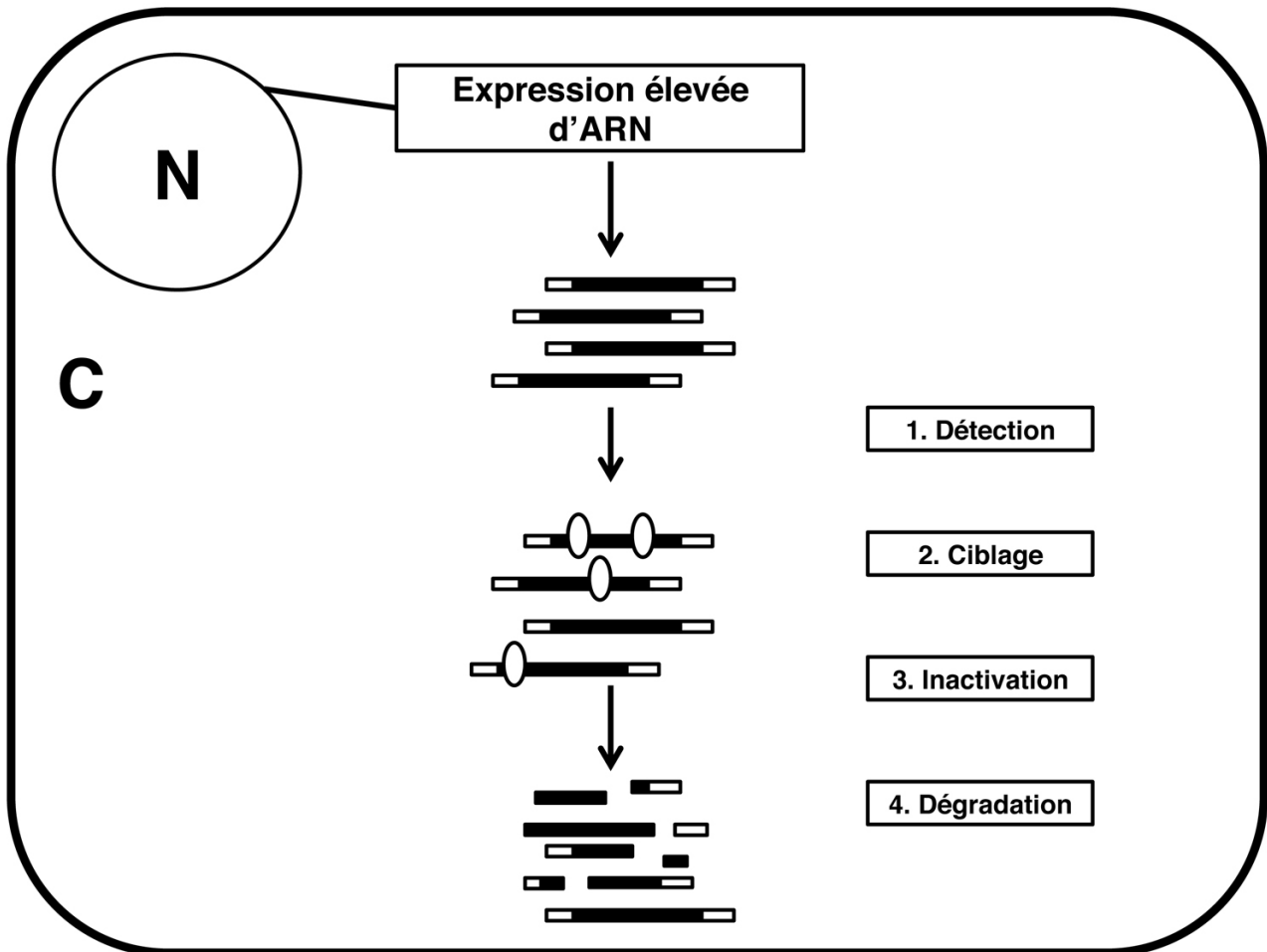
La découverte accidentelle du mécanisme du RNA silencing à la fin du 20<sup>ème</sup> siècle a mis en lumière un niveau de régulation jusqu'à lors inconnu qui a significativement augmenté la complexité de notre compréhension des mécanismes de régulation de l'expression des gènes. Un des éléments centraux de ce mécanisme, conservé chez la majorité des eucaryotes, est la présence de petites molécules d'ARN double brin, de 20 à 25 nucléotides (nt) de long (small RNA, sRNA), qui confèrent la spécificité de séquence à ce mécanisme de régulation agissant tant au niveau transcriptionnel que post-transcriptionnel.

Depuis la découverte et la mise au point de la technique de transformation stable des plantes par *Agrobacterium tumefaciens* (Chilton *et al*, 1977; Zambryski *et al*, 1983; Fraley *et al*, 1983; Hooykaas-Van Slogteren *et al*, 1984), cette technique a été massivement utilisée pour l'étude de l'expression des gènes ou dans des approches biotechnologiques. C'est dans ce contexte qu'en 1989, Matzke et ses collègues font une première observation, chez *Nicotiana tabacum*, du phénomène que l'on appellera plus tard le "transcriptional gene silencing" (TGS) (Matzke *et al*, 1989). Dans cette étude, les auteurs observèrent que lors de l'introduction consécutive dans le génome de la plante de deux ADN de transfert (T-DNA), possédant des marqueurs de sélections différents, le gène de sélection du premier T-DNA était inactivé et ce, après l'introduction du second T-DNA. Ils mirent en évidence que cette inactivation était corrélée à une méthylation de la séquence promotrice du premier transgène et que l'initiation de cette méthylation était dépendante du contexte génomique du second. Les auteurs suggérèrent que l'homologie des deux T-DNA, présente notamment au niveau des copies du promoteur de la nopaline synthase, pouvait être à l'origine de

l'initiation de la méthylation sur le premier T-DNA (Matzke *et al*, 1989). L'année suivante, des études sur des pétunias (*Petunia hybrida*) transgéniques portant une copie additionnelle d'un gène de la pigmentation des fleurs, la chalcone synthase (CHS), apportèrent quelques informations supplémentaires. Le but de ces études était de sur-exprimer la CHS pour intensifier la couleur violette des fleurs. Cependant, à la grande surprise des chercheurs, les plantes transformées arboraient une pigmentation variée allant du violet intense au blanc. Des analyses moléculaires sur ces plantes transgéniques révélèrent que dans certains cas, les ARN messagers (ARNm) provenant à la fois de la copie transgénique et de la copie endogène du gène de la CHS n'étaient plus détectables. Les auteurs nommèrent ce phénomène "co-suppression" (Napoli *et al*, 1990; van der Krol *et al*, 1990). Un phénomène similaire fut également observé chez *Neurospora crassa* et décrit sous le terme de "quelling" (Romano & Macino, 1992) mais également chez *Caenorhabditis elegans* (Fire *et al*, 1991) laissant présager du caractère central et essentiel de ce mécanisme.

Une première hypothèse concernant l'initiation et la séquence-spécificité du phénomène de co-suppression fut apportée par Lindbo et ses collègues en 1993 (Lindbo *et al*, 1993). En générant des plantes transgéniques exprimant le gène de la capsid (CP) du *Tobacco etch virus* (TEV) dans le but d'obtenir des tabacs résistants à ce virus, ils observèrent que, dans les plantes transformées, l'infection systémique par le TEV était fortement inhibée 3-5 semaines après inoculation. De plus, les feuilles systémiques asymptomatiques étaient devenues résistantes à une nouvelle infection par le TEV, mais elles restaient susceptibles à une infection par un virus d'une autre famille tel que le PVX (*Potato X virus*). Par des expériences de nuclear run off et de northern blot sur les feuilles résistantes, ils purent montrer que le transgène codant pour la CP du TEV était toujours fortement transcrit mais que son ARNm demeurait sous le seuil de détection, suggérant que ce phénomène agissait à un niveau post-transcriptionnel. Sur la base de ces résultats, les auteurs proposèrent que ce processus pouvait être initié par des niveaux élevés d'ARNm dans le cytoplasme qui, au-delà d'un certain





**Figure I: Premier modèle du PTGS (adapté de Lindbo *et al*, 1993)**

Modèle de la répression post-transcriptionnelle médiée par l'ARN proposé par Lindbo et ses collègues en 1993. (1) Une quantité élevée d'ARN au sein de la cellule est perçue comme aberrante. (2) Le brin complémentaire à cet ARN va être synthétisé par une RDR endogène et (3) servira à reconnaître tout ARN présentant une complémentarité de séquence, conduisant *in fine* à sa dégradation (4).

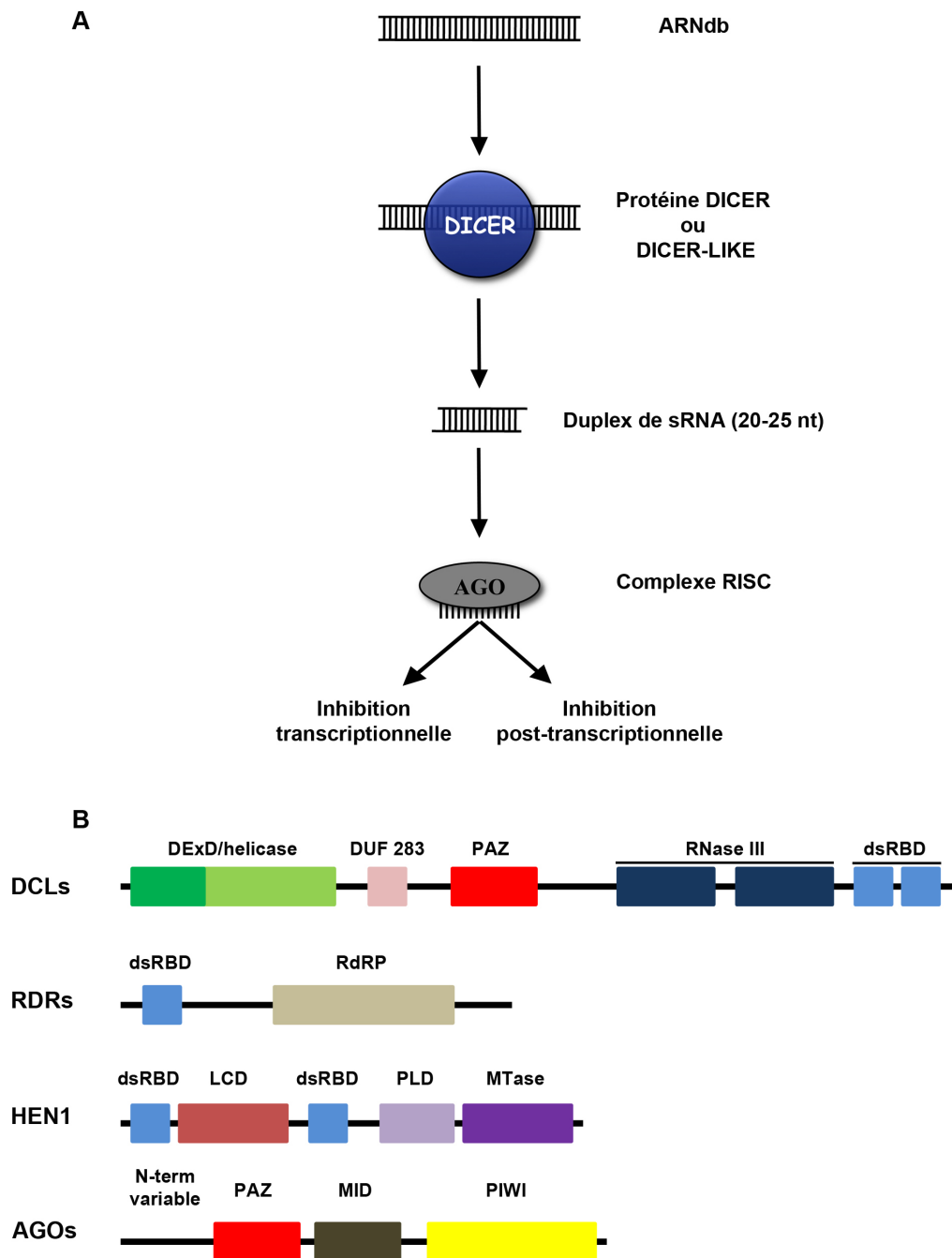
seuil, étaient perçus comme aberrant par la cellule. Ces ARN aberrants seraient alors pris en charge par une ARN polymérase ARN-dépendante (RDR) qui générerait un brin d'ARN complémentaire pouvant s'hybrider avec les ARN de séquences homologues (ARNm ou ARN viral), entraînant leur dégradation (**Figure I**).

En 1997, plusieurs études menées dans le laboratoire de David Baulcombe apportèrent un poids supplémentaire à la théorie de Lindbo, en montrant que le RNA silencing post-transcriptionnel (PTGS) était un mécanisme de défense antivirale chez les plantes (Angell & Baulcombe, 1997; Ratcliff *et al*, 1997). Toutefois, la nature exacte de la molécule initiatrice (ARN simple-brin (ARNsb) sens ou anti-sens, ou ARN double brin (ARNdb)) demeurait évasive (Baulcombe, 1996; Dougherty & Parks, 1995; Metzloff *et al*, 1997). C'est en 1998 que le mystère sur l'initiation du RNA silencing fut résolu lorsque qu'Andrew Fire et Craig Mello démontrèrent, chez *C. elegans*, que l'ARN double brin (ARNdb) était l'initiateur de ce mécanisme (Fire *et al*, 1998). En effet, ils purent mettre en évidence que l'injection de long ARNdb était 10 à 100 fois plus efficace pour induire la dégradation d'un ARNm de séquence homologue, que l'injection d'ARNsb sens ou anti-sens. Ils nommèrent ce mécanisme "interférence par l'ARN" (RNAi). Au cours de la même année, ces observations furent confirmées chez les insectes (Kennerdell & Carthew, 1998), les protozoaires (Ngô *et al*, 1998) et les plantes (Waterhouse *et al*, 1998). L'année suivante, toujours chez les plantes, Baulcombe et Hamilton démontrèrent que le PTGS était opéré par des ARN de petite taille, environ 25 nt, servant de guide pour induire la dégradation séquence-spécifique de l'ARN cible (Hamilton & Baulcombe, 1999). Par la suite, plusieurs groupes confirmèrent ces résultats en cellules d'insectes et de mammifères (Hammond *et al*, 2000; Zamore *et al*, 2000; Elbashir *et al*, 2001a, 2001b) et nommèrent ces sRNAs "small interfering" RNA (siRNA) (Elbashir *et al*, 2001b).

Au cours de ces mêmes années, des travaux pionniers dans de nombreux laboratoires ont permis d'isoler les enzymes responsables d'une part de la production des siRNAs et d'autre part du clivage de l'ARN cible.

C'est ainsi que la protéine Dicer, une endoribonucléase de type III (RNase III) conservée chez la plupart des eucaryotes, a pu être caractérisée comme l'enzyme responsable de la biogénèse des sRNAs (Bernstein *et al*, 2001). Le nombre de Dicer varie selon les organismes, de un chez les vertébrés et les nématodes à quatre chez la plante modèle *Arabidopsis thaliana* (voire six chez le riz) où ils portent le nom de Dicer-like protein (DCL) (Schauer *et al*, 2002; Lee *et al*, 2004; Finnegan *et al*, 2003; Xie *et al*, 2004; Margis *et al*, 2006). En parallèle, le complexe ribonucléoprotéique (nommé RNA-induced silencing complex, RISC) responsable du clivage de l'ARN cible fut isolé (Hammond *et al*, 2000) et sa purification permit d'identifier les protéines de la famille ARGONAUTE (AGO) comme un des constituants majeurs de ce complexe (Hammond *et al*, 2001). Comme pour les protéines Dicer, les protéines AGO sont extrêmement conservées chez les eucaryotes et leur nombre varie d'un organisme à l'autre (de 2 chez la Drosophile à 26 chez *C. elegans* (Meister, 2013; Swarts *et al*, 2014; Youngman & Claycomb, 2014).

Depuis, le RNA silencing est devenu rapidement un des sujets majeurs de la recherche fondamentale et appliquée grâce, à son implication dans un grand nombre de processus biologiques chez de nombreux organismes, à son utilité pour l'étude de la fonction des gènes et pour son fort potentiel en terme d'application agronomique et médicale. C'est pour l'ensemble de ces raisons que la découverte du RNA silencing donna lieu à l'obtention en 2006 du prix Nobel de médecine à Andrew Fire et Craig Mello.



**Figure II: Modèle général et domaines fonctionnels des enzymes clés du RNA silencing.**

(A) Mécanisme général du RNA silencing. La molécule initiatrice de ce mécanisme est un ARN double brin (ARNdb) qui sera reconnu par une enzyme de type RNase III appelée Dicer chez les animaux et Dicer-like chez les plantes. Cette enzyme va générer, à partir de cet ARNdb, des duplex de sRNA de 20 à 25 nt de long. Un des brins de ce duplex sera incorporé dans le complexe RISC (RNA-induced silencing complex), où il servira de guide pour lui permettre de reconnaître tout ARN présentant une complémentarité de séquence. Cette reconnaissance entrainera la répression de l'ARN cible soit, au niveau transcriptionnel, soit au niveau post-transcriptionnel. (B) Organisation des différents domaines fonctionnels conservés des enzymes clés du RNA silencing chez *Arabidopsis thaliana*. Abréviations: AGO, Argonaute; DCL, Dicer-like; dsRBD, double-stranded RNA binding domain; DUF, Domain of unknown function; LCD, La-motif-containing domain; MTase, Methyltransferase; PAZ, PAZ/ARGONAUTE/ZWILLE; PLD, PPIase-like domain; RDR, RNA-dependent RNA polymerase.

# LE RNA SILENCING CHEZ LES PLANTES

Chez les eucaryotes, le RNA silencing est un mécanisme conduisant à la suppression de l'expression des gènes. Il est impliqué dans un grand nombre de processus biologiques comme le maintien de la stabilité du génome, le développement et la défense antivirale (Chapman & Carrington, 2007; Baulcombe, 2004; Maillard *et al*, 2013).

Après reconnaissance de l'ARNdb initiateur par une protéine Dicer ou DCL, l'ARNdb sera clivé pour générer un ou plusieurs duplex de sRNA possédant 2nt 3' sortant et un groupement phosphate en 5' (Elbashir *et al*, 2001b). Un des brins de ce duplex sera incorporé dans le complexe RISC où, il servira de guide pour permettre à celui-ci de reconnaître dans la cellule tout ARN présentant une complémentarité de séquence. Cette reconnaissance entraîne alors une inhibition de l'expression des gènes ciblés, soit au niveau transcriptionnel, soit au niveau post-transcriptionnel (Eamens *et al*, 2008; Axtell, 2013; Bologna & Voinnet, 2014) (**Figure II A**). Chez les plantes, les sRNAs impliqués dans cette régulation peuvent être classés en deux grandes catégories selon la nature de leurs précurseurs et leur mode de biogenèse: Les micro-ARNs (miRNAs) et les small-interfering ARNs (siRNAs). Les miRNAs, interviennent majoritairement dans la régulation de l'expression des gènes, notamment au cours du développement, alors que les siRNAs sont principalement impliqués dans la défense antivirale et la répression d'éléments génétiques mobiles (tel que les transposons, TE) (Ghildiyal & Zamore, 2009; Matzke & Birchler, 2005). Un des déterminants majeurs dans la classification des sRNAs est la structure de leurs précurseurs (Axtell, 2013). En effet la majorité des siRNAs ont pour origine des ARNdb qui peuvent être générés soit par l'action de RDR, soit par appariement intramoléculaire d'ARNb présentant une structure en tige-boucle parfaite ou quasi-parfaite. Les miRNAs sont eux générés à partir d'ARNdb relativement

court possédant une structure en tige-boucle présentant plusieurs mésappariement.

# 1 LES ENZYMES CLES DE LA BIOGENESE DES SRNA CHEZ

## LES PLANTES

### 1.1 LES PROTEINES DCLS

Chez la plante modèle, *Arabidopsis thaliana*, on trouve quatre homologues de Dicer nommés DCL1 à DCL4 (Margis *et al*, 2006; Henderson *et al*, 2006; Mukherjee *et al*, 2013). Ces protéines possèdent, comme chez les animaux, six domaines fonctionnels et produisent des sRNAs de différentes tailles remplissant des fonctions distinctes ou redondantes. Ces domaines sont: Le domaine DExD-box, hélicase-C, DUF283 (Domain of unknown function), PAZ (PIWI/ARGONAUTE/ZWILLE), RNase III, et dsRBD (dsRNA-binding domain) (**Figure II B**) (Mukherjee *et al*, 2013). Le rôle du domaine PAZ est de fixer l'extrémité 3' des précurseurs, permettant de positionner les deux domaines RNase III le long de l'ARNdb qu'ils vont cliver (Zhang *et al*, 2004). Le domaine PAZ et les deux domaines catalytiques sont connectés par une hélice alpha dont la longueur détermine la taille du sRNA produit (MacRae *et al*, 2006). Les domaines DExD et hélicase sont requis pour la translocation de la protéine le long de longue molécule d'ARNdb de façon ATP-dépendante. Enfin, le domaine DUF283 a été décrit comme un domaine potentiel de fixation à l'ARNdb et une plateforme d'interaction protéine-protéine, notamment avec certains co-facteurs (Qin *et al*, 2010).

La majorité des miRNAs est produite par DCL1 (Kurihara & Watanabe, 2004; Rajagopalan *et al*, 2006). Le rôle crucial de cette protéine, et des miRNAs dans le développement de la plante (cf. §2.1), est illustré par l'embryo-létalité de mutant nul de DCL1 (Vaucheret *et al*, 2004). DCL4, DCL2 et DCL3 sont responsables respectivement de la production des siRNAs de 21nt, 22nt, et 24nt de long (Chapman & Carrington, 2007). DCL4 est la

protéine DICER la plus polyvalente chez *A. thaliana*. Cette protéine intervient dans la production des trans-acting siRNAs (tasiRNAs, cf. §2.2) impliqués notamment dans la mise en place de la polarité des feuilles au cours du développement (Peragine *et al*, 2004; Vazquez *et al*, 2004b; Yoshikawa *et al*, 2005), mais également dans la production de certains miRNAs (Rajagopalan *et al*, 2006). De plus, DCL4 participe activement à la défense antivirale en étant le principal producteur de siRNAs antiviraux (viral-derived siRNAs, vsiRNAs) (Blevins *et al*, 2006; Bouché *et al*, 2006; Deleris *et al*, 2006). DCL3 est responsable de la production des siRNAs dit hétérochomatiques ou p4-siRNAs (Xie *et al*, 2004; Péliissier *et al*, 2011; Li *et al*, 2015; Blevins *et al*, 2015) représentant la classe de sRNA la plus abondante chez les plantes. Ils interviennent dans le maintien de la stabilité du génome, un mécanisme appelé RdDM (RNA directed DNA Methylation) (cf. §2.3) (Castel & Martienssen, 2013). DCL2, quant à lui, est considéré comme un substitut de DCL3 et DCL4 quand ceux-ci sont absents. Cette protéine joue également un rôle au cours de la défense antivirale et stimule la production de siRNAs secondaires lors de l'amplification du RNA silencing (Blevins *et al*, 2006; Moissiard & Voinnet, 2006; Voinnet, 2008; Parent *et al*, 2015). Enfin, DCL4, DCL3 et DCL2 interviennent dans la production de siRNAs provenant de séquences répétées et inversées (inverted-repeat, IR) présentes naturellement dans le génome d'*A. thaliana* (endoIR-siRNA, cf. §2.4) (Henderson *et al*, 2006).

## 1.2 LES PROTEINES RDRS

A l'exception des miRNAs et endoIR-siRNAs, la production des précurseurs ARNdb de la majorité des siRNAs endogènes ou exogènes nécessite l'action d'une RDR. Le génome d'*Arabidopsis* possède six gènes codant pour des RDRs (Wassenegger & Krczal, 2006) qui sont définies par un domaine catalytique en partie C-terminale (**Figure II B**), leur permettant de convertir un ARNb en ARNdb. A l'heure actuelle, seul RDR1, RDR2 et RDR6 ont une fonction connue.

RDR2 est essentielle à la biogenèse des p4-siRNAs (Xie *et al*, 2004; Jia *et al*, 2009), en produisant de courts ARNdb, généralement de 27 à 30 paires de bases, à partir de transcrits synthétisés par l'ARN polymérase ADN-dépendante IV (PolIV, cf. §2.3) (Blevins *et al*, 2015; Pikaard *et al*, 2008). RDR6 produit des ARNdb plus longs et intervient dans la défense antivirale et au cours de la production des tasiRNAs (cf. §2.2). Au cours de la défense antivirale, RDR6 participe à l'amplification de la réponse RNA silencing en convertissant les ARN viraux clivés par les vsiRNAs dit "primaires" en nouveaux ARNdb (Mallory & Vaucheret, 2009; Voinnet, 2008). Ces ARNdb serviront à produire de nouveaux vsiRNAs, dits "secondaires" qui permettront un renforcement de la défense antivirale en amplifiant le pool de sRNAs ciblant le virus (Mourrain *et al*, 2000; Donaire *et al*, 2008; Qu *et al*, 2008). Avec RDR6, RDR1 joue un rôle important dans la défense antivirale et l'inactivation simultanée de ces deux RDRs conduit à une hyper-susceptibilité contre de nombreux virus tels que le CMV (*Cucumber Mosaic Virus*) et le TMV (*Tobacco mosaic virus*) (Diaz-pendon *et al*, 2007; Qi *et al*, 2009; Wang *et al*, 2010). Il est intéressant de noter que l'expression de RDR1 est d'ailleurs fortement induite en réponse à la production d'acide salicylique (SA), une hormone de défense produite au cours de l'infection par de très nombreux pathogènes (Yu *et al*, 2003).

### 1.3 HEN1

Après leur production par les protéines DCLs, les duplex de sRNAs sont méthylés sur chacun des brins aux extrémités 3'. Cette action est catalysée par HEN1 (HUA ENHANCER 1), une 2'-O-méthyltransférase qui ajoute un groupement méthyle sur le 2' OH du ribose du dernier des deux nucléotides 3'sortants. Cette modification se fait séquentiellement sur chaque brin du duplex de sRNA, ce qui permet leur stabilisation en les protégeant de la dégradation (Yu *et al*, 2005; Li *et al*, 2005; Yang *et al*, 2006; Tkaczuk *et al*, 2006). De ce fait, HEN1 est essentielle à l'accumulation des miRNAs et des siRNAs *in vivo* (Chen *et al*, 2002; Park *et al*, 2002; Boutet *et al*,



2003). HEN1 possède cinq domaines fonctionnels dont quatre interagissant directement avec l'ARNdb (Huang *et al*, 2009; Baranauskė *et al*, 2015) (**Figure II B**). Les deux dsRBDs séparés par le domaine LCD (La motif-containing domain) reconnaissent et stabilisent le complexe HEN1-sRNAs. Le domaine LCD, *via* le motif La, fixe spécifiquement les 2 nucléotides 3'sortants du brin non méthylé, ainsi que l'extrémité 5' de l'autre brin, permettant le bon positionnement de l'extrémité 3' à méthyler, dans la poche formée par le domaine méthyltransférase (MT) (Huang *et al*, 2009). Le domaine PLD (PPIase-like domain) n'interagit pas avec l'ARNdb mais a été récemment montré comme requis pour l'association avec HYL1 (HYPONASTIC LEAVE 1), une protéine fixant l'ARNdb (dsRNA-binding protein, DRB) et impliquée dans la biogenèse des miRNAs (cf.§2.1) (Baranauskė *et al*, 2015).

Du fait qu'HEN1 soit requis pour la stabilisation des miRNAs, seuls les mutants hypomorphes sont viables. Des analyses sur ces mutants ont permis de révéler que la 2'-O-méthylation protège les sRNAs de l'ajout de groupement uridyl à leurs extrémités 3'. Cette modification est réalisée par au moins deux nucléotidyl transférases, HESO1 (HEN1 SUPPRESSOR 1) et URT1 (UTP:RNA URYDILTRANSFERASE 1), et constitue un signal de dégradation des sRNAs (Zhao *et al*, 2012; Ren *et al*, 2012; Tu *et al*, 2015; Wang *et al*, 2015). De plus une nucléase spécifique, SDN1 (SMALL DEGRADING NUCLEASE 1), intervient également dans la dégradation des miRNAs (Ramachandran & Chen, 2008).

## 1.4 LES PROTEINES AGOS

Les protéines AGOs sont les effecteurs majeurs du RNA silencing et, constituent un élément central dans le fonctionnement du complexe RISC (Meister, 2013). Les protéines AGOs eucaryotiques contiennent quatre domaines majeurs: le domaine N-terminal composé d'une région variable et les domaines MID (middle), PAZ et PIWI (P-element-induced wimpy testes) (**Figure II B**) (Song *et al*, 2004; Wang *et al*, 2008). Le domaine PAZ fixe le

nucléotide 3' du sRNA guide alors que la jonction des domaines MID et PIWI forme une poche pour fixer son nucléotide 5' (Wang *et al*, 2008). Ces trois domaines participent ensemble au bon positionnement des sRNAs par rapport à leurs cibles (Wang *et al*, 2009). Le domaine N-terminal semble quant à lui, faciliter la séparation du complexe sRNA-ARN cible (Mallory & Vaucheret, 2010). Le domaine PIWI possède également une activité endonucleolytique en adoptant une structure semblable à la RNase H de *Bacillus holodurans* (Song *et al*, 2004). Le génome d'*Arabidopsis thaliana* contient dix gènes codant pour des protéines AGOs (Morel *et al*, 2002; Vaucheret, 2008). Des études de séquençage des sRNAs associés aux différentes protéines AGOs ont montrées que l'identité du nucléotide 5', mais également la taille des sRNAs, influencent leur incorporation dans une AGO spécifique. Ainsi AGO1, -2, -5, -7 et AGO10 fixent les sRNAs de 21 et 22nt de long alors qu'AGO3, -4, -6 et AGO9 fixent ceux de 24 nt. De plus, AGO1 et AGO10 fixent préférentiellement les sRNAs possédant une uracyle en 5', AGO2, -3, -4, -6, -7 et -9 une adénine et AGO5 une cytosine (Mi *et al*, 2008; Takeda *et al*, 2008; Montgomery *et al*, 2008a; Zhu *et al*, 2011; Havecker *et al*, 2010).

AGO1 est la protéine AGO la plus importante chez les plantes puisqu'elle est majoritairement impliquée dans la voie des miRNAs, mais aussi dans la défense antivirale et la voie des tasiRNAs (Vaucheret *et al*, 2004; Bortolamiol *et al*, 2008; Vazquez *et al*, 2004b). Après reconnaissance, l'expression du gène ciblé sera réprimée soit par clivage, soit par inhibition de la traduction (Baumberger & Baulcombe, 2005; Brodersen *et al*, 2008). AGO2 joue également un rôle dans la voie des miRNAs (Maunoury & Vaucheret, 2011) mais semble être principalement impliquée dans la défense antivirale de manière additive à AGO1 (Harvey *et al*, 2011; Carbonell *et al*, 2012; Garcia-Ruiz *et al*, 2015). AGO4, AGO6 et AGO9 sont impliquées dans le maintien de la stabilité du génome en induisant la méthylation de l'ADN et/ou la modification d'histones au niveau des séquences ciblées, permettant entre autre la répression des TEs (Zilberman, 2003; Zheng *et al*, 2007; Havecker *et al*, 2010; Olmedo-Monfil *et al*, 2010). Très récemment, il a été

montré que AGO3 pouvait également agir de manière redondante à AGO4 et AGO6 et pourrait jouer un rôle important au cours de stress abiotiques (Zhang *et al*, 2016). AGO7 intervient dans la biogenèse de certains tasiRNAs, *via* son association avec miR390, et participe à la régulation de la morphologie foliaire au cours du développement (Montgomery *et al*, 2008a; Axtell *et al*, 2006). AGO10 intervient dans la régulation du méristème apical caulinaire et la mise en place de la polarité des feuilles *via* son association avec miR165/166 (Mallory *et al*, 2009; Liu *et al*, 2009; Zhu *et al*, 2011). AGO5 serait, quant à elle, impliquée dans le développement des gamètes mâles et femelles et dans la défense antivirale (Takeda *et al*, 2008; Tucker *et al*, 2012; Borges *et al*, 2011; Brosseau & Moffett, 2015). Enfin AGO8 est considérée comme un pseudogène au regard de sa faible expression tout au long du développement d'*Arabidopsis thaliana*. De ce fait aucune fonction n'a pu être jusqu'à lors attribuée à cette protéine (Bologna & Voinnet, 2014).

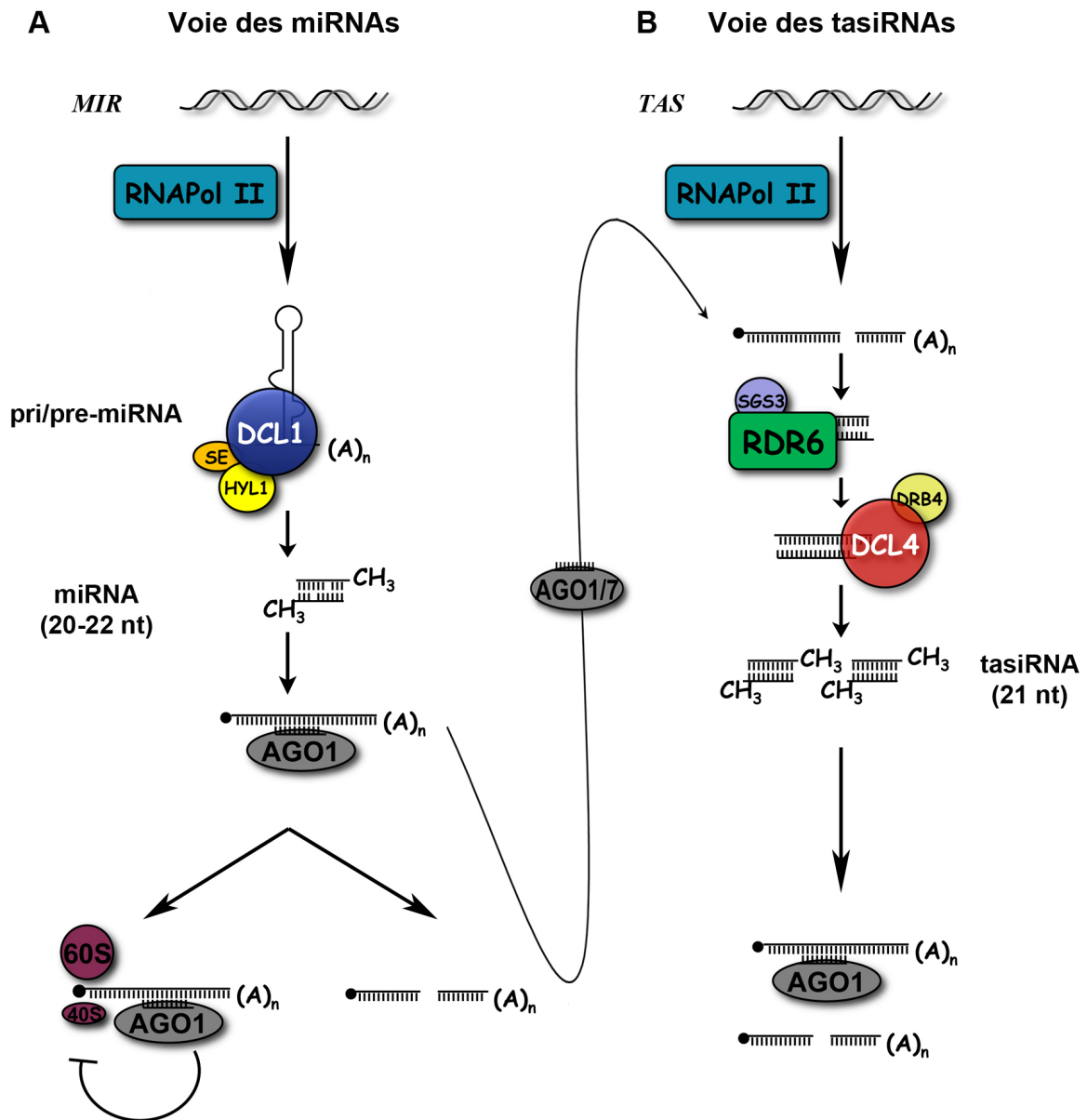
## **2 LES DIFFERENTES VOIES ENDOGENES DU RNA SILENCING**

### **CHEZ ARABIDOPSIS THALIANA**

Chez *Arabidopsis* plusieurs classes fonctionnelles de sRNAs sont présentes et peuvent être classées en fonction de leur biogenèse, de leur mode d'action et de leur fonction au sein de la cellule (Axtell, 2013; Borges & Martienssen, 2015).

#### **2.1 LA VOIE DES MICRO-RNAs**

Les miRNAs sont les sRNAs les plus étudiés à l'heure actuelle et jouent un rôle central dans de nombreux processus biologiques tels que le



**Figure III: Voie des miRNAs et des tasiRNAs chez *Arabidopsis thaliana***

(A) Les gènes codant pour les précurseurs de miRNAs (*MIR*) sont transcrits en ARNs par Pol II. Après repliement intra-moléculaire, ces ARN vont adopter une structure en tige-boucle pour former les pri-miRNAs. Ces pri-miRNAs vont être reconnus et clivés par DCL1, assisté de HYL1 et SE, pour produire les miRNAs matures. Un des brins du duplex de miRNAs sera incorporé dans AGO1 pour induire l'inhibition post-transcriptionnelle du gène cible soit, par répression traductionnelle, soit, par clivage de l'ARNm ciblé. Cette voie est également à la base de la voie des tasiRNAs (B) Dans cette voie certains miRNAs incorporés dans AGO1 ou AGO7 vont induire le clivage de transcrits généralement non codants. Un des fragments issu du clivage de ces transcrits va ensuite être convertis en ARNdb par RDR6, en complexe avec SGS3. Cet ARNdb sera clivé par DCL4, assisté de DRB4, pour former les tasiRNAs de 21 nt de long. Comme dans le cas des miRNAs un des brins du duplex de tasiRNAs sera incorporé dans AGO1 pour induire la répression en *trans* des gènes cibles.

développement ou la résistance aux stress (Kozomara & Griffiths-Jones, 2011). La majorité des gènes de miRNA (*miR*) forment des unités de transcription propres et sont dépendants de l'ARN polymérase ADN-dépendante II (PolII) pour leur transcription (**Figure III**) (Xie *et al*, 2005a; Zhao *et al*, 2013). Plusieurs autres composants généraux de la machinerie de transcription des ARNm (p.e. le complexe Mediator ; Kim *et al*, 2011), mais également des facteurs spécifiques (p.e. FUS3; Wang & Perry, 2013) sont nécessaires à l'expression des *miRNAs*.

Une fois transcrits, les précurseurs de miRNAs sont coiffés en 5' et polyadénylés en 3' pour former les pri-miRNAs. Ces pri-miRNAs vont intrinsèquement adopter une structure en tige-boucle présentant une structure simple brin aux extrémités 5' et 3' (Xie *et al*, 2005a). La maturation de la majorité des pri-miRNAs en duplex de miRNAs matures nécessite la présence de DCL1 et de ses cofacteurs, tels que HYL1 et SE, pour former le complexe de maturation des miRNAs. Ce complexe semble être localisé exclusivement dans le noyau où il apparaît sous forme de foyers discrets appelés "Dicing bodies" (D-bodies) (Reinhart *et al*, 2002; Han *et al*, 2004; Vazquez *et al*, 2004a; Lobbes *et al*, 2006; Song *et al*, 2007; Fang & Spector, 2007). Cette maturation se réalise en deux étapes séquentielles impliquant le clivage du pri-miRNA par DCL1, générant successivement le pré-miRNA et le duplex de miRNA mature, majoritairement de 21 nt de long (Bartel, 2004; Kim, 2005; Carthew & Sontheimer, 2009). Les protéines HYL1 et SE sont nécessaires au bon fonctionnement de DCL1 en augmentant l'efficacité et la précision du clivage de cette dernière (Kurihara *et al*, 2006; Dong *et al*, 2008). De nombreuses études ont montré qu'en plus des facteurs impliqués dans la maturation des miRNAs présentés ci-dessus, plusieurs déterminants structuraux des pré-miRNAs sont reconnus par DCL1. Ceux-ci influencent et initient les deux événements séquentiels de clivage générant le duplex de miRNA mature (Bologna *et al*, 2009; Mateos *et al*, 2010; Bologna *et al*, 2013).

Une fois le duplex de miRNA mature produit, celui-ci va être 2'-O-méthylé aux extrémités 3' par HEN1 (**Figure III**) afin de les stabiliser et d'empêcher sa dégradation, notamment par SND1 (Ramachandran & Chen,

2008). Les miRNAs matures sont ensuite, *a priori*, exportés dans le cytoplasme par la protéine HASTY (HST; Bollman *et al*, 2003; Park *et al*, 2005), l'orthologue de l'exportin-5 chez les animaux (Yi *et al*, 2003; Lund *et al*, 2004). Toutefois, la fonction exacte d'HST dans la biogénèse des miRNAs demeure, à l'heure actuelle, mal définie et nécessite de plus amples investigations. Une fois dans le cytoplasme, le duplex de miRNA est chargé dans le complexe RISC contenant une protéine AGO, généralement AGO1. Un des deux brins est sélectionné pour servir de guide tandis que l'autre (le passager) est généralement dégradé (Iki *et al*, 2010; Endo *et al*, 2013). De manière intrigante, HYL1, une protéine localisée exclusivement dans le noyau, semble participer à la sélection du brin guide incorporé dans AGO1 (Eamens *et al*, 2009), suggérant que cette étape pourrait également avoir lieu dans ce compartiment cellulaire. Une fois sélectionné, le brin guide permettra au complexe RISC de reconnaître tout ARNm présentant une complémentarité de séquence (Llave *et al*, 2002; Rhoades *et al*, 2002). Cette reconnaissance entraînera soit, le clivage de l'ARNm ciblé, soit l'inhibition de sa traduction. Dans le premier cas, une fois clivé, les parties 5' et 3' du produit de clivage seront respectivement dégradées par l'exosome et par l'exoribonuclease 5'-3' XRN4, (**Figure III**) (German *et al*, 2008; Bologna & Voinnet, 2014). Ce mode de répression semble crucial à la plante car l'expression d'un allèle d'*AGO1* déficient pour son activité de clivage, amplifie les défauts développementaux observés dans un mutant *agol* hypomorphe (Carbonell *et al*, 2012). L'inhibition traductionnelle des ARNm ciblés a probablement lieu au niveau des polysomes car un certain nombre de miRNAs a été retrouvé associé à ceux-ci de façon AGO1-dépendante (**Figure III**) (Brodersen *et al*, 2008, 2012; Lanet *et al*, 2009). Chez les animaux, la répression de la traduction initiée par les miRNAs est également associée à une séquestration des ARNm ciblés, dans des foyers cytoplasmiques appelés "Processing bodies" (P-bodies) liés à la dégradation des ARN (Bologna & Voinnet, 2014). Ce phénomène peut potentiellement être également vrai chez les plantes où il a été montré que SUO et VARICOSE (VCS), deux protéines associées aux P-bodies et interagissant avec des protéines impliquées dans la dégradation de l'ARN chez *Arabidopsis*, jouent un rôle dans la répression de

la traduction miRNA-dépendante (Brodersen *et al*, 2008; Yang *et al*, 2012). De plus une fraction de AGO1 semble être localisée au niveau de ces foyers (Xu & Chua, 2011).

## 2.2 LA VOIE DES *TRANS-ACTING* siRNAs

Une autre voie du RNA silencing endogène chez *A. thaliana* est la voie des tasiRNAs. L'accumulation de ces siRNAs est sous la dépendance de la voie des miRNAs, car leur production dépend du clivage de leur précurseur par certains miRNAs spécifiques (Yoshikawa *et al*, 2005; Allen *et al*, 2005). Ces ARN, transcrits par PolIII, sont généralement non codants bien que, récemment, des tasiRNAs issus d'ARNm ont pu être mis à jour (**Figure III**) (Howell *et al*, 2007; Zhai *et al*, 2011; Shivaprasad *et al*, 2012; Boccara *et al*, 2014). Chez *A. thaliana* huit loci correspondants à des transcrits de *TAS* ont été identifiés (*TAS1a-c*, *TAS2*, *TAS3a-c* et *TAS4*) (Allen *et al*, 2005; Rajagopalan *et al*, 2006). Ces transcrits une fois clivés par un complexe AGO/miRNA spécifique, vont être convertis en ARNdb par RDR6 et reconnus principalement par DCL4, assisté de DRB4 une DRB, pour produire des siRNAs de 21 nt, en phase à partir du site de clivage reconnu par le complexe AGO/miRNA (**Figure III**) (Gascioli *et al*, 2005; Peragine *et al*, 2004; Ronemus *et al*, 2006; Fei *et al*, 2013).

La production des tasiRNAs matures se fait suivant deux grands mécanismes selon le locus d'origine (Bologna & Voinnet, 2014). Dans le premier mécanisme, dit du "one hit", un site unique présent en position 5' des transcrits *TAS1/TAS2* et *TAS4* est reconnu, respectivement, par les miRNAs miR173 et miR828 (Yoshikawa *et al*, 2005; Allen & Howell, 2008). Ces miRNAs de 22 nt, toujours en complexe avec AGO1, vont induire le clivage des transcrits *TAS* correspondants et le fragment 3' libéré sera pris en charge successivement par RDR6 et DCL4 (Peragine *et al*, 2004; Gascioli *et al*, 2005). Le second mécanisme, dit du "two hits", concerne exclusivement les tasiRNAs provenant des loci *TAS3a-c* (Axtell *et al*, 2006). Dans ce mécanisme

deux sites, un 5' et un 3', sont reconnus par le complexe AGO7/miR390 dont l'association est extrêmement spécifique (Montgomery *et al*, 2008a; Endo *et al*, 2013). Ce complexe va ainsi reconnaître ces sites et induire le clivage du transcrit *TAS3* uniquement au niveau du site 3'. Bien que non clivé, la reconnaissance du site 5' par le complexe AGO7/miR390 est indispensable à la production des tasiRNAs issus de ce précurseur (Montgomery *et al*, 2008a). Une fois clivé, le fragment 5' est converti par RDR6 en ARNdb jusqu'au site de reconnaissance 5' où le complexe RISC, fixé, stoppe la progression de RDR6, définissant ainsi la taille du précurseur ARNdb (Rajeswaran & Pooggin, 2012).

En plus des facteurs énoncés précédemment, un certain nombre d'autres protéines impliquées dans la biogénèse des tasiRNAs ont pu être identifiées. C'est le cas par exemple de la protéine SGS3 (SUPPRESSOR OF GENE SILENCING 3) qui participe à la biogénèse des tasiRNAs en stabilisant le produit 3' de clivage favorisant ainsi la production de l'ARNdb par RDR6 (**Figure III**) (Yoshikawa *et al*, 2013). D'autres facteurs plus généraux du RNA silencing, comme HEN1, participent également à la biogénèse des tasiRNAs (Vazquez *et al*, 2004b). Une fois produits ces sRNAs sont chargés dans AGO1 (**Figure III**) pour réguler par clivage l'expression d'ARNm codants notamment pour des facteurs de transcription de type ARFs (AUXINE RESPONSE FACTOR). Les tasiRNAs participent entre autre, à la mise en place de la polarité et de la morphologie foliaire, ou encore dans l'adaptation de la plante aux stress environnementaux (Peragine *et al*, 2004; Vazquez *et al*, 2004b; Allen *et al*, 2005; Garcia *et al*, 2006; Fahlgren *et al*, 2006; Luo *et al*, 2012).

## 2.3 LA VOIE DES siRNAs HÉTÉROCHROMATIQUES

La deuxième classe de siRNAs endogènes présente chez *A. thaliana* est la classe des siRNAs hétérochromatiques, ou p4-siRNAs, impliquée dans la voie dite du RdDM. Contrairement aux deux autres classes de sRNAs



## Voie des P4-siRNAs

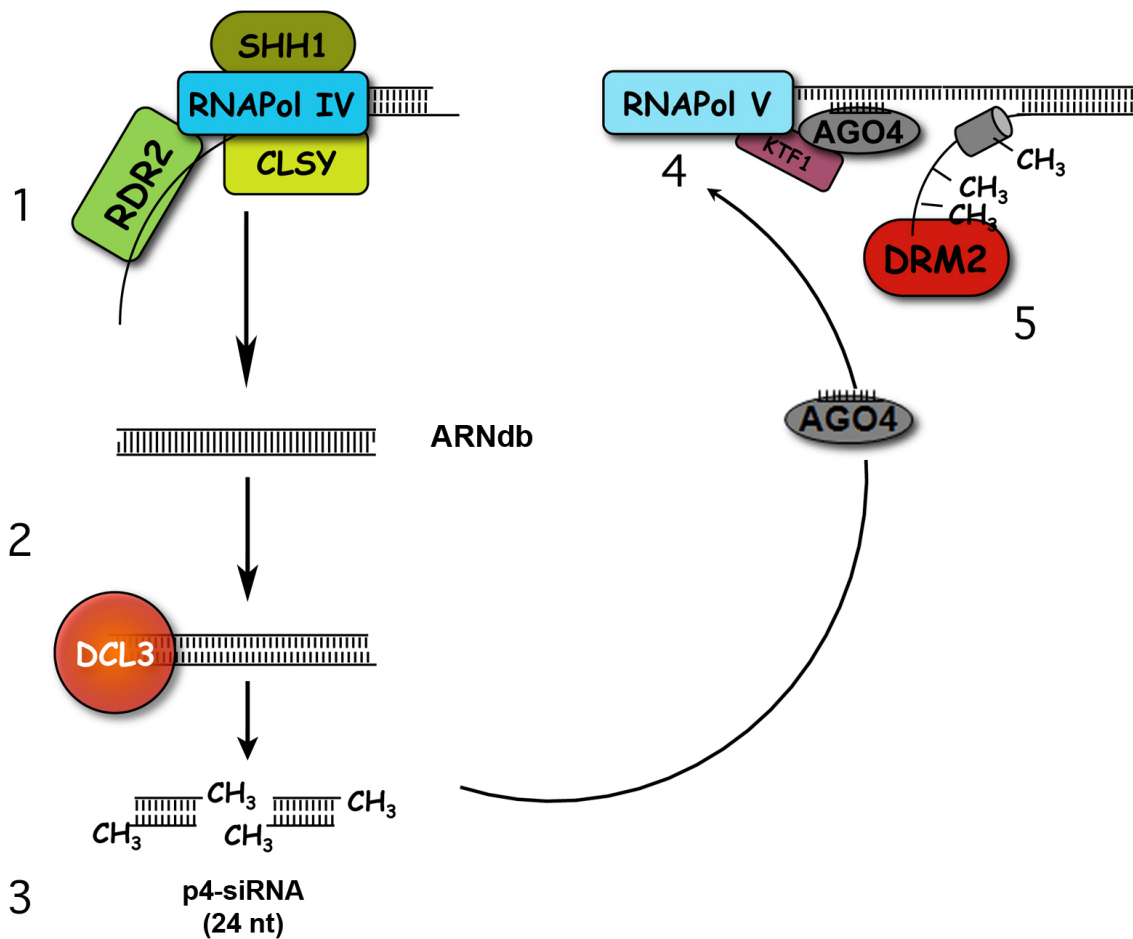


Figure IV: Voie des siRNAs hétérochromatiques (p4-siRNAs) chez *Arabidopsis thaliana*

(1) Les précurseurs de p4-siRNAs sont transcrits par l'ARN polymérase ADN-dépendante Pol IV, associée à SHH1 et CLSY. (2) Ces ARN seront convertis en ARNdb par RDR2 puis (3) clivés par DCL3 pour produire des sRNAs de 24 nt de long (3). (4) Ces sRNAs seront incorporés majoritairement dans AGO4, pour cibler des transcrits produits par l'ARN polymérase ADN-dépendante Pol V. (5) Cette reconnaissance, aidée par KTF1, va recruter DRM2 pour induire la méthylation de l'ADN aux niveaux des loci ciblés.

présentées précédemment (cf. §2.1 et 2.2), les précurseurs de ces siRNAs ne sont pas transcrits par PolII mais par des ARN polymérase ADN-dépendantes présentes uniquement chez les plantes, PolIV et PolV (**Figure IV**) (Onodera *et al*, 2005; Herr *et al*, 2005; Kanno *et al*, 2005; Pontier *et al*, 2005).

Bien que le recrutement de PolIV, au niveau des loci qu'elle transcrit, ne soit pas encore complètement élucidé à ce jour, le modèle actuel propose que les protéines SHH1 (SAWADEE HOMEODOMAIN HOMOLOGUE 1) et CLSY1 (CLASSY 1) participent à la reconnaissance des loci ciblés par cette polymérase, principalement au niveau de transposons et d'éléments répétés (**Figure IV 1**) (Law *et al*, 2013; Zhang *et al*, 2013; Smith *et al*, 2007; Law *et al*, 2011). De part son interaction avec PolIV, RDR2 va ensuite convertir ces transcrits en ARNdb de petites tailles, 30-50 nt de long (**Figure IV 2**) (Xie *et al*, 2004; Law *et al*, 2011; Haag *et al*, 2012; Blevins *et al*, 2015; Zhai *et al*, 2015; Li *et al*, 2015), qui seront reconnus et clivés par DCL3 pour générer des siRNAs de 24 nt (p4-siRNAs) (**Figure IV 3**). Après méthylation par HEN1 (Xie *et al*, 2004), ces siRNAs seront incorporés dans, au moins, trois des dix protéines AGO présentes chez *A. thaliana*, AGO4, AGO6 et AGO9 (Zilberman, 2003; Zheng *et al*, 2007; Havecker *et al*, 2010; Olmedo-Monfil *et al*, 2010; Tucker *et al*, 2012). Bien qu'AGO4 semble être l'effecteur principal de la voie RdDM, AGO6 et AGO9 semblent jouer un rôle redondant à AGO4 mais aussi spécifique selon les tissus analysés (Havecker *et al*, 2010; Eun *et al*, 2011; Tucker *et al*, 2012).

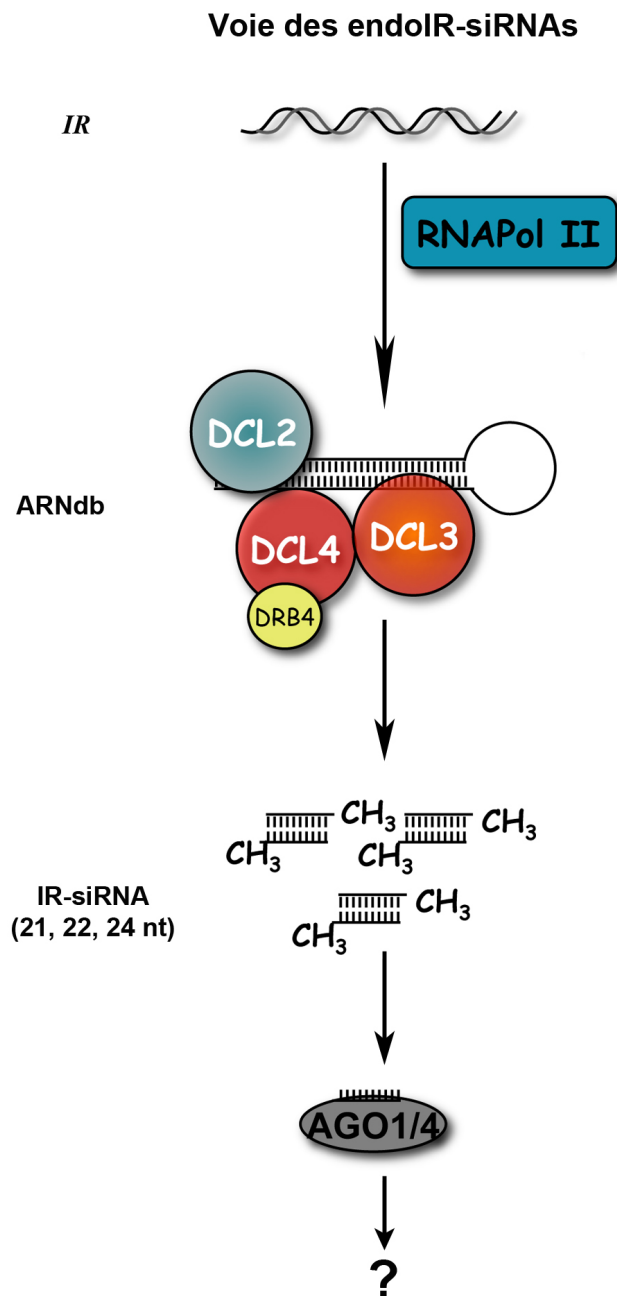
Le complexe AGO4/p4-siRNA va ensuite reconnaître les transcrits produits par PolV afin d'induire l'apposition de marques répressives au niveau des loci dont ils proviennent (**Figure IV 4**). Cette reconnaissance sera facilitée par l'interaction du complexe AGO4/p4-siRNA, avec PolV, *via* le domaine C-terminal de sa grande sous-unité, au niveau d'une région appelée "AGO hook" (Li *et al*, 2006; El-Shami *et al*, 2007). La protéine KTF1 (KOW-DOMAIN CONTAINING TRANSCRIPTION FACTOR 1) participe également au recrutement et à la fixation d'AGO4 au niveau des transcrits PolV en fixant directement AGO4 *via* son domaine AGO-hook, assurant ainsi un ancrage solide du complexe RISC au niveau des loci ciblés (He *et al*, 2009; Bies-Etheve

*et al*, 2009). La fixation d'AGO4 va permettre le recrutement de l'ADN méthyltransférase DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE 2) qui va catalyser la méthylation des cytosines au niveau des loci transcrits par PolV (**Figure IV 5**)(Cao & Jacobsen, 2002). Le RdDM peut également être impliqué dans la modification d'histones, notamment par méthylation de la Lysine 9 de l'histone H3 (H3K9). Environ 70 % des loci cibles du RdDM sont modifiés par cette méthylation renforçant ainsi leur répression (Bernatavichute *et al*, 2008). La méthylation de l'ADN est requise pour le recrutement de l'histone méthyltransférase SUVH4/KYP (KRYPTONITE) au niveau des H3K9 (Johnson *et al*, 2007, 2014) .

Contrairement aux miRNAs et aux tasiRNAs qui jouent des rôles cruciaux dans le développement et la physiologie de la plante, la plupart des mutants de la voie RdDM ne présente pas de défauts développementaux particuliers. Cependant plusieurs études montrent que cette voie est impliquée dans un grand nombre de processus biologiques notamment au niveau de la reproduction et de l'adaptation des plantes aux stress biotiques et abiotiques (Matzke & Mosher, 2014). Le rôle principal du RdDM est le maintien de la stabilité du génome en participant à la répression de l'expression des transposons, de manière complémentaire à la voie de maintenance de la méthylation des cytosines, opérée notamment par MET1 (METHYLTRANSFERASE 1) et CMT3 (CHROMOMETHYLTRANSFERASE 3) (Aufsatz *et al*, 2004; Cao *et al*, 2003). De plus, le RdDM participe également à la défense antivirale contre les virus à ADN, tels que les Geminivirus (Blevins *et al*, 2006; Raja *et al*, 2008; Buchmann *et al*, 2009; Incarbone & Dunoyer, 2013).

## 2.4 LA VOIES DES ENDOIR-SiRNAs

Diverses analyses par séquençage à haut-débit des sRNAs présents chez *A. thaliana* ont permis de mettre en évidence l'existence de loci discrets présentant des séquences répétées et inversées (IR) et pouvant adopter une



**Figure V: Voie des endoIR-siRNA chez *Arabidopsis thaliana***

Certains ARNs produits par Pol II à partir de loci présentant une séquence répétée et inversée vont adopter une structure en tige-boucle quasi parfaite une fois transcrits. Ces structures en ARNdb seront clivées majoritairement par DCL2, DCL3 et DCL4 pour produire, respectivement, des endoIR-siRNAs de 22, 24 et 21 nt de long. Un des brins de ces siRNAs sera incorporé dans AGO1 (21,22 nt) ou dans AGO4 (24 nt) pour induire la répression de cibles encore inconnues à l'heure actuelle.

structure en tige boucle une fois transcrit, mais ne remplissant pas les critères d'annotation des miRNAs (**Figure V**) (Lindow et al, 2007; Kasschau et al, 2007). Certains de ces loci, produisant de longs ARNdb parfaits ou quasi-parfaits, sont la source de la troisième catégorie de siRNAs endogènes: les endoIR-siRNAs (Henderson et al, 2006; Zhang et al, 2007; Vazquez et al, 2008; Yelina et al, 2010; Shamandi et al, 2015). Ces ARNdb, comme les précurseurs de miRNAs et tasiRNAs, sont transcrits par PolIII (**Figure V**) (Zhang et al, 2007; Yelina et al, 2010). Cependant, contrairement aux précurseurs de miRNAs et tasiRNAs, ces ARNdb vont être reconnus et clivés majoritairement par DCL3, DCL2 et DCL4, et, dans une moindre mesure, par DCL1 (**Figure V**) (Henderson et al, 2006; Vazquez et al, 2008; Shamandi et al, 2015). Le rôle biologique des endoIR-siRNAs n'étant pas connu à l'heure actuelle, les protéines AGOs responsables de leur fonction au sein de la plante n'ont pu être déterminées avec exactitude. Cependant, par analogie aux siRNAs produits à partir d'IRs exogènes (exoIR-siRNAs) couramment utilisées au laboratoire, on peut postuler que les endoIR-siRNAs de 21/22nt seront probablement chargés dans AGO1, et ceux de 24nt dans AGO4, afin d'induire, respectivement, le clivage de l'ARNm ou la méthylation du locus ciblé (**Figure V**) (Himber et al, 2003; Dunoyer et al, 2007; Smith et al, 2007). Alternativement, ces endoIR-siRNAs peuvent n'avoir aucune fonction biologique à proprement parlé et n'être que des proto-miRNAs en cours d'évolution (Vazquez et al, 2008).

# OBJECTIFS DE LA THESE

En plus des protéines DICERs, des protéines DRBs jouent un rôle crucial dans la production des diverses catégories de sRNAs. En effet chez l'insecte *Drosophila melanogaster* la DRB R2D2 assiste DCR2 pour la biogénèse des siRNAs (Tomari *et al*, 2004). Le génome d'*Arabidopsis* code pour 5 DRBs (DRB1 à DRB5) qui sont strictement composées de deux domaines dsRBD sans aucun autre domaine catalytique (Hiraguri *et al*, 2005; Curtin *et al*, 2008). Parmi ces 5 DRBs, DRB1 ou HYL1 est la plus étudiée et joue un rôle crucial dans la biogénèse des miRNAs en augmentant l'efficacité ainsi que la précision du clivage de leurs précurseurs par DCL1 (cf.§2.1), avec qui elle interagit directement (Hiraguri *et al*, 2005; Dong *et al*, 2008). De plus, DRB1 est requise pour la sélection du brin guide et son incorporation dans AGO1 (Eamens *et al*, 2009). DRB2 joue également un rôle dans la voie de biogénèse des miRNAs mais seulement au niveau de certains tissus spécifiques tels que les fleurs (Reis *et al*, 2015). De manière surprenante, les miRNAs produits par DCL1/DRB2 semblent inhiber l'expression des ARNm ciblés majoritairement par répression traductionnelle alors que ceux produits par DCL1/DRB1 semblent agir principalement par clivage (Reis *et al*, 2015). Si DRB3 et DRB5 n'ont pas de fonction connue à l'heure actuelle, DRB4 est requise à l'activité de DCL4, en augmentant son efficacité, notamment pour la production des tasiRNAs et des miRNAs DCL4-dépendants tel que miR822. (Nakazawa *et al*, 2007; Pélissier *et al*, 2011; Clavel *et al*, 2016).

Récemment, une nouvelle famille de DRB a été identifiée chez *Arabidopsis*, mais leur rôle dans les différentes voies du RNA silencing n'était pas encore défini. Un des objectifs de ma thèse a donc été de caractériser la fonction d'une de ces nouvelles DRB, la protéine DRB7.2.

Dans un deuxième temps, mon travail de thèse a consisté à développer et valider des outils moléculaires dans le but de caractériser le mode d'action

et de reconnaissance des différents substrats de DCL4 et d'identifier ses déterminants de spécificités. En effet, les protéines DCLs sont des éléments clés du mécanisme du RNA silencing. Bien que DCL4 soit, avec DCL1, la protéine DCL la plus étudiée chez les plantes, son mode d'action et son mode de reconnaissances des ARNdb restent encore mal compris chez *A. thaliana*. De plus, malgré la grande polyvalence de DCL4 (cf. §1.1), DRB4 est, à l'heure actuelle, le seul cofacteur connu de cette protéine. De façon surprenante, plusieurs mutations faux-sens de DCL4 ont permis de mettre en évidence un découplage dans la production des différents sRNAs DCL4-dépendant suggérant l'existence de déterminants spécifiques des différentes voies du RNA silencing dans lesquelles cette protéine intervient. Mon objectif a donc été de déterminer si ces spécificités d'action dépendaient de cofacteur(s) non identifié(s) ou de déterminants structuraux intrinsèques à DCL4.

# *RÉSULTATS*



**CHAPITRE 1: ÉTUDE FONCTIONNELLE DE  
DRB7.2, UNE NOUVELLE PROTEINE DE  
FIXATION A L'ARN DOUBLE BRIN**


# 1.1 HISTOIRE EVOLUTIVE DES PROTEINES DRBS CHEZ LES PLANTES

Comme indiqué précédemment (cf. Introduction §2.1, 2.2, objectif de la thèse), certaines voies du RNA silencing requièrent la présence de protéines DRB pour un fonctionnement optimal (Vazquez *et al*, 2004a; Hiraguri *et al*, 2005; Fukudome *et al*, 2011). Parmi les 5 DRB identifiées au préalable, deux sont requises dans la voie des miRNAs DCL1-dépendant, et une intervient dans la production de divers sRNAs DCL4-dépendant. Toutefois, à l'heure actuelle, aucune DRB n'a de fonction connue dans la voie des p4-siRNAs et des endoIR-siRNAs. Au cours d'une collaboration avec l'équipe de J.M. Deragon au LGDP (Laboratoire Génome et Développement de Plantes, Perpignan), j'ai participé à l'identification et la caractérisation de deux nouvelles familles de DRB (DRB6 et DRB7), isolées au cours d'une étude phylogénétique basée sur la recherche de séquences correspondant au dsRBD exclusivement. DRB6, bien que conservée chez la quasi-totalité des plantes vasculaires, est absente des brassicacées, dont *Arabidopsis* fait partie, rendant impossible l'étude de son rôle dans cette plante modèle. La seconde famille de DRB nouvellement identifiée est composée de deux membres (DRB7.1 et DRB7.2). Elles se caractérisent par la présence d'un domaine dsRBD unique ayant évolué de manière concertée avec le second dsRBD de DCL4. Ces protéines étant présentes dans l'ensemble des plantes vasculaires, je me suis attaché à caractériser en détail la fonction de DRB7.2. En utilisant la méthode du BiFC (Bimolecular Fluorescence Complementation), j'ai ainsi pu mettre en évidence que cette protéine ne semblait pas interagir avec DCL4 malgré la co-évolution de leur domaine dsRBD. De manière surprenante, DRB7.2 semble plutôt interagir avec DRB4, le cofacteur de DCL4, mais nous n'avons, dans un premier temps, pas pu mettre en évidence d'effet notable de l'absence de DRB7.2 sur la biogénèse des différents sRNAs DCL4/DRB4-dépendant. Cependant, en observant l'accumulation de certains easiRNAs (epigenetically-activated siRNAs), une nouvelle classe de sRNAs impliqués

dans la répression de transposons au niveau des tissus reproducteurs et principalement produits par DCL4/DCL2 (McCue *et al*, 2012; Creasey *et al*, 2014), nous avons pu constater qu'en l'absence de DRB7.2, les niveaux de easiRNAs de 24nt DCL3-dépendant étaient augmentés de façon significative. Toutefois, la fonction de ces easiRNAs de 24 nt n'étant pas encore connue, le rôle de DRB7.2 dans cette voie particulière nécessite de plus amples investigations.

L'ensemble des résultats obtenus au cours de cette collaboration a été réuni au sein d'un manuscrit publié en Février 2016 dans le journal *Plant Molecular Biology*.

# Evolutionary history of double-stranded RNA binding proteins in plants: identification of new cofactors involved in easiRNA biogenesis

Marion Clavel<sup>1,2</sup> · Thierry Pélissier<sup>3</sup> · Thomas Montavon<sup>4</sup> · Marie-Aude Tschopp<sup>5</sup> · Marie-Noëlle Pouch-Pélissier<sup>3</sup> · Julie Descombin<sup>1,2</sup> · Viviane Jean<sup>1,2</sup> · Patrice Dunoyer<sup>4</sup> · Cécile Bousquet-Antonelli<sup>1,2</sup> · Jean-Marc Deragon<sup>1,2</sup> 

Received: 31 August 2015 / Accepted: 3 February 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** In this work, we retrace the evolutionary history of plant double-stranded RNA binding proteins (DRBs), a group of non-catalytic factors containing one or more double-stranded RNA binding motif (dsRBM) that play important roles in small RNA biogenesis and functions. Using a phylogenetic approach, we show that multiple dsRBM DRBs are systematically composed of two different types of dsRBMs evolving under different constraints and likely fulfilling complementary functions. In vascular plants, four distinct clades of multiple dsRBM DRBs are always present with the exception of Brassicaceae species, that do not possess member of the newly identified clade we named DRB6. We also identified a second new and highly conserved DRB family (we named DRB7) whose members possess a single dsRBM that shows concerted evolution with the most C-terminal dsRBM domain of the Dicer-like 4 (DCL4) proteins. Using

a BiFC approach, we observed that *Arabidopsis thaliana* DRB7.2 (AtDRB7.2) can directly interact with AtDRB4 but not with AtDCL4 and we provide evidence that both AtDRB7.2 and AtDRB4 participate in the epigenetically activated siRNAs pathway.

**Keywords** RNAi · siRNA · easiRNAs · Double-stranded RNA binding protein · *Arabidopsis* · Evolution

## Introduction

The double-stranded RNA binding motif (dsRBM) is evolutionarily conserved to mediate specific interaction with double-stranded RNAs (dsRNAs) of various origin and structure, and plays critical roles in a wealth of cellular processes in prokaryotes, eukaryotes and viruses. This motif was first characterized as a 65–70 amino acids domain (St Johnston et al. 1992) that adopts an  $\alpha\beta\beta\beta\alpha$  fold and contains conserved residues important for both structure and function (reviewed in Chang and Ramos 2005; Masliah et al. 2013). Several dsRBMs can also mediate protein–protein interaction allowing intramolecular packing (Sohn et al. 2007), homodimerization (Hitti et al. 2004; Yang et al. 2010), heterodimerization involving two dsRBMs (Laraki et al. 2008; Qin et al. 2010) or involving a dsRBM and other functional domains (Daniels et al. 2009; Nanduri et al. 2000; Schuldt et al. 1998).

DsRBM-containing proteins are involved in numerous cellular mechanisms ranging from localization and transport of messenger RNAs, through maturation and degradation of RNAs, to viral response and signal transduction. The animal Staufen protein is a striking example of the versatility achieved by a single dsRBM-containing protein. This protein contains no additional domain but five

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-016-0448-9) contains supplementary material, which is available to authorized users.

✉ Jean-Marc Deragon  
jean-marc.deragon@univ-perp.fr

<sup>1</sup> UMR5096 LGDP, Université de Perpignan Via Domitia, 58 Avenue Paul Alduy, 66860 Perpignan Cedex, France

<sup>2</sup> CNRS UMR5096 LGDP, Perpignan Cedex, France

<sup>3</sup> UMR 6293 CNRS - INSERM U1103 – GreD, Clermont Université, 24 avenue des Landais, B.P. 80026, 63171 Aubière Cedex, France

<sup>4</sup> Institut de Biologie Moléculaire des Plantes du CNRS, UPR2357, Université de Strasbourg, Strasbourg Cedex, France

<sup>5</sup> Department of Biology LFW D17/D18, ETH Zürich, Universitätsstrasse 2, 8092 Zurich, Switzerland

dsRBMs with varying affinities towards dsRNA and proteins. Staufen is essential for establishing the correct polarity of the *Drosophila* embryo by anchoring two specific maternal mRNAs, *oskar* and *bicoid*, at the opposite poles of the oocyte (St Johnston et al. 1991). The association of Staufen with the 3'UTR of *bicoid* is essential for the transport of the resulting ribonucleoprotein (RNP) complex along microtubules (Ferrandon et al. 1994), while the association of Staufen with *oskar* is more specifically involved in regulating translation after mRNA deposition (Micklem et al. 2000). More recently, STAU1 (human Staufen) was shown to bind the 3'UTRs of messengers base paired with *Alu* transcripts and to induce, with the help of UPF1 a known nonsense mediated decay factor, a specific degradation pathway (Gleghorn et al. 2013; Kim et al. 2007). STAU1 association to 5'UTR hairpins has also been shown to enhance translation of other targeted mRNAs (Dugre-Brisson et al. 2005).

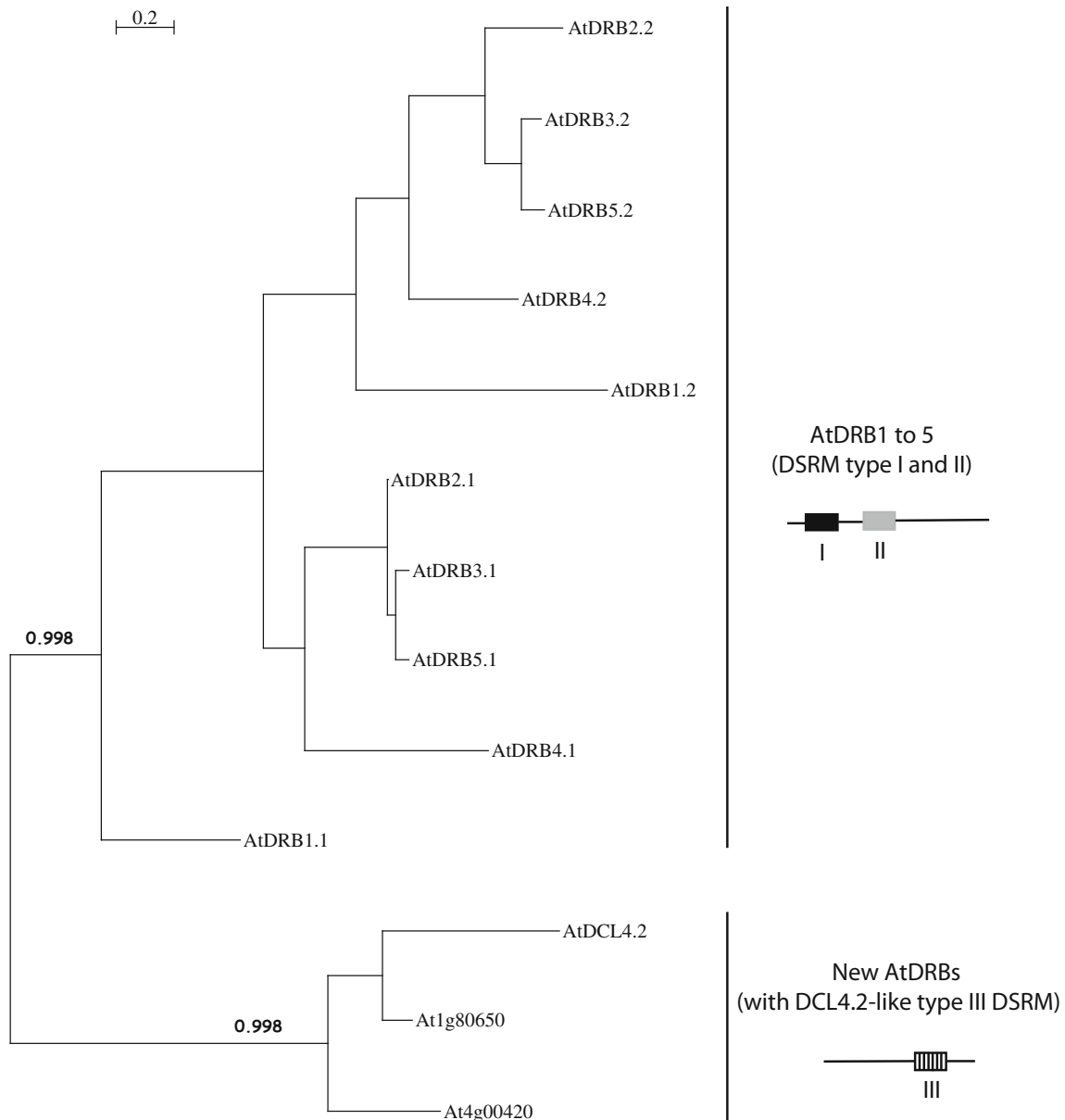
DsRBM-containing proteins were also found to be tightly linked to RNA interference mechanisms. Animal Drosha and Dicer, which are part of the RNase III family, have been extensively studied for their ability to generate from dsRNA precursors, small RNA molecules involved in transcriptional or post-transcriptional gene regulation. In a pathway that is conserved in human, flies and nematodes, Drosha is responsible for the first nuclear cleavage of microRNA precursor hairpins (Lee et al. 2003) while Dicer acts in the cytoplasm on this cleavage product to generate small RNA duplexes (Bernstein et al. 2001; Hutvagner et al. 2001). Despite the presence of a dsRBM in each of these proteins, their action requires the presence of other dsRBM-containing protein acting as cofactors. For instance, the dsRBM of Drosha seems to retain little ability to bind RNA while both dsRBMs of its cofactor, DGCR8/Pasha, are necessary to bind dsRNA while mediating the interaction with Drosha (Denli et al. 2004; Gregory et al. 2004; Han et al. 2004; Han et al. 2006; Yeom et al. 2006). Similarly, Dicer proteins are accompanied by non catalytic dsRBM-containing protein, Loquacious and R2D2 in flies (Saito et al. 2005; Tomari et al. 2004), RDE-4 in worms (Tabara et al. 2002), and TRBP and PACT in human cells (Chendrimada et al. 2005; Lee et al. 2006), that assist the processing and/or strand selection of small RNA duplexes therefore promoting the formation of the RNA induced silencing complex (RISC).

dsRBM-containing proteins are also found in plants, and most of these proteins have been linked to RNA interference or related phenomenon (Kiyota et al. 2011; Manavella et al. 2012). In the model species, *Arabidopsis thaliana*, the most studied are the four DICER-LIKE (AtDCL) proteins, which exhibit a domain organization similar to the animal Dicers, with one (for AtDCL2) or two (for AtDCL1, 3 and 4) C-terminal dsRBMs. AtDCL1 is responsible for the production of 21-nt microRNAs from Pol II precursor transcripts (Kurihara

and Watanabe 2004) as well as for the production of phased cis natural antisense siRNAs, while AtDCL2 cleaves the primary convergent transcript into 24-nt duplex in this pathway (Borsani et al. 2005). AtDCL2 is also implicated in gene silencing induced by exogenous dsRNAs, as is AtDCL4 (Bouche et al. 2006; Dunoyer et al. 2007). AtDCL4 also generates phased trans-acting siRNAs (Gascioli et al. 2005) or transposon-derived 21-nt epigenetically activated small interfering RNAs (easiRNAs) (Creasey et al. 2014; McCue et al. 2012; Sarazin and Voinnet 2014) using dsRNA provided by the action of RNA-dependent RNA polymerase 6 (AtRDR6), and is also responsible for the formation of some microRNAs (Rajagopalan et al. 2006). Finally, AtDCL3 acts with RNA-dependent RNA polymerase 2 (AtRDR2) in the RNA-dependent DNA methylation (RdDM) pathway, to produce 24-nt siRNAs that guide DNA methylation, mostly to repeated sequences and transposable elements, thus participating in genome defense (Xie et al. 2004).

As for animals, some plant DCLs necessitate dsRBM-containing non-catalytic cofactors named DRB (for double stranded RNA binding). *Arabidopsis* possesses five known DRBs (AtDRB1–5) (Hiraguri et al. 2005), each containing two N-terminal dsRBMs. AtDRB1, also known as HYL1, is required for AtDCL1-mediated processing of miRNA precursors (Kurihara et al. 2006). AtDRB1 seems to act as a dimer and to interact with AtDCL1 via its second dsRBM (Yang et al. 2010), while the first dsRBM binds miRNA precursors as well as mature miRNA duplexes (Eamens et al. 2009; Song et al. 2007), assisting in the cleavage and in the miRNA strand selection. AtDCL4 is assisted by AtDRB4 (Nakazawa et al. 2007) and this protein seems to be essential for the in vitro activity of AtDCL4 (Fukudome et al. 2011). Similarly, in case of geminivirus infection, AtDRB3 seems to interact with AtDCL3, impacting the methylation of the viral genome (Raja et al. 2014). Furthermore, AtDRB2, AtDRB3 and AtDRB5 have been implicated in an atypical miRNA biogenesis pathway (Eamens et al. 2012a, b), and AtDRB2 and AtDRB4 affect the quantity of siRNAs acting in RdDM, suggesting a role for these two proteins in this particular pathway (Pélicier et al. 2011). More recent data, however, suggest that AtDRB2 modulates transposon element derived siRNA accumulation mostly through an indirect pathway (Clavel et al. 2015). Finally, AtDRB4 seems to have a role in resistance against pathogens, distinct from its action alongside AtDCL4 (Jakubiec et al. 2012; Zhu et al. 2013).

In this work, we reconstruct the evolutionary history of the DRB family in the Viridiplantae lineage. We discovered the presence of a new DRB family, we name DRB6, absent from Brassicaceae but present in all other vascular plant species tested. DRB1–6 are systematically composed of two different types of dsRBM (type I and II) that have evolved under different constraints and likely fulfill complementary functions. We also uncovered a second new



**Fig. 1** Phylogenetic analysis of double stranded RNA binding motifs (dsRBMs) of *Arabidopsis thaliana* double stranded RNA binding proteins (DRB) and of the Dicer-like 4 protein (DCL4) (see Table S1 for protein sequences and Figure S1 for the alignment). Selected

DRB family (named DRB7) conserved in all vascular plants and harboring a single dsRBM that evolved in a concerted manner with the second dsRBM of DCL4. We present evidence suggesting that DRB7 and DRB4 are both involved in easiRNA biogenesis.

## Results and discussion

As a first approach, we collected the sequence of all *Arabidopsis* proteins containing at least one dsRBM and no other conserved domains (here after defined as DRBs).

informative statistical supports [approximate likelihood-ratio test (aLRT) data] are indicated. A schematic representation of the position of the different conserved domains in each DRB family is presented (see also Fig. 3 and S3)

Based on these criteria, in addition to the already described five AtDRBs (AtDRB1–5), two other proteins encoded by the At1g80650 and At4g00420 loci and presenting a single dsRBM, were identified. The dsRBM from these seven proteins were aligned (see Supplemental Figure S1) and used to construct a phylogenetic tree (Fig. 1). This initial analysis suggests that the AtDRBs1–5 are systematically composed of two types of dsRBM (type I and II) that form a distinct clade from the type III dsRBM present on the At1g80650 and At4g00420 proteins. This type III motif is surprisingly closely related to the second dsRBM present in the C-terminal region of the AtDCL4 enzyme (Fig. 1). This

second dsRBM of AtDCL4 (AtDCL4.2 in Fig. 1) is the only case where a dsRBM present in a DCL protein is closely related to a dsRBM present in a DRB protein (not shown).

To test if the situation in *Arabidopsis* is representative of species in the Viridiplantae lineage, we looked for the presence of DRBs in the completely sequenced genomes of five Chlorophyte (green algae), a bryophyte, a lycophyte, four monocotyledon and twelve dicotyledon species representing the diversity of Viridiplantae (see Materials and methods for a list of species used and supplementary Table 1 for the sequence of all DRBs used in this analysis). DRBs containing at least two dsRBMs are present in all plant species while DRBs containing a single dsRBM are present in all Angiosperm but absent from the bryophyte and lycophyte species tested. No DRB could be detected in the five Chlorophyte genomes tested strongly suggesting that they are completely absent from this monophyletic phylum. This is surprising since DRBs are largely represented in animal species (Fierro-Monti and Mathews 2000; Tian et al. 2004) and a complex system of small RNAs exists in at least some Chlorophytes (Zhao et al. 2007). Based on these observations, the likely hypothesis is that RNAi machinery and DRBs were present in the common ancestor of the Viridiplantae but that DRBs were lost specifically in the Chlorophyte phylum and that the overall small RNA biogenesis process was either conserved (for *Chlamydomonas reinhardtii* and *Volvox carteri* Zhao et al. 2007) or lost (for the microalgae *Ostreococcus lucimarinus* Cerutti et al. 2011) depending on the specific Chlorophyte lineage. Based on this, one may speculate that the Chlorophyte DICER machinery evolved to generate small RNAs without the need of DRB cofactors or that another RNA binding protein fulfill this role.

### Evolutionary history of multiple dsRBM DRBs

We first decided to study the evolutionary history of DRBs containing multiple dsRBMs. To do so we fused and aligned dsRBM sequences from our eighteen selected species (see above) and used this alignment (see Supplemental Figure S2) to build a phylogenetic tree (Fig. 2). We observed from this analysis that plant DRBs can be grouped in four major clades that we named according to the *A. thaliana* DRB families.

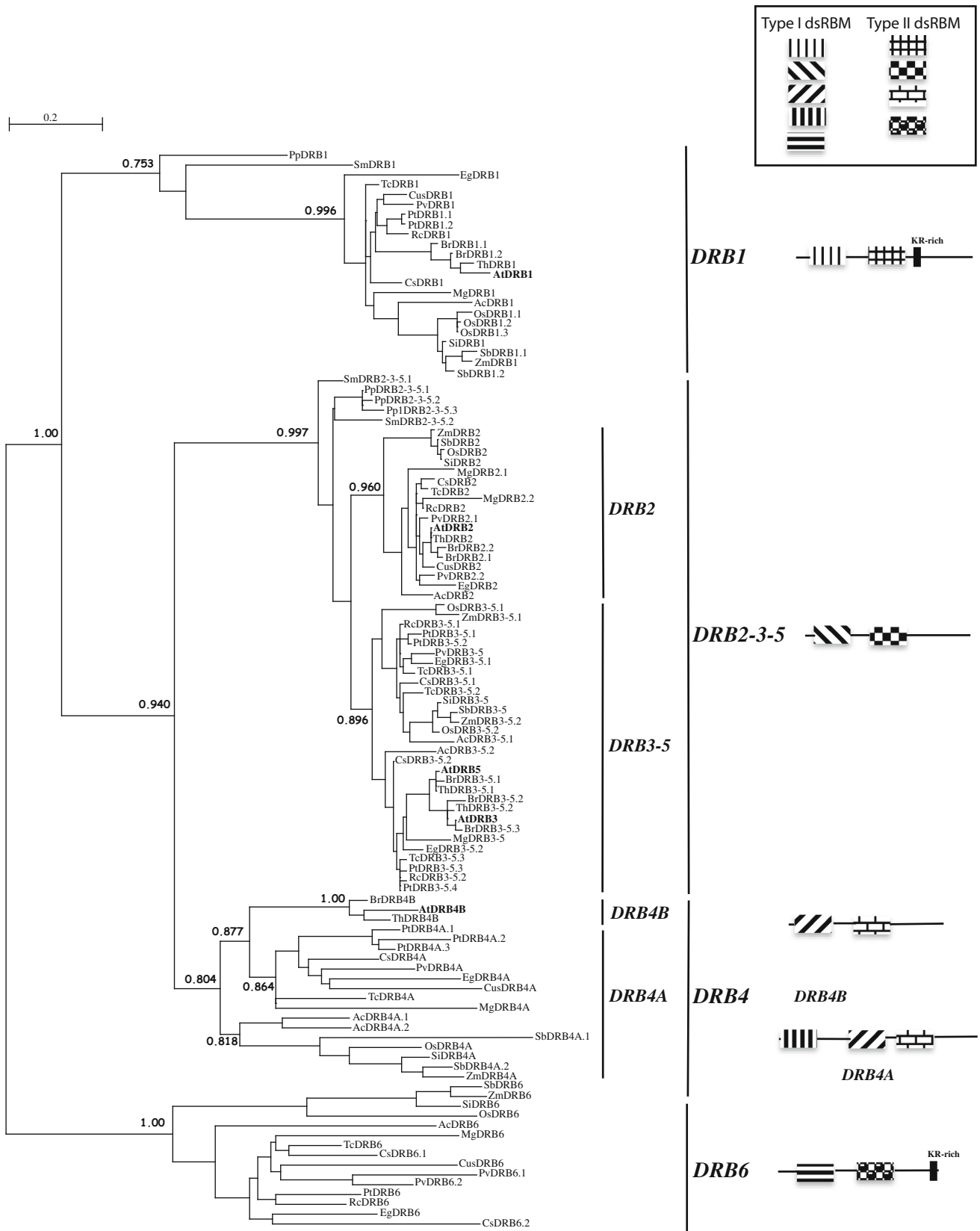
One clade represents the DRB1 family and members from this family are found in all species tested, in one or several versions. DRB1 proteins possess two dsRBMs separated by a linker of around 20 amino acids (Figure S3). The position of the first dsRBM related to the N-terminal amino acid is also constrained and 15 amino acids precede the first dsRBM in bryophyte, lycophyte and dicotyledon species while this region is of 35 amino acids for

**Fig. 2** Phylogenetic analysis of the DRB proteins having multiple dsRBMs. The phylogenetic tree was obtained by fusing and aligning the dsRBMs of 110 DRB proteins from 18 plant species (see Table S1 for protein sequences and Figure S2 for the alignment). Selected informative statistical supports [approximate likelihood-ratio test (aLRT) data] are indicated. Names of the different DRB families are indicated with a schematic representation of the position of the different conserved domains in each family (see also Figure S3). The classification as type I or type II dsRBM was obtained by analyzing the individual motifs (see Fig. 3). Species codes are the following: Pp: *Physcomitrella patens*, Sm: *Selaginella moellendorffii*, Sb: *Sorghum bicolor*, Zm: *Zea mays*, Si: *Setaria italic*, Os: *Oryza sativa*, Ac: *Aquilegia coerulea*, Mg: *Mimulus guttatus*, Eg: *Eucalyptus grandis*, Cs: *Citrus sinensis*, Tc: *Theobroma cacao*, Th: *Thellungiella halophila*, Br: *Brassica rapa*, At: *Arabidopsis thaliana*, Cus: *Cucumis sativus*, Pv: *Phaseolus vulgaris*, Pt: *Populus trichocarpa*, Rc: *Ricinus communis*

monocotyledon species. These results suggest that the position of the two dsRBMs in DRB1 proteins is important for their function. DRB1s also have, closely after the second dsRBM (<40 amino acids), a well-conserved arginine and lysine-rich region (KR-rich) (see Supplemental Figure S4).

A second, more complex, clade is composed of members of the DRB2–3–5 families. All members of this clade possess two dsRBMs precisely separated by a linker of 19 amino acids (Figure S3). The two dsRBMs are located at the very N-terminal region of the protein as only a single amino acid is present before the first dsRBM (Figure S3). The very high positional conservation of the dsRBMs strongly suggests that this is a critical factor for function. *Physcomitrella patens* and *Selaginella moellendorffii* DRBs are found as outgroup of this clade while vascular plants DRBs can be further divided in two subclades, one of which corresponds to the DRB2 family. The *A. thaliana* DRB3 and DRB5 proteins result from a recent segmental duplication shared by all three Brassicaceae species tested (Blanc et al. 2003) and our phylogenetic analysis supports the fact that they belong to the same family. Therefore, for simplicity and to conserve as much as possible the previous nomenclature, we suggest naming the second subclade, the DRB3–5 family. All vascular plants possess at least one DRB2 and one DRB3–5 protein, but some species have several versions of each.

The third clade is composed of members of the DRB4 family. This family is present in all vascular plants but is absent from the two bryophyte and lycophyte genomes tested. In all vascular plants, DRB4 proteins are composed of three dsRBMs (named the DRB4A subfamily), except for Brassicaceae species (including *Arabidopsis*), where DRB4 proteins are composed of only two dsRBMs (named the DRB4B subfamily). A likely scenario to explain this situation is that the coding capacity for the first dsRBM of DRB4 was lost in the common ancestor of





Brassicaceae species. A direct consequence of this observation is that the many functional analyses conducted using the *A. thaliana* DRB4 (Fukudome et al. 2011; Nakazawa et al. 2007) may not be fully representative of the situation in most vascular plants for this family. Here again the position of the three dsRBMs of DRB4A proteins is highly conserved with the exception of the linker between the first and second dsRBM of dicot DRB4A that is much more variable in length (from 35 to 299 amino acids, see Figure S3). The position of the two dsRBMs of Brassicaceae DRB4B is also well conserved (Figure S3).

The fourth clade is composed of DRB proteins, with two dsRBMs, present in all vascular plants but absent in bryophyte, lycophyte and from Brassicaceae species. We named this family DRB6. Here again a likely scenario to explain this situation is that the gene coding for DRB6 emerged in the common ancestor of vascular plants but was lost much later in the common ancestor of Brassicaceae species. DRB6 proteins show two structural similarities with DRB1 proteins: a more internal positioning of the two dsRBMs compared to DRB2, DRB3–5 and DRB4 and the presence of a well-conserved arginine and lysine-rich region (KR-rich) localized after the second dsRBM (see Figure S3 and S4). Since DRB6 is absent from *A. thaliana*, nothing is known on the function of this protein at the moment. Therefore, one important conclusion from this analysis is that DRB families in *Arabidopsis* are quite atypical compared to most plant species and that conclusions reached using this species may not always be applicable to all plant lineages.

### Evolution of dsRBMs present in multiple dsRBM DRBs

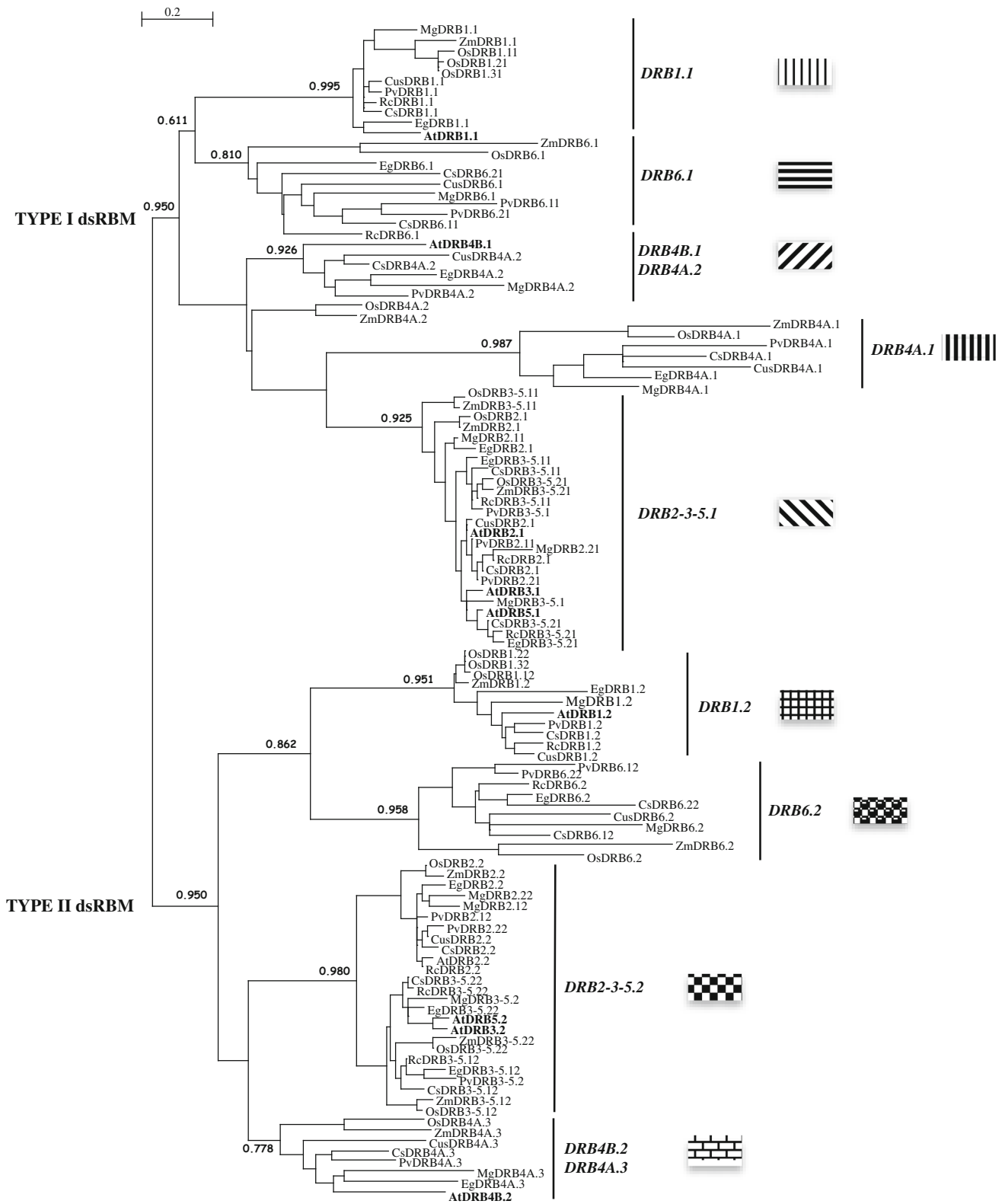
We next decided to address questions concerning the origin and evolution of dsRBMs present in multiple dsRBM DRBs. One of our objectives is to determine if these dsRBMs can be classified in two distinct types as suggested by our initial analysis using *A. thaliana* proteins (Fig. 1) and if one dsRBM of each type is always associated with DRBs of the different families. We therefore, this time, collected and aligned the different dsRBMs individually, without taking into account their protein of origin and build a phylogenetic tree using this alignment (see Supplemental Figure S5 and Fig. 3). We observed that the dsRBMs indeed form two well-supported clades. The first clade (type I dsRBM) contains all dsRBMs present in the first (more N-terminal) region of DRB1, DRB2–3–5, DRB4B and DRB6 proteins and in the first and second position of the DRB4A proteins. The second clade (type II dsRBM) contains all dsRBMs present in the second (more C-terminal) region of DRB1, DRB2–3–5, DRB4B and DRB6 proteins and in the third position of the

DRB4A proteins. This result suggests that two different types of dsRBM evolving under different selection pressure are always associated with the different multiple dsRBM DRBs and that a DRB with type I and II dsRBMs was already present in the plant common ancestor. Based on this result, it is likely that the RNA-binding properties of these DRBs depend on the combination of the two non-redundant types of dsRBMs that form the true RNA-binding module. In support to this conclusion, several *in vitro* studies clearly demonstrated the different RNA-binding properties of the two individual dsRBMs of plant and animal DRBs and their cooperative nature to bind efficiently RNA *in vivo* (Han et al. 2006; Micklem et al. 2000; Qin et al. 2010; Yang et al. 2010).

When analyzed separately, the DRB1 and DRB6 type I dsRBMs group in the same clade and the same is true for the type II dsRBM of these proteins (Fig. 3). This result is in agreement with the structural similarities observed previously among DRB1 and DRB6 proteins (Fig. 2 and S3) and support the notion that these two families, although clearly distinct, are closely related.

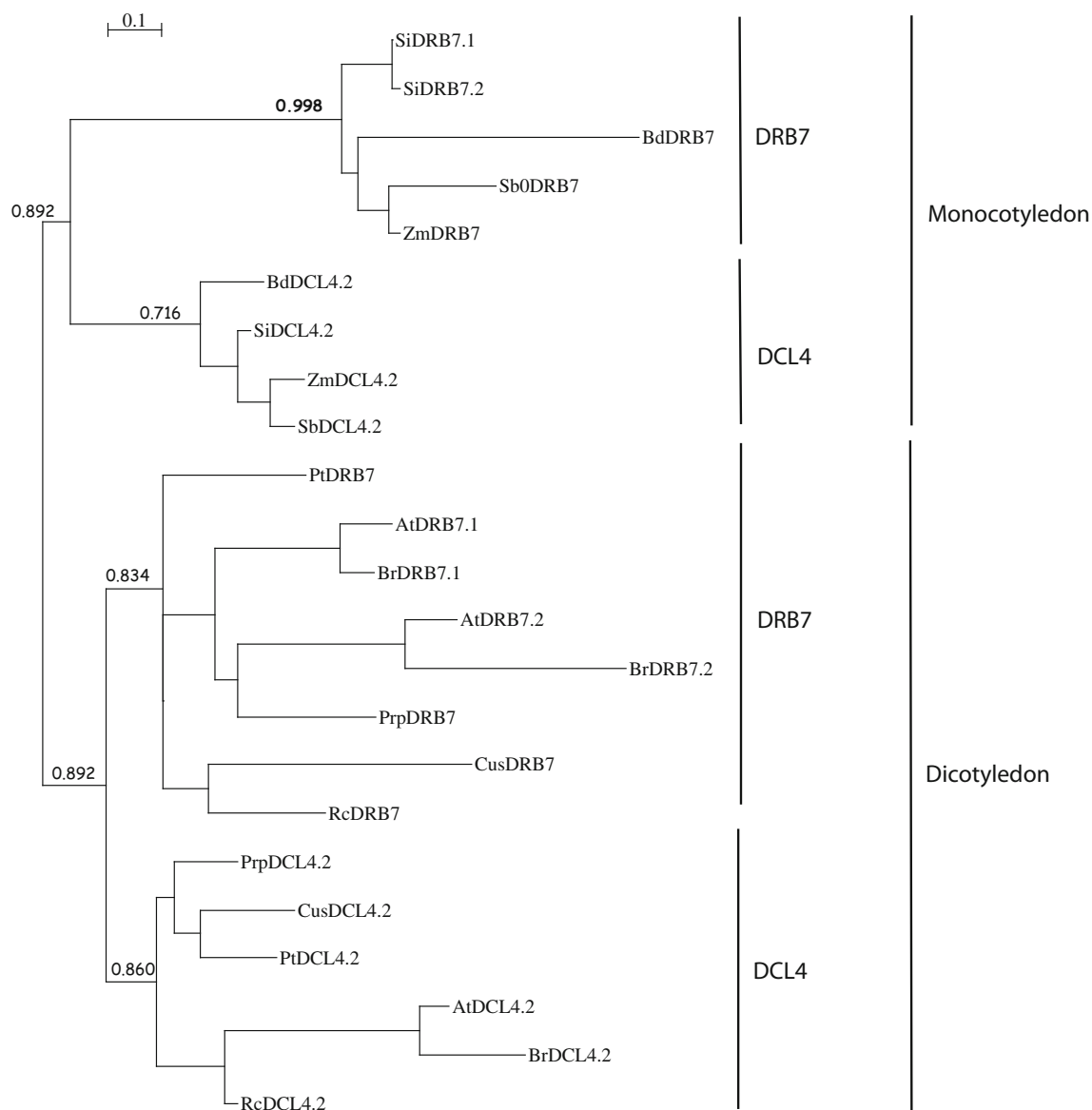
### Evolutionary history of single dsRBM DRBs

We next searched our collection of DRBs for proteins having a single dsRBM. This search was complicated by the fact that some of these proteins may merely be non-functional truncated versions of multiple dsRBM DRBs and/or result of incorrect *in silico* definition of genetic organization. Therefore, to circumvent these problems, we only retained single dsRBM DRBs that were conserved at least in two different plant species. Using this criterion, a single family of mono dsRBM DRBs was identified that we named DRB7 (Fig. 4). This family includes proteins coded by the At1g80650 and At4g00420 (renamed AtDRB7.1 and AtDRB7.2) that possess a dsRBM closely related to the second (closest to the C-terminal end) dsRBM of DCL4 (Fig. 1). DRB7 proteins are found in all vascular plants but are absent from the bryophyte and lycophyte species tested. To precise the evolutionary link between DRB7 and DCL4, we collected and aligned their dsRBMs (see Supplemental Figure S6) and used this alignment to build a phylogenetic tree (Fig. 4, see also Supplementary Table 1 for a list of DRB7 and DCL4 sequences used). This analysis strongly suggests that the single dsRBM of DRB7 and the second dsRBM of DCL4 evolved in a concerted manner. This can be inferred from the fact that all plant DRB7 sequences are not found in a single clade (that would be distinct from a second DCL4 clade) but rather that DRB7 and DCL4 sequences from monocotyledon and DRB7 and DCL4 sequences from dicotyledon species form two distinct clades (Fig. 4). Therefore, it is likely that both dsRBMs have a common origin and that both proteins were present



**Fig. 3** Phylogenetic analysis of individual dsRBMs present in multiple dsRBM DRBs. Selected informative statistical supports [approximate likelihood-ratio test (aLRT) data] are indicated (see Table S1 for protein sequences and Figure S5 for the alignment). Names of the different DRBs are indicated in which the last number

after the dot (1, 2 or 3) denotes the position of the dsRBM in the protein (1 being the most N-terminal). The two different clades define the two types of dsRBM (I or II). The color code defined here was used in the schematic presentation of the DRB families in Fig. 2. Species codes are as in Fig. 2



**Fig. 4** Phylogenetic analysis of individual dsRBMs present in the single dsRBM DRB7 family and in the most C-terminal region of the DCL4 family. Selected informative statistical supports [approximate

in the common ancestor of vascular plants but also that DRB7 and DCL4 sequences evolved from the common ancestor in a concerted manner in the mono- and dicotyledon lineages. Although the DRB7 function is for the moment unknown, our results can best be explained if DRB7 proteins cooperates with DCL4 in a yet to define pathway.

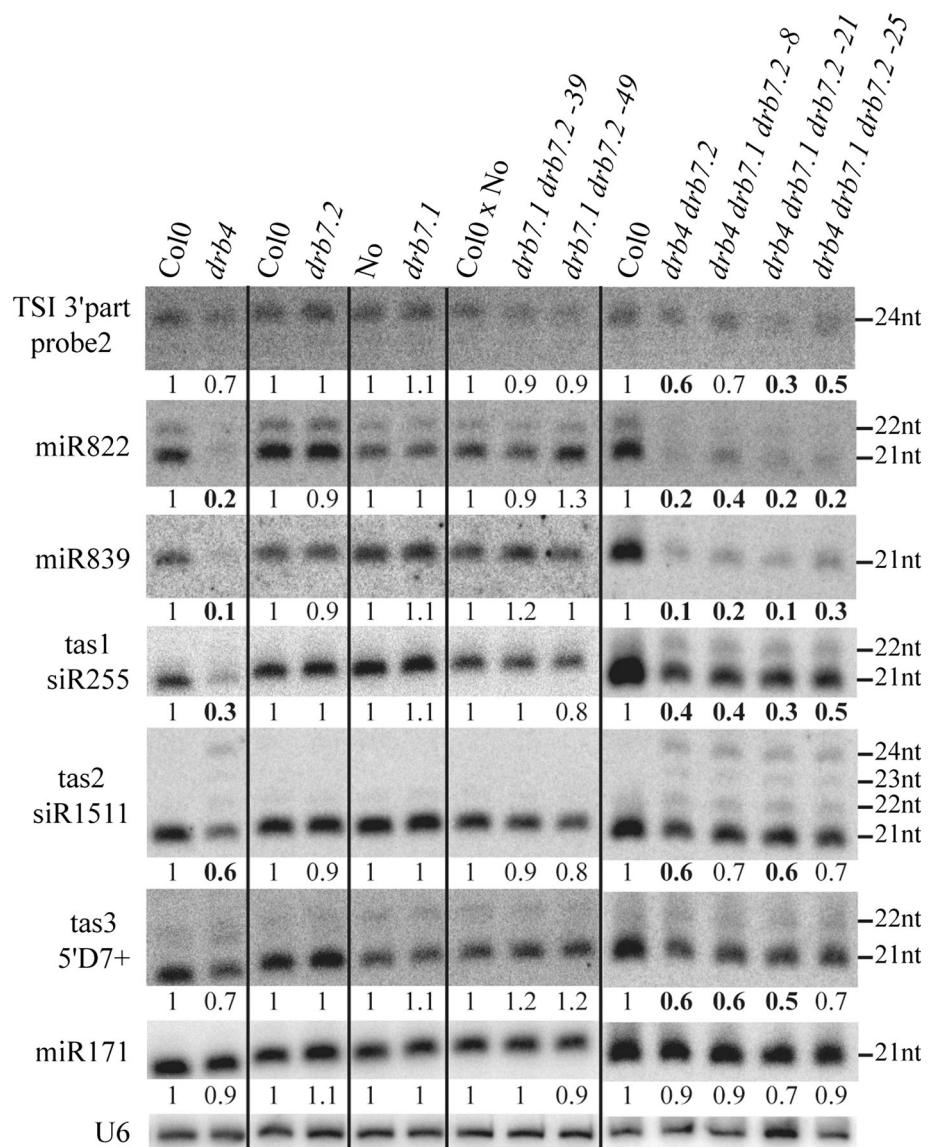
#### Impact of AtDRB7 loss of function on known AtDCL4 *in vivo* substrates

AtDRB4 is a well-characterized cofactor of AtDCL4 and is important for the biogenesis of tasiRNAs and some miRNAs (Adenot et al. 2006; Rajagopalan et al. 2006). To

likelihood-ratio test (aLRT) data] are indicated (see Table S1 for protein sequences and Figure S6 for the alignment). Bd: *Brachypodium distachyon*

determine if one or both of the two AtDRB7 proteins could also participate as AtDCL4 cofactor in these two pathways, we selected *Arabidopsis* mutants defective in the production of AtDRB7.1, AtDRB7.2 or both proteins. Figure 5 shows that, as expected, several AtDRB4/AtDCL4-dependent small RNAs (tasiRNAs, miRNA822, miRNA839) underaccumulate in *drb4* but are present at a level similar to wild-type in the single (*drb7.1* or *drb7.2*) or double (*drb7.1/drb7.2*) mutant lines. Similarly, polIV-dependent siRNA accumulation, that was previously shown to be decreased in *drb4* and *dcl4* mutants (Pélessier et al. 2011), is not affected in *drb7* mutants (Fig. 5 and data not shown). The introduction of *drb7* mutations into the *drb4* background does not enhance the *drb4* molecular phenotype

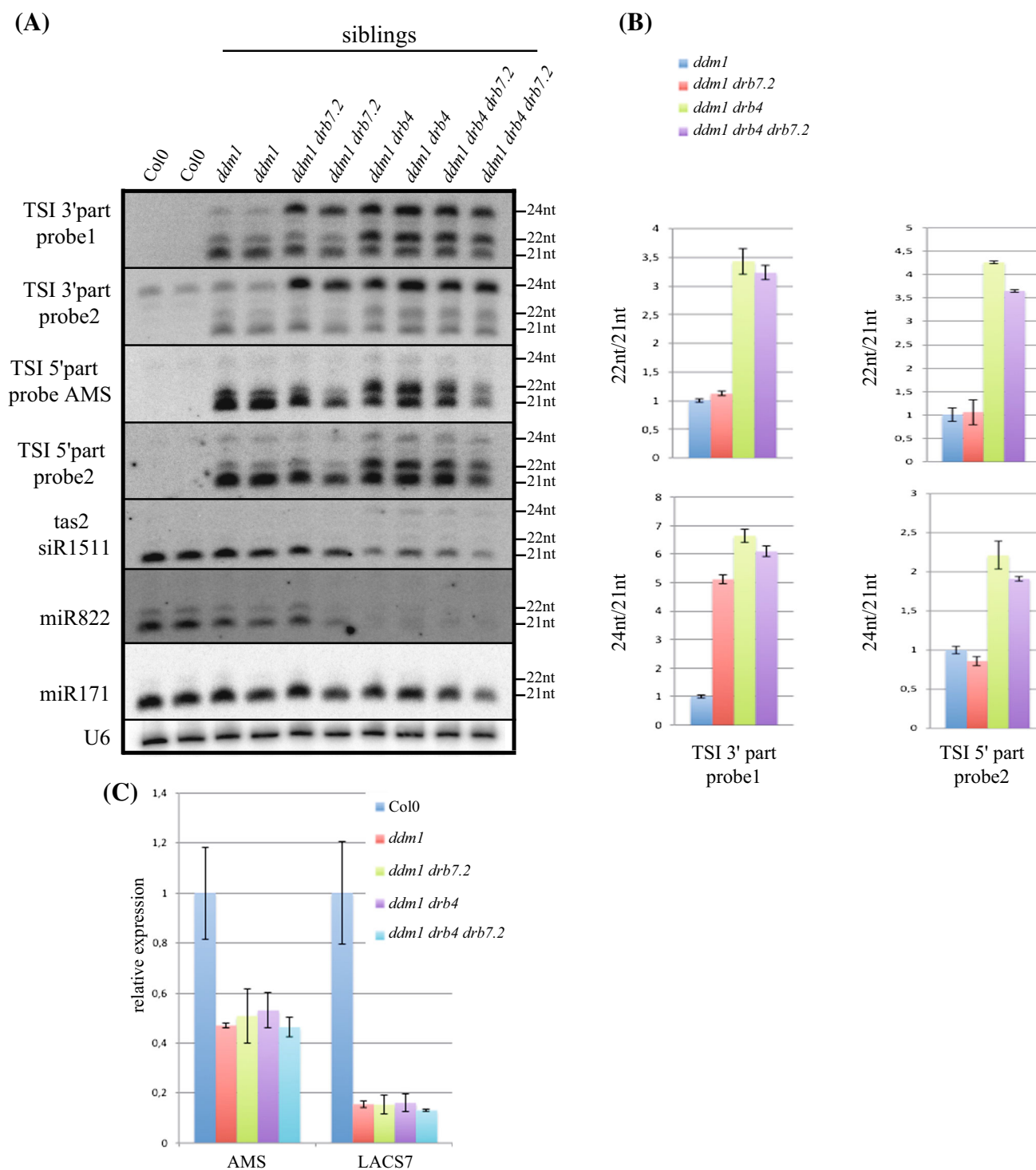
**Fig. 5** Impact of AtDRB7 loss of function on known AtDCL4 *in vivo* substrates. DRB4, but not DRB7 proteins, is required for proper accumulation of polIV-dependent siRNAs (TSI), tasiRNAs and DCL4-dependent miRNAs (miR822 and 839). Total RNA extracts from immature flowers were blotted and hybridized sequentially with the indicated probes. While DCL1-dependent miRNA171 is unaffected in all mutant combinations, significant change in small RNA level/pattern is observed in the *drb4* mutant background only; no additional impact is visible when *dr7.1* and *drb7.2* mutations are combined with *drb4*. Because Col0 was used as control following introgression of the *drb7.1* Nössen (No) mutant allele into *drb4 drb7.2* line, three individual plants of the progeny were analyzed to minimize potential effects of mixed genetic backgrounds. Values for 21-nt RNA accumulation (or 24-nt for TSI 3' part probe 2) are normalized to U6 RNA and are expressed as a ratio relative to the wild-type control plants. Variation of more than 1.5-fold are highlighted in bold



(Fig. 5) further suggesting that if AtDRB7 is indeed a cofactor of AtDCL4, it is not involved in these pathways.

Recently, AtDCL4 has been implicated in transcription termination of an endogenous *Arabidopsis* gene named FCA (Liu et al. 2012). To determine if *Arabidopsis* AtDRB7 proteins could be involved as a AtDCL4 cofactor in this process, we performed QRT-PCR to monitor the levels of FCA read-through products in single and double *drb7* mutant lines compared to two *dcl4* null mutant lines (*dcl4-2* and *dcl4-5*). However, in contrast to published results (Liu et al. 2012), we were not able to observe a significant increase of FCA read-through products in the two *dcl4* mutant lines nor in single and double *drb7* mutant lines (Figure S7). We were therefore unable to document a role for AtDRB7 in these putative AtDCL4-dependent transcription termination events.

Recently, a new class of epigenetically activated siRNAs from transposable elements (easiRNAs), whose production depends on AtDCL4, AtDCL2 and AtRDR6, has been described (Creasey et al. 2014; McCue et al. 2012; Sarazin and Voinnet 2014). These 21–22 nucleotides easiRNAs are abundant in the DECREASED DNA METHYLATION 1 (*dcl4-1*) mutants (Creasey et al. 2014; McCue et al. 2013, 2014; Nuthikattu et al. 2013), as well as in the vegetative nucleus of pollen grains and dedifferentiated plant cell cultures (Slotkin et al. 2009; Tanurdzic et al. 2008). We therefore tested if one or both AtDRB7 proteins could be involved in the biogenesis of easiRNAs. By far, the major contributor of easiRNAs is represented by the *Athila* family of retrotransposons. The internal domain downstream of the *Athila* gag-pol-encoding ORF, also known as Transcriptionally Silent Information (TSI), is



particularly prone to produce 21–22 nt easiRNAs (McCue et al. 2014; Slotkin 2010). While Athila full-length transcripts initiate within the 5′LTR region, specific TSI transcription originates from internal sequences and was associated with three different types of transcripts (Steimer et al. 2000). Using two different probes located in the TSI 3′ part (within the TSI-A region as defined in Steimer et al.

2000), we observed that 21-nt easiRNAs were not significantly affected by the loss of AtDRB7.2 or AtDRB4 (Fig. 6a, b). However, a strong increase in AtDCL3-dependent 24 nucleotides siRNAs was noted in *ddm1 drb7.2* and *ddm1 drb4* double mutants, and in the *ddm1 drb4 drb7.2* triple mutant compared to the *ddm1* mutant (Fig. 6a, b). Such a situation is analogous to what was observed

**Fig. 6** Impact of AtDRB7.2 loss of function on the easiRNA pathway and on targets of the easiRNA pathway. **a** To properly compare easiRNA accumulation patterns, total RNAs of immature floral inflorescences were extracted from Col0 or siblings segregating from a single *ddm-/- drb4+/- drb7.2+/-* parental plant. Blotted RNAs were probed with the indicated probes. AtDRB4 and AtDRB7.2 both impact the processing of easiRNA precursors. As expected AtDRB4 also impact the production of tasiRNA (*tas2*) and a AtDCL4-AtDRB4 dependent miRNA (*mir822*) but none of the two proteins impact a AtDCL1-AtDRB1 dependent miRNA (*mir171*). For *tas2*, 21 nt AtDCL4-dependent tasiRNAs are reduced in *drb4* and low levels of 22–24 nt species accumulate due to compensatory functions of AtDCL2/AtDCL3, as expected when AtDCL4 activity is reduced on *tas2* precursor (Gascioli et al. 2005). **b**. Accumulation of 22- and 24-nt siRNAs relative to the 21-nt level. Hybridization results from only one of the two TSI 5' or 3' probes are shown as both probes in each case, gave similar results. Values are means derived from two individual siblings (A.)  $\pm$ SEM and are given relative to *ddm1* set arbitrarily to 1. **c**. qRT-PCR analyses of AMS and LACS7 expression in different genetic backgrounds. Targeting of AMS and LACS7 mRNAs by AGO1-loaded 21 nt-siRNAs in *ddm1* plants is not affected by the additional *drb4* and *drb7.2* mutations. Values normalized to *actin2* are means derived from three individual siblings (only two for triple mutants)  $\pm$ SEM and are represented relative to Col0 set arbitrarily to 1

when the easiRNA pathway is perturbed in a *ddm1 rdr6* double mutant and could reflect a switch from AtRDR6/AtDCL4 towards AtRDR2/AtDCL3 processing of the templates (Creasey et al. 2014; Sarazin and Voinnet 2014). Interestingly, a substantial overaccumulation of AtDCL2-dependent 22-nt easiRNAs in the *drb4*, but not *drb7.2*, mutated plants was also observed suggesting that AtDCL2 loading on AtRDR6-derived templates becomes facilitated only when the AtDRB4 cofactor is lost. Monitoring the accumulation of easiRNAs targeting the 5' part of TSI, we observed a similar situation for 21- and 22-nt but not 24-nt easiRNAs (Fig. 6a, b). Indeed, no massive shift to 24-nt siRNAs was observed; only a modest increase of 24-nt easiRNAs occurs in *ddm1 drb4* and *ddm1 drb4 drb7.2* mutants compared to *ddm1* (Fig. 6a, b). To explain this discrepancy, it should be noted that TSI transcription is rather complex as three main TSI transcripts accumulate, two polyA<sup>+</sup> RNAs of 5000 and 2500 nucleotides and a polyA<sup>-</sup> RNA of 1250 nucleotides. While these transcripts overlap within the TSI 3' region, only the 5000-nt RNAs can be detected with a 5' specific-probe (Steimer et al. 2000). Additionally, the Athila2 and 6A subfamilies are the main suppliers of easiRNAs (Nuthikattu et al. 2013), and the TSI 5' part probes can only detect Athila6A members-derived easiRNAs while the 3' probes can detect a wider range of Athila2 and 6A subfamilies loci-derived easiRNAs. These differences in targeting specificity for the 5' and 3' probes are likely to be at the origin of the differences in easiRNAs populations we observed. Elucidating the complex mechanisms involved in easiRNA biogenesis is a challenging topic (Sarazin and Voinnet 2014) and would certainly benefit from the identification reported here of

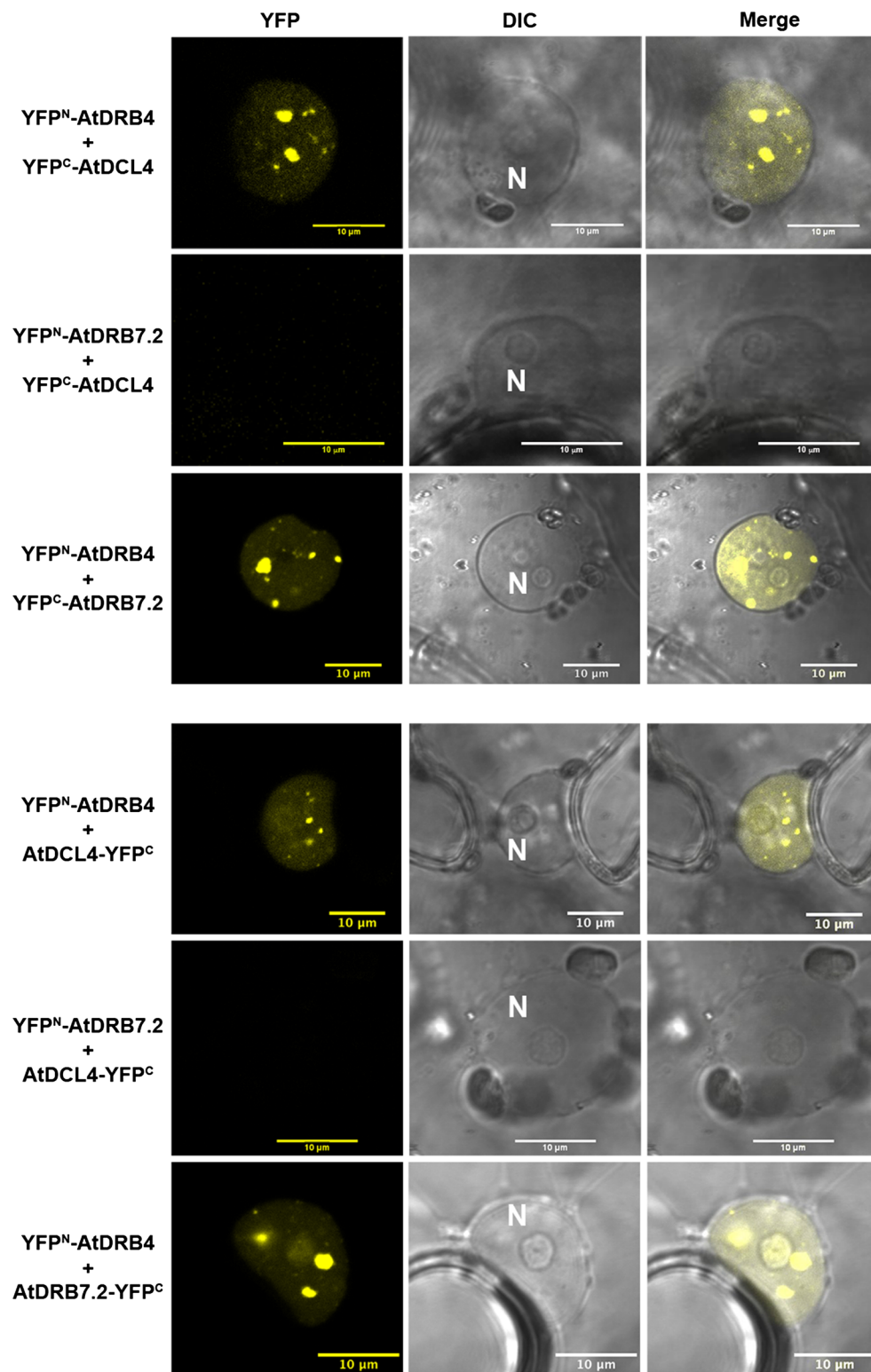
new cofactors impacting this pathway, such as the AtDRB7.2 and AtDRB4 proteins.

TSI-specific easiRNAs are preferentially incorporated into AGO1 and regulate target genes displaying complementary sequences (McCue et al. 2013, 2014). This is the case for the Athila6A derived AMS-specific easiRNAs (Fig. 6a) that strongly downregulate the AMS gene in *ddm1* plants (McCue et al. 2013). Figure 6c confirms such impact on AMS mRNA levels, and also shows that this downregulation was not changed in our double or triple mutant plants. A similar result was obtained for another known easiRNA target, the LACS7 transcript (McCue et al. 2013) indicating that modification of small RNAs pattern in *ddm1 drb4*, *ddm1 drb7.2* or *ddm1 drb4 drb7.2* mutants does not impact targets of easiRNAs.

Using the same probes, no variation in small RNA populations was observed when using *drb7.1* instead of *drb7.2* background (Figure S8), suggesting that in contrast to AtDRB7.2, AtDRB7.1 is not involved in easiRNA biogenesis but possibly in another yet to define DCL4-dependent pathway.

#### AtDRB7.2 is interacting with AtDRB4 but not with AtDCL4

To test if AtDRB7.2 could interact directly with AtDCL4, a situation that could explain the concerted evolution signature of Fig. 4, we used a BiFC strategy. Using as positive control the expected AtDRB4-AtDCL4 interaction, we observed that AtDRB7.2 interacts strongly with AtDRB4 but not directly with AtDCL4 (Fig. 7, for additional positive and negative controls of the BiFC experiment see Figure S9 and S10). This result, combined to our inability to pull-down AtDCL4 using a tagged version of AtDRB7.2 (data not shown), suggest that AtDRB7.2 exists mainly, in vivo, in complex with AtDRB4 and not with AtDCL4. Therefore, since the concerted evolution signature is apparently not based on a direct contact between the two proteins, it is tempting to propose that AtDRB7.2 and AtDCL4 cooperate to bind a common ligand. However, our genetic results (see Figs. 5, 6) suggest that none of the tested AtDCL4 end-products (including 21-nt easiRNAs) are affected by the loss of AtDRB7.2. Surprisingly, the absence of AtDRB4 does not modify the accumulation of 21-nt easiRNAs either (Fig. 6), suggesting that AtDCL4 is able to generate 21-nt easiRNAs without involving those DRBs and that the concerted evolution signature we observed in Fig. 4 cannot be explained by the cooperative binding of AtDCL4 and AtDRB7.2 to easiRNA precursor molecules. Therefore, the molecular origin of this signature is yet to be determined and could result from the cooperative or competitive binding of both proteins to a common RNA ligand yet to identify or occurring in specific and/or discrete cell types.



**Fig. 7** BiFC visualization of AtDRB7.2 interaction with AtDRB4. Shown are confocal images of *Nicotiana benthamiana* epidermal cells coinfiltrated with *Agrobacterium* cultures harboring different combination of expression vectors (indicated on the left of the panels). *Scale bars* and the position of nucleus are shown. Positive control for the BiFC experiment corresponds to coexpression of AtDRB4 and AtDCL4, fused to the indicated YFP moieties (one example of two different combinations is shown). As previously reported (Hiraguri et al. 2005; Zhu et al. 2013), AtDRB4 was found to be associated with AtDCL4 in the nucleus. YFP complementation was also observed when AtDRB7.2 and AtDRB4 were coexpressed (one example of two different combinations is shown) and showed similar nuclear localization as the one observed with AtDRB4 and AtDCL4. By contrast, none of the combinations between AtDRB7.2 and AtDCL4, nor any of the negative controls involving GST fusion as one of the partner (not shown), gives a visible fluorescent signal. For additional positive and negative controls of the BiFC experiment see Figure S9 and S10

## Conclusion

By retracing the evolutionary history of DRB proteins in plants we were able to identify two new families we named DRB6 and DRB7. The earlier discovery of the DRB6 family was impaired by its absence in the model species *Arabidopsis*. DRB6 share structural similarities with DRB1 and it is tempting to speculate that it could also be involved in the production of miRNAs, except in Brassicaceae where, other DRBs (such as AtDRB2, AtDRB3 and AtDRB5; Eamens et al. 2012a, b) could play its part.

Determining precisely the role of DRB7 proteins will require more work. We show here that one of the two *Arabidopsis* DRB7 proteins (AtDRB7.2) is involved, in combination with AtDRB4, in the easiRNA pathway, by inhibiting the accumulation of DCL3-dependent 24 nt siRNAs from easiRNA-generating loci. However, the increased accumulation of those 24 nt siRNAs in *drb4* or *drb7* mutant background has no impact on the regulation of easiRNA target accumulation. How AtDRB7 and AtDRB4, in association or not with DCL4, negatively regulate the accumulation of those 24 nt siRNAs remains to be determined. One possibility is that AtDRB4/AtDRB7 may be inhibitors of DCL3 activity. This inhibitory effect can be mediated through a direct interaction with DCL3 that will either compete for binding of DCL3 cofactors or directly inhibit its processing on particular substrate. Alternatively, AtDRB4/AtDRB7 may directly sequester specific dsRNA precursors, thereby preventing their access and processing by DCL3. Finally, given that this inhibitory effect is only observed in a *ddm1* mutant background, it is also possible that this mutation triggers a change in the homeostasis of either DCL3, its cofactor(s), or of factor(s) required for the production of these particular polIV-dependent dsRNAs.

This change may, in turn, makes DCL3-mediated processing more susceptible to the presence or absence of AtDRB4/AtDRB7. This is particularly striking in the *drb4* mutant background where opposite molecular phenotypes are observed for 24 nt TSI 3' part-derived siRNA in presence or absence of DDM1 (Figs. 5, 6). Although the reason for this difference is currently unknown, one interesting factor to test would be DRB2 as increased accumulation of several polIV-dependent siRNA were previously observed in *drb2* mutant background (Clavel et al. 2015; Pélissier et al. 2011). Although closely related to AtDRB7.2, AtDRB7.1 does not seem to be involved in any small RNA pathway tested so far and its function will also need to be further investigated in the future.

## Methods

### Sequence selection, multiple sequences alignments and phylogenetic reconstruction

Twenty-three species representing the diversity of the Viridiplantae lineage were searched for the presence of DRBs. At the root of the lineage, we have selected five chlorophytes (*C. reinhardtii*, *V. carteri*, *Coccomyxa subellipsoidea*, *Micromonas pusilla* and *O. lucimarinus*), one bryophyte (*P. patens*), and one lycophyte (*S. moellendorffii*). The four monocotyledons species we have selected (*Sorghum bicolor*, *Zea mays*, *Setaria italica* and *Oryza sativa*) are representative of the three major subclades of this lineage. For the analysis of the DRB7 family, we used the *Brachypodium distachyon* instead of the *O. sativa* protein due to a sequence quality problem for this gene in *O. sativa*. To represent the dicotyledons, we used twelve species, one at the root of this lineage (*Aquilegia coerulea*), one at the root of the Pentapetalae (*Mimulus guttatus*) an outgroup of the Malvaceae (*Eucalyptus grandis*), seven Malvaceae species representing the four major subclades of this group (*Citrus sinensis*, *Theobroma cacao*, *Thellungiella halophila*, *Brassica rapa*, *A. thaliana*, *Populus trichocarpa* and *Ricinus communis*) and two Fabaceae species representing the two subclades of this group (*Cucumis sativus*, *Phaseolus vulgaris*). Blast searches were performed starting from known *Arabidopsis* dsRBMs. However, to make sure we did not miss any DRBs, each time a new dsRBM was found in a given species it was itself used as a probe in a new BLAST search. By using this iterated BLAST procedure we are very confident that we did not miss any dsRBM containing



proteins and that we have collected all DRBs in species we have studied.

Sequences were aligned using the multiple sequence comparison by log-expectation (MUSCLE v3.7) software (Edgar 2004). Alignment shading was obtained using BOXSHADE v3.21. Trees were reconstructed using the fast maximum likelihood tree estimation program PHYML (Guindon and Gascuel 2003) using the LG amino acids replacement matrix (Le and Gascuel 2008). Statistical support for the major clusters were obtained using the approximate likelihood-ratio test (aLRT) (Anisimova and Gascuel 2006).

### Plant material and plant growth conditions

The *drb4-1* and *ddm1-2* mutant lines are in Col0 background and were described previously (Curtin et al. 2008; Jeddeloh et al. 1999). The *drb7.2* mutant corresponds to a GABI-Kat line (GK-525B11; background Col0) bearing a tDNA insertion into the 3rd intron of the At4g00420 gene. The *drb7.1* mutant in the Nössen (No) genetic background was obtained from the “RIKEN *Arabidopsis* transposon mutants” resources (PST 12013) and contains a Ds insertion into the 2nd exon of the At1g80650 gene. Absence of DRB7.1 and 7.2 transcripts was ascertained by semi-qRT-PCR using exon-specific primers surrounding the integration sites. Double *drb7.1 drb7.2* mutant lines were obtained by crossing and a Col x No crossed line was used as control. Plant seeds were stratified for 2–4 days at 4 °C before growth in chambers on soil at 23 °C under a 16 h-light/8 h-dark cycle.

### RNA isolation and hybridization

Total RNA was extracted as described in (Pélissier et al. 2004), using immature floral inflorescences (stages 1–12) or aerial parts from 2 weeks-old seedlings. For the detection of small RNAs, 15 µg of total RNA samples were heat-treated in 1.5 volume of standard formamide buffer and loaded on a 15 % polyacrylamide (19 :1 acrylamide :bis-acrylamide)—8.3 M urea—0.5X TBE gel and separated by electrophoresis. The samples were electroblotted to hybond-NX membranes (GE healthcare) and fixed following a carbodiimide-mediated cross-linking procedure (Pall et al. 2007). Pre-hybridization and hybridization was carried out in 5X SSC, 20 mM Na<sub>2</sub>HPO<sub>4</sub> pH 7.2, 7 % SDS, 2X Denhardt solution, 50 mg/ml herring DNA at 50 °C. Filters were washed twice with 3X SSC, 5 % SDS at 50 °C for 10 min, followed by one to two washes with 1X SSC, 1 % SDS at 50 °C for 10 min. Signals were visualized by autoradiography or by using a phosphorimager (Molecular Imager FX; Bio-Rad) for

quantification. Sequences of the different probes used are given in Table S2.

### QRT-PCR analysis

2 µg of total RNA were DNase treated using the Promega RQ1 kit following manufacturer’s recommendations. 0.2 µg of treated RNA was reverse transcribed using PrimeScript RT reagent kit (Perfect real time, Takara) in a final volume of 10 µl, using either a mix of oligodT and random hexanucleotides or a mixture of gene specific primers for *fca* 3’ end analysis experiments (*rt7-R* and *act2-R*). One µl of cDNA was used for amplification, using the Eco Real-Time system (Illumina) and SYBR Premix Ex Taq II (Tli RnaseH Plus) (Takara) in a final volume of 15 µl. Actin2 RNA was used as input control to normalize the expression levels, with a minimum of two biological repeats per experiments. Primers used can be found in Table S2.

### BiFC constructs

DRB4 and DRB7.2 open reading frames flanked by AttB1 and AttB2 sites (see Table S2 for primer sequences) were amplified from *Arabidopsis* cDNA clones (Columbia ecotype), cloned into Gateway vector pDONR221 using BP recombination (Invitrogen), and sequenced. An LR reaction between the entry vector and the complete set of four pBiFP vectors (Azimzadeh et al. 2008) produced the final expression vectors, where coding sequences are cloned in fusion with the N- and C-terminal parts of YFP (Hu et al. 2002), either as N-terminal or C-terminal fusions, under the control of the cauliflower mosaic virus 35S promoter. cDNA of DCL4 was cloned, in fusion with the N- and C-terminal parts of YFP, either as N-terminal or C-terminal fusions under the control of its endogenous promoter, in pB7GW34 vector using the “MultiSite Gateway Three-Fragment Vector Construction Kit” (Invitrogen). For fluorescence complementation tests, all eight compatible combinations between protein pairs (i.e., providing both parts of the YFP) were assayed in transient expression. A C-terminal fusion of the Glutathion S-transferase (GST) with the C-terminal part of the YFP (kindly provided by N. Lukhovitskaya, IBMP, France) was used a negative control with the appropriate compatible combination of DCL4, DRB4 or DRB7.2 YFP fusions.

### Transient assay in *Nicotiana benthamiana* leaves

Each expression vector was introduced in *Agrobacterium tumefaciens* strain GV3101 by electroporation. *Agrobacterium* bacterial cultures were incubated overnight at 28 °C with agitation. Each culture was pelleted, and resuspended

in 10 mM MgCl<sub>2</sub>, 200 μM Acetosyringone buffer to an OD<sub>600</sub> of 1.5 and put 2 h under gentle agitation at room temperature. To enhance transient expression of BiFC fusion proteins, the P19 viral suppressor of RNA silencing was coexpressed (Voinnet et al. 2003). Equal volumes of each bacterial culture were mixed before infiltration and YFP fluorescence was scored 3–4 days post-infiltration with a confocal microscope (LSM780, ZEISS).

**Acknowledgments** This work was supported by l'Agence Nationale de la Recherche (ANR-06-BLAN-0203-02), by the CNRS, by the Université de Perpignan (UPVD). It was also published under the framework of the LABEX: ANR-10-LABX-0036\_NETRINA and benefits from a funding from the state managed by the French National Research Agency as part of the Investments for the future program. M.C. and T.M. were supported by a grant from the French Ministry of Research and M-A.T. by a core grant from ETH-Z.

**Author contribution** MC and TP carried out the molecular biology studies and JMD the molecular evolution studies. VJ and JD provided general technical help. TM, PD and MAT made the BiFC work. CBA and MNPP helped with the immunoprecipitation experiments. MC, TP, PD and JMD conceived the study, and participated in its design. JMD coordinated the writing of the manuscript. All authors read and approved the final manuscript.

## References

- Adenot X, Elmayan T, Laussergues D, Boutet S, Bouche N, Gasciolli V, Vaucheret H (2006) DRB4-dependent TAS3 transacting siRNAs control leaf morphology through AGO7. *Curr Biol* 16:927–932. doi:10.1016/j.cub.2006.03.035
- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55:539–552
- Azimzadeh J, Nacry P, Christodoulidou A, Drevensek S, Camilleri C, Amieur N, Parcy F, Pastuglia M, Bouchez D (2008) Arabidopsis TONNEAU1 proteins are essential for preprophase band formation and interact with centrin. *Plant Cell* 20:2146–2159. doi:10.1105/tpc.107.056812
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409:363–366. doi:10.1038/35053110
- Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Res* 13:137–144. doi:10.1101/gr.751803
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 123:1279–1291
- Bouche N, Laussergues D, Gasciolli V, Vaucheret H (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J* 25:3347–3356. doi:10.1038/sj.emboj.7601217
- Cerutti H, Ma X, Msanne J, Repas T (2011) RNA-mediated silencing in algae: biological roles and tools for analysis of gene function. *Eukaryot Cell* 10:1164–1172. doi:10.1128/EC.05106-11
- Chang KY, Ramos A (2005) The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J* 272:2109–2117. doi:10.1111/j.1742-4658.2005.04652.x
- Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436:740–744. doi:10.1038/nature03868
- Clavel M et al (2015) Parallel action of AtDRB2 and RdDM in the control of transposable element expression. *BMC Plant Biol* 15:70. doi:10.1186/s12870-015-0455-z
- Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* 508:411–415. doi:10.1038/nature13069
- Curtin SJ, Watson JM, Smith NA, Eamens AL, Blanchard CL, Waterhouse PM (2008) The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS Lett* 582:2753–2760. doi:10.1016/j.febslet.2008.07.004
- Daniels SM et al (2009) Characterization of the TRBP domain required for dicer interaction and function in RNA interference. *BMC Mol Biol* 10:38. doi:10.1186/1471-2199-10-38
- Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* 432:231–235. doi:10.1038/nature03049
- Dugre-Brisson S, Elvira G, Boulay K, Chatel-Chaix L, Moulard AJ, DesGroseillers L (2005) Interaction of Staufe1 with the 5' end of mRNA facilitates translation of these RNAs. *Nucleic Acids Res* 33:4797–4812. doi:10.1093/nar/gki794
- Dunoyer P, Himber C, Ruiz-Ferrer V, Alioua A, Voinnet O (2007) Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nat Genet* 39:848–856. doi:10.1038/ng2081
- Eamens AL, Smith NA, Curtin SJ, Wang MB, Waterhouse PM (2009) The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* 15:2219–2235. doi:10.1261/rna.1646909
- Eamens AL, Kim KW, Curtin SJ, Waterhouse PM (2012a) DRB2 is required for microRNA biogenesis in *Arabidopsis thaliana*. *PLoS One* 7:e35933. doi:10.1371/journal.pone.0035933
- Eamens AL, Wook Kim K, Waterhouse PM (2012b) DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in *Arabidopsis thaliana*. *Plant Signal Behav* 7:1224–1229. doi:10.4161/psb.21518
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Ferrandon D, Elphick L, Nusslein-Volhard C, St Johnston D (1994) Staufe protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* 79:1221–1232
- Fierro-Monti I, Mathews MB (2000) Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem Sci* 25:241–246
- Fukudome A, Kanaya A, Egami M, Nakazawa Y, Hiraguri A, Moriyama H, Fukuhara T (2011) Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of Arabidopsis Dicer-like 4. *RNA* 17:750–760. doi:10.1261/rna.2455411
- Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* 15:1494–1500. doi:10.1016/j.cub.2005.07.024
- Gleghorn ML, Gong C, Kielkopf CL, Maquat LE (2013) Staufe1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat Struct Mol Biol* 20:515–524. doi:10.1038/nsmb.2528
- Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432:235–240. doi:10.1038/nature03120
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704

- Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN (2004) The Droscha–DGCR8 complex in primary microRNA processing. *Genes Dev* 18:3016–3027. doi:10.1101/gad.1262504
- Han J et al (2006) Molecular basis for the recognition of primary microRNAs by the Droscha–DGCR8 complex. *Cell* 125:887–901. doi:10.1016/j.cell.2006.03.043
- Hiraguri A et al (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Mol Biol* 57:173–188. doi:10.1007/s11103-004-6853-5
- Hitti EG, Sallacz NB, Schoft VK, Jantsch MF (2004) Oligomerization activity of a double-stranded RNA-binding domain. *FEBS Lett* 574:25–30. doi:10.1016/j.febslet.2004.07.080
- Hu CD, Chinenov Y, Kerppola TK (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* 9:789–798. doi:10.1016/S1097-2765(02)00496-3
- Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293:834–838. doi:10.1126/science.1062961
- Jakubiec A, Yang SW, Chua NH (2012) Arabidopsis DRB4 protein in antiviral defense against Turnip yellow mosaic virus infection. *Plant J* 69:14–25. doi:10.1111/j.1365-313X.2011.04765.x
- Jeddellouh JA, Stokes TL, Richards EJ (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet* 22:94–97. doi:10.1038/8803
- Kim YK, Furic L, Parisien M, Major F, DesGroseillers L, Maquat LE (2007) Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J* 26:2670–2681. doi:10.1038/sj.emboj.7601712
- Kiyota E, Okada R, Kondo N, Hiraguri A, Moriyama H, Fukuhara T (2011) An Arabidopsis RNase III-like protein, AtrTL2, cleaves double-stranded RNA in vitro. *J Plant Res* 124:405–414. doi:10.1007/s10265-010-0382-x
- Kurihara Y, Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci USA* 101:12753–12758. doi:10.1073/pnas.0403115101
- Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna* 12:206–212
- Laraki G, Clerzius G, Daher A, Melendez-Pena C, Daniels S, Gatignol A (2008) Interactions between the double-stranded RNA-binding proteins TRBP and PACT define the Medial domain that mediates protein-protein interactions. *RNA Biol* 5:92–103
- Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320. doi:10.1093/molbev/msn067
- Lee Y et al (2003) The nuclear RNase III Droscha initiates microRNA processing. *Nature* 425:415–419. doi:10.1038/nature01957
- Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN (2006) The role of PACT in the RNA silencing pathway. *EMBO J* 25:522–532. doi:10.1038/sj.emboj.7600942
- Liu F, Bakht S, Dean C (2012) Cotranscriptional role for Arabidopsis DICER-LIKE 4 in transcription termination. *Science* 335:1621–1623. doi:10.1126/science.1214402
- Manavella PA, Hagmann J, Ott F, Laubinger S, Franz M, Macek B, Weigel D (2012) Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* 151:859–870. doi:10.1016/j.cell.2012.09.039
- Masliah G, Barraud P, Allain FH (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci CMLS* 70:1875–1895. doi:10.1007/s00018-012-1119-x
- McCue AD, Nuthikattu S, Reeder SH, Slotkin RK (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet* 8:e1002474. doi:10.1371/journal.pgen.1002474
- McCue AD, Nuthikattu S, Slotkin RK (2013) Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. *RNA Biol* 10:1379–1395. doi:10.4161/rna.25555
- McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK (2014) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J*. doi:10.15252/embj.201489499
- Micklem DR, Adams J, Grunert S (2000) St Johnston D. Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation *EMBO J* 19:1366–1377. doi:10.1093/emboj/19.6.1366
- Nakazawa Y, Hiraguri A, Moriyama H, Fukuhara T (2007) The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. *Plant Mol Biol* 63:777–785. doi:10.1007/s11103-006-9125-8
- Nanduri S, Rahman F, Williams BR, Qin J (2000) A dynamically tuned double-stranded RNA binding mechanism for the activation of antiviral kinase PKR. *EMBO J* 19:5567–5574. doi:10.1093/emboj/19.20.5567
- Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol* 162:116–131. doi:10.1104/pp.113.216481
- Pall GS, Codony-Servat C, Byrne J, Ritchie L, Hamilton A (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res* 35:e60
- Péllissier T, Bousquet-Antonelli C, Lavie L, Deragon JM (2004) Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res* 32:3957–3966
- Péllissier T, Clavel M, Chaparro C, Pouch-Péllissier MN, Vaucheret H, Deragon JM (2011) Double-stranded RNA binding proteins DRB2 and DRB4 have an antagonistic impact on polymerase IV-dependent siRNA levels in Arabidopsis. *RNA* 17:1502–1510. doi:10.1261/rna.2680711
- Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA (2010) Structure of the *Arabidopsis thaliana* DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA* 16:474–481. doi:10.1261/rna.1965310
- Raja P, Jackel JN, Li S, Heard IM, Bisaro DM (2014) Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. *J Virol* 88:2611–2622. doi:10.1128/JVI.02305-13
- Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* 20:3407–3425. doi:10.1101/gad.1476406
- Saito K, Ishizuka A, Siomi H, Siomi MC (2005) Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol* 3:e235. doi:10.1371/journal.pbio.0030235
- Sarazin A, Voinnet O (2014) Exploring new models of easiRNA biogenesis. *Nat Genet* 46:530–531. doi:10.1038/ng.2993
- Schuldt AJ, Adams JH, Davidson CM, Micklem DR, Haseloff J, Johnston D, Brand AH (1998) Miranda mediates asymmetric protein and RNA localization in the developing nervous system. *Genes Dev* 12:1847–1857
- Slotkin RK (2010) The epigenetic control of the Athila family of retrotransposons in Arabidopsis. *Epigenetics* 5:483–490

- Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136:461–472. doi:[10.1016/j.cell.2008.12.038](https://doi.org/10.1016/j.cell.2008.12.038)
- Sohn SY, Bae WJ, Kim JJ, Yeom KH, Kim VN, Cho Y (2007) Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* 14:847–853. doi:[10.1038/nsmb1294](https://doi.org/10.1038/nsmb1294)
- Song L, Han MH, Lesicka J, Fedoroff N (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci USA* 104:5437–5442. doi:[10.1073/pnas.0701061104](https://doi.org/10.1073/pnas.0701061104)
- St Johnston D, Beuchle D, Nusslein-Volhard C (1991) *Staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell* 66:51–63
- St Johnston D, Brown NH, Gall JG, Jantsch M (1992) A conserved double-stranded RNA-binding domain. *Proc Natl Acad Sci USA* 89:10979–10983
- Steimer A, Amedeo P, Afsar K, Fransz P, Scheid OM, Paszkowski J (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. *Plant Cell* 12:1165–1178
- Tabara H, Yigit E, Siomi H, Mello CC (2002) The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DEXH-box helicase to direct RNAi in *C. elegans*. *Cell* 109:861–871
- Tanurdzic M et al (2008) Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol* 6:2880–2895. doi:[10.1371/journal.pbio.0060302](https://doi.org/10.1371/journal.pbio.0060302)
- Tian B, Bevilacqua PC, Diegelman-Parente A, Mathews MB (2004) The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol* 5:1013–1023. doi:[10.1038/nrm1528](https://doi.org/10.1038/nrm1528)
- Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD (2004) A protein sensor for siRNA asymmetry. *Science* 306:1377–1380. doi:[10.1126/science.1102755](https://doi.org/10.1126/science.1102755)
- Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* 33:949–956. doi:[10.1046/j.1365-3113.2003.01676.x](https://doi.org/10.1046/j.1365-3113.2003.01676.x)
- Xie Z et al (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* 2:E104. doi:[10.1371/journal.pbio.0020104](https://doi.org/10.1371/journal.pbio.0020104)
- Yang SW, Chen HY, Yang J, Machida S, Chua NH, Yuan YA (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* 18:594–605. doi:[10.1016/j.str.2010.02.006](https://doi.org/10.1016/j.str.2010.02.006)
- Yeom KH, Lee Y, Han J, Suh MR, Kim VN (2006) Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res* 34:4622–4629. doi:[10.1093/nar/gkl458](https://doi.org/10.1093/nar/gkl458)
- Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang XJ, Qi Y (2007) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev* 21:1190–1203. doi:[10.1101/gad.1543507](https://doi.org/10.1101/gad.1543507)
- Zhu S et al (2013) Double-stranded RNA-binding protein 4 is required for resistance signaling against viral and bacterial pathogens. *Cell Rep* 4:1168–1184. doi:[10.1016/j.celrep.2013.08.018](https://doi.org/10.1016/j.celrep.2013.08.018)

## 1.2 IDENTIFICATION D'UN COMPLEXE SPECIFIQUE DE DRBs

### SEQUESTRANT LES PRECURSEURS D'ENDOIR-siRNAs

N'ayant pu (i) ni mettre en évidence d'effet majeur de l'absence de DRB7.2 sur les principales classes de sRNAs endogènes (miRNAs, tasiRNAs, p4-siRNAs), (ii) ni constater de phénotypes développementaux évidents chez les plantes *drb7.2* mutantes, nous nous sommes alors demandés si cette protéine pouvait être impliquée dans la biogénèse de sRNAs endogènes discrets. Afin d'avoir directement une vue exhaustive de l'impact de DRB7.2, si toutefois il y en avait un, sur une des voies du RNA silencing, nous avons alors décidé de réaliser un séquençage à haut-débit des sRNAs présents dans les fleurs de plantes sauvages (Col-0) ou mutantes (*drb7.2*). Ce tissu a été choisi car c'est celui possédant la plus grande variété en sRNAs.

L'analyse bioinformatique de ces banques a alors permis de révéler un effet majeur de DRB7.2 dans la biogénèse des endoIR-siRNAs. Cet effet se traduisait principalement par une forte augmentation de la quantité d'endoIR-siRNAs présents dans le mutant *drb7.2* mais aussi par un changement dans le pattern d'accumulation des différentes tailles de siRNAs qui dérivent de ces loci endogènes, avec principalement, une augmentation de la quantité des endoIR-siRNAs de 24nt. De manière intéressante, un pattern similaire d'accumulation des endoIR-siRNAs avait déjà été observé par le passé dans des plantes mutées pour DRB4 (*drb4*), le cofacteur de DRB7.2, comme révélé précédemment par BiFC (Pélissier *et al*, 2011; Clavel *et al*, 2016). Grâce à une combinaison d'approches génétiques et biochimiques, nous avons pu mettre en évidence que DRB7.2 régulait négativement la production des endoIR-siRNAs en séquestrant spécifiquement leur précurseur ARNdb et que cette répression nécessitait la présence de DRB4. Simultanément, ces travaux ont révélé que DRB4 était partitionné en deux pools cellulaires distincts, remplissant des fonctions différentes, à travers son interaction mutuellement exclusive avec DCL4 et DRB7.2. Collectivement,

ces résultats ont d'une part permis d'attribuer une fonction importante à DRB7.2 et de définir les bases moléculaires d'un nouveau rôle de DRB4 dans une des voies du RNA silencing endogène, en mettant en évidence l'existence d'un complexe de DRBs (DRB4/DRB7.2) modulant spécifiquement la production des endoIR-siRNAs. D'autre part, l'existence d'un tel complexe chez *A. thaliana*, ainsi que la conservation de DRB7.2 et DRB4 chez la totalité des plantes vasculaires, suggèrent fortement que, bien qu'encore inconnue à l'heure actuelle, les endoIR-siRNAs doivent probablement avoir une fonction biologique importante chez les plantes.

L'ensemble des résultats obtenus lors de ces travaux a été réuni au sein d'un manuscrit qui a été soumis en Juillet 2016 dans le journal *Nucleic Acids Research*.

# A specific dsRNA-binding protein complex selectively sequesters endogenous inverted-repeat siRNA precursors and inhibits their processing

Thomas Montavon<sup>1</sup>, Yerim Kwon<sup>1</sup>, Aude Zimmermann<sup>1</sup>, Philippe Hamann<sup>2</sup>,  
Timothée Vincent<sup>1</sup>, Valérie Cognat<sup>1</sup>, Fabrice Michel<sup>1</sup> and Patrice Dunoyer<sup>1,\*</sup>

<sup>1</sup>Université de Strasbourg, CNRS, IBMP UPR 2357, F-67000 Strasbourg, France and <sup>2</sup>Université de Strasbourg, CNRS, IBMC FRC1589, Plateforme Protéomique Strasbourg - Esplanade, F-67000 Strasbourg, France

Received July 25, 2016; Revised December 01, 2016; Editorial Decision December 02, 2016; Accepted December 06, 2016

## ABSTRACT

In plants, several dsRNA-binding proteins (DRBs) have been shown to play important roles in various RNA silencing pathways, mostly by promoting the efficiency and/or accuracy of Dicer-like proteins (DCL)-mediated small RNA production. Among the DRBs encoded by the *Arabidopsis* genome, we recently identified DRB7.2 whose function in RNA silencing was unknown. Here, we show that DRB7.2 is specifically involved in siRNA production from endogenous inverted-repeat (endoIR) loci. This function requires its interacting partner DRB4, the main cofactor of DCL4 and is achieved through specific sequestration of endoIR dsRNA precursors, thereby repressing their access and processing by the siRNA-generating DCLs. The present study also provides multiple lines of evidence showing that DRB4 is partitioned into, at least, two distinct cellular pools fulfilling different functions, through mutually exclusive binding with either DCL4 or DRB7.2. Collectively, these findings revealed that plants have evolved a specific DRB complex that modulates selectively the production of endoIR-siRNAs. The existence of such a complex and its implication regarding the still elusive biological function of plant endoIR-siRNA will be discussed.

## INTRODUCTION

In eukaryotes, RNA silencing is a conserved mechanism that plays essential roles in many biological processes such as maintenance of genome stability, development or antiviral defense. The various classes of endogenous or exogenous 21–24 nucleotide (nt) small RNA (sRNA), which confer the sequence specificity of this mechanism, are produced from structurally distinct double-stranded RNA (dsRNA) precursors by RNaseIII-like enzymes called Dicers, or Dicer-

like (DCL) in plants (1–3). The plant model *Arabidopsis thaliana* encodes four DCL proteins with specialized functions. DCL1 produces the majority of micro RNAs (miRNAs) from relatively short imperfect stem-loop RNA precursors, whereas populations of 21, 22 and 24 nt short-interfering RNAs (siRNAs) are generated through the action of DCL4, DCL2 and DCL3, respectively, on various dsRNA substrates. For instance, DCL4-dependent 21 nt trans-acting (ta-)siRNAs are produced by sequential processing of long dsRNA precursors generated by the action of RNA-dependent RNA polymerase 6 (RDR6) on single-stranded RNA (4–6). By contrast, DCL3-dependent 24 nt siRNAs, the most abundant class of sRNAs, are produced from short dsRNA precursors, 27–50 nt in length, generated by PolIV and RDR2 and are usually referred to as p4-siRNAs (7,8). Finally, long and RDR-independent dsRNA precursors, originating from several endogenous loci configured as inverted-repeat (IR) transcripts are processed by the three siRNA-producing DCLs to generate 21, 22 and 24 nt endogenous inverted-repeat-derived (endoIR-)siRNAs (3,9–11).

Upon processing, sRNAs are incorporated into an RNA-induced silencing complex containing 1 of the 10 Argonaute (AGO) proteins that effect RNA silencing in *Arabidopsis*. Most 21 and 22 nt sRNAs load into AGO1 to promote cleavage or translational inhibition of target transcripts (12–14), whereas 24 nt siRNAs associate with AGO4, AGO6 or AGO9 to guide heterochromatin formation by DNA methylation and histone modification (15). Infection by viruses also leads to production of, mostly, 21 and 22 nt virus-derived (v)siRNAs through processing by DCL4, or its surrogate DCL2, of dsRNA replication intermediates or intramolecular fold-back structures within viral genomes (16–18). These v-siRNA are mainly loaded into AGO1 and AGO2, which then target single-stranded viral RNA for cleavage (19–24).

In addition to Dicer or DCL proteins, several dsRNA-binding proteins (DRBs) have been shown to play impor-

\*To whom correspondence should be addressed. Tel: +33 3 67155368; Fax: +33 3 67155300; Email: patrice.dunoyer@ibmp-cnrs.unistra.fr

tant roles in plant and animal RNA silencing pathways (25–28). The *Arabidopsis* genome encodes five DRBs (DRB1–5) that are strictly composed of two dsRNA-binding motif (dsRBM) with no other catalytic domain. Among those five, DRB1, also known as HYL1, is the best studied and was shown to be required for precise and efficient processing of miRNA precursors (27,29–31) and for selection of the miRNA guide strand loaded into RNA-induced silencing complex. These functions are achieved through DRB1 interaction with DCL1 *via* its second dsRBM (32–35), while the first dsRBM binds miRNA precursors as well as mature miRNA duplexes (30,36,37). DRB2 is also involved in processing miRNAs but only in the shoot apical meristem where it represses DRB1 transcription (38,39). Interestingly, DRB1 seems to be specifically required for miRNA-guided cleavage whereas DRB2 is required for miRNA-mediated translational inhibition, suggesting that the miRNA mode of action is, at least partly, defined by those two DRBs (39,40). The roles of DRB3 and DRB5 are more elusive. They have been shown to be dispensable for sRNA production but seem to be required for translational repression of DRB2-associated miRNA target transcripts (40,41). DRB3 was also found to interact with DCL3 and AGO4 and impact the methylation of a viral genome without being required for the processing of viral dsRNA by DCL3 (42). Finally, DRB4 is essential for DCL4 activity *in vitro* (43) and was shown to physically and functionally interact with DCL4 *in vivo* (35,44), where it is required for accurate and efficient processing of ta-siRNA precursors (44,45) and of the few DCL4-dependent miRNAs (46,47). DRB4 plays also a role in antiviral defense either by promoting DCL4-dependent vsiRNA production (41,48) or by regulating resistance (*R*) gene-mediated immunity (49). More intriguingly, DRB4 was also shown to affect the processing by DCL3 of p4-siRNA and IR-siRNA precursors, where the former is decreased whereas the latter is strongly increased in *drb4* mutant (47,50). However, this effect is most likely indirect given that DRB4 does not interact *in vivo* with DCL3 (47).

Besides those five DRBs containing two dsRBMs, we have recently identified a new DRB family (named DRB7) conserved in all vascular plants and harboring a single dsRBM that shows concerted evolution with the most C-terminal dsRBM of DCL4 (51). We showed that one of the two *Arabidopsis* DRB7 proteins (DRB7.2) interacts with DRB4 but does not seem to be required for the production of any DCL4-dependent sRNAs, including the newly identified class of epigenetically activated (ea)siRNAs that accumulate mostly in the vegetative nucleus of pollen grains or in Decreased DNA Methylation 1 (*ddm1*) mutants (52–56). By contrast, loss of DRB7.2 triggered an increase in the accumulation of DCL3-dependent 24 nt siRNAs from easiRNA-generating loci. However, given that this observation was only made in a *ddm1* mutant background, and considering the pleiotropic nature of this mutation, the reason of this change is still unclear, as is the function of DRB7.2 (51).

Using genetic, biochemical and small RNA profiling approaches, we show, here, that DRB7.2 negatively regulates the production of endogenous IR-derived siRNA through specific sequestration of their dsRNA precursors. This se-

questration requires its interacting partner DRB4 and represses their processing by the siRNA-generating DCLs, chiefly DCL3. The present study also reveals that DRB4 is partitioned into two distinct cellular pools fulfilling different functions, through mutually exclusive binding with either DCL4 or DRB7.2 that accumulate in distinct subcellular compartments. Collectively, these results uncover the existence of a specific DRB complex that selectively modulates processing of endogenous IRs, unravelling further layers of complexity in the plant RNA silencing pathways.

## MATERIALS AND METHODS

### Plant materials, transformation and virus inoculation

Knock-out T-DNA mutant lines *dcl2-1* (SALK\_064627), *dcl3-1* (SALK\_005512), *dcl4-2* (GABI\_160G05), *dcl2/dcl3*, *dcl2/dcl4*, *dcl3/dcl4*, *drb4-1* (SALK\_000736) and missense mutant *dcl4-8* were described previously (44,50,57–60). The 35S-GFP line in Columbia ecotype (Col-0) was kindly provided by M. Jean Molinier (IBMP). The mutant *drb7.2* (GABI 525B11) was obtained from the Arabidopsis Biological Resource Center (ABRC) and was described previously (51). *Arabidopsis thaliana* reference ecotype used was Columbia. Genotyping of the Transfer DNA (T-DNA) insertion lines was performed by polymerase chain reaction (PCR), using allele-specific primers. PCR primers are listed in Supplementary Table S1. The *drb7.2/dcl2*, *drb7.2/dcl3*, *drb7.2/dcl4*, *drb4/dcl2*, *drb4/dcl3* and *drb4/dcl4* mutant lines were generated by standard genetic crosses. Homozygous mutant genotypes were confirmed by allele-specific PCR assays after two generations.

Plant seeds were stratified for 2–4 days at 4°C before growth under standard conditions at 22°C with a 16 h light/8 h dark cycle, either in soil or on Murashige and Skoog (MS) agar plates supplemented with kanamycin (50 µg/ml), hygromycin (34 µg/ml) or with phosphinotricin (10 µg/ml). *Arabidopsis* were transformed using the floral dip method (61).

Binary vectors carrying the infectious clone of *Tobacco rattle virus*-PDS or *Turnip crinkle virus* were described previously (62,63). These vectors were mobilized into *Agrobacterium* strain GV3101 and used for virus infection in *Arabidopsis*. Infected systemic leaves were collected at 14 dpi.

### DNA constructs

For the DRB7.2:GFP fusion protein, the genomic sequence of DRB7.2 (At4g00420), either alone or with a 3.5 kb region upstream of the ATG, containing its endogenous promoter, was amplified with DNA PHUSION polymerase (Thermo Scientific), cloned into pGEMT-easy vector (Promega) for sequencing and sub-cloned into pCTL1300 binary vector for transient expression to generate 35S-DRB7.2:GFP or pDRB7.2-DRB7.2:GFP constructs. For the DCL4:FHA/GFP constructs, the genomic coding sequence of DCL4 was cloned, in fusion with GFP or double-epitope tag (2xFlag 2xHA), as C-terminal fusions under the control of its endogenous promoter (1.5 kb upstream of the ATG), in pB7GW34 vector using the ‘MultiSite Gateway Three-Fragment Vector Construction Kit’ (Invitrogen).



## RNA analysis

Total RNA was extracted from *Arabidopsis* tissues with Tri-Reagent (Sigma, St Louis, MO, USA) according to manufacturer's instructions. RNA gel blot analysis of high and low molecular weight RNA was on 5 and 45  $\mu\text{g}$  of total RNA, respectively, and was conducted as described previously (64). Radiolabeled probes for detection of the IR71 and IR2039 siRNAs were made by random priming reactions (Promega) in the presence of  $\alpha\text{-}^{32}\text{P}\text{-dCTP}$ . The template used was a 650-bp-long (for IR71) and 670-bp-long (for IR2039) PCR product amplified from the *Arabidopsis* gDNA. DNA oligonucleotides complementary to miRNAs, trans-acting siRNAs or heterochromatic siRNAs (Supplementary Table S1) were end-labeled with  $\gamma\text{-}^{32}\text{P}\text{-ATP}$  using T4 PNK (Thermo Scientific). Detection of the *Arabidopsis* U6 small nuclear RNA was used to confirm equal loading. Each result was, on average, confirmed on four independent biological replicates.

## Protein extraction and analysis

Total proteins were extracted from *Arabidopsis* flower buds as previously described (65) and were resolved on SDS-PAGE. After electroblotting onto Immobilon-P membrane (Millipore), protein gel blot analysis was carried out using the appropriate antiserum. The specificity of DRB4, DCL4, DCL3, AGO1 and AGO4 antibodies used in this study has been verified by Western blot analysis on protein extracts from wild type and the corresponding T-DNA mutant plants (64). Each result was, on average, confirmed on three independent biological replicates.

## Protein and RNA immunoprecipitation

For immunoprecipitation, 0.1 g of flower buds was ground in liquid nitrogen and homogenized in 1 ml of lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100) containing 1 tablet/50 ml of protease inhibitor cocktail (Roche) for 15 min at 4°C. Cell debris was removed by two successive centrifugations at 13 000 rpm at 4°C for 10 min. After the second centrifugation an aliquot of supernatants was taken for input fraction. The remaining extracts were incubated with magnetic microparticles coated with monoclonal GFP or HA antibodies (MACS purification system, Miltenyi Biotec) at 4°C for 30 min. Samples were passed through Mcolumn (MACS purification system, Miltenyi Biotec) and an aliquot of the flow-through fraction was taken. The Mcolumn were then washed 2 times with 500  $\mu\text{l}$  of lysis buffer and 1 time with 100  $\mu\text{l}$  of washing buffer (20 mM Tris-HCl, pH 7.5). To elute the immunoprecipitated proteins, 95°C pre-warmed Western blot loading buffer (10% glycerol, 4% SDS, 62.5 mM Tris-HCl pH 6.8, 5% (v/v) 2- $\beta$ -mercaptoethanol, Bromophenol Blue) was passed through the Mcolumn. Proteins were analyzed by Western blotting. For RNA extraction from immunoprecipitated proteins, magnetic beads were eluted with 50°C pre-warmed Tri-Reagent (Sigma, St Louis, MO, USA) and RNA were extracted according to manufacturer's instructions. RNA were analyzed by Northern blotting.

## Real-time RT-qPCR

In all, 2  $\mu\text{g}$  of total RNA samples were reverse-transcribed into cDNA using SuperScript III reverse transcriptase (Invitrogen) with a mix of oligo(dT) and random hexamers. The cDNA was quantified using a SYBR Green qPCR kit (Eurogentec) and DRB4 specific primers (Supplementary Table S1). PCR was performed in triplicate in 384-well optical reaction plates heated for 10 min at 95°C, followed by 45 cycles of denaturation for 15 s at 95°C, annealing for 20 s at 60°C and elongation for 40 s at 72°C. A melting curve was performed at the end of the amplification by steps of 1°C (from 53°C to 95°C). The number of cycles after which fluorescence reached a set threshold (Ct value) was averaged for each triplicate and expressed as a ratio to the actin-2 mRNA.

## Sequencing and bioinformatic analyses

Total RNA from two independent biological replicate of Col-0 and *drb7.2* flowers was extracted with Tri-Reagent (Sigma). A 6  $\mu\text{g}$  weight of total RNA was used for preparation of small-RNA libraries. Briefly, a first acrylamide gel purification of small RNA between 18–30 nt was performed. The libraries were then constructed using the TruSeq Small RNA preparation kit from illumina following manufacturer's instructions and the libraries were sequenced using a Illumina Genome Analyser (Fasteris, Switzerland). After removal of the 3' adapters from the reads by Fasteris using an in-house developed script (detailed in the Supplementary Table S2), sequences (18–26 nt in length) were mapped using Bowtie (66) to the TAIR10 assembly. Only unique reads with no mismatch to the genome were kept using an in-house python script (available upon request). Counts were made with intersectBed, from bedtools suite (v.2.25.0) (67). Data were normalized with their own conserved miRNA counts (miR156, 157, 159, 160, 162, 164, 165, 166, 167, 168, 169, 170, 171, 172, 319, 390, 391, 393, 394, 395, 396, 397, 398, 399, 403, 408, 472) (68) under the assumption that they were globally unaffected (Figure 2 and (51)). miRNA annotation was downloaded from miRBase v.21 (69). EndoIR-siRNAs were annotated using the sRNA-producing loci predicted by (10), which did not match miRBase miRNAs. Loci-producing siRNAs in a PolIV-dependent manner were retrieved from (70). Those data have been converted from TAIR6 to TAIR10 coordinates by a perl script provided by the 1001 genomes project ([http://1001genomes.org/data/software/translate\\_tair8/](http://1001genomes.org/data/software/translate_tair8/)) and manually validated. tasiRNA producing loci were given by tasiRNADB (71). The data reported in this paper will be deposited in the Gene Expression Omnibus (GEO) database, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accession no. GSE92309).

## Imaging

Imaging of fluorescent protein fusions was performed in roots of 2 weeks old stably-transformed plants using confocal microscope LSM700 or LSM780 from Carl Zeiss with a 40x objective and a 488-nm laser for excitation. Images were converted in .tif and scale bars added using ImageJ software (NIH). At least 30 roots of each genotype were analyzed.

## RESULTS

### DRB7.2 is involved in endogenous IR-siRNA production

In order to gain insights into the function of DRB7.2, we performed genome-wide small RNA profiling on wild-type Col-0 and *drb7.2* mutant flowers, by small RNA sequencing of two independent biological replicates. We first examined the distribution of sRNAs that perfectly match the *Arabidopsis* genome, excluding rRNA and tRNA, and found no major differences in the global size distribution of total reads between *drb7.2* and Col-0 plants (Figure 1A and Supplementary Table S2), apart from a slight decrease in the amount of 22 nt-long sRNAs that might be correlated with the slight increase in the amount of 24 nt-long siRNA.

Next, we sorted sRNAs into their major functional categories and normalized their abundance to the amount of the evolutionarily conserved DCL1-dependent miRNAs (68), whose accumulation is not affected by the *drb7.2* mutation (Figure 2A; (51)). We found that lack of DRB7.2 does not significantly impact the accumulation of ta-siRNAs or p4-siRNAs (Figure 1B). However, and in sharp contrast, the accumulation of endogenous IR-siRNAs was strongly altered in *drb7.2* mutant plants. This was characterized by (i) a global increase in the total amount of endoIR-siRNAs produced and (ii) a clear change in their accumulation pattern with, chiefly, a strong increase in the accumulation of DCL3-dependent 24 nt endoIR-siRNAs, compare to wild-type plants (Figure 1B and Supplementary Table S2). This specific effect of the *drb7.2* mutation was also evident when size distribution of sRNAs corresponding to three endogenous IR representative loci were analyzed (Figure 1C and Supplementary Table S2). This effect was also accompanied by a robust increase in DCL4-dependent 21 nt endoIR-siRNAs (Figure 1B), although not on all endogenous IR loci (Supplementary Table S2). Collectively, these observations suggest that DRB7.2 is, somehow, specifically involved in endoIR-siRNA production.

Northern analysis of various representatives of the different classes of endogenous sRNAs confirmed the results of sRNA sequencing. Indeed, in the *drb7.2* mutant background, the steady-state levels of DCL1-dependent miRNAs, DCL4-dependent ta-siRNAs and miRNAs, and DCL3-dependent p4-siRNAs were similar to the WT control Col-0 (Figure 2A). By contrast, the accumulation pattern of endogenous IR-siRNAs was clearly altered in *drb7.2* mutant plants, with a strong increase in DCL3-dependent 24 nt IR-siRNA levels and a more modest, but visible, increase in DCL4-dependent 21 nt IR-siRNA (Figure 2A and Supplementary Figure S1). Interestingly, a similar pattern of endoIR-siRNA accumulation was also observed in the *drb4* mutant background (Figure 2A and Supplementary Figure S1), indicating that both DRB4 and DRB7.2 regulate the processing of endogenous IR transcripts.

### Increased 24 nt IR-siRNA accumulation is not caused by a change in the homeostasis of DCL3

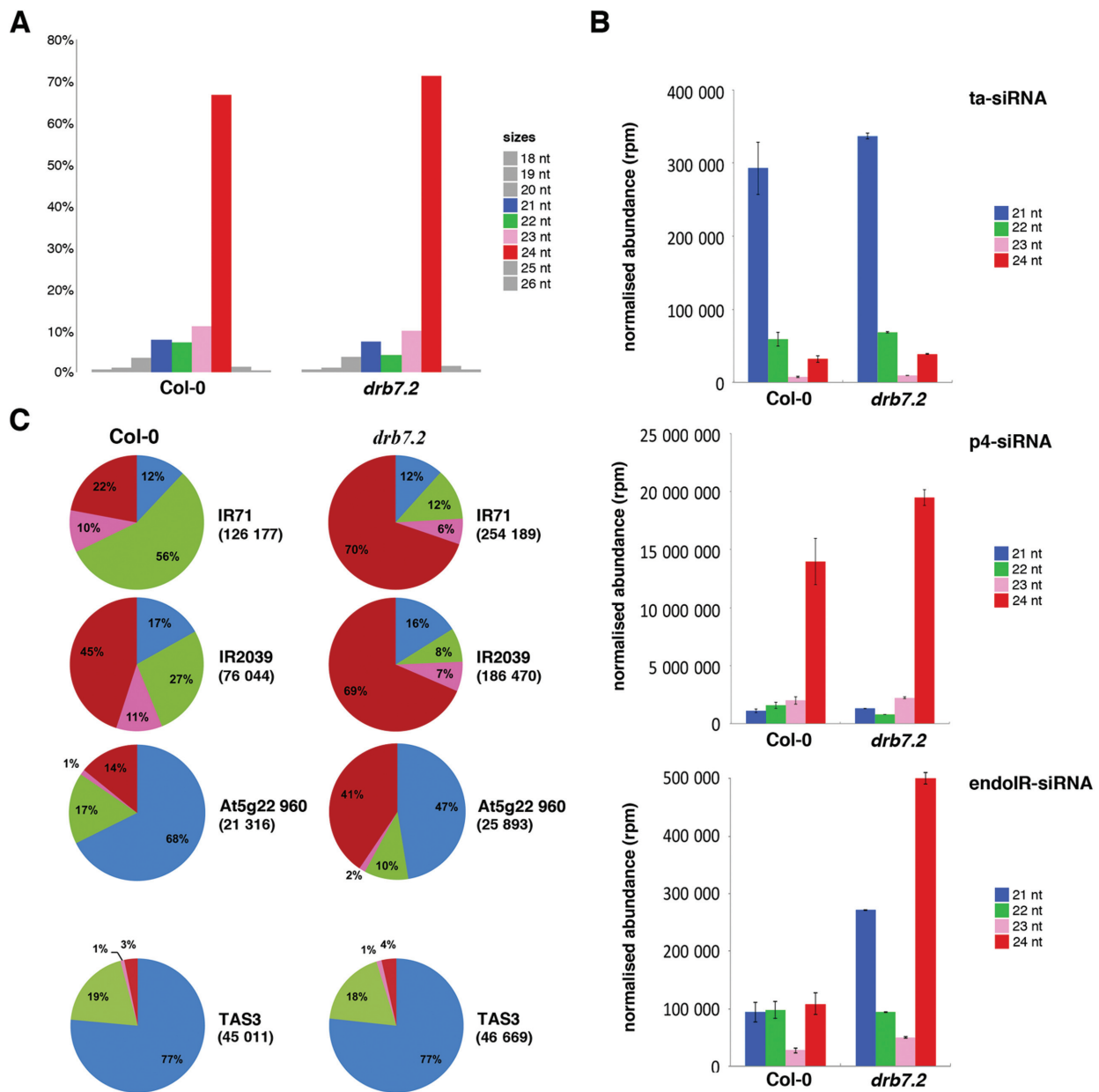
The above results prompted us to investigate whether the global increase in endoIR-derived siRNA accumulation observed in *drb7.2* (Figure 1B) and *drb4* (47), and the change in their accumulation pattern (Figures 1B, C and 2A), could

result from a modification in the steady-state level of one, or more, of the sRNA-producing DCLs. To test this hypothesis, we assessed protein levels of DCL3 and DCL4, which are responsible for the production of the 24 nt and 21 nt siRNAs from endogenous IR loci, respectively. Unfortunately, despite several attempts, we were not able to obtain specific antibodies directed against DCL2 and, therefore, its accumulation could not be tested. We also decided to assess the accumulation of DCL1 because DCL1 has been previously shown to optimize the processing of IRs, presumably by facilitating access and subsequent processing by the other DCLs (50). Western blot analysis revealed that DCL1, DCL3 and DCL4 steady-state levels were similar in *drb7.2* and *drb4* to those found in control plants (Figure 2B), indicating that the altered accumulation of endoIR-derived siRNA observed in *drb7.2* or *drb4* does not result from a change in the homeostasis of those processing factors.

Similar observations were made for AGO1 and AGO4 accumulation, the two main Argonaute proteins that load and potentially stabilize, 21/22 nt and 24 nt endoIR-siRNAs, respectively (Figure 2B). Finally, based on its interaction with DRB7.2 (51) and their similar pattern of IR-siRNA accumulation, we also assessed the level of DRB4 and found that its accumulation was slightly decreased in *drb7.2* mutant plants (Figure 2B). This effect does not result from decreased level of the DRB4 mRNA as assessed by qRT-PCR analyses (Figure 2C), and is most likely due to destabilization and/or increased turn-over of the pool of DRB4 normally in complex with DRB7.2 (see below). Of note, this lower amount of DRB4 observed in *drb7.2* does not impact the production of DCL4/DRB4-dependent sRNAs such as *TAS1* or *TAS3* ta-siRNAs and miR822 (Figure 2A), indicating that DRB4 is not a limiting factor for optimal DCL4 activity on those precursors.

### DRB4 forms specific and mutually exclusive complexes with DRB7.2 or DCL4

Based on the concerted evolution of DCL4 and DRB7.2 dsRBM, we previously tested a potential interaction between these two proteins by bimolecular fluorescence complementation. In these experiments, we observed that DRB7.2 interacts strongly with DRB4 but not with DCL4, whereas DRB4 was found to interact with both DRB7.2 and DCL4 (51). These observations, together with the change in endoIR-siRNA production, but not of other DCL4/DRB4-dependent sRNAs, found in *drb7.2* mutant plants (Figure 2A), suggested that DRB4 might be partitioned into two distinct cellular pools, one specifically interacting with DCL4 and the other with DRB7.2, where it fulfills other/specific function(s). However, an equally plausible explanation was that DCL4 specifically requires DRB7.2 as a cofactor, in addition to DRB4, for efficient processing of IR transcripts. Loss of either DRBs would then alter DCL4 activity on IR precursors and concurrently stimulate DCL3-mediated production of the 24 nt endoIR-siRNAs. This latter possibility entails that DRB7.2 interacts indirectly with DCL4, most likely through its association with DRB4, a possibility that was not previously addressed. Indeed, the bimolecular fluorescence complementation strategy is not the best suited approach to detect in-

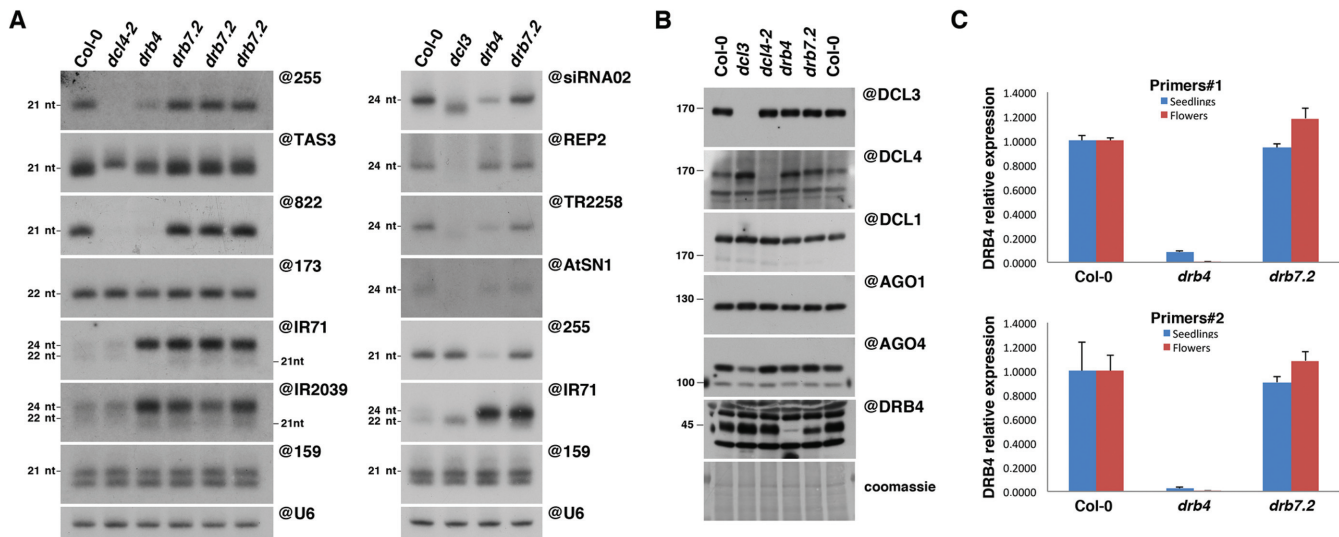


**Figure 1.** *drb7.2* mutant plants display altered accumulation of endoIR-siRNA. Wild-type (Col-0) and *drb7.2* mutant plants were subjected to high throughput sequencing. (A) Size distribution of small RNA reads that perfectly match the Arabidopsis nuclear genome, excluding rRNA and tRNA. The proportion of each size of small RNA is indicated by a color code: 21 nt (blue), 22 nt (green), 23 nt (pink), 24 nt (red) and grey for 18 to 20 nt and 25 to 26 nt. (B) Normalized small RNA abundance of the three major classes of endogenous siRNAs. Small RNAs from Col-0 and *drb7.2* mutant plants were classified as ta-siRNAs (upper panel), p4-siRNAs (middle panel) or endoIR-siRNAs (lower panel) based on published annotation and their abundance was normalized to the total amount of conserved miRNAs. The color code of small RNA size is the same as in (A). (C) Normalized size distribution of endoIR-siRNAs from three IR loci (IR71, IR2039 and AT5G22960). Size distribution of TAS3-derived ta-siRNAs is shown here as a control. Numbers of normalized small RNA reads that perfectly match the aforementioned loci are represented under bracket. The color code of small RNA size is the same as in (A). Graphs in (B) were obtained using the mean value of the two biological replicates and error bars indicate the variation observed between the two replicates.

direct protein–protein interactions as it requires very close proximity of the N- and C-terminal YFP moieties fused to the proteins tested, in order to reconstitute a functional YFP and fluorescent signal.

Therefore, in order to discriminate between these two possibilities we decided to perform co-immunoprecipitation (co-IP) experiments. For that purpose, we first generated transgenic *Arabidopsis* lines expressing functional,

epitope-tagged versions of DRB7.2 or DCL4. Transgenes expressing, under the control of the 35S promoter or its own promoter, the genomic sequence of DRB7.2 fused in C-terminal to GFP (35S:DRB7.2:GFP/*drb7.2* or *pDRB7.2:DRB7.2:GFP/dr7.2*) rescued the *drb7.2* mutation and restored IR-siRNA production back to WT levels (Figures 3A and 6E). Similarly, a transgene expressing, under its own promoter, the genomic sequence of



**Figure 2.** Effect of the *drb7.2* mutation on endogenous small RNAs or RNA silencing factor accumulation. (A) Northern blot analysis of trans-acting siRNA (@255, @TAS3), DCL4-dependent miRNA (@822), DCL1-dependent miRNA (@173, @159), endoIR-derived siRNA (@IR71, @IR2039) and DCL3-dependent p4-siRNA (@siRNA02, @REP2, @TR2558, @AtSN1) accumulation in wild-type (Col-0), *dcl4-2*, *drb4*, *drb7.2* or *dcl3* mutant plants was performed by sequential rounds of probing and stripping the same membranes. Please note that the three lanes labelled as *drb7.2* on the left blots correspond to three independent biological replicates. Accumulation of small RNA U6 (@U6) is used as loading control. (B) Accumulation of endogenous DCL3, DCL4, DCL1, AGO1, AGO4 and DRB4 was assessed by protein blot analysis of wild type, *dcl4-2*, *dcl3*, *drb4* and *drb7.2* plants. Equal loading was verified by Coomassie staining of the membrane after Western blotting. (C) Quantitative real-time PCR of the DRB4 mRNA accumulation in wild type, *drb4* and *drb7.2* plants. The mRNA level was normalized to that of *Actin2* (At3g18780) and then to the WT plants that was arbitrarily set to 1. Error bars represent standard deviation from two independent experiments in which triplicate PCRs were performed. Figure source data can be found with the Supplementary Data.

DCL4 fused in C-terminal to two Flag and two HA tags (pDCL4:*DCL4:FHA/dcl4*) complemented the phenotypic and molecular defects of the *dcl4* mutation (Figure 3A). Importantly, in these plants, DRB4 was efficiently immunoprecipitated with both DRB7.2:GFP and DCL4:FHA (Figure 3B and C). Collectively, these results indicate that both fusion proteins retain their biological functions and could, therefore, be used for co-IP experiments.

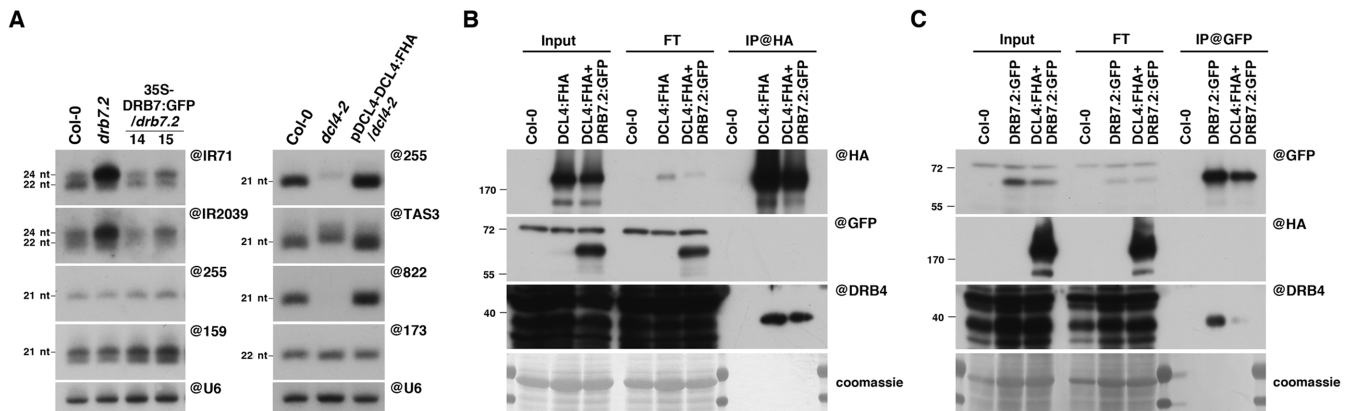
We, thus, generated transgenic plants expressing both DRB7.2:GFP and DCL4:FHA (*DRB7.2:GFP+DCL4:FHA/dcl4*). For that purpose, we introduced into our pDCL4:*DCL4:FHA/dcl4* transgenic line the DRB7.2:GFP transgene by agrotransformation and selected transformants that accumulate similar amount of DRB7.2:GFP than the one found in our transgenic line only expressing the DRB7.2:GFP transgene. Despite efficient immunoprecipitation of either recombinant protein, no DRB7.2:GFP signal could be detected in the IP fraction of DCL4:FHA (Figure 3B) and, conversely, no DCL4:FHA signal could be detected in the IP fraction of DRB7.2:GFP (Figure 3C), indicating that DRB7.2 and DCL4 are not able to interact *in planta*, even in the presence of DRB4. These results support the existence of specific and mutually exclusive DRB4-containing complexes that differ by their association with either DRB7.2 or DCL4. Further support to this hypothesis was provided when we compared the amount of DRB4 immunoprecipitated by DRB7.2:GFP, in the presence or absence of DCL4:FHA. Despite similar levels of immunoprecipitated DRB7.2:GFP, a strong reduction of the DRB4 signal was observed in the co-IP fraction of plants co-expressing

DCL4:FHA (Figure 3C), suggesting that DCL4 and DRB7.2 compete for interaction with DRB4 and that DRB4 has a higher affinity for DCL4 than for DRB7.2. In line with this assumption, similar levels of DRB4 were detected in IP fractions of DCL4:FHA, in presence or absence of DRB7.2:GFP (Figure 3B), further indicating that DCL4 outcompetes DRB7.2 for the interaction with DRB4. Finally, IP fractions of DRB7.2:GFP were also sent to mass spectrometry analysis in order to identify the protein partners of DRB7.2. DRB4 was found to be the top scoring protein with very high coverage (Supplementary Table S2) and spectral count whereas no peptide corresponding to DCL4 (or of any of the other DCLs) was retrieved (Montavon and Dunoyer, unpublished observations). Collectively, these findings indicate that DRB7.2 interacts with DRB4 in a complex that does not contain DCL4 and strongly suggest the existence of, at least, two distinct cellular pools of DRB4.

### The DRB4/DRB7.2 complex represses IR processing by the siRNA-generating DCLs

The strong and specific increase of DCL3-dependent 24 nt endoIR-siRNAs levels observed in both *drb4* and *drb7.2* mutants (Figure 2A) suggests that the DRB4/DRB7.2 complex may act as a specific repressor of DCL3 processing on IR substrates.

To address this genetically, we compared the accumulation of endoIR-siRNAs in mutants carrying genetic lesions in each of the three siRNA-generating DCL genes (*DCL2-4*), or combinations thereof, with the one found in *drb7.2* single mutant, or *drb7.2/dcl2*, *drb7.2/dcl3* and *drb7.2/dcl4*



**Figure 3.** DRB4 interacts in a mutually exclusive manner with DCL4 and DRB7.2. (A) Northern analysis of endoIR-siRNA (@IR71, @IR2039), trans-acting siRNA (@255, @TAS3), DCL4-dependent miRNA (@822) and DCL1-dependent miRNA (@173, @159) accumulation in wild-type, *dcl4-2*, *drb7.2*, DRB7.2:GFP or DCL4:FHA-expressing plants was performed by sequential rounds of probing and stripping the same membranes. Accumulation of small RNA U6 (@U6) is used as loading control. (B) Immunoprecipitation experiments were conducted in wild-type, DCL4:FHA or DRB7.2:GFP+DCL4:FHA-expressing plant using a HA-specific antibody. Co-IP of DRB7.2:GFP and DRB4 was examined by Western blot analyses using appropriate antibodies. Coomassie staining of the same membrane was used as a loading control. (C) Immunoprecipitation experiments were conducted in wild-type, DRB7.2:GFP or DRB7.2:GFP+DCL4:FHA-expressing plant using a GFP-specific antibody. Co-IP of DCL4:FHA and DRB4 was examined by Western blot analyses using appropriate antibodies. Coomassie staining of the same membrane was used as a loading control. Figure source data can be found with the Supplementary Data.

double mutants (Figure 4A). A similar comparison was also performed using *drb4* single or *drb4/dcl2*, *drb4/dcl3* and *drb4/dcl4* double mutants (Figure 4B). Northern analysis of endogenous IR-siRNAs in these various mutant backgrounds revealed that the amount of 24 nt IR-siRNA produced in *dcl2* (Figure 4A and B) or *dcl2/4* mutants (Figure 4C), where DCL3 is, respectively, the prevalent or the only remaining DCL that processes endogenous IR, were similar to the one found in WT control plants. By contrast, loss-of-function mutation in DRB4 or DRB7.2, either alone or in combination with the *dcl2* or *dcl4* mutations, triggers a huge increase in DCL3-dependent 24 nt IR-siRNA production (Figure 4A and B). Moreover, no additive effect on the accumulation of those sRNAs was observed in the *drb4/dr7.2* double mutant (Figure 4D), supporting the hypothesis that DRB4 and DRB7.2 act together, as part of the same complex, to repress DCL3 processing.

Most likely, this effect does not entail direct interaction of the DRB4/DRB7.2 complex with DCL3 given that DRB4 does not interact *in vivo* with DCL3 (47), and that no peptides corresponding to DCL3 were retrieved in the mass spectrometry analysis of DRB7.2 immunoprecipitates (Montavon and Dunoyer, unpublished observations). Yet, when either component of the DRB4/DRB7.2 complex is absent, DCL3 becomes, somehow, much more efficient in processing those long dsRNA precursors and outcompetes DCL2/DCL4 for endogenous IR-siRNA production (Figures 2A and 4).

Of note, this inhibitory effect only occurs on endoIR-siRNA production, as DCL3-dependent p4-siRNA were unaffected in *drb7.2* or *drb7.2/dcl2* mutant backgrounds (Figures 2A and 4A). This comparison cannot be made for all p4-siRNAs with plants containing the *drb4* and/or *dcl4* mutation as it was shown previously (47), and confirmed in this study (Figure 4), that these mutants exhibit, for currently unknown reason, reduced accumulation of some of

those sRNAs. However, similar conclusions can be reached for REP2-derived p4-siRNAs that are not affected by the *dcl4* or *drb4* mutations, confirming the specificity of the inhibitory effect mediated by the DRB4/DRB7.2 complex.

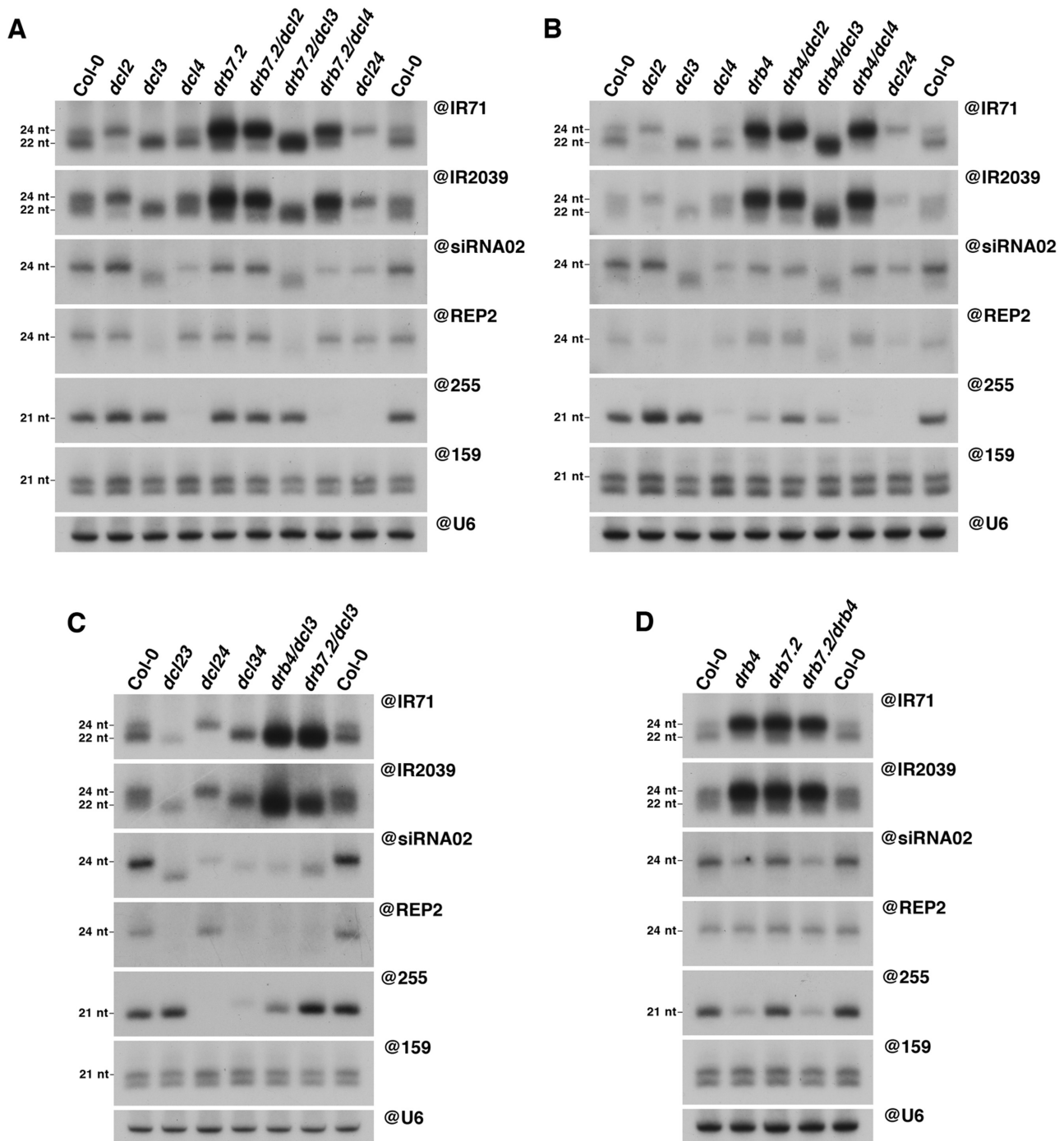
Interestingly, we also noticed a strong increase of the 22 nt endoIR-siRNA accumulation in *drb7.2/dcl3* and *drb4/dcl3* double mutants compared to the one found in *dcl3* or *dcl34* mutants where DCL2 is, respectively, the prevalent or the only remaining DCL that processes endogenous IR (Figure 4A–C). These results indicate that the DRB4/DRB7.2 complex is also acting as a repressor of DCL2 processing on IR substrates, when DCL3 is absent.

In order to assess the effect of DRB7.2 on exogenous siRNA accumulation, wild-type Col-0 and *drb7.2* mutant plants were also infected with either *Turnip crinkle virus* or *Tobacco rattle virus*. Northern analysis revealed that neither siRNAs production nor viral RNA accumulation were affected by the *drb7.2* mutation (Supplementary Figure S2), indicating that DRB7.2 is not involved in the plant antiviral RNA silencing response.

### The DRB4/DRB7.2 complex represses IR-siRNA production by specifically sequestering their long dsRNA precursors

Next, we sought to determine how the DRB4/DRB7.2 complex represses DCL3 and DCL2 processing on IR substrates. A direct interaction of this complex with DCL3 and DCL2 is rather unlikely based on the mass spectrometry analyses of DRB7.2 immunoprecipitated fraction, and on the documented lack of interaction between DRB4 and DCL3 (47). Therefore, based on their intrinsic property to bind dsRNA, we rather reasoned that the DRB4/DRB7.2 complex might directly sequester specific dsRNA precursors, thereby preventing their access and processing by DCL3 and/or DCL2.

To test this hypothesis, we performed RNA immunoprecipitation (RIP) experiments using *drb7.2* transgenic lines



**Figure 4.** The DRB4/DRB7.2 complex specifically represses DCL3 and DCL2-dependent production of endoIR-siRNAs. (A) RNA gel blot analysis of endoIR-siRNA (@IR71, @IR2039), p4-siRNA (@REP2, @siRNA02) and trans-acting siRNA (@255) accumulation in wild-type, *dcl2*, *dcl3*, *dcl4*, *drb7.2*, *drb7.2/dcl2*, *drb7.2/dcl3*, *drb7.2/dcl4* and *dcl2/4* mutant plants was performed by sequential rounds of probing and stripping the same membranes. (B) Accumulation of the same small RNAs depicted in (A) was assessed in wild-type, *dcl2*, *dcl3*, *dcl4*, *drb4*, *drb4/dcl2*, *drb4/dcl3*, *drb4/dcl4* and *dcl2/4* mutant plants. (C) Similar analysis to the one depicted in (A) was performed in wild-type, *dcl2/3*, *dcl2/4*, *dcl3/4*, *drb4/dcl3* and *drb7.2/dcl3* plants. (D) Similar analysis to the one depicted in (A) was performed in wild-type, *drb4*, *drb7.2* and *drb7.2/drab4* plants. Accumulation of the DCL1-dependent miR159 (@159) and small RNA U6 (@U6) were used as loading control. Figure source data can be found with the Supplementary Data.

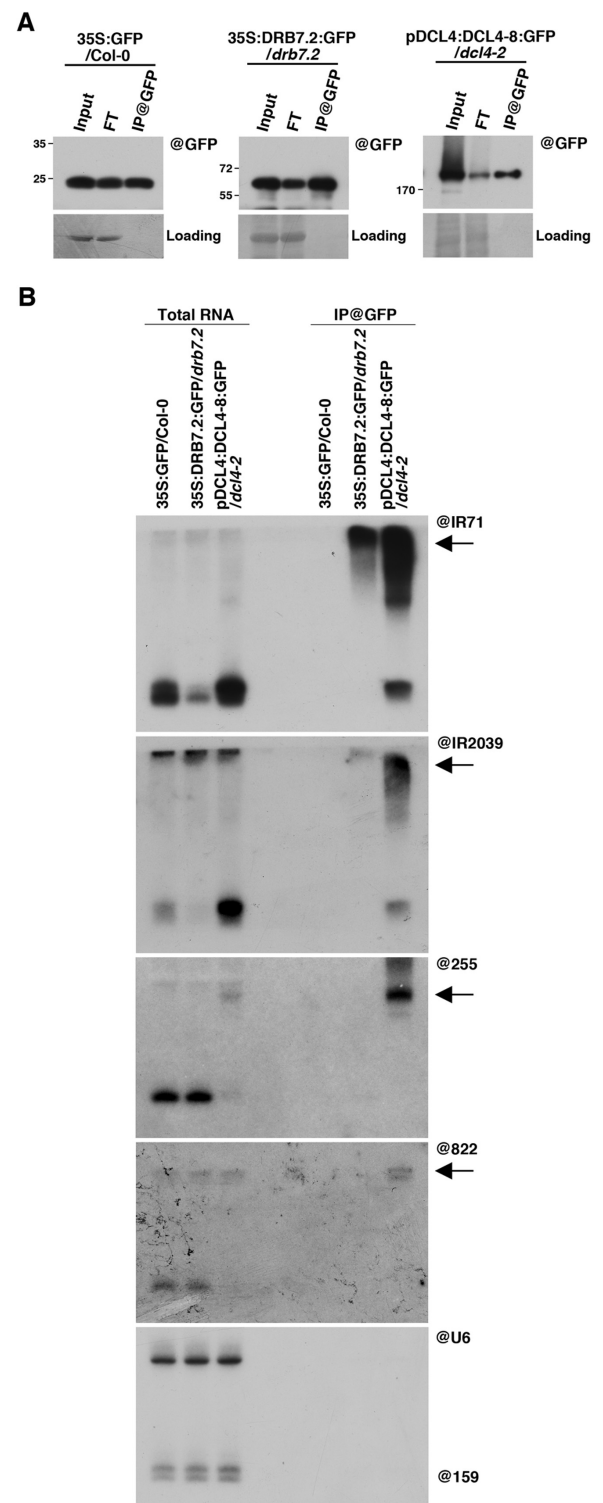
expressing DRB7.2:GFP (Figure 3A). As a positive control for these RIP experiments, we decided to use a *dcl4* transgenic line expressing a mutated version of DCL4, DCL4-8, fused in C-terminal to GFP (pDCL4:DCL4-8:GFP/*dcl4*). This allele carries a G to A transition within the DEAD helicase domain of DCL4 (leading to G610D mutation; (50)) that impairs production of all the DCL4-dependent sRNAs, despite producing a stable protein (Supplementary Figure S3). Importantly, this mutant allele is still able to bind and stabilize the DCL4 dsRNA substrates such as *TASI*, miR822 or endoIR-siRNA precursors (Figure 5). In parallel, 35S-GFP expressing plants were used as negative control.

Immunoprecipitation of DRB7.2:GFP followed by Northern blot analysis allowed us to detect a specific signal corresponding to IR71 and IR2039 RNA precursors in the IP fraction, indicating that DRB7.2 is indeed able to bind endoIR-siRNAs precursors (Figure 5). Importantly, and by contrast with DCL4-8, no signal corresponding to the precursors of *TASI* ta-siRNA or miRNA822 were found associated with DRB7.2 (Figure 5), in agreement with the unaltered accumulation of their associated sRNAs in the *drb7.2* mutant. In parallel, we also noticed that the signal detected for endoIR precursors in the DCL4-8:GFP IP fraction was more heterogeneous in size than the one detected in the DRB7.2:GFP IP fraction (Figure 5). Given that (i) endo-IR precursors are processed by the three siRNA-generating DCLs, chiefly DCL2 and DCL3, and (ii) that this heterogeneity was not observed for the other precursors bound by DCL4-8, whose processing is purely DCL4-dependent (such as *TASI* ta-siRNA and miRNA822 precursors, Figure 5), this most likely reflects binding by DCL4-8 to IR dsRNAs that have been already partially processed by the two other siRNA-generating DCLs. Finally, neither DCL1-dependent miRNA precursors nor DCL3-dependent p4-siRNA precursors could be detected in the IP fraction of DRB7.2, indicating that this protein specifically binds precursors of endoIR-siRNAs but not of other sRNAs (Figure 5 and Supplementary Figure S4). Therefore, the strict correlation between the altered accumulation of endoIR-siRNAs observed in the *drb7.2* mutant and the specificity of the precursors bound by DRB7.2, strongly support our hypothesis that DRB7.2 selectively sequesters endoIR precursors.

Moreover, band-shift experiments (Supplementary Figure S5A), as well as RIP experiments followed by RNase A/T1 treatment, that selectively cleave single-stranded RNAs (ssRNAs) but do not digest dsRNAs (Supplementary Figure S5B), revealed that DRB7.2 has a higher affinity for the latter *in vitro* and binds them *in vivo*. This supports that the endoIR-siRNA precursors bound by DRB7.2 are sequestered under their dsRNA form.

### DCL4 and DRB7.2 compete for DRB4 binding

The above results also suggest that DRB4 is partitioned between DCL4, where it acts as a cofactor for efficient production of DCL4-dependent sRNAs, and DRB7.2, where it functions to specifically sequester endogenous IR precursors and inhibit DCL3 and DCL2 processing of these substrates. Therefore, we reasoned that expressing a higher



**Figure 5.** DRB7.2 specifically binds endoIR-siRNAs precursors. (A) Immunoprecipitation experiments were conducted in GFP-, DRB7.2:GFP- or DCL4-8:GFP-expressing plants using a GFP-specific antibody. The presence of the corresponding proteins in each IP was confirmed by protein blot analysis. (B) Total RNA extracted from the respective IPs was subjected to northern analysis using the indicated probes by sequential rounds of probing and stripping the same membrane. The signal corresponding to the sRNA precursors is indicated by an arrow. @U6: loading control. Figure source data can be found with the Supplementary Data.

amount than the endogenous level of one of the two DRB4 partners should result in displacing the equilibrium toward one or the other complex.

In agreement with this hypothesis, Northern analysis of endogenous IR-siRNA accumulation in *pDCL4:DCL4-8:GFP/dcl4* transgenic plants revealed a similar pattern to the one observed in the *drb7.2* mutant background, with a strong increase in the DCL3-dependent 24 nt siRNA species (Figures 5B and 6A). Using our DCL4 antibody, we confirmed by Western blot analysis that the amount of DCL4-8:GFP accumulating in this transgenic line was higher than the amount of endogenous DCL4 present in wild-type Col-0 plants (Figure 6B). To discard the possibility that this observation was, somehow, linked to the mutation carried by the DCL4-8 transgene, we repeated this analysis in transgenic plants expressing a wild-type functional version of DCL4 (*pDCL4:DCL4WT:GFP/dcl4*) that rescued the *dcl4* mutation and restored production of DCL4-dependent ta-siRNAs and miRNAs back to WT levels (Figure 6A). Similar results were obtained in these transgenic plants, where higher levels of DCL4 than those found in wild-type Col-0 trigger an endoIR-siRNA accumulation pattern analogous to the one observed in *drb7.2* mutant (Figure 6A and B). This effect was further confirmed in a third, independent, transgenic line expressing, under the 35S promoter, a genomic copy of DCL4 (35S:DCL4) (Figure 6C and D). Collectively, these results suggest that, by expressing DCL4 at a higher level, the DRB4 equilibrium is displaced toward an interaction with this protein, to the detriment of the DRB4/DRB7.2 complex formation. This results in a defect of IR dsRNA precursors sequestration leading to an increase in DCL3-dependent endoIR-siRNA production, similar to the one observed in *drb4* or *drb7.2* mutant plants. In agreement with this hypothesis, this altered endoIR-siRNA accumulation was not observed in *dcl4-8* mutant plants, where DCL4 accumulates to similar level as in wild-type plants (Supplementary Figure S3).

Conversely, expressing increasing amount of DRB7.2 should favor the formation of the DRB4/DRB7.2 complex. This should, in turn, promote the sequestration of more IR dsRNA precursors and translate into less endoIR-siRNAs produced. In agreement with this hypothesis, Northern analysis revealed that the accumulation of both DCL3-dependent 24 nt and DCL2-dependent 22 nt endoIR-siRNAs was indeed reduced in *35S:DRB7.2:GFP/drb7.2* transgenic lines compared to the level detected in *pDRB7.2:DRB7.2:GFP/drb7.2* (where, comparatively, DRB7.2:GFP accumulates to much lower level) or Col-0 plants (Figure 6E and F). As expected, based on the apparent higher affinity of DCL4, over DRB7.2, for DRB4 interaction (Figure 3), and on the observation that optimal DCL4 activity can still be obtained, even when lower amounts of DRB4 are available (Figure 2), no obvious effect on the accumulation of the purely DCL4/DRB4-dependent sRNAs could be detected in these transgenic lines (Figure 6E).

Collectively, the above results further substantiate the existence of two distinct pools of DRB4 that differ by their interaction with either DCL4 or DRB7.2.

### Partitioning of DRB4 between DCL4 and DRB7.2 occurs in distinct subcellular compartments

Given that DRB4 was previously shown to accumulate in both nucleus and cytoplasm (48), we next sought to address if the partitioning of DRB4 between DRB7.2 and DCL4 occurred in the same, or distinct, subcellular compartment(s). For this purpose, we analyzed by confocal microscopy the subcellular localization of DRB7.2 and DCL4, in transgenic plants expressing functional GFP fusion of either protein.

DRB7.2 localization was found to be purely nuclear in both *35S:DRB7.2:GFP/drb7.2* and *pDRB7.2:DRB7.2:GFP/drb7.2* transgenic lines. As shown in Figure 7A–D, DRB7.2 is uniformly distributed throughout the nucleoplasm and found enriched in discrete nuclear foci that may correspond to previously described ‘siRNA-processing centers’ or ‘Dicing bodies’ (72,73). Strikingly, and in sharp contrast to the DRB7.2 nuclear localization, DCL4 was found to be mostly cytoplasmic in *pDCL4:DCL4WT:GFP/dcl4* transgenic plants (Figure 7E and F). In addition, and as previously described when expressed under the control of its native endogenous promoter (74,75), DCL4 was, on occasions, also observed in the nucleus, almost exclusively in discrete nuclear foci. A similar localization pattern was also observed for DCL4-8, indicating that this mutation does not affect DCL4 localization (Figure 7G and H).

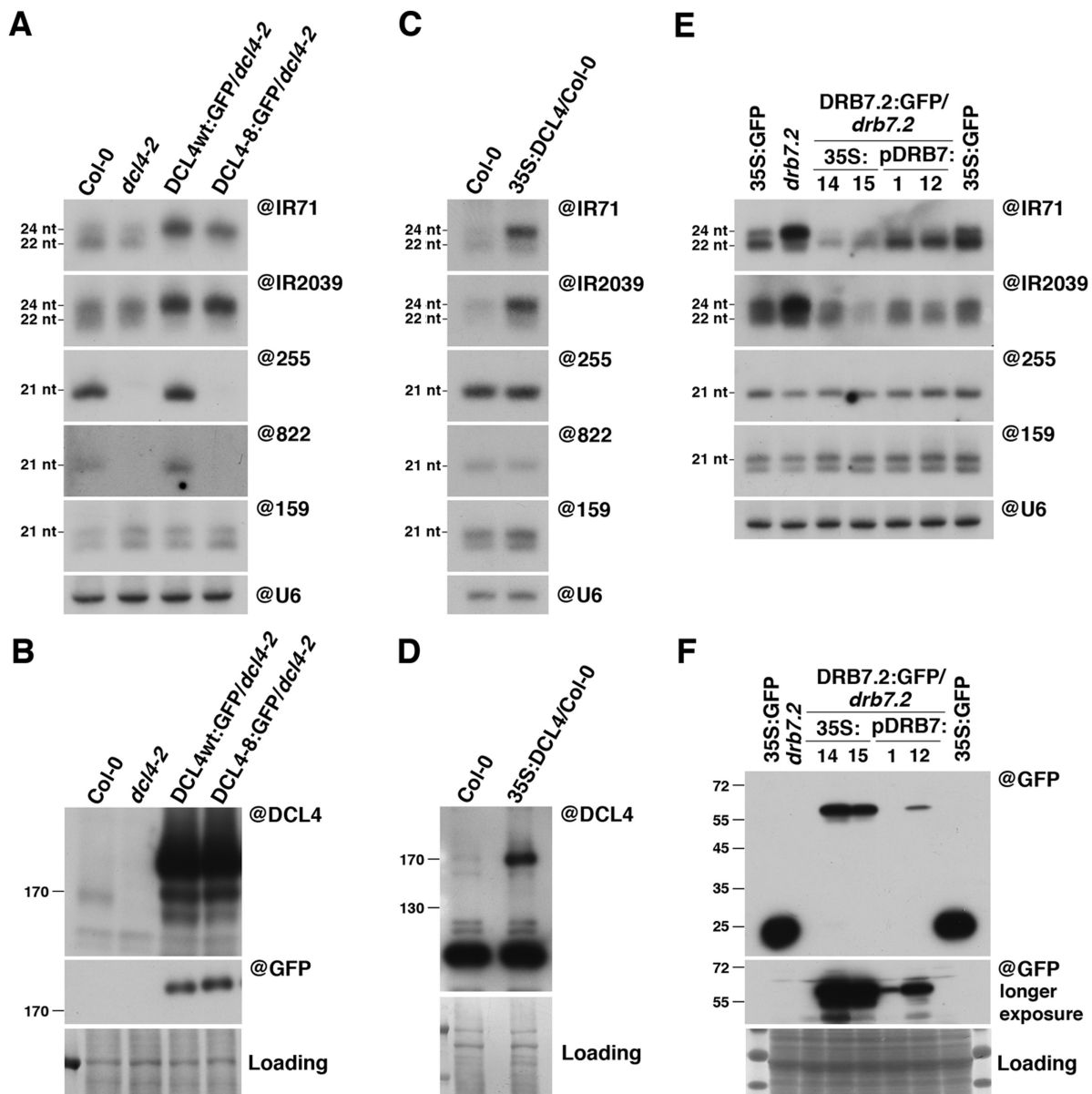
Collectively, these observations indicate that DCL4 and DRB7.2 are mostly located in distinct subcellular compartments and strongly suggest that DRB4 is partitioned between the nucleus and the cytoplasm to form specific complexes. Of note, the DCL4 localization observed here differs from the purely nuclear localization reported earlier, where DCL4 was expressed under the control of strong constitutive promoter (32,76), suggesting that genetic information regulating DCL4 localization might be embedded within its native promoter sequence and emphasizing the importance of using an endogenous promoter to accurately assess protein localization. In line with these assumptions, it has indeed recently been shown that DCL4 endogenous promoter exhibits a discrete methylation patch that influences the transcriptional start site of this gene (77). In a wild-type situation, this methylated promoter leads to the production of the cytoplasmic DCL4 observed here (Figure 7E and F), whereas removal of this methylation patch (e.g. in mutants of the RNA-directed DNA methylation pathway, or through the use of an alternative promoter) leads to the production of a DCL4 isoform extended by a 61 amino acids, containing a nuclear localization signal (77).

### DISCUSSION

This study not only unravels the function of DRB7.2, a new player in plant RNA silencing pathways, it also provides additional information regarding DRB4, one of the best-characterized plant DRB, by uncovering a new cellular function of this protein.

DRB4 was previously seen as a mere cofactor of DCL4, mainly required to promote efficient production of various DCL4-dependent sRNAs, whereas DRB7.2, one of the two members of a newly identified DRB family conserved

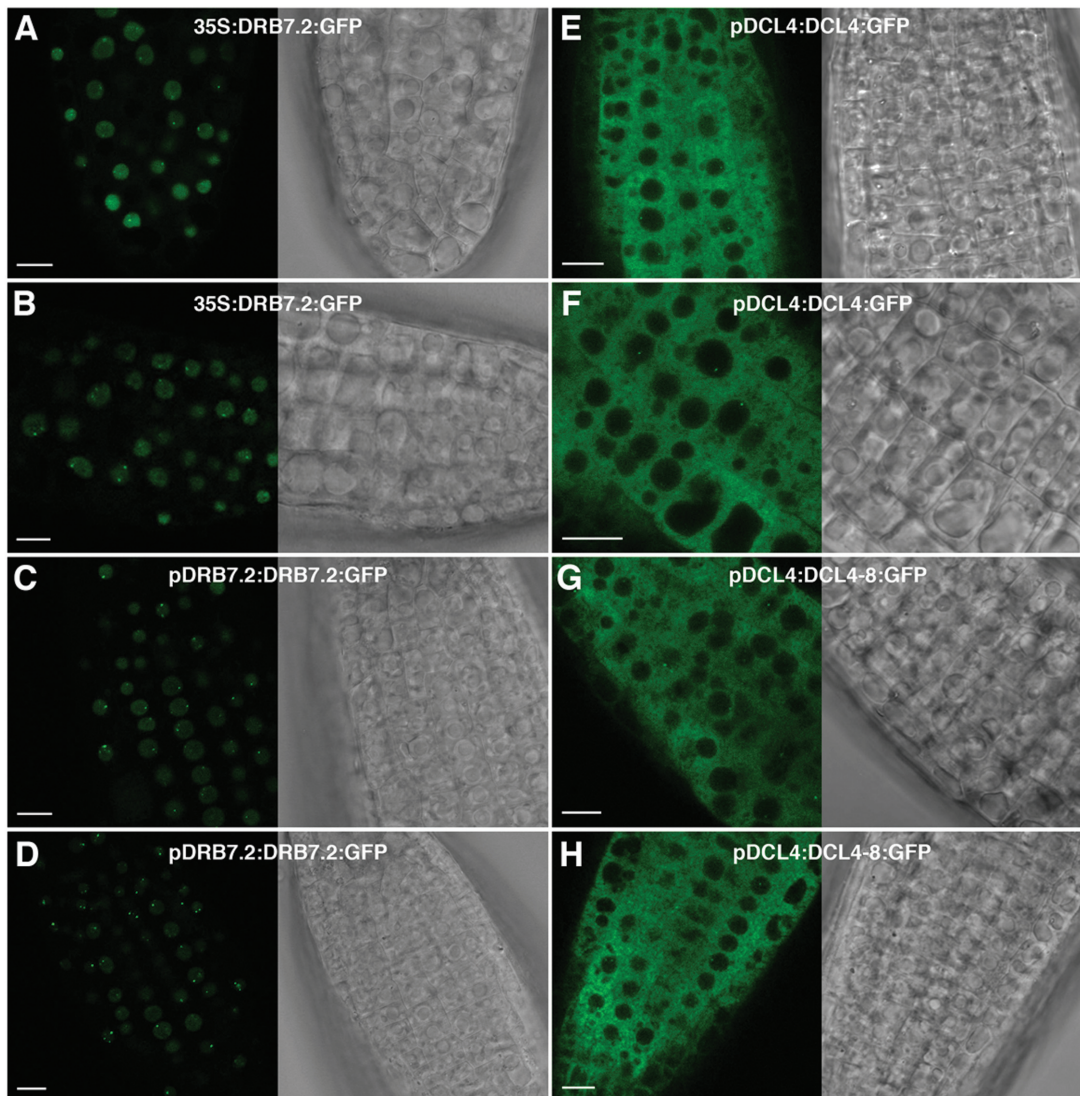




**Figure 6.** Effect of DCL4 or DRB7.2 overaccumulation on endogenous small RNA levels. (A) Northern analysis of endoIR-derived siRNA (@IR71, @IR2039), trans acting siRNA (@255) and DCL4-dependent miRNA (@822) accumulation in wild-type, *dcl4-2* and transgenic plants expressing DCL4wt:GFP or DCL4-8:GFP was performed by sequential rounds of probing and stripping the same membranes. (B) DCL4 accumulation was visualized by Western blotting using a DCL4- or GFP-specific antibody. (C) RNA gel analysis in wild-type or 35S:DCL4 transgenic plants using the same probes as in (A). (D) Over accumulation of the DCL4 protein was confirmed by protein blot analysis using a DCL4-specific antibody. (E) Accumulation of endo-IR siRNA and trans-acting siRNA was assessed by Northern blot in 35S:GFP-, 35S:DRB7.2:GFP- (two independent transgenic lines, #14 and #15) or pDRB7.2:DRB7.2:GFP- (two independent transgenic lines, #1 and #12) expressing plants by sequential rounds of probing and stripping the same membranes. (F) Accumulation of GFP or DRB7.2:GFP protein was verified by Western blot. For RNA blot, miR159 and U6 were used as loading control. For protein blot, Coomassie staining of membranes were used to verify equal loading. Figure source data can be found with the Supplementary Data.

in all vascular plants, had no clear role in RNA silencing. Here, we show that these two DRBs play together an important role in endoIR-siRNA production by negatively regulating the processing of their dsRNA precursors. All together, the findings that (i) DRB4 and DRB7.2 physically associate in a complex that does not contain DCL4 (Figure 3), (ii) that both *drb4* and/or *drb7.2* mutants display the same change in the accumulation pattern of endoIR-siRNA (Figure 4), and (iii) that DRB7.2 specifically binds

the precursors of endoIR-siRNA but not of other sRNAs (Figure 5), indicate that this regulation is achieved through the formation of a DRB4/DRB7.2 complex that selectively binds to, and most likely sequesters, endoIR-siRNA precursors, thereby preventing their access and processing by the siRNA-generating DCLs, chiefly DCL3. Although not assessed in the present study, the increased accumulation of 24 nt easiRNA species previously observed in the *drb7.2/dcl3* double mutant (51) may also suggest that, at



**Figure 7.** DRB7.2 and DCL4 localize in distinct subcellular compartments. Confocal microscope images of GFP in root tips of (A and B) 35S-DRB7.2:GFP-, (C and D) pDRB7.2-DRB7.2:GFP-, (E and F) pDCL4-DCL4:GFP- or (G and H) pDCL4-DCL4-8:GFP- transgenic lines. Scale bar = 50  $\mu$ m.

least, some of the easiRNA precursors can be sequestered by the DRB4/DRB7.2 complex.

While the reason(s) for the selectivity of the DRB4/DRB7.2 complex is still unknown, differences in terms of dsRNA size, structure and/or subcellular localization can potentially explain its specific effect. For instance, although both are produced in the nucleus, endoIR-siRNA precursors are long perfect or near perfect, dsRNA molecules whereas p4-siRNAs are generated from short dsRNA precursors, 27–50 nt in length. Interestingly, band-shift assays, performed with purified DRB7.2, showed that this protein displays stronger affinity for long (>150 bp) dsRNA as compared to smaller species (<80 bp) (Supplementary Figure S6). This observation may, therefore, explain the lack of binding of p4-siRNA precursors by DRB7.2 (Supplementary Figure S4) and, consequently, the unaltered accumulation of p4-siRNAs in the *drb7.2* mutant (Figure 2). A similar rationale may also explain the

lack of DRB7.2 binding to miRNA precursors, which have short imperfect stem-loop structures. Moreover, miRNA precursors contain several mismatches and bulges that may also affect the affinity of DRB7.2 for a given dsRNA. Alternatively, DRB1 affinity for miRNA precursors may outcompete DRB7.2 for their binding. Finally, precursors of ta-siRNAs are long perfect dsRNA, structurally similar in essence to those of endoIR-siRNA. However, *TAS* dsRNA precursors being, most likely, generated in cytoplasmic foci containing the different key factors required for their synthesis, such as SGS3 and RDR6 (76,78), this localization may preclude a potential binding by DRB7.2, which is exclusively nuclear (Figure 6).

In line with this latter observation, the findings that DRB7.2 and DCL4 (i) are localized in different subcellular compartments (Figure 7), (ii) associate with DRB4 in a mutually exclusive manner (Figure 3) and (iii), when over-expressed, compete for DRB4 binding (Figure 6), strongly

support the existence of, at least, two distinct cellular pools of DRB4. These two pools fulfil different and specific functions, e.g. promoting DCL4-dependent sRNA production in the cytoplasm and repressing endoIR-siRNA production in the nucleus, which agrees with the documented nucleocytoplasmic localization of DRB4 (48). In addition, the recent report that DRB4 seems to be required for the formation of the discrete nuclear foci of DCL4 (77) may either suggest the existence of a third cellular pool of DRB4 of currently unknown function, or that a small proportion of the DRB4 partitioning between DCL4 and DRB7.2 described here, may also occur in the nucleus.

These findings are reminiscent to the distinct cellular pools of AGO1 that were shown to specifically interact with either siRNA or miRNA and that were differentially affected by viral suppressors of RNA silencing (64). Although in the case of AGO1, the reason for this specificity is still unknown, these observations, together with the findings made in the present study, stress the importance of addressing the properties of a given RNA silencing factor with respect to its subcellular localization and/or complex composition in order to reveal specific and/or discrete cellular functions. In this respect, characterizing the function of the other DRB7.2-interacting partners, identified in our mass spectrometry analysis, may provide additional information or reveal new function for this protein. Moreover, this study, together with the recent role ascribed to non-DCL RNase three-like (RTL) proteins, RTL1 and RTL2, in modulation of sRNA production (79,80), shed light on new layers of regulation and the ever-growing complexity, of plant RNA silencing pathways. In connection with this, it will be interesting to assess the function and biological role of the other member of the DRB7 family (DRB7.1) that still awaits characterization. Based on its nuclear localization, it might be also important to assess if DRB7.2 affect somehow the accumulation of DNA viruses that replicate in this subcellular compartment.

Finally, our results also reveal that plants have evolved a specific DRB complex to modulate selectively the production of endoIR-siRNAs (Figures 1 and 2, Supplementary Table S2 and Supplementary Figure S7). In addition to illustrating how endogenous IR loci constitute useful molecular probes of the mechanisms of RNA silencing, the existence of such a complex put into question the function of those particular sRNA precursors, particularly in the light of the absence of any obvious developmental defects observed in *drb7.2* mutant plants. Indeed, so far, endogenous IRs with an extended fold-back structure have been mostly considered as relatively ill-defined, primary steps in the evolution of young MIRNA loci (68,81). They have also been regarded as having little, if any, regulatory potential of their own, notably because they were thought to be expressed at low or very low levels to avoid the off-targeting effects of their associated siRNA populations (2,81).

Although these assumptions might be true in some cases, small RNA deep-sequencing analysis revealed that at least some specific IR loci are in fact transcribed at high levels (Figure 1). The function of the DRB4/DRB7.2 complex could, therefore, be to dampen the production of endoIR-siRNAs from those particular loci in order to minimize potential off-targeting effects with deleterious consequences.

In addition, the DRB4/DRB7.2 complex could help prevent highly expressed endoIRs from titrating away all the DCL2 and DCL3 available in the cell, leaving these DCLs free to perform their conventional roles in siRNA biogenesis. Alternatively, and perhaps more appealingly, some of the sRNAs produced from those endogenous IR loci may have regulatory functions that might help plants to respond and cope more efficiently to changes in their environment. Any external stimuli, such as biotic or abiotic stresses, leading to a change in the accumulation, or the availability, of either components of the DRB4/DRB7.2 complex will then translate into the rapid release of the sequestered precursors and quick production of their associated sRNAs, allowing the plants to potentially respond more efficiently to the perceived stress. In that case, the role of the DRB4/DRB7.2 complex could then be seen as a safety reservoir for precursors of sRNAs involved in adapting sequence-specific plant responses to stress, which might be particularly important at specific developmental stages or in discrete cell types. This hypothesis will undoubtedly deserve in-depth investigation in the future.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank N. Pumplin for sharing his pDCL4:DCL4:GFP constructs before publication; R. Wagner's team for plant care; Baptiste Monsion and the P3P platform for help in protein production and members of the IBMP for critical reading of the manuscript.

## FUNDING

Agence Nationale pour la Recherche [ANR-14-CE19-0014-01 to P.D.]; LABEX NetRNA [ANR-10-LABX-0036\_NETRNA] from the French National Research Agency as part of the Investments for the future program; French Ministry of Research [T.M.]. Funding for open access charge: ANR [ANR-14-CE19-0014-01].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Baulcombe, D.C. (2004) RNA silencing in plants. *Nature*, **431**, 356–363.
2. Voinnet, O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669–687.
3. Chapman, E.J. and Carrington, J.C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.*, **8**, 884–896.
4. Allen, E., Xie, Z., Gustafson, A.M. and Carrington, J.C. (2005) microRNA-Directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.
5. Yoshikawa, M., Peragine, A., Park, M.Y. and Poethig, R.S. (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.*, **19**, 2164–2175.
6. Gascioli, V., Mallory, A.C., Bartel, D.P. and Vaucheret, H. (2005) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.*, **15**, 1494–1500.

7. Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Tang, H. and Pikaard, C.S. (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *eLife*, **4**, e09591.
8. Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T.F., Teng, C., Chen, X., Park, S.Y., Liu, L., Gallego-Bartolome, J. *et al.* (2015) A one precursor one siRNA Model for Pol IV-dependent siRNA biogenesis. *Cell*, **163**, 445–455.
9. Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumble, J.S., Givan, S.A. and Carrington, J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol.*, **5**, e57.
10. Lindow, M., Jacobsen, A., Nygaard, S., Mang, Y. and Krogh, A. (2007) Intragenomic matching reveals a huge potential for miRNA-mediated regulation in plants. *PLoS Comput. Biol.*, **3**, e238.
11. Bologna, N.G. and Voinnet, O. (2014) The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Ann. Rev. Plant Biol.*, **65**, 473–503.
12. Baumberg, N. and Baulcombe, D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 11928–11933.
13. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L.E. and Voinnet, O. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, **320**, 1185–1190.
14. Mallory, A. and Vaucheret, H. (2010) Form, function, and regulation of ARGONAUTE proteins. *Plant Cell*, **22**, 3879–3889.
15. Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
16. Szittya, G., Moxon, S., Pantaleo, V., Toth, G., Rusholme Pilcher, R.L., Moulton, V., Burgyn, J. and Dalmay, T. (2010) Structural and functional analysis of viral siRNAs. *PLoS Pathog.*, **6**, e1000838.
17. Qi, X., Bao, F.S. and Xie, Z. (2009) Small RNA deep sequencing reveals role for *Arabidopsis thaliana* RNA-Dependent RNA Polymerases in Viral siRNA Biogenesis. *PLoS One*, **4**, e4971.
18. Donaire, L., Wang, Y., Gonzalez-Ibeas, D., Mayer, K.F., Aranda, M.A. and Llave, C. (2009) Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology*, **392**, 203–214.
19. Morel, J.B., Godon, C., Mourrain, P., Beclin, C., Boutet, S., Feuerbach, F., Proux, F. and Vaucheret, H. (2002) Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. *Plant Cell*, **14**, 629–639.
20. Qu, F., Ye, X. and Morris, T.J. (2008) Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14732–14737.
21. Jaubert, M., Bhattacharjee, S., Mello, A.F., Perry, K.L. and Moffett, P. (2011) ARGONAUTE2 mediates RNA-silencing antiviral defenses against Potato virus X in Arabidopsis. *Plant Physiol.*, **156**, 1556–1564.
22. Wang, X.-B., Jovel, J., Udamporn, P., Wang, Y., Wu, Q., Li, W.-X., Gascioli, V., Vaucheret, H. and Ding, S.-W. (2011) The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in *Arabidopsis thaliana*. *Plant Cell*, **23**, 1625–1638.
23. Harvey, J.J.W., Lewsey, M.G., Patel, K., Westwood, J.H., Heimstädt, S., Carr, J.P. and Baulcombe, D.C. (2011) An antiviral defense role of AGO2 in plants. *PLoS One*, **6**, e14639.
24. Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K.B., Montgomery, T.A., Nguyen, T., Cuperus, J.T. and Carrington, J.C. (2012) Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell*, **24**, 3613–3629.
25. Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P. and Wang, X. (2003) R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science*, **301**, 1921–1925.
26. Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D. and Liu, Q. (2005) Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev.*, **19**, 1674–1679.
27. Han, M.H., Goud, S., Song, L. and Fedoroff, N. (2004) The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 1093–1098.
28. Czech, B. and Hannon, G.J. (2011) Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.*, **12**, 19–31.
29. Vazquez, F., Gascioli, V., Crete, P. and Vaucheret, H. (2004) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.*, **14**, 346–351.
30. Eamens, A.L., Smith, N.A., Curtin, S.J., Wang, M.-B. and Waterhouse, P.M. (2009) The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA*, **15**, 2219–2235.
31. Manavella, P.A., Hagmann, J., Ott, F., Laubinger, S., Franz, M., Macek, B. and Weigel, D. (2012) Fast-Forward Genetics Identifies Plant CPL Phosphatases as Regulators of miRNA Processing Factor HYL1. *Cell*, **151**, 859–870.
32. Hiraguri, A., Itoh, R., Kondo, N., Nomura, Y., Aizawa, D., Murai, Y., Koiwa, H., Seki, M., Shinozaki, K. and Fukuhara, T. (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. *Plant Mol. Biol.*, **57**, 173–188.
33. Kurihara, Y., Takashi, Y. and Watanabe, Y. (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA*, **12**, 206–212.
34. Yang, S.W., Chen, H.-Y., Yang, J., Machida, S., Chua, N.-H. and Yuan, Y.A. (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and Its Molecular Implications for miRNA Processing. *Structure*, **18**, 594–605.
35. Qin, H., Chen, F., Huan, X., Machida, S., Song, J. and Yuan, Y.A. (2010) Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA*, **16**, 474–481.
36. Song, L., Han, M.-H., Lesicka, J. and Fedoroff, N. (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 5437–5442.
37. Yang, X., Ren, W., Zhao, Q., Zhang, P., Wu, F. and He, Y. (2014) Homodimerization of HYL1 ensures the correct selection of cleavage sites in primary miRNA. *Nucleic Acids Res.*, **42**, 12224–12236.
38. Eamens, A.L., Kim, K.W., Curtin, S.J. and Waterhouse, P.M. (2012) DRB2 is required for microRNA biogenesis in Arabidopsis thaliana. *PLoS One*, **7**, e35933.
39. Reis, R.S., Hart-Smith, G., Eamens, A.L., Wilkins, M.R. and Waterhouse, P.M. (2015) Gene regulation by translational inhibition is determined by Dicer partnering proteins. *Nat. Plants*, **1**, 14027.
40. Eamens, A.L., Wook Kim, K. and Waterhouse, P.M. (2012) DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in Arabidopsis thaliana. *Plant Sign. Behav.*, **7**, 1224–1229.
41. Curtin, S.J., Watson, J.M., Smith, N.A., Eamens, A.L., Blanchard, C.L. and Waterhouse, P.M. (2008) The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS Lett.*, **582**, 2753–2760.
42. Raja, P., Jackel, J.N., Li, S., Heard, I.M. and Bisaro, D.M. (2014) Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. *J. Virol.*, **88**, 2611–2622.
43. Fukudome, A., Kanaya, A., Egami, M., Nakazawa, Y., Hiraguri, A., Moriyama, H. and Fukuhara, T. (2011) Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of Arabidopsis Dicer-like 4. *RNA*, **17**, 750–760.
44. Nakazawa, Y., Hiraguri, A., Moriyama, H. and Fukuhara, T. (2007) The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. *Plant Mol. Biol.*, **63**, 777–785.
45. Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gascioli, V. and Vaucheret, H. (2006) DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr. Biol.*, **16**, 927–932.
46. Rajagopalan, R., Vaucheret, H., Trejo, J. and Bartel, D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.*, **20**, 3407–3425.
47. Pelissier, T., Clavel, M., Chaparro, C., Pouch-Pelissier, M.N., Vaucheret, H. and Deragon, J.M. (2011) Double-stranded RNA binding proteins DRB2 and DRB4 have an antagonistic impact on polymerase IV-dependent siRNA levels in Arabidopsis. *RNA*, **17**, 1502–1510.

48. Jakubiec, A., Yang, S.W. and Chua, N.H. (2012) Arabidopsis DRB4 protein in antiviral defense against Turnip yellow mosaic virus infection. *Plant J.*, **69**, 14–25.
49. Zhu, S., Jeong, R.D., Lim, G.H., Yu, K., Wang, C., Chandra-Shekara, A.C., Navarre, D., Klessig, D.F., Kachroo, A. and Kachroo, P. (2013) Double-stranded RNA-binding protein 4 is required for resistance signaling against viral and bacterial pathogens. *Cell Rep.*, **4**, 1168–1184.
50. Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A. and Voinnet, O. (2007) Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.*, **39**, 848–856.
51. Clavel, M., Pelissier, T., Montavon, T., Tschopp, M.A., Pouch-Pelissier, M.N., Descamin, J., Jean, V., Dunoyer, P., Bousquet-Antonelli, C. and Deragon, J.M. (2016) Evolutionary history of double-stranded RNA binding proteins in plants: identification of new cofactors involved in easiRNA biogenesis. *Plant Mol. Biol.*, **91**, 131–147.
52. Slotkin, R.K., Vaughn, M.W., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A. and Martienssen, R.A. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, **136**, 461–472.
53. McCue, A.D., Nuthikattu, S. and Slotkin, R.K. (2013) Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. *RNA Biol.*, **10**, 1379–1395.
54. McCue, A.D., Panda, K., Nuthikattu, S., Choudury, S.G., Thomas, E.N. and Slotkin, R.K. (2015) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J.*, **34**, 20–35.
55. Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N. and Slotkin, R.K. (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol.*, **162**, 116–131.
56. Creasey, K.M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B.C. and Martienssen, R.A. (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature*, **508**, 411–415.
57. Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.*, **2**, e104.
58. Xie, Z., Allen, E., Wilken, A. and Carrington, J.C. (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12984–12989.
59. Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C. and Voinnet, O. (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science*, **313**, 68–71.
60. Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C. and Voinnet, O. (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J.*, **22**, 4523–4533.
61. Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.*, **16**, 735–743.
62. Yu, D., Fan, B., MacFarlane, S.A. and Chen, Z. (2003) Analysis of the involvement of an inducible Arabidopsis RNA-dependent RNA polymerase in antiviral defense. *Mol. Plant Microbe Interact.*, **16**, 206–216.
63. Thomas, C.L., Leh, V., Lederer, C. and Maule, A.J. (2003) Turnip crinkle virus coat protein mediates suppression of RNA silencing in Nicotiana benthamiana. *Virology*, **306**, 33–41.
64. Schott, G., Mari-Ordonez, A., Himber, C., Alioua, A., Voinnet, O. and Dunoyer, P. (2012) Differential effects of viral silencing suppressors on siRNA and miRNA loading support the existence of two distinct cellular pools of ARGONAUTE1. *EMBO J.*, **31**, 2553–2565.
65. Hurkman, W.J. and Tanaka, C.K. (1986) Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiol.*, **81**, 802–806.
66. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
67. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
68. Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangel, J.L. et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One*, **2**, e219.
69. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
70. Mosher, R.A., Schwach, F., Studholme, D.J. and Baulcombe, D.C. (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3145–3150.
71. Zhang, C., Li, G., Zhu, S., Zhang, S. and Fang, J. (2014) tasiRNAdb: a database of ta-siRNA regulatory pathways. *Bioinformatics*, **30**, 1045–1046.
72. Pontes, O., Li, C.F., Costa Nunes, P., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E. and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, **126**, 79–92.
73. Fang, Y. and Spector, D.L. (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr. Biol.*, **17**, 818–823.
74. Pontes, O., Vitins, A., Ream, T.S., Hong, E., Pikaard, C.S. and Costa-Nunes, P. (2013) Intersection of small RNA pathways in Arabidopsis thaliana sub-nuclear domains. *PLoS One*, **8**, e65652.
75. Hoffer, P., Ivashuta, S., Pontes, O., Vitins, A., Pikaard, C.S., Mroczka, A., Wagner, N. and Voelker, T. (2011) Posttranscriptional gene silencing in nuclei. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 409–414.
76. Kumakura, N., Takeda, A., Fujioka, Y., Motose, H., Takano, R. and Watanabe, Y. (2009) SGS3 and RDR6 interact and colocalize in cytoplasmic SGS3/RDR6-bodies. *FEBS Lett.*, **583**, 1261–1266.
77. Pumplin, N., Sarazin, A., Jullien, P., Bologna, N.G., Oberlin, S. and Voinnet, O. (2016) DNA methylation influences the expression of DICER-LIKE4 isoforms, which encode proteins of alternative localization and function. *Plant Cell*, doi:10.1105/tpc.16.00554.
78. Jouannet, V., Moreno, A.B., Elmayan, T., Vaucheret, H., Crespi, M.D. and Maizel, A. (2012) Cytoplasmic Arabidopsis AGO7 accumulates in membrane-associated siRNA bodies and is required for ta-siRNA biogenesis. *EMBO J.*, **31**, 1704–1713.
79. Shamandi, N., Zytynicki, M., Charbonnel, C., Elvira-Matelot, E., Bochnakian, A., Comella, P., Mallory, A.C., Lepere, G., Saez-Vasquez, J. and Vaucheret, H. (2015) Plants encode a general siRNA suppressor that is induced and suppressed by viruses. *PLoS Biol.*, **13**, e1002326.
80. Elvira-Matelot, E., Hachet, M., Shamandi, N., Comella, P., Saez-Vasquez, J., Zytynicki, M. and Vaucheret, H. (2016) Arabidopsis RNASE THREE LIKE2 modulates the expression of protein-coding genes via 24-nucleotide small interfering RNA-directed DNA methylation. *Plant Cell*, **28**, 406–425.
81. Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W. and Carrington, J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat. Genet.*, **36**, 1282–1290.

## SUPPLEMENTARY MATERIAL & METHODS

### Protein expression and purification

Full-length DRB7.2 was expressed from a pSMT3 vector in *E.coli* strain BL21. Cells were grown at 37°C in 250 ml of Luria–Bertani medium containing 50 µg/ml Kanamycin until an OD600 of 0.6 was achieved. Protein expression was induced with 0.5 mM isopropyl-β-d-thiogalactopyranoside and cells were further grown overnight at 20°C before harvesting by centrifugation. Cells were resuspended in 50 ml of PBS pH7.4, 0,1% TritonX100 and lysed by sonication. The protein was loaded onto a 1 ml HisTrap FF crude column (Pharmacia) pre-equilibrated in PBS pH7.4, 25 mM imidazole. After washing with 20 ml of PBS pH7.4, 1M NaCl and followed by 20 ml of PBS pH7.4, 50mM imidazole, the protein was eluted using a 50–500 mM linear gradient of imidazole. Fractions containing DRB7.2 were pooled and incubated overnight at 4°C in 1 mM DTT, 10% glycerol and 13 µg of home made Ubiquitin-like-specific protease 1 (ULP1) to remove the His6-SUMO tag. The cleaved DRB7.2 protein was again passed through Ni-NTA beads (Quiagen) equilibrated in PBS pH7.4, 5 mM imidazole to remove the tag and non-cleaved protein, and concentrated using an Ultra4 MWCO 3kDa concentrator (Amicon). Protein concentration was determined using the Bradford colorimetric assay (Bio-Rad).

### Nucleic acids

Four distinct primer pairs (Table S1) were used to amplify either the 400 first nucleotides or the 50 first nucleotides of the mGFP5 cDNA (“gf” or sense part, “fg” or antisense part). Transcription reactions (T7 RIBOMAX kit, Promega) were performed according to the manufacturer’s instructions directly on amplified products to give RNAs of 400 nt or 50 nt long. Internally labelled RNAs were obtained by addition of [ $\alpha$ -<sup>32</sup>P]-UTP (Hartmann Analytic) in the transcription reactions. The sense and antisense products of the same size were mixed in a 1:1 molar ratio in annealing buffer (10mM Tris-HCl and 20mM NaCl), heated at 95°C for 5 minutes, and annealed 5 hours to overnight at room temperature to give dsRNA.

### Binding assays

dsRNA ladder (N0363S, Biolabs) was incubated with the recombinant DRB7.2 protein in a final volume of 20 µl in 70 mM NaCl, 1.3 mM KCl, 5 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.9 mM KH<sub>2</sub>PO<sub>4</sub>, 89 mM Tris, 89 mM Borate, 0.2 mM EDTA for 20 min on ice. As control, dsRNA ladder was incubated in the same buffer without recombinant DRB7.2 protein. The reaction products were analysed directly by electrophoretic mobility-shift assay (EMSA) on 5% and 12% polyacrylamide (19/1) gels in 1X Tris-Gly at 4°C. Products were visualised by UV light after incubation of the gel in 1x TBE, 5 µg/ml Ethyidium bromide.

EMSA with <sup>32</sup>P-labelled dsRNA (400bp) was made in the same conditions except for competition experiments in which the indicated amount of cold dsRNA (400bp- or 50bp-long) was mixed with labelled dsRNA before addition of recombinant DRB7.2 protein and incubated 30 min at room temperature. Products were analysed on 5% polyacrylamide (19/1) gel in 1X Tris-Gly at 4°C and were visualised by autoradiography.

### **RNaseA/T1 treatment**

RNA immunoprecipitation was performed as described in the Material & Methods section with the following modifications. After extraction, the immunoprecipitated RNA were resuspended in sterile RNase-free water and digested for 30min at 37°C with 4 µg/ml of RNase A and 10 U/ml of RNaseT1 (Thermo Fischer Scientific) in 10 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM EDTA. After phenol extraction, RNA were analysed by northern blotting.

## **SUPPLEMENTARY FIGURE LEGENDS**

**Figure S1:** Effect of the *drb7.2* mutation on endogenous small RNAs. **(A)** Longer exposure of the northern blot analysis of IR71-derived siRNAs accumulation depicted in Figure 2A to visualize more easily the increase of the 21 nt size class. **(B)** Northern blot analysis of endoIR-siRNAs (@IR71, @IR2039), trans-acting siRNA (@255) and p4-siRNA (@REP2) accumulation was performed by sequential rounds of probing and stripping the same membranes obtained with an higher resolution gel. Accumulation of small RNA U6 (@U6) is used as loading control. Figure source data can be found with the Supplementary data.

**Figure S2:** Effect of the *drb7.2* mutation on viral RNA and viral-derived siRNA. **(A)** RNA gel analysis of TCV viral RNA accumulation in Col-0, *drb7.2* and *dcl2* mutant plants. **(B)** Accumulation of TCV-derived siRNA, IR71-derived siRNA, TAS1 trans-acting siRNA255 and miRNA159 in the samples depicted in (A). **(C)** Northern blot analysis of TRV-PDS viral RNAs accumulation in Col-0, *drb7.2* and *dcl4* plants (two independent replicates each). **(D)** Accumulation of TRV-derived siRNA, IR71-derived siRNA, TAS1 trans-acting siRNA255 and miR159 accumulation were analysed in the same plants as in (C). Ethidium bromide staining of gel (A and C), and snRNA U6 (B and D) were used as a loading control. Figure source data can be found with the Supplementary data.

**Figure S3:** DCL4-8 is stable but impaired in processing of DCL4-dependent small RNA production. **(A)** Accumulation of DCL4 was assessed by protein blot analysis in wild-type SUC-SUL (SS), SS*xdcl4-2*, and SS*xdcl4-8* plants. Equal loading was verified by Coomassie staining of the membrane after western blotting. **(B)** Northern blot analysis of trans-acting siRNA (@255, @TAS3), DCL4-

dependent miRNA (@822), DCL3-dependent siRNA (@REP2), endogenous IR-derived siRNA (@IR71, @2039) and miR159 accumulation in the same plants as in (A). Accumulation of snRNA U6 (@U6) was used as loading control. Figure source data can be found with the Supplementary data.

**Figure S4:** DRB7.2 does not bind precursors of DCL3-dependant siRNAs nor DCL1 dependant miRNAs. (A) Immunoprecipitation experiments were conducted in GFP- or DRB7:GFP-expressing plants using a specific GFP antibody. Total RNA extracted from the respective IPs was subjected to northern analysis using the indicated probes by sequential rounds of probing and stripping the same membrane. (B) The presence of GFP-tagged proteins in each IP was confirmed by protein blot analysis. Accumulation of snRNA U6 (@U6) (A) and coomassie staining of the membrane (B) was used as loading control. Figure source data can be found with the Supplementary data.

**Figure S5:** DRB7.2 binds preferentially double-stranded RNA *in vitro* and *in vivo*. (A) Band-shift experiments with <sup>32</sup>P-labelled (\*) 400bp-long dsRNA or ssRNA (0.2 nM) in the absence (0) or presence of increasing amount of recombinant DRB7.2 protein (0.18 μM to 5.8 μM). The protein-RNA complexes were separated from free RNA by 5% non-denaturing gel electrophoresis at 4°C (B) Immunoprecipitation experiments were conducted in GFP- or DRB7:GFP-expressing plants using a specific GFP antibody. Total RNA extracted from the respective IPs was subjected to northern analysis after RNase A/T1 treatment, using the indicated probes by sequential rounds of probing and stripping the same membrane. (C) The presence of GFP-tagged proteins in each IP was confirmed by protein blot analysis. Accumulation of snRNA U6 (@U6) (B) and coomassie staining of the membrane (C) was used as loading control. Figure source data can be found with the Supplementary data.

**Figure S6:** DRB7.2 binds preferentially high molecular weight double-stranded RNA molecules *in vitro*. (A-B) Purified recombinant DRB7.2 protein was incubated with commercial dsRNA ladder as described in the Supplementary Materials and Methods. Increasing amount of DRB7.2 protein, indicated above each lane, was used for the binding reaction. The protein-RNA complexes were separated from free dsRNA by 12% (A) or 5% (B) non-denaturing gel electrophoresis at 4°C. (C) Band-shift experiments with <sup>32</sup>P-labelled (\*) 400bp-long dsRNA (0.2 nM) and DRB7.2 (2.9 μM) in the absence (0) or presence of either 400bp- or 50bp-long cold dsRNAs as competitors. Competitors were added in increasing molar excess (from 1 nM to 200 nM which corresponds to an excess of 5- to 1000-fold). Protein concentration was determined experimentally to give near complete binding of dsRNA. Figure source data can be found with the Supplementary data.

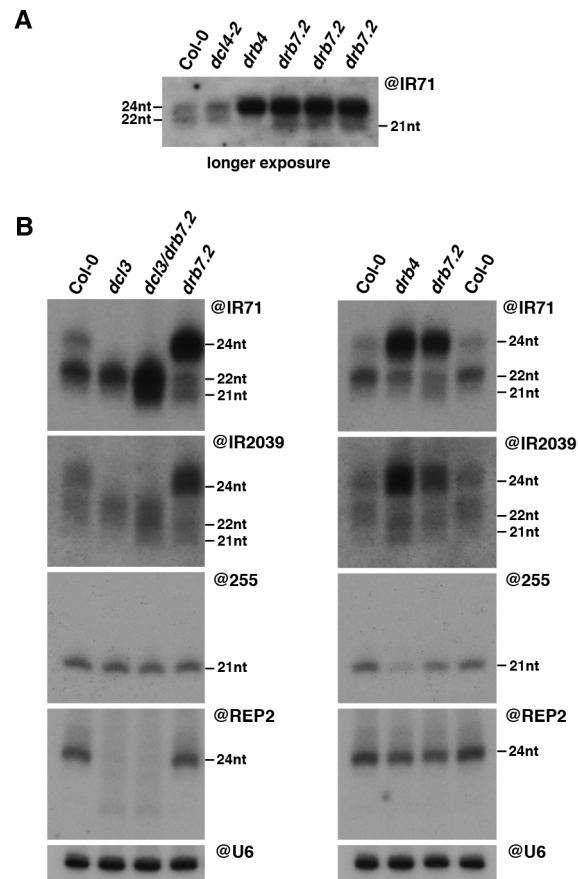
**Figure S7:** Effect of the *drb7.2* mutation on other endogenous IR-derived small RNAs. Northern blot analysis of three other endoIR-siRNA accumulation (@IRAtGP10, @At5g22960 and @At4g21366)



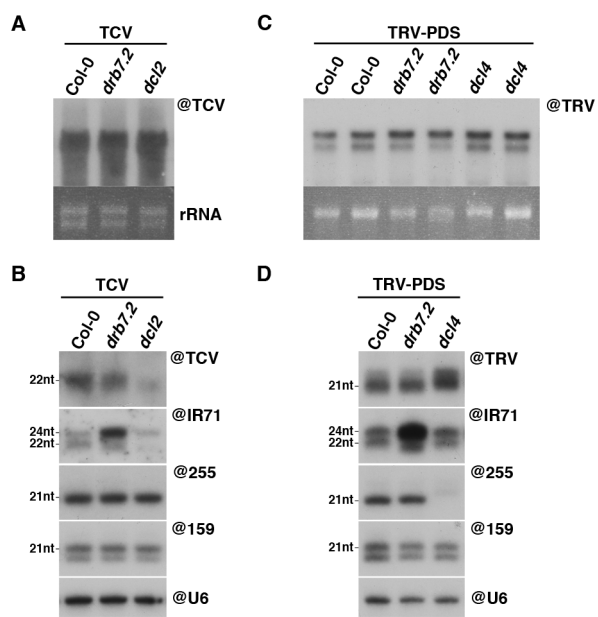
in wild-type, *dcl2*, *dcl3*, *dcl4*, *drb7.2*, *drb7.2/dcl2*, *drb7.2/dcl3*, *drb7.2/dcl4*, and *dcl2/4* mutant plants was performed by sequential rounds of probing and stripping the same membranes. Accumulation of small RNA U6 (@U6) is used as loading control. Figure source data can be found with the Supplementary data.

**Table S1:** Sequences of primers used in this study

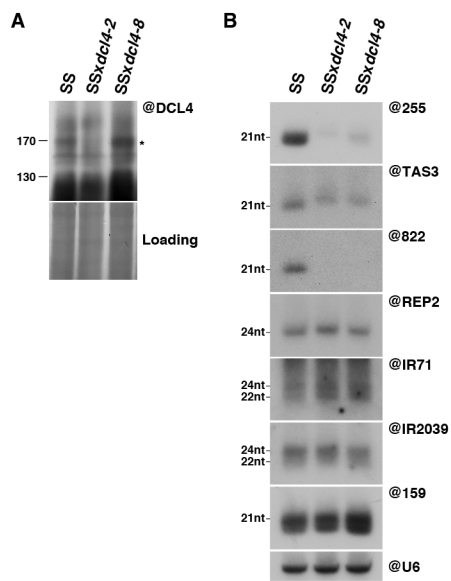
**Table S2:** Excel spreadsheet containing, in separate sheets, the characteristics of our whole genome small RNA sequencing libraries, the underlying numerical data for Figure 1A, 1B, 1C, quantification of band intensities on northern blots displayed in Figure 2A and sequence coverage of DRB4 obtained by mass spectrometry analysis of DRB7.2:GFP IP fractions.



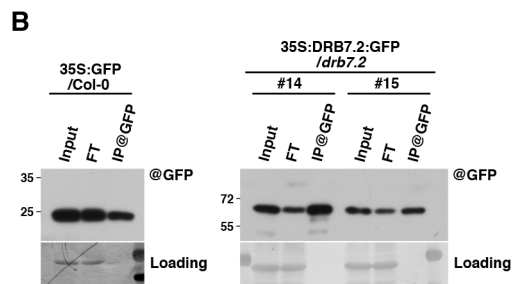
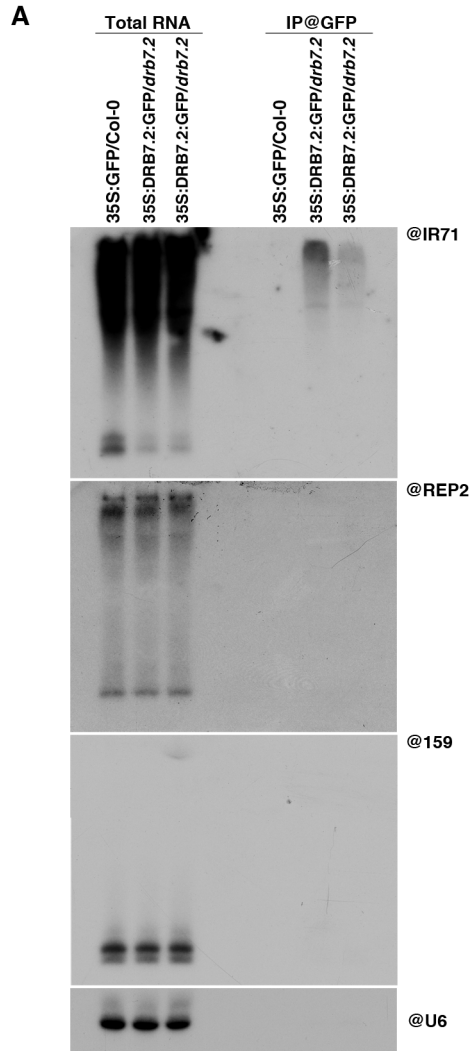
Montavon et al., Figure S1



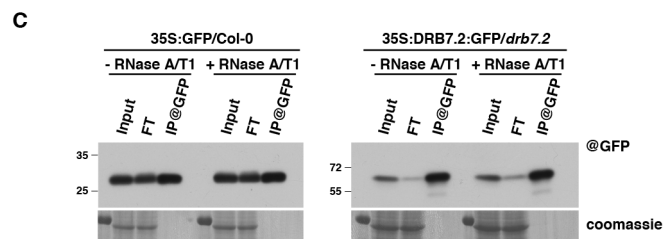
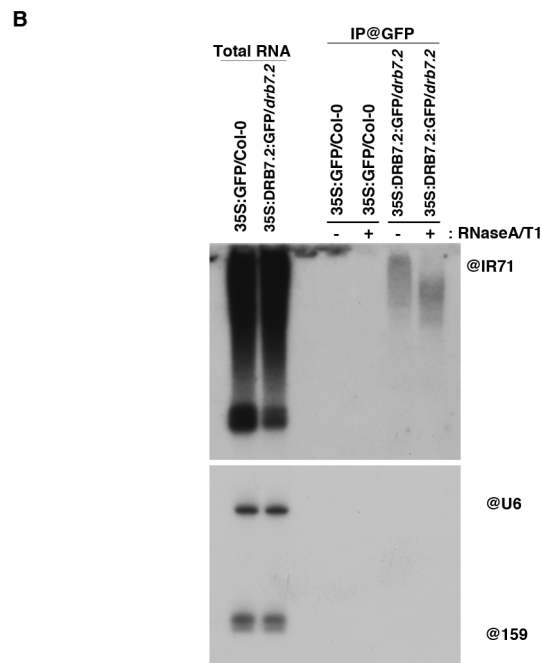
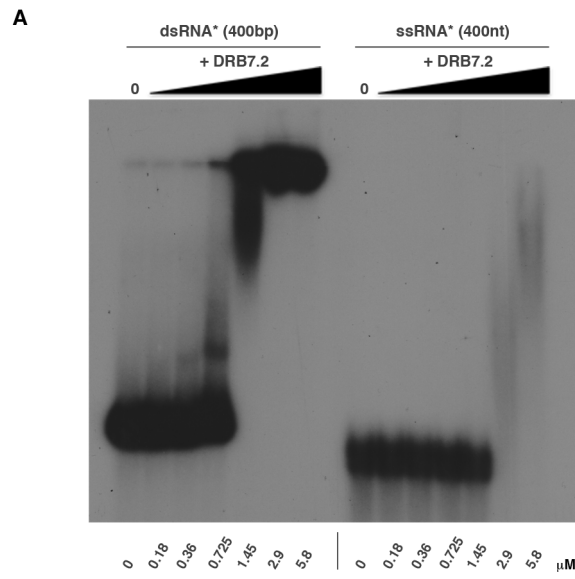
Montavon et al., Figure S2



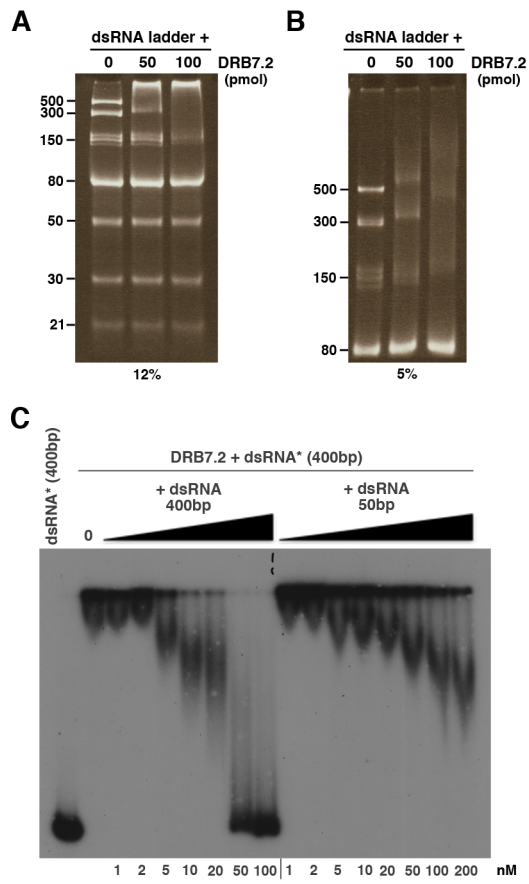
Montavon et al., Figure S3



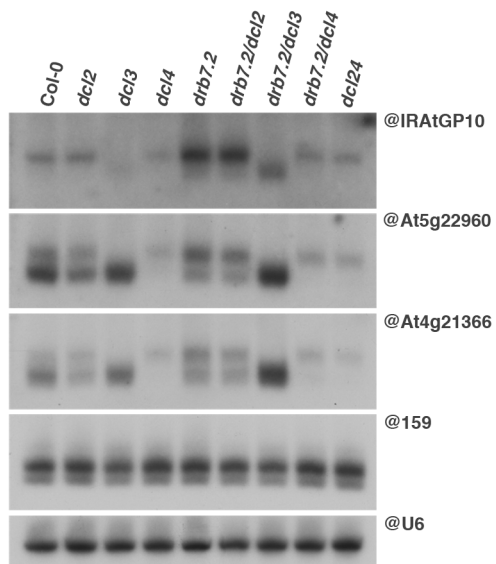
Montavon et al., Figure S4



Montavon et al., Figure S5



Montavon et al., Figure S6



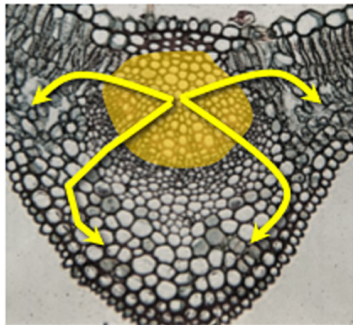
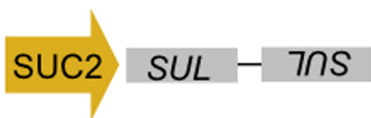
Montavon et al., Figure S7



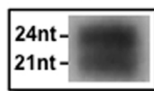
Supplementary Table S1: sequence of primers used in this study

Primer name	Sequence (5'-3')	Purpose
<i>drb4</i> -LP	AGGCGATTCTCTCGAATTC	Genotyping
<i>drb4</i> -RP	TTGTAGGCAACATCAATTCCC	Genotyping
<i>drb7</i> -LP	GGGTCTTGCCATATGAAAAT	Genotyping
<i>drb7</i> -RP	AAGAGTGATTATGCGTCGTGG	Genotyping
<i>dcl2</i> -Lower	CCTCTCCGGAAGTCTCCACAATT	Genotyping
<i>dcl2</i> -Upper	GGCTGAGATACCTCAAGGTGGTTT	Genotyping
<i>dcl3</i> -LP	CCTGAAGAGCGTGAAAGGAGTGAAAA	Genotyping
<i>dcl3</i> -RP	GACTAGAAGAGGACTCAATGCAAT	Genotyping
<i>dcl4</i> -Lower	GCCGCTCGAGATCATCAGCAAAGGAAT	Genotyping
<i>dcl4</i> -Upper	GGCTGCCACAGCTGATGATTACA	Genotyping
miR159	TCGGCAAGTCATCCTTGGCTG	Northern
miR173	GTGATTTCTCTGCAAGCGAA	Northern
miR822	CATGTGCAAATGCTTCCCGCA	Northern
siRNA TAS1 255	TACGCTATGTTGGACTTAGAA	Northern
siRNA TAS3	TGGGGTCTTACAAGGTCAAGA	Northern
snU6	AGGGGCCATGCTAATCTTCTC	Northern
siRNA02	GTTGACCAGTCCGCCAGCCGA	Northern
siRNA AtRep2	GCGGGACGGGTTTGGCAGGAC	Northern
siRNA TR2558 #1	TTCATCAGATGACCAGATAGCTTA	Northern
siRNA TR2558 #2	TTTGTTCAATCATTCAATGGCCAT	Northern
siRNA AtSN1 #1	CACCAACGTGTTGTTGGCCAGTGTAATCTCTCAGATAGAGGTGCTGGATTGAGACA	Northern
siRNA AtSN1 #2	CTCTCAGATAGAGGTGCTGGATTGAGACACGTTGGGAAGGATCTCTTTCCAAAAGAAGA	Northern
siRNA AtSN1 #3	CGTTGGGAAGGATCTCTTTCCAAAAGAAGATGAATTTCTGGTATGGGTCCCGCTCTGGG	Northern
siRNA AtSN1 #4	ATTTCTGGTATGGGTCCCGCTCTGGGAGATGAAGGCCTTGGGCCTGAACCTCCAG	Northern
siRNA IR71fw	AAATGACCGCTACACTGCTTAT	Northern
siRNA IR71rev	TCTCTCGTCAATGGACAATGAATC	Northern
siRNA 2039 fw	CTTCCAGACCTTCTCTACATGT	Northern
siRNA 2039 rev	ATTCGCTCTCGTTTAATGGATG	Northern
PDS rev	ACCATATATGAACATTAATAAC	Northern
PDS fw	AGATTGACTTCCCAGATGTC	Northern
TCV P38 fw	ATGGAAAATGATCCTAGAGTCC	Northern
TCV P38 rev	CTAAATTCTGAGTGTGCAATTT	Northern
35s-DRB7.2-Sall fw	GTCGCATGGATCTTCTCCTCCAGAG	Cloning
pDRB7.2-Sall fw	GTCGACCACAGAAAGATCCACGG	Cloning
DRB7.2-XmaI rev	CCCCGGCACGGGTTACCTGTTACC	Cloning
attB1-pDCL4 fw	GGGG ACA ACT TTG TAT AGA AAA GTT G gtcgaaccgagtgcagatca	Cloning
attB2-DCL4 rev	GGGG AC TGC TTT TTT GTA CAA ACT TG aattcacgatctcagctctccctct	Cloning
DRB4 qPCR #1 rev	CACTAGGCAAAGTAGGCGATC	q-PCR
DRB4 qPCR #1 fw	AAAAGTTGGCTGAAATGAGCG	q-PCR
DRB4 qPCR #2 rev	TTTGTAGGATTGTGAGGTATAGGC	q-PCR
DRB4 qPCR #2 fw	GCTAAGGAGATTAACGTGCTTTG	q-PCR
T7gf-fwd	TAATACGACTCACTATAGGGAGTAAAGGAGAAGAAGCTTTTCACT	PCR template for EMSA
gf400-rev	GGGTTCCTCCTCCTTGAATTCGA	PCR template for EMSA
T7fg400-fwd	TAATACGACTCACTATAGGGTCCGTCCTCCTTGAATTCGA	PCR template for EMSA
fg-rev	GGGAGTAAAGGAGAAGAAGCTTTTCACT	PCR template for EMSA
gf50-rev	GGGAATTCAACAAGAATTGGGACAA	PCR template for EMSA
T7fg50-fwd	TAATACGACTCACTATAGGGAAATCAACAAGAATTGGGACAA	PCR template for EMSA

### Initiateur spécifique du phloème



DCL4 (21 nt)  
et  
DCL3 (24 nt)



@ siRNA SUL



### Gène endogène



**Figure 1.1: Principe du système SUC:SUL**

Une séquence répétée et inversée du gène *SULPHUR* (*SUL*) sous contrôle du promoteur SUCROSE 2 (*SUC*), spécifique des cellules-compagnes (CC) est introduite dans des plantes sauvages. Une fois transcrit ce transgène va adopter une structure en tige-boucle qui sera reconnue par la machinerie du RNA silencing pour produire des siRNAs de 21 et 24 nt (@ siRNA SUL). Les siRNAs de 21 nt, une fois dans le complexe RISC, réprimeront par clivage de l'ARNm l'expression du gène *SUL* dans les CC mais également dans les 10 à 15 cellules adjacentes, grâce au mouvement des siRNAs. Cette répression conduit à l'apparition de chlorose suivant le réseau de veines des feuilles.

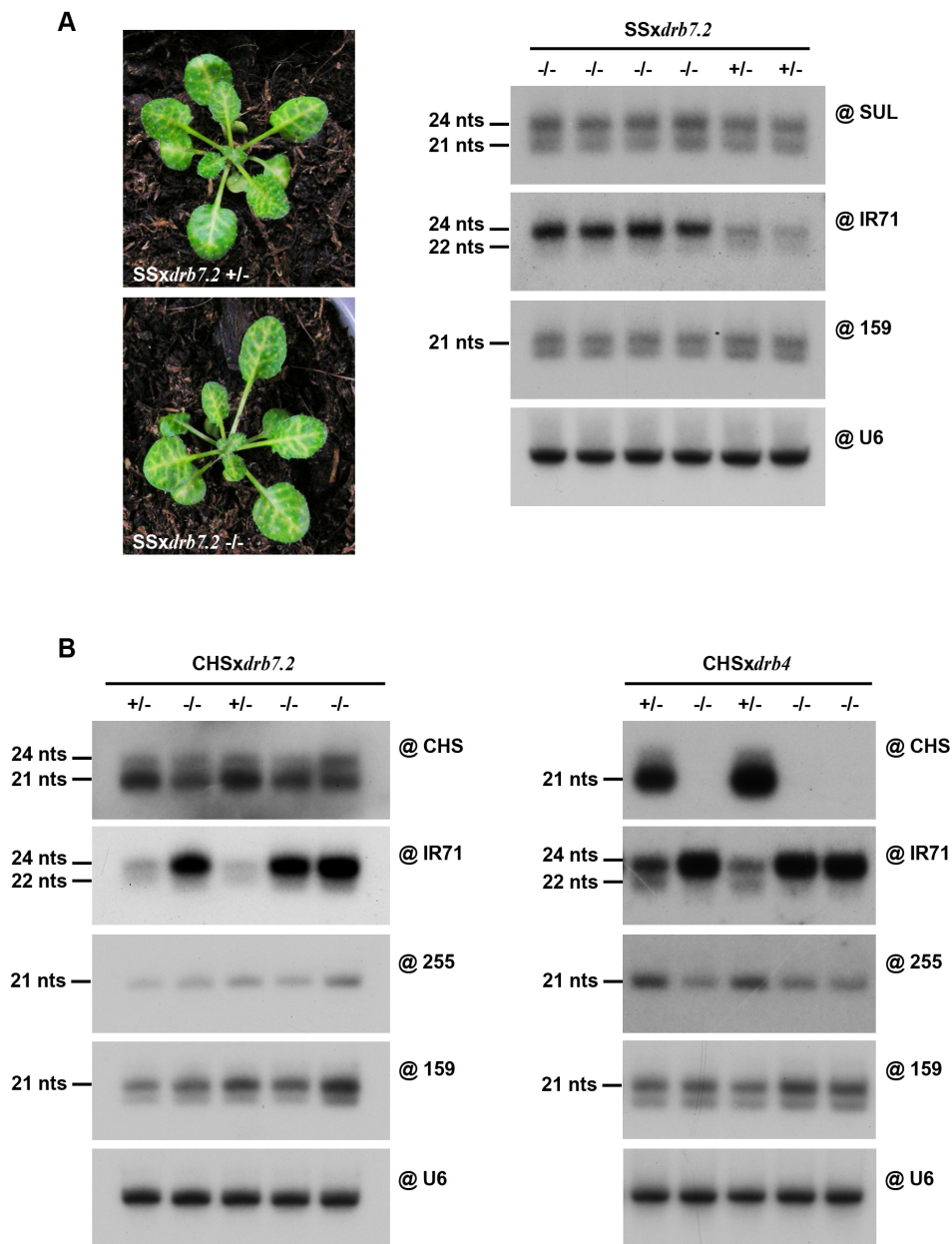
## 1.3 RESULTATS COMPLEMENTAIRES

En plus des résultats présentés précédemment, d'autres analyses sur DRB7.2 ont été, ou sont réalisées au laboratoire afin de mieux caractériser la ou les fonctions de cette protéine.

### 1.3.1 DRB7.2 ET LES EXOIR-siRNA

Ayant pu mettre en évidence le rôle crucial de DRB7.2 dans la biogénèse des siRNAs à partir de loci IR endogènes, nous nous sommes alors demandés si cet effet s'exerçait également sur la production de siRNAs provenant d'IR exogènes, similaires à celles couramment utilisées en recherche fondamentale, dans le but d'étudier le rôle d'un gène d'intérêt en réprimant son expression (IR-RNAi) (Smith *et al*, 2000; Dunoyer *et al*, 2005).

Pour cela, la mutation *drb7.2* a été introduite dans notre lignée de référence SUC:*SUL* (SS) par croisement, et des plantes hétérozygotes (+/-) ou homozygotes (-/-) pour la mutation *drb7.2* ont été sélectionnées par génotypage à la génération F2. L'effet de la mutation *drb7.2* sur les endoIR-siRNAs étant récessive, les plantes hétérozygotes pour la mutation nous servent ici de contrôle. Le système transgénique SUC:*SUL* repose sur l'expression d'une séquence répétée et inversée (IR) correspondant au gène *SULPHUR* (At4g18480), sous contrôle du promoteur cellules compagnes (CC)-spécifiques SUC2. Cette IR est clivée principalement par DCL3 et DCL4 pour générer des siRNAs @SUL de 21 et 24 nt au niveau des CC. Les siRNAs de 21 nt, une fois chargés dans AGO1, vont permettre le clivage de l'ARNm *SUL* dans les CC, mais aussi 10-15 cellules au delà, grâce au mouvement de ces siRNAs (Dunoyer *et al*, 2005, 2007). Ceci entraîne l'apparition d'une chlorose caractéristique qui suit les réseaux de veines des feuilles (**Figure 1.1**).



**Figure 1.2 : Effet des mutations *drb7.2* et *drb4* sur l'accumulation de différents exoIR-siRNAs**

(A) Phénotype d'une plante hétérozygote (+/-) ou homozygote (-/-) pour la mutation *drb7.2* exprimant le transgène *SUC:SUL* (SS) (gauche). Analyse par northern blot de l'accumulation des exoIR-siRNAs *SUL* (@*SUL*), et des endoIR-siRNAs *IR71* (@*IR71*), dans des plantes *SSxdrb7.2* hétérozygotes (+/-) ou homozygotes (-/-) (droite). L'accumulation des différents siRNAs a été détectée par hybridations successives de la même membrane à l'aide de sondes spécifiques. (B) L'accumulation des exoIR-siRNAs *CHS* (@*CHS*), des endoIR-siRNAs *IR71* (@*IR71*), et du tasiRNA *TAS1* (@255) a été analysée par northern blot, dans des plantes *CHSxdrb7.2* (gauche) ou *CHSxdrb4* (droite) hétérozygotes (+/-) ou homozygotes (-/-) pour la mutation. La visualisation du snRNA *U6* (@*U6*) et du miRNA *159* (@159) a été utilisé comme contrôle de charge.

L'analyse phénotypique et moléculaire de ces plantes a permis de mettre en évidence que la mutation *drb7.2* ne semblait ni influencer le phénotype de chlorose, ni l'accumulation des siRNAs @SUL (**Figure 1.2A**), suggérant que contrairement aux endoIR-siRNAs, les exo-IR-siRNAs ne semblent pas être affectés par cette mutation. Ces observations pourraient indiquer que l'effet de DRB7.2 dans la régulation de la biogénèse des IR-siRNAs est spécifique aux IR endogènes. Cependant, une autre possibilité serait que cette protéine soit peu ou pas exprimée au niveau des CC, où sont générés les siRNAs @SUL.

Afin de discriminer entre ces deux possibilités, nous avons décidé d'utiliser un autre système transgénique, dans lequel une IR exogène serait exprimée sous contrôle d'un promoteur ubiquitaire. Pour cela, la mutation *drb7.2* a été introduite dans la lignée CHS-RNAi (Smith *et al*, 2000), où une séquence répétée et inversée du gène de la *CHALCONE SYNTHASE* (At5g13930), exprimée sous contrôle du promoteur 35S, induit le silencing de ce gène dans l'ensemble des tissus de la plante. Comme dans le cas du système SS, l'analyse moléculaire des siRNAs @CHS n'a pas permis de détecter de différence d'accumulation notable chez les plantes mutées pour *drb7.2* (**Figure 1.2B**), à l'exception d'une légère diminution des siRNAs de 21 nt. La diminution de cette espèce de siRNAs DCL4-dépendant s'explique probablement par le fait que l'accumulation de DRB4 est elle aussi réduite dans un mutant *drb7.2* (cf. §1.2; **Figure 2**, Montavon *et al*, soumis). En effet, l'introduction de la mutation *drb4* dans le système CHS-RNAi a permis de mettre évidence la forte dépendance de ces siRNAs @CHS à la présence de cette protéine (**Figure 1.2 B**).

Ainsi, dans aucun des deux systèmes transgéniques testés, DRB7.2 n'influence la biogénèse des siRNAs à partir d'IR exogènes. Ces résultats suggèrent donc que DRB7.2 semble capable de discriminer ces précurseurs ARNdb pour sa fonction de séquestration, ou qu'un élément intrinsèque à l'un de ces deux types d'ARNdb entrainerait sa non-reconnaissance par DRB7.2. Un de ces éléments pourrait être la présence, ou l'absence, d'intron. En effet, dans chacun des deux systèmes transgéniques testés, les séquences

répétées et inversées des IR exogènes sont interrompues par un intron qui doit être épissé, afin d'augmenter l'efficacité du RNA silencing (Smith *et al*, 2000). A l'inverse, les IR endogènes, tels qu'IR71 et IR2039, ne possèdent pas de séquences introniques entre les deux parties répétées et inversées. Il est donc possible que l'étape d'épissage subis par les transcrits des exo-IR empêche la fixation de DRB7.2. Afin de tester cette hypothèse, un système contenant une exo-IR sans intron pourrait être introduit dans le mutant *drb7.2* afin de voir si cette étape d'épissage est potentiellement limitante pour la fixation de son précurseur par DRB7.2.

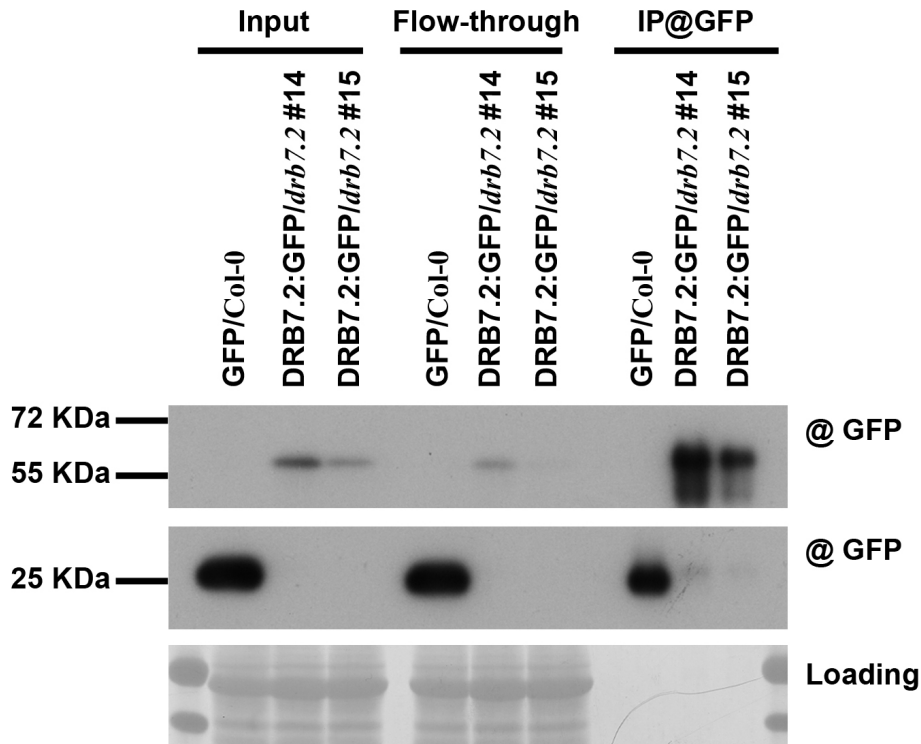
Une autre possibilité (non mutuellement exclusive) serait la quantité de DRB7.2 au sein de la cellule. En effet si celle-ci est limitante, la présence d'une grande quantité d'IR endogènes pourrait entraîner la titration de DRB7.2, de sorte que les IR exogènes ne seraient jamais séquestrées. Enfin la taille des IR exogènes pourrait également être un facteur déterminant. Les IR produites dans les systèmes SS et CHS-RNAi ont une taille moyenne d'environ 400 pb, alors que la majorité des IR endogènes affectées par DRB7.2 ont une taille bien supérieure (cf. §1.2; Supplemental Table S2 Montavon *et al*, soumis) (**Annexe 1**). Si DRB7.2 a une affinité supérieure pour les très longs ARNdb, là encore, la présence des IR endogènes pourrait conduire à une titration de DRB7.2.

### **1.3.2 RECHERCHE DE NOUVEAUX INTERACTANTS DE DRB7.2, PREMIERS RESULTATS**

Afin d'approfondir nos connaissances sur DRB7.2, et d'identifier éventuellement d'autres fonctions de cette protéine, nous avons également entrepris de définir son interactome.

Pour cela, des expériences d'immunoprécipitation (IP) couplées à des analyses par spectrométrie de masse (MS), ont été réalisées sur les lignées transgéniques 35S-DRB7.2:GFP/*drb7.2* générées précédemment (cf. §1.2;

**A**



**B**

AGI/Nom	GFP			DRB7.2:GFP#14			DRB7.2:GFP#15		
	1	2	3	1	2	3	1	2	3
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT4G00420./DRB7.2				2849	2741	2541	2228	2114	1690
AT3G62800.1/DRB4				96	104	98	147	122	144
AT3G57990.1/D7IP1 (OEP40)				24	28	27	31	31	23

**Figure1.3: Analyse de l'interactome de DRB7.2**

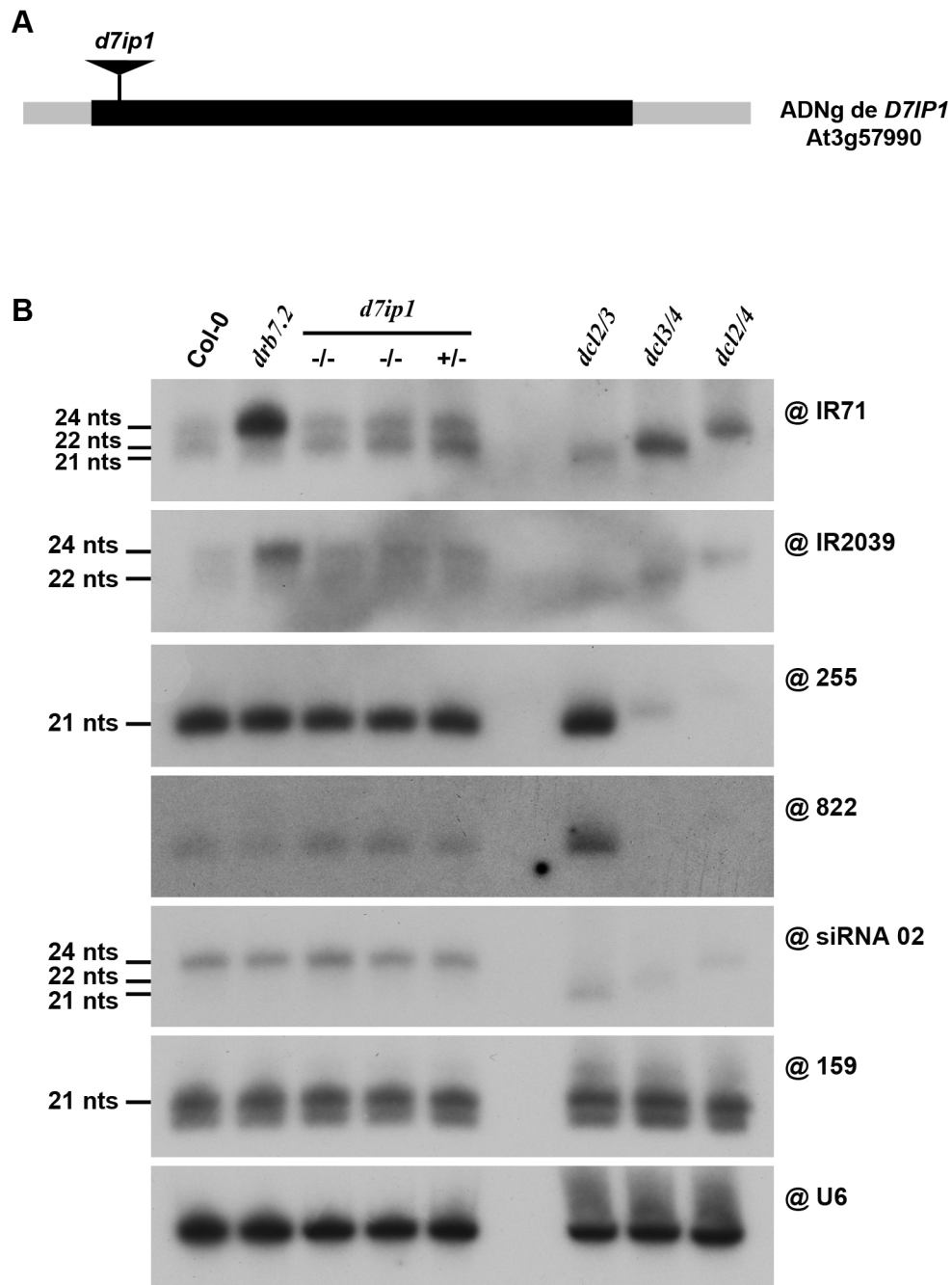
(A) Analyse de l'efficacité d'immunoprécipitation réalisée à partir des plantes contrôles (GFP/Col 0) ou des lignées transgéniques (35S-DRB7.2:GFP/drbb7.2 #14 et 15). Le niveau des protéines GFP et DRB7.2:GFP (@GFP), présentes dans la fraction totale (Input), le surnageant après immunoprécipitation (Flow-through) et la fraction immunoprécipitée (IP@GFP) a été analysé par western blot. La coloration au bleu de coomassie des protéines présentes sur la membrane sert ici de témoin de charge. (B) Liste des protéines majoritairement retrouvées dans nos analyses IP-MS de l'interactome de DRB7.2. Ce tableau constitue une liste non exhaustive des protéines identifiées au cours de ces analyses et présente uniquement les deux protéines retrouvées dans tous les réplicas réalisés. #Spectres: Nombre de spectre de fragmentation MS/MS.

Montavon *et al*, soumis), et comparées aux lignées transgéniques 35S-GFP servant de contrôle négatif. Chaque expérience indépendante d'IP-MS a été effectuée au minimum en *triplica*, à partir de plantules ou de fleurs. Pour chaque échantillon utilisé, un aliquot de la fraction immunoprécipitée a été gardé pour des analyses par western blot afin de vérifier l'efficacité d'IP (**Figure 1.3A**). Comme observé précédemment par BiFC (Clavel *et al*, 2016) et co-IP (cf. §1.2; **Figure 3**, Montavon *et al*, soumis), de nombreux spectres correspondant à DRB4 (59% de couverture; Supplemental Table S2, Montavon *et al*, soumis) (**Annexe1**) ont pu être identifiés spécifiquement dans les analyses MS des IP DRB7.2 (**Figure 1.3B**), confirmant que DRB4 est le partenaire majeur de cette protéine. De plus, aucun spectre correspondant à des peptides de DCL1, DCL2, DCL3 ou DCL4 n'a pu être détecté dans ces mêmes analyses.

Parmi les différentes protéines spécifiquement associées à DRB7.2:GFP, une protéine possédant un nombre de spectre de fragmentation MS/MS relativement élevé a été retrouvée dans tous les échantillons testés (**Figure 1.3B**). Cette protéine que nous avons nommée D7IP1 (DRB7.2 Interacting Protein 1, At3g57990), est une protéine de fonction inconnue, ne possédant pas de domaine référencé. Afin de déterminer si cette protéine pouvait jouer un rôle dans une des voies du RNA silencing en général, et sur la fonction de DRB7.2 sur les endoIR-siRNAs en particulier, l'accumulation des divers sRNAs endogènes a été testée chez des mutants knock-out (KO) de *d7ip1* (**Figure 1.4A**). Ces analyses n'ont malheureusement pas permis de mettre en évidence une différence d'accumulation des divers sRNAs testés entre *d7ip1* et les plantes contrôles. (**Figure 1.4B**).

Ces premiers résultats ne permettant pas de nous aiguiller quant à la fonction de cette protéine et de son interaction avec DRB7.2, des analyses complémentaires vont être prochainement entreprises. Elles consisteront dans un premier temps à confirmer l'interaction observée entre DRB7.2 et D7IP1 par des approches complémentaires (BiFC, co-IP). Si cette interaction est confirmée, l'analyse par RNA-seq des ARN associés à DRB7.2, que nous prévoyons de réaliser (cf. §1.4.3), pourra alors s'avérer informatif en mettant





**Figure 1.4: Effet de la mutation *d7ip1* sur l'accumulation de divers sRNAs endogènes**

(A) Organisation schématique de l'ADN génomique de *D7IP1*. La séquence codante et les 5' et 3' UTR sont représentées respectivement en noir et en gris. Le mutant d'insertion T-DNA utilisé durant cette étude est indiqué par un triangle. (B) Analyse par northern blot de l'accumulation des endoIR-siRNAs (@IR71, @IR2039), du tasiRNA TAS1 (@255), du miRNA DCL4-dépendant miR822 (@822), du p4-siRNA siRNA 02 (@siRNA 02), et du miRNA DCL1-dépendant miR159 (@159) à partir d'ARN de plantes sauvages (Col-0), *drb7.2*, *dcl2/3*, *dcl3/4*, *dcl2/4* et de plantes homozygotes (-/-) ou hétérozygotes pour la mutation *d7ip1*. Le snRNA U6 (@U6) est utilisé comme contrôle de charge. La visualisation des différents sRNAs a été réalisée comme en figure 1.2.

en évidence des cibles additionnelles de cette protéine et pour lesquels D7IP1 pourrait jouer un rôle dans leur régulation.

**Note:** Au cours de la rédaction de cette thèse, l'identification et la caractérisation de D7IP1 a été publiée par Harsman *et al*, dans la revue *J. Biol. Chem* (Harsman *et al*, 2016). Cette protéine nommée OEP40 semble former un canal perméable au glucose au niveau de la membrane externe des chloroplastes. Ces propriétés, ainsi que la localisation de cette protéine, rendent donc peu probable son implication dans la fonction de DRB7.2 et suggèrent que son identification spécifique dans nos expériences d'IP-MS de DRB7.2 soit un exemple de faux positif.

## **1.4 DISCUSSION, PERSPECTIVES**

Les résultats présentés dans ce chapitre ont simultanément permis de définir la fonction de la protéine DRB7.2 dans la voie de biogénèse des endoIR-siRNAs et de caractériser un nouveau rôle de la protéine DRB4. En plus d'illustrer comment les IR endogènes peuvent constituer des outils moléculaires utiles (bien que souvent non considérées) dans l'étude des mécanismes du RNA silencing, l'existence d'un complexe de deux protéines DRB qui semble spécifiquement dévolu à la régulation de la production de cette classe particulière de sRNAs endogènes pose également la question de leur(s) rôle(s) biologique(s).

### **1.4.1 ROLE BIOLOGIQUE DES ENDO-IR siRNA**

Un grand nombre de loci dans le génome d'*Arabidopsis* ont été annoté comme sites potentiels de production d'endoIR-siRNAs (Lindow *et al*, 2007).

Ces loci n'ont, pour l'heure, été considérés que comme les étapes précoces de précurseurs de jeunes miRNAs en cours d'évolution (proto-*MIR*). De plus, ces loci sont considérés comme ayant peu, voire pas, de rôle biologique du fait de leurs faibles niveaux d'expression (Allen *et al*, 2004; Fahlgren *et al*, 2007; Vazquez *et al*, 2008). Cependant, notre analyse par séquençage à haut-débit a permis de révéler que parmi ceux-ci, certains loci sont capables de générer d'importantes quantités de sRNAs (cf. §1.2; Supplemental Table S2, Montavon *et al*, soumis), (**Annexe 1**).

Ainsi, le rôle de séquestration des précurseurs d'endoIR-siRNAs par le complexe DRB4/DRB7.2 pourrait être justement d'empêcher la production trop importante de sRNAs à partir de ces loci, soit pour limiter leurs potentiels effets "off-target", soit pour prévenir la titration des DCLs de leurs cibles principales. Cependant, une hypothèse plus attractive serait que ces sRNAs possèdent une fonction biologique propre. L'absence totale de défauts morphologiques visibles chez les mutants *drb7.2* écarte un rôle de ces sRNAs au cours du développement. De même, l'absence d'hyper-susceptibilité de ces plantes en réponse à l'infection par divers virus à ARN semble indiquer qu'ils ne participent pas à la défense antivirale, à l'inverse de ce qui est suggéré pour les virus-activated siRNAs (Cao *et al*, 2014). Un des rôles possible de ces siRNA serait de participer à l'adaptation de la plante en réponse à des stress environnementaux. Ainsi, n'importe quel stimulus biotique ou abiotique entraînant un changement dans l'accumulation d'un des composants du complexe DRB4/DRB7.2 conduirait à la libération rapide des précurseurs d'IR endogènes séquestrés et la production accrue de leur sRNAs, permettant à la plante de répondre plus efficacement au stress perçu. Dans ce cas, le rôle du complexe DRB4/DRB7.2 pourrait être vu comme un réservoir de secours de précurseurs de sRNAs impliqués dans l'adaptation des plantes à un stress particulier. Tester la résistance des plantes *drb7.2* à différents stress abiotiques tels que la chaleur, le froid, la sécheresse ou l'hypersalinité permettrait d'apporter quelques éléments de réponse quant à la validité de cette hypothèse.

Une autre possibilité envisageable concernant les fonctions biologiques des loci IR endogènes est qu'ils soient source potentielle d'épiallèles. Bien que relativement peu d'exemples naturels d'épiallélisme soient disponibles dans la littérature, la réversibilité et la flexibilité de ce processus ont amené à spéculer qu'ils pourraient constituer une forme de régulation de l'expression des gènes particulièrement présente chez les organismes sessiles tels que les plantes (Finnegan, 2002; Vaillant & Paszkowski, 2007). Le rôle d'une IR dérivant d'un transposon a ainsi été mis en évidence chez le maïs, avec la caractérisation du locus IR *Mu Killer* (*Muk*), capable de déclencher le silencing en *trans* et de façon héritable de toute la famille de transposon *MuDR* (Slotkin *et al*, 2003, 2005; Lisch, 2009). Le contrôle épigénétique de l'expression de certains gènes étant assuré par la présence proximale de transposons, les résultats observés avec le locus *Muk* illustre le concept que l'état épigénétique à l'échelle du génome est en partie contrôlé par un locus d'IR endogène. Sur ce point, il est intéressant de noter que parmi les loci d'IR endogènes fortement exprimés, et identifiés au cours de notre séquençage à haut débit, l'un d'entre eux correspond à un transposon de la famille AtGP10 (cf. §1.2; Supplemental Table S2, Montavon *et al*, soumis) (**Annexe 1**). Etant donné que ces loci d'IR apparaissent ou disparaissent d'un point de vue évolutif à l'échelle de l'écotype, il serait intéressant d'identifier un écotype d'*A. thaliana* ne possédant pas encore ce locus IR-AtGP10 afin de déterminer l'impact des endoIR-siRNAs qui en découlent sur l'activité des transposons issus de cette famille. Enfin, de manière plus générale, les IR endogènes se trouvant à proximité de gènes, pourraient également réguler leur expression en modifiant l'état de méthylation des promoteurs présents à proximité. L'utilisation du mutant *drb7.2* pourrait s'avérer un outil de choix dans l'étude de ces mécanismes.

### 1.4.2 DRB7.1 vs DRB7.2

Lors de l'étude phylogénique menée sur les différentes protéines DRBs, une autre protéine de la même famille que DRB7.2 a également été identifiée:

DRB7.1 (Clavel *et al*, 2016). La fonction de DRB7.1 n'a pas été caractérisée jusqu'à présent, mais cette protéine ne semble pas être impliquée dans la biogénèse des différents types de sRNAs endogènes, exception faite des endoIR-siRNAs pour lesquels aucune analyse n'a à ce jour été réalisée (Clavel *et al*, 2016). Ainsi, il est toujours possible que DRB7.1 puisse remplir une fonction similaire à DRB7.2. Cependant, si tel est le cas, ceci implique que DRB7.1 soit exprimée dans des assises cellulaires spécifiques où, DRB7.2 n'est pas présente. En effet, la forte augmentation des endoIR-siRNAs observée dans le mutant *drb7.2* suggère que DRB7.1 n'est pas en mesure de compenser l'absence de cette protéine (cf. §1.2; Montavon *et al*, soumis). Une autre possibilité serait que ces deux protéines agissent de manière cumulative. Si tel est le cas, le même genre d'effet sur les endoIR-siRNAs devrait être observé dans un mutant *drb7.1*, et décuplé dans le double mutant *drb7.1/drb7.2*. Un séquençage à haut débit des sRNAs endogènes dans ce mutant devrait nous permettre de répondre à ces questions.

### 1.4.3 AUTRE FONCTION DE DRB7.2

Bien que nos analyses indiquent que la régulation de la biogénèse des endoIR-siRNAs par DRB7.2 dépende de la séquestration de leurs précurseurs, il n'est, à l'heure actuelle, pas encore possible d'exclure formellement que l'augmentation globale de l'accumulation de ces sRNAs endogènes soit également dépendante, en partie, d'une augmentation de la transcription des loci d'IR. Ainsi, il a récemment été montré que la protéine CDF2 (cycling DOF transcription factor) était capable d'une part de se fixer sur certains pri-miRNAs pour réguler négativement leur maturation par DCL1 (agissant, de ce fait, de manière similaire au complexe DRB4/DRB7.2), et d'autre part sur le promoteur des gènes de ces mêmes miRNAs pour induire ou réprimer leur transcription (Sun *et al*, 2015). Bien que cette hypothèse ne puisse pour le moment être clairement écartée, il est peu probable que DRB7.2 agisse de manière similaire à CDF2 au niveau des loci d'IR. En effet, à la différence de CDF2 (Fornara *et al*, 2009; Moreno-Risueno *et al*, 2007), DRB7.2 ne possède

pas d'homologie avec des facteurs de transcription ou de domaine de fixation à l'ADN, et des expériences préliminaires réalisées au laboratoire semblent indiquer que cette protéine ne présente pas d'affinité pour ce type d'acide nucléique.

Bien qu'informatives, les expériences de séquençage à haut-débit de sRNAs réalisées sur le mutant *drb7.2* ne permettent pas forcément la découverte exhaustive de l'ensemble des cibles de cette protéine (associée ou non à DRB4). En effet, les cibles qui ne seraient, soit pas exprimées dans les tissus analysés (fleurs), soit qui ne généreraient pas de sRNAs, peuvent ainsi nous avoir échappées. Afin d'avoir une vue globale de l'ensemble des ARN fixés par cette protéine, il serait donc intéressant de réaliser des expériences de RNA-seq sur les acides nucléiques immunoprécipités en association avec DRB7.2. Ces analyses nous permettraient d'identifier potentiellement d'autres fonctions de cette protéine.

#### 1.4.4 DRB7.2 ET LES EASI RNAs

Un autre rôle de DRB7.2, que nous avons pu mettre en évidence, est son effet sur la production des easiRNAs de 24 nt. Les easiRNAs sont une classe de sRNAs endogènes produites suite à la réactivation de TE dans des mutants *ddm1* (Decreased Dna Methylation 1) ou *met1*, dans les noyaux végétatifs des grains de pollen, dans les cellules différenciées ou encore en réponse à un stress (Slotkin *et al*, 2009; McCue *et al*, 2012; Creasey *et al*, 2014; Tanurdzic *et al*, 2008). Lors de cette réactivation, un changement dans la taille des siRNAs provenant de ces TE est observé. En effet, des siRNAs de 21-22 nt sont produits à la place de ceux de 24 nt normalement générés quand ces TE sont épigénétiquement inactivés (McCue *et al*, 2012; Creasey *et al*, 2014). La biogénèse de ces siRNA de 21-22 nt dépend de RDR6, DCL4 et DCL2 (McCue *et al*, 2012). La diminution des easiRNAs de 24 nt durant cette phase de réactivation est probablement liée à l'action de RDR6 sur les

transcrits de TE qui sont normalement convertis en ARNdb par RDR2 (Creasey *et al*, 2014)

Etant donné que DRB7.2 semble avoir une affinité supérieure pour les ARNdb de haut poids moléculaire (Montavon *et al*, soumis), l'absence d'effet sur la production des tasiRNAs s'explique probablement par la localisation subcellulaire distincte de leurs précurseurs (générés dans le cytoplasme, Kumakura *et al*, 2009) et de DRB7.2 (exclusivement nucléaire Montavon *et al*, soumis). Cependant, une certaine proportion de RDR6 semble également être localisée dans le noyau (Pontes *et al*, 2013) et c'est probablement cette fraction nucléaire qui est responsable de la production des précurseurs d'easiRNAs. De ce fait, DRB7.2 pourrait être capable de fixer ces précurseurs ARNdb lorsqu'ils sont générés (p.e. dans un mutant *ddm1*). Comme l'effet d'inhibition du complexe DRB4/DRB7.2 sur le clivage des précurseurs d'endoIR-siRNAs semble principalement agir sur DCL3, ceci pourrait expliquer pourquoi ceux sont majoritairement les easiRNAs de 24 nt qui se suraccumulent dans le double mutant *ddm1/drb7.2* comparé au simple mutant *ddm1* (Clavel *et al*, 2016). Ceci implique toutefois, que ces précurseurs RDR6-dépendant soient, en plus de leur clivage par DCL4/DCL2, en partie attaqués par DCL3.

Bien que l'accumulation des easiRNAs de 24 nt soit clairement augmentée en absence de DRB7.2 ou de DRB4, cette augmentation ne semble pas se traduire par un changement significatif du niveau d'accumulation des cibles des easiRNAs de 21-22 nt. Cependant, si les easiRNAs de 21-22 nt sont principalement chargés dans AGO1 pour réguler post-transcriptionnellement l'accumulation de ces cibles, il est peu probable que ce soit également le cas des easiRNAs de 24nt. Ainsi, il est toujours possible que ces easiRNAs de 24nt entraînent un changement dans le niveau de méthylation de leurs cibles qui restent à définir.

**CHAPITRE 2: ETUDE DU MODE  
D'ACTION DE DCL4 DANS LA  
BIOGENESE DES sRNAs ENDOGENES**

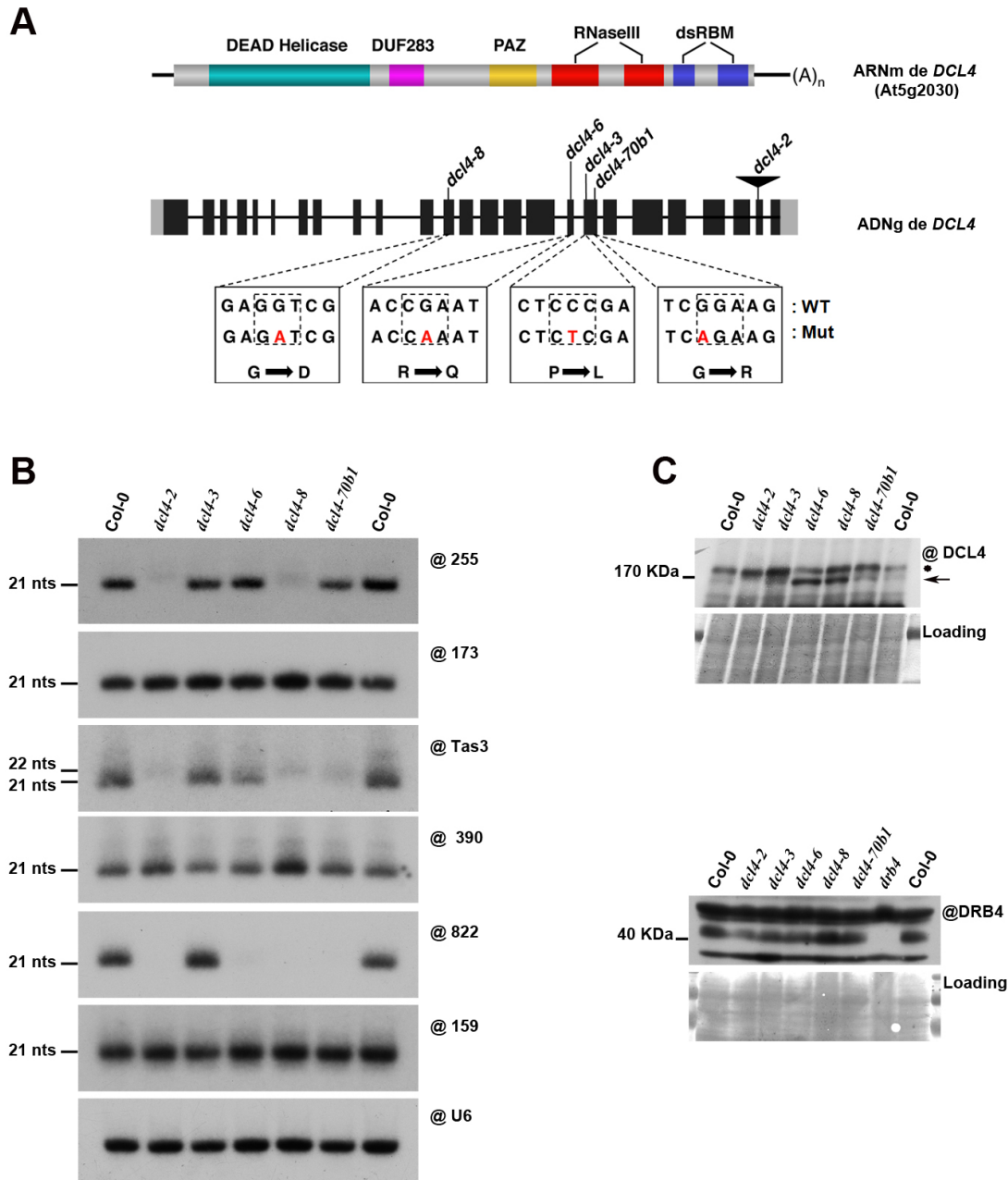


## 2.1 INTRODUCTION

Chez *A. thaliana*, DCL4 est l'enzyme DCL la plus polyvalente. En effet, elle participe à la production de plusieurs classes distinctes de sRNAs de 21 nt à partir de différents substrats ARNdb. Ainsi DCL4 est capable de reconnaître et cliver les intermédiaires de réplication viral (Blevins *et al*, 2006; Deleris *et al*, 2006), les précurseurs RDR6-dépendant de tasiRNAs (Xie *et al*, 2005; Gascioli *et al*, 2005), certains pri-miRNAs (Rajagopalan *et al*, 2006; Péliissier *et al*, 2011), les précurseurs d'IR endogène (ou exogène) (Dunoyer *et al*, 2005, 2007; Henderson *et al*, 2006), ainsi que ceux des easiRNAs (McCue *et al*, 2012; Creasey *et al*, 2014).

Cependant, bien que cette protéine soit le DCL le plus polyvalent chez *Arabidopsis*, son seul cofacteur connu à l'heure actuelle est la protéine DRB4 (Nakazawa *et al*, 2007; Fukudome *et al*, 2011). De manière intéressante, au cours de précédents cribles génétiques, un certain nombre d'allèles de *DCL4*, capable de découpler la production des ces différents sRNAs, a pu être identifié (Dunoyer *et al*, 2005, 2007; Cuperus *et al*, 2010). Ce découplage génétique, lié à la présence de différentes mutations ponctuelles au sein de *DCL4*, pourrait suggérer l'existence (i) de cofacteurs spécifiques à chacune des voies DCL4-dépendant et/ou (ii) de modes de reconnaissance distincts et spécifiques de chacun des différents substrats de cette protéine.

Afin de tenter de répondre à ces questions et ainsi de mieux comprendre le mode de fonctionnement de DCL4, une partie des travaux réalisés au cours de ma thèse a consisté à développer et tester des outils moléculaires basés sur certains des allèles découplants préalablement identifiés.



**Figure 2.1: Caractérisation des allèles découplant de DCL4**

(A) Représentation schématique de l'ARNm (haut) et de l'ADNg (bas) de DCL4. Sur cette dernière, les exons et les introns sont représentés, respectivement, par des rectangles noirs et des traits fins. Les rectangles gris représentent les séquences 5' et 3' non traduites. La localisation du T-DNA dans le mutant KO *dcl4-2* est indiquée par un triangle noir. Les transitions nucléotidiques ainsi que le changement d'acide aminé induit par le traitement EMS sur la séquence de DCL4 sont indiqués pour chacun des allèles étudiés. (B) Analyse par northern blot de l'accumulation des tasiRNAs (@255, @TAS3), du miRNA DCL4-dépendant miR822 (@822), et des miRNAs DCL1-dépendant miR173, miR390 et miR159 (@173, @390, @159) dans des plantes sauvages (Col-0) et les allèles *dcl4-3*, *dcl4-6*, *dcl4-8*, *dcl4-70b1*. Le snRNA U6 (@U6) est utilisé comme contrôle de charge. La visualisation des différents sRNAs a été réalisée par hybridations successives de la même membrane. (C) L'accumulation des protéines DCL4 et DRB4 a été visualisée par western blot dans les mêmes plantes qu'en (B). L'étoile indique la position d'une bande aspécifique et la flèche indique celle correspondant à la protéine DCL4. Loading : La coloration au bleu de coomassie des protéines sert ici de témoin de charge.

## 2.2 PRESENTATION ET CARACTERISATION DES ALLELES

### UTILISES

#### 2.2.1 PRESENTATION DES ALLELES DECOUPLANTS SELECTIONNES

Afin d'étudier le mode de reconnaissance de ses différents substrats, nous avons décidé d'utiliser quatre allèles distincts de *DCL4* (*dcl4-3*, *dcl4-6*, *dcl4-8*, et *dcl4-70b1*) (**Figure 2.1A**). Les allèles *dcl4-3*, *dcl4-6*, *dcl4-8* ont été identifiés au laboratoire au cours d'un crible génétique sur la lignée SUC:*SUL*, et sont tous incapables de générer les siRNAs de 21 nt provenant de l'IR exogène *SUL* (Dunoyer *et al*, 2005, 2007). L'allèle *dcl4-70b1* a, quant à lui, été isolé au laboratoire de J. Carrington au cours d'une mutagénèse visant à identifier de nouveaux facteurs impliqués dans la biogénèse des tasiRNAs jouant un rôle dans la régulation de la morphologie foliaire (TAS3) (Cuperus *et al*, 2010).

L'allèle *dcl4-3* présente une substitution d'une proline (P) en leucine (L) dans le domaine PAZ. La mutation portée par l'allèle *dcl4-6* conduit à la substitution d'une arginine (R) en glutamine (Q) également dans ce domaine. L'allèle mutant *dcl4-8* possède une mutation dans le domaine DEAD-Helicase conduisant au changement d'une glycine (G) en acide aspartique (D). Enfin, l'allèle *dcl4-70b1* présente une substitution d'une glycine (G) en arginine (R) située entre le domaine PAZ et le premier domaine RNase III (**Figure 2.1A**). Ainsi, on peut noter qu'à l'exception de *dcl4-8*, les trois autres allèles portent une mutation dans ou à proximité du domaine PAZ de DCL4. En parallèle, le mutant nul d'insertion T-DNA *dcl4-2* a été utilisé comme contrôle négatif (Xie *et al*, 2005b).

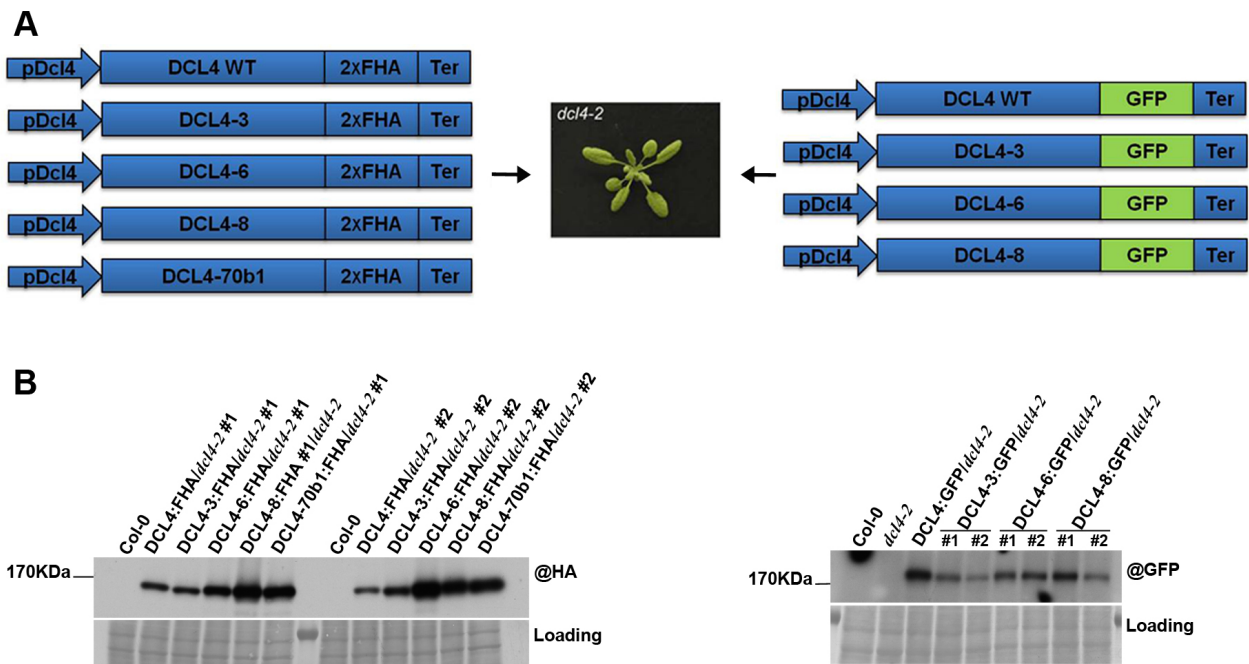
## 2.2.2 CARACTERISATION MOLECULAIRE DE CES ALLELES

Dans le but de caractériser moléculairement l'impact de ces différentes mutations sur l'accumulation des sRNAs endogènes DCL4-dépendant, les ARN totaux de feuilles de ces mutants ont été extraits et analysés par northern blot à l'aide de sondes spécifiques de sRNAs représentatifs de chacune de ces catégories (**Figure 2.1B**).

Dans le cas du mutant *dcl4-3*, seuls les siRNAs provenant de l'IR exogène *SUL* ne sont pas générés. En effet, des niveaux similaires d'accumulation à ceux d'une plante sauvage sont observés pour les tasiRNAs TAS1 (@255), TAS3 et le miRNA DCL4-dépendant miR822. A l'inverse, le mutant *dcl4-8* semble se comporter comme un mutant nul puisqu'aucun des sRNAs DCL4-dépendant ne semble être produit dans ces plantes (**Figure 2.1 B**; Dunoyer *et al*, 2007). Le mutant *dcl4-6*, quant à lui, est toujours compétent pour la biogenèse des tasiRNAs TAS1 (@255) et TAS3, mais n'est plus capable de produire miR822. Enfin le mutant *dcl4-70b1*, en plus d'être déficient dans la biogénèse des miRNAs DCL4-dépendant (@822), présente la particularité de ne plus être capable de produire les tasiRNAs TAS3 (AGO7-dépendant) mais de rester parfaitement compétent pour la biogénèse des tasiRNAs TAS1 (AGO1-dépendant) (**Figure 2.1 B**; Cuperus *et al*, 2010). Ces défauts d'accumulation des tasiRNAs TAS1 et TAS3, observés dans ces différents mutants, ne sont pas liés à un effet indirect de ces mutations sur la biogénèse des miRNAs dont dépendent ces sRNAs pour leur production (miR173-TAS1, miR390-TAS3) (Allen *et al*, 2005). En parallèle, la mutation *dcl4-70b1* a été introduite dans le système rapporteur SUC:*SUL* afin de déterminer l'effet de cette mutation sur la biogénèse des siRNAs DCL4-dépendant provenant d'IR exogène. Comme dans le cas des autres allèles, les plantes SS*dcl4-70b1* ne présentent pas le phénotype caractéristique lié au silencing du gène *SUL* de cette lignée transgénique. Cette absence de phénotype découle, là aussi, de l'absence des siRNAs de 21 nt provenant de l'IR *SUL*, indiquant que DCL4-70b1 est également déficient dans la biogénèse de cette catégorie de sRNAs.

Ces différences de production des sRNAs DCL4-dépendant, observés chez ces allèles, pourraient être dues à une variation dans l'accumulation de DCL4 ou de son cofacteur, DRB4. En effet, DCL4 pourrait avoir une affinité plus importante pour un ou plusieurs types de substrat donné. Ainsi, lorsqu'elle serait présente en plus faible quantité, cette différence d'affinité pourrait toujours lui permettre de produire efficacement les sRNAs provenant de ces substrats, mais plus ceux découlant de précurseurs pour lesquelles elle aurait une plus faible affinité. De même, la présence de DRB4 est plus ou moins requise pour la production des différents types de sRNAs DCL4-dépendant. Ainsi, l'accumulation des miRNAs DCL4-dépendant semble être plus sensible à l'absence de DRB4 que celle des tasiRNA (**Figure 2A**, Montavon *et al*, soumis). Afin d'écartier cette possibilité, les niveaux d'accumulation de DCL4 et de DRB4 dans les plantes exprimant ces différents allèles ont donc été détectés par western blot à partir d'extrait protéique total de fleurs (**Figure 2.1C**). Aucune diminution des niveaux de DCL4 ou DRB4 n'a pu être observée chez ces mutants (**Figure 2.1C**), indiquant que le découplage dans la production des différents sRNAs DCL4-dépendant n'est pas dû à une déstabilisation de ces deux protéines.

Ainsi, à l'exception du mutant *dcl4-8* dont la mutation dans le domaine DEAD-Helicase semble totalement abolir la capacité de clivage des ARNdb, chacun des autres allèles de *DCL4* semblent avoir sa propre spécificité de substrat. De plus, aucune de ces mutations dans, ou à proximité du domaine PAZ, ne semblent affecter l'activité RNase III de DCL4 *per se*. En effet, bien que les allèles DCL4-3, DCL4-6 et DCL4-70b1 ne soient plus capables de produire certains sRNAs DCL4-dépendant, ceux pour lesquels ils sont toujours compétents s'accumulent à des niveaux identiques à ceux observés dans une plante sauvage (**Figure 2.1B**). Cette caractérisation moléculaire permet donc d'émettre l'hypothèse que le découplage observé dans la biogénèse de ces différents sRNAs pourrait, soit provenir de la perte d'une interaction entre DCL4 et un cofacteur spécifiquement impliqué dans l'une de ces voies, soit d'un défaut de reconnaissance d'un ou plusieurs types de précurseur par ces mutants.



**Figure 2.2: Stratégie expérimentale pour l'étude du mode d'action de DCL4**

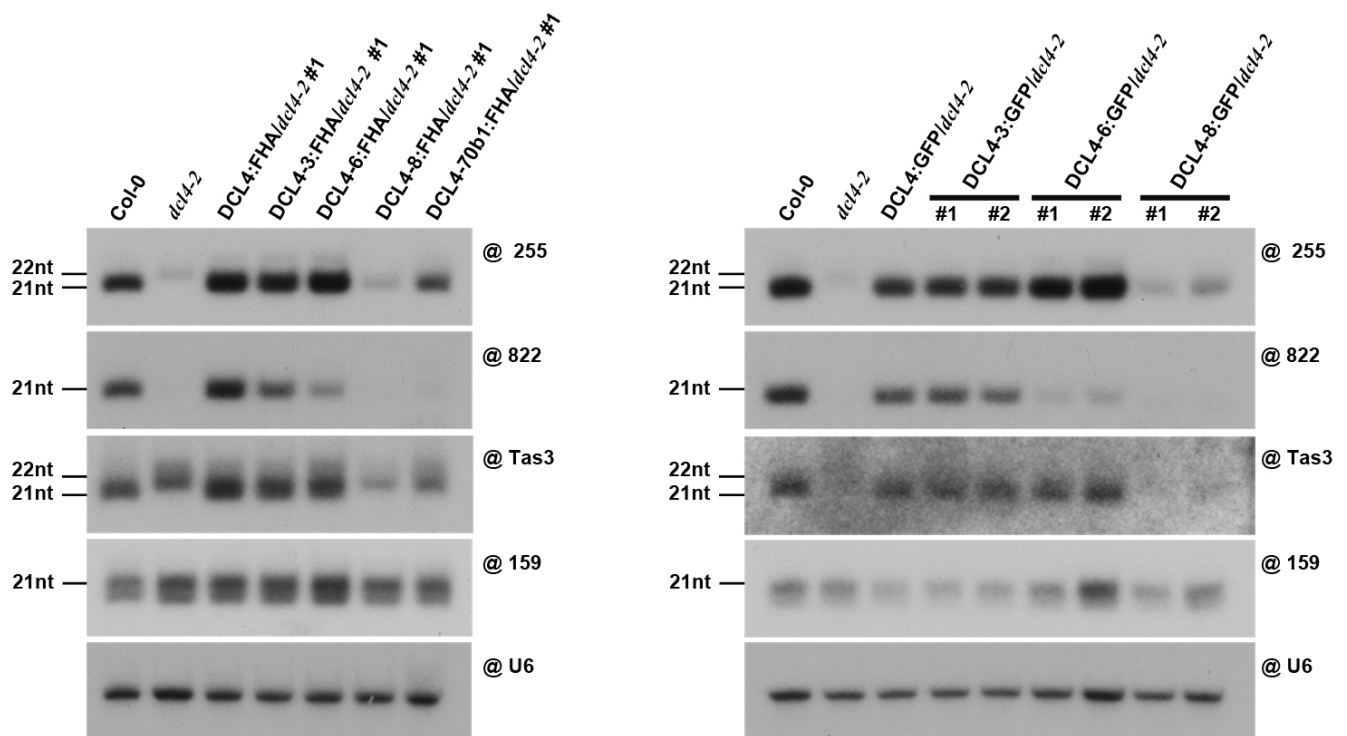
(A) Des plantes *dcl4-2* ont été transformées avec un transgène exprimant sous contrôle de son propre promoteur, une version étiquetée (FHA ou GFP) de la protéine DCL4 sauvage ou mutée (DCL4-3, DCL4-6, DCL4-8, DCL4-70B1). (B) A l'exception des plantes exprimant la protéine DCL4:GFP, deux lignées transgéniques ont été sélectionnées pour chaque construction et l'expression des protéines étiquetées a été analysée par western blot. Loading : La coloration au bleu de coomassie des protéines sert ici de témoin de charge.

## 2.3 DEVELOPPEMENT DES OUTILS MOLECULAIRES

### 2.3.1 STRATEGIE EXPERIMENTALE

Chez les protéines Dicer, le domaine PAZ intervient dans la fixation de l'extrémité 3' des précurseurs ARNdb (Zhang *et al*, 2004). En parallèle, une étude cristallographique du Dicer humain a révélé qu'une partie de ce domaine semblait être exposée à la surface de la protéine (Lau *et al*, 2012), ouvrant ainsi la possibilité qu'il puisse également intervenir dans des interactions de type protéine-protéine. La conservation des séquences en acides aminés entre le Dicer humain et DCL4 d'*Arabidopsis* étant faible, il ne nous a malheureusement pas été possible de modéliser, sur la structure tridimensionnelle disponible, la localisation des résidus mutés dans notre série allélique. De ce fait, ne pouvant écarter d'emblée aucune des deux hypothèses expliquant le découplage observé dans la production des sRNAs DCL4-dépendant, nous avons tenté de développer des outils nous permettant de les confirmer ou de les infirmer.

Ainsi pour tenter de comprendre les bases moléculaires du découplage observé chez ces mutants de *DCL4*, et ne disposant pas d'un anticorps @DCL4 natif de suffisamment bonne qualité pour réaliser des expériences d'immunoprécipitation (IP), nous avons introduit dans les plantes *dcl4-2*, des transgènes exprimant, sous contrôle de son propre promoteur, une version étiquetée (2xFLAG-2xHA (FHA) ou GFP) des protéines DCL4 sauvages ou mutantes (**Figure 2.2A**). Entre 6 (fusion GFP) et 15 (fusion FHA) lignées transgéniques ont été sélectionnées et les niveaux d'accumulation des protéines de fusion analysés par western blot afin d'isoler les transformants produisant les différents allèles de DCL4 étiquetés dans des proportions similaires à celles obtenues dans la lignée transgénique exprimant la version sauvage de DCL4 (**Figure 2.2B**).



**Figure 2.3: Caractérisation des outils moléculaires**

L'accumulation des tasiRNAs (@255, @TAS3) et des miRNAs (@822,@159) a été analysée par northern blot à partir d'ARN totaux de plantes sauvages (Col-0), *dcl4-2* et des différentes plantes transgéniques exprimant les protéines (DCL4, DCL4-3, DCL4-6, DCL4-8, DCL4-70b1) étiquetées avec FHA (gauche) ou GFP (droite). Le snRNA U6 (@U6) est utilisé comme contrôle de charge. Les sRNAs ont été visualisé comme dans la figure 2.1.



### 2.3.2 VALIDATION DES OUTILS MIS EN PLACE

Afin de vérifier que nos protéines étiquetées présentaient la même spécificité de substrat que les mutants correspondants, l'accumulation des différents sRNAs DCL4-dépendant, dans ces lignées, a été analysée par northern blot (**Figure 2.3**).

Cette analyse révèle que, quelque soit l'étiquette (FHA ou GFP), l'expression de la version sauvage de DCL4 complémente parfaitement les défauts moléculaires observés dans le mutant nul *dcl4-2*, produisant tous les sRNAs DCL4-dépendant à des niveaux similaires à ceux observés dans les plantes sauvages (Col-0) (**Figure 2.3**). D'autre part, cette analyse a également permis de mettre en évidence que l'accumulation des différents sRNAs DCL4-dépendant dans les lignées exprimant les allèles mutants de DCL4, mimait parfaitement celle observée dans les mutants correspondants (**Figure 2.1B et 2.3**). Ainsi, l'expression de DCL4-6:FHA/GFP permet de restaurer la production des tasiRNAs mais pas celle des miRNAs DCL4-dépendant. Les plantes exprimant DCL4-70b1:FHA/GFP n'accumulent toujours pas miR822 ou les tasiRNAs TAS3 mais produisent à nouveau les tasiRNAs TAS1 (@255), à des niveaux similaires à ceux obtenus dans les plantes sauvages. Enfin, si l'expression de DCL4-3:FHA/GFP restaure la production de l'ensemble des sRNAs endogènes DCL4-dépendant, à l'inverse, celle de DCL4-8:FHA/GFP ne permet pas de compléter les défauts d'accumulation de sRNAs observés dans le mutant *dcl4-2* (**Figure 2.3**).

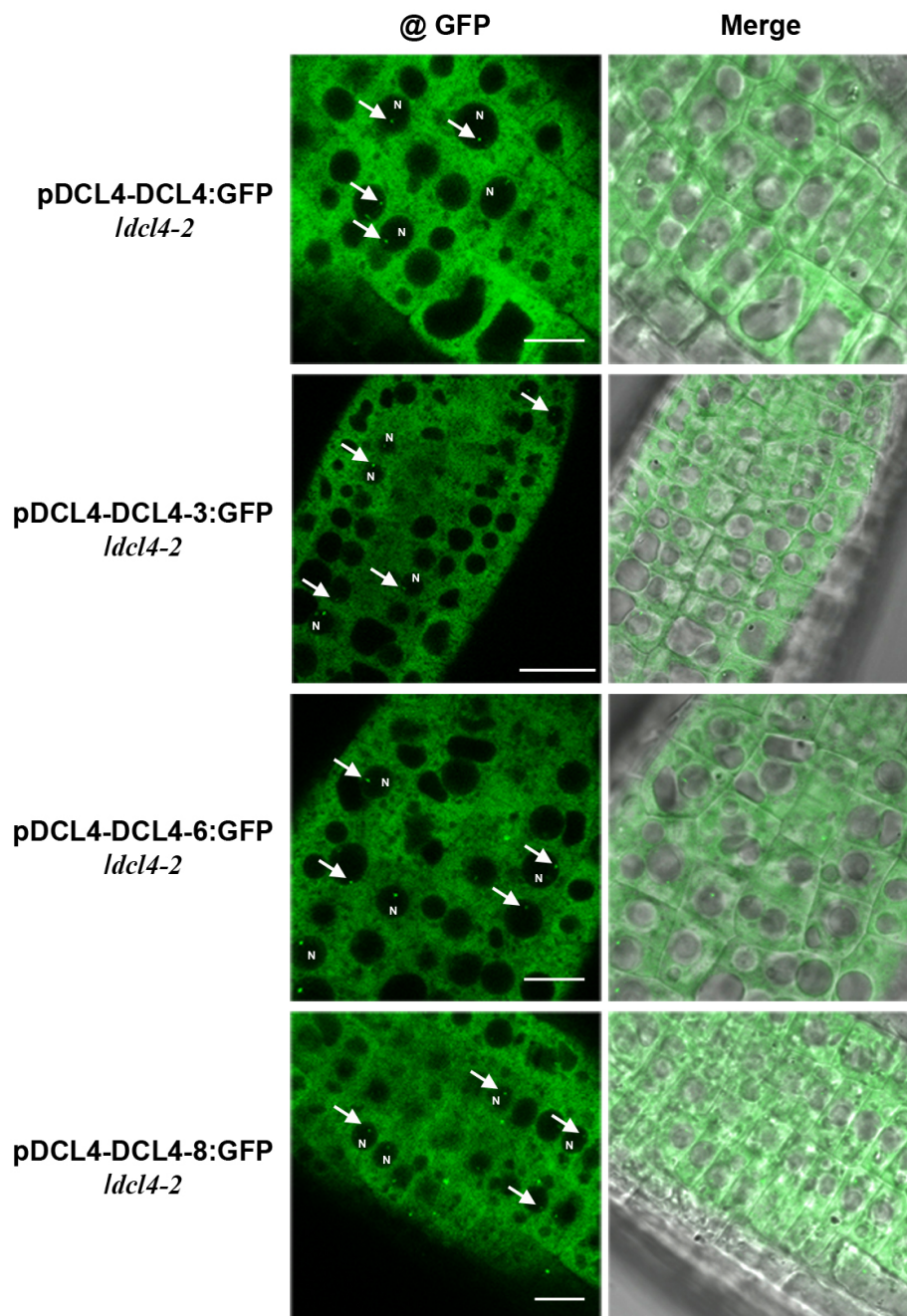
L'ensemble de ces résultats indique donc que chacune des protéines DCL4 étiquetées présente la même spécificité de substrat que les mutants dont ils découlent, validant ainsi ces outils moléculaires.

## 2.4 UTILISATION DES OUTILS MOLECULAIRES

### **2.4.1 LOCALISATION SUBCELLULAIRE DES DIFFERENTS ALLELES DE DCL4**

Jusqu'à présent, la localisation de DCL4 a été décrite comme étant purement nucléaire (Hiraguri *et al*, 2005). Cependant, étant donné le rôle de DCL4 dans la défense antivirale comme principal producteur de vsiRNAs à partir de nombreux virus se répliquant exclusivement dans le cytoplasme (Deleris *et al*, 2006; Bouché *et al*, 2006; Qu *et al*, 2008), et son rôle dans la production des tasiRNAs dont les précurseurs sont *a priori* générés dans ce même compartiment subcellulaire (Kumakura *et al*, 2009), cette localisation exclusivement nucléaire était relativement surprenante et contre-intuitive. Cependant, cette localisation a été observée dans des plantes transgéniques exprimant, sous contrôle du promoteur 35S, une version étiquetée de DCL4 (DCL4:GFP) dont la fonctionnalité n'a jamais été vérifiée (complémentation des défauts moléculaires d'un mutant nul *dcl4*), suggérant que cette localisation puisse ne pas être adéquat du point de vue de sa fonction biologique. De plus, le pattern de localisation nucléaire obtenu au cours de ces expériences (où DCL4:GFP est décrit comme étant présent de manière diffuse dans l'ensemble du nucléoplasme; Hiraguri *et al*, 2005) était totalement différent de celui observé par immunolocalisation à l'aide d'un anticorps @DCL4 natif (où DCL4 s'accumule sous forme de foyers discrets; Pontes *et al*, 2013).

Sur la base de ces observations, et de la fonctionnalité démontrée de nos protéines étiquetées, nous avons alors décidé (i) de vérifier la localisation subcellulaire de DCL4 et (ii) de déterminer si le découplage observé avec nos allèles pouvait s'expliquer par une différence de localisation par rapport à la protéine sauvage. Pour cela, des racines provenant de nos différentes lignées transgéniques ont été observées au microscope confocal.

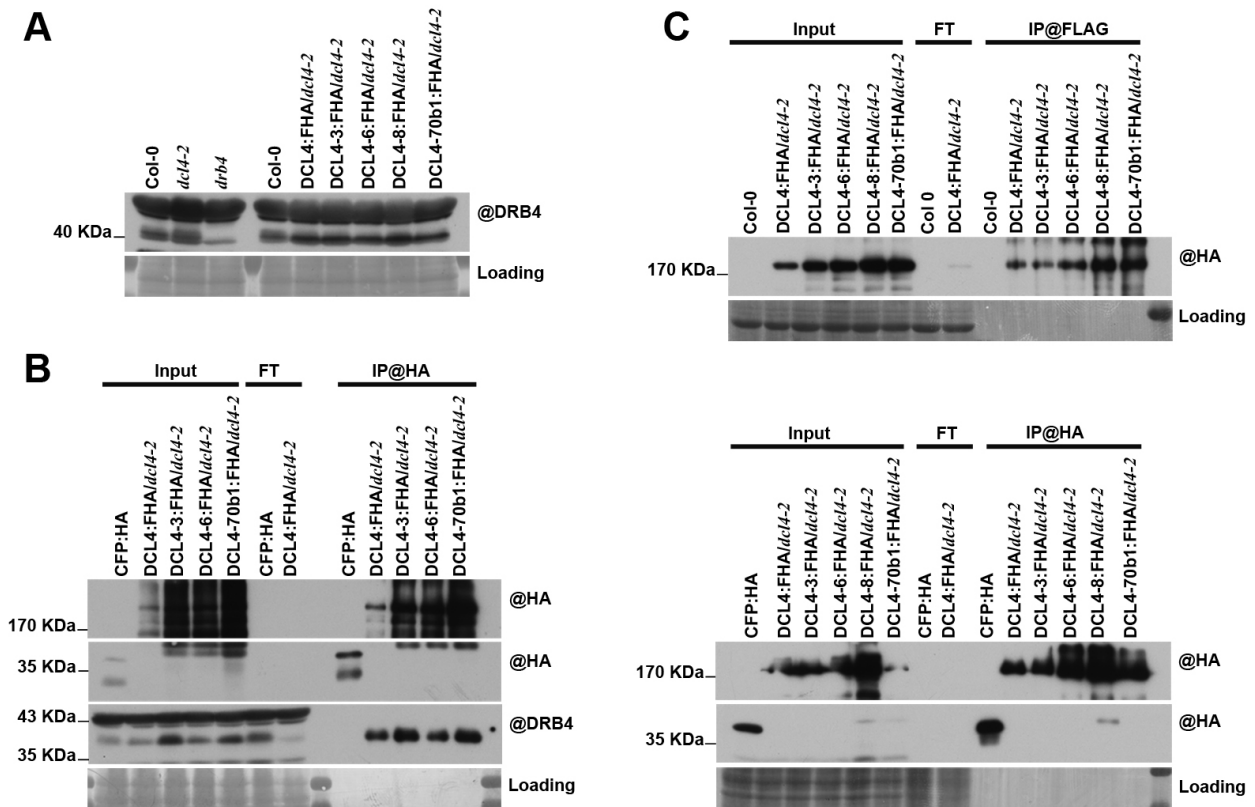


**Figure 2.4: Localisation subcellulaire des différentes allèles de DCL4**

Visualisation de la GFP au microscope confocal dans des pointes racinaires de plantes transgéniques exprimant la protéine sauvage (pDCL4-DCL4:GFP) ou les différentes protéines mutantes (pDCL4-DCL4-3/-6/-8:GFP). Les flèches indiquent quelques foyers nucléaires. N: Noyau. Barre d'échelle: 20  $\mu\text{m}$  (DCL4/4-3:GFP) et 10  $\mu\text{m}$  (DCL4-6/-8:GFP).

De manière surprenante, nos observations ont révélé que la localisation de la protéine DCL4 sauvage était très majoritairement cytoplasmique (**Figure 2.4**). En certaines occasions, cette protéine de fusion a également été détectée dans le noyau, mais exclusivement sous forme de foyers discrets (**Figure 2.4**), similaire à ceux décrits précédemment lors d'expériences d'immunolocalisation (Pontes *et al*, 2013). Ces expériences ayant été réalisées sur des noyaux purifiés, elles ne permettaient malheureusement pas de mettre en évidence la localisation cytoplasmique de DCL4 révélée au cours de nos observations. Ainsi, en accord avec sa fonction dans la défense antivirale et la production des tasiRNAs, DCL4 semble être localisé principalement dans le cytoplasme. De même, DCL4-3, DCL4-6 et DCL4-8 présentent tous une localisation subcellulaire identique à celle de la protéine sauvage, indiquant que le découplage dans la production des sRNAs DCL4-dépendant ne provient pas d'une différence de localisation de ces différents allèles. Cette conclusion ne peut être faite pour la protéine DCL4-70b1, bien qu'il soit probable que ce soit également le cas, n'ayant pas réussi à obtenir le transgène exprimant sa version étiquetée avec la GFP.

La différence de localisation subcellulaire observée entre nos lignées transgéniques et celle préalablement utilisée (Hiraguri *et al*, 2005) s'explique vraisemblablement par le type de promoteur utilisé pour l'expression de DCL4. En effet, dans notre cas, le promoteur endogène de DCL4, par opposition au promoteur constitutif 35S, a été utilisé. Nos résultats suggèrent donc qu'un ou plusieurs élément(s) présent(s) en *cis* sur le promoteur endogène de DCL4 seraient importants pour déterminer sa localisation subcellulaire. Bien que ces éléments *cis*, ainsi que leur mode de régulation, ne soient pas identifiés, ces résultats illustrent l'importance d'utiliser le promoteur endogène d'une protéine donnée, au cours d'expériences visant à caractériser sa fonction.



**Figure 2.5: Expériences de Co-IP et d'IP-MS à partir des différentes protéines DCL4**

(A). L'accumulation de DRB4 a été analysée par western blot à l'aide d'un anticorps reconnaissant la protéine native (@DRB4) dans des plantes sauvages (Col-0), *dcl4-2*, *drb4* ou les lignées transgéniques exprimant les différentes version de DCL4 étiquetées avec la FHA (DCL4 sauvage ou DCL4-3/-6/-8/-70b1:FHA). (B) Analyse par western blot des niveaux de CFP-HA (contrôle négatif) et des différentes versions de DCL4 étiquetées avec la FHA (@HA) dans la fraction totale (Input), le surnageant après immunoprécipitation (FT) et la fraction immunoprécipitée, à partir de plantes contrôles (CFP-HA) ou des lignées transgéniques présentées en (A). En parallèle, dans ces mêmes fractions, la présence de DRB4 a également été visualisée. (C) L'efficacité des expériences d'IP-MS (IP@FLAG haut et IP@HA bas) a été vérifiée par western blot à l'aide d'un anticorps spécifique (@HA). Deux expériences sur les six réalisées sont représentées ici. Loading : La coloration au bleu de coomassie des protéines sert ici de témoin de charge.

## 2.4.2 ETUDE DE L'INTERCATOME DES DIFFERENTS ALLELES DE DCL4

Afin de déterminer si le découplage observé découlait d'une perte d'interaction de certains de ces allèles de *DCL4* avec des cofacteurs spécifiques des voies de biogénèse des sRNAs pour lesquelles ces protéines ne sont plus compétentes, des expériences d'immunoprécipitation (IP) couplées à des analyses par spectrométrie de masse (MS) ont été réalisées. Pour cela, l'interactome de ces allèles a été défini et comparé à celui obtenu pour la protéine sauvage à partir des lignées transgéniques exprimant les versions étiquetées FHA.

En préalable à cette analyse, les niveaux de DRB4 dans les fractions IP de ces différents allèles ont été testés. En effet, comme décrit précédemment, la présence de DRB4 étant plus ou moins requise pour la production des différents types de sRNAs DCL4-dépendant (**Figure 2A**, Montavon *et al*, soumis), une première hypothèse était que les mutations portées par ces allèles pouvaient réduire (ou abolir) l'efficacité d'interaction entre DCL4 et son cofacteur. Ainsi, si les niveaux d'accumulation de DRB4 ne sont pas affectés par l'expression de ces versions mutées de DCL4 (**Figure 2.2B et 2.5A**), le découplage observé pourrait découler d'une diminution de la formation du complexe DCL4/DRB4. Afin de tester cette hypothèse, bien que peu probable étant donné la localisation de ces mutations sur la protéine DCL4, distinctes du domaine d'interaction avec DRB4 (Qin *et al*, 2010), les quantités de DRB4 immunoprécipitées par les différentes versions étiquetées de DCL4 ont été analysées. Si DRB4 n'est jamais retrouvée dans notre contrôle négatif (35S-CFP:HA), les quantités immunoprécipitées de ce cofacteur ne semblent pas varier entre les différents allèles de DCL4 (**Figure 2.5B**). Ces résultats excluent ainsi que le découplage observé découle d'un défaut d'interaction de nos allèles avec ce cofacteur. De plus ces résultats indiquent que l'interaction entre DCL4 et DRB4 n'est pas, à elle seule, suffisante pour permettre le clivage par DCL4 des différents précurseurs qu'elle reconnaît.

Expériences #1						
Fleurs						
AGI/Nom	CFP	DCL4:FHA	DCL4-3:FHA	DCL4-6:FHA	DCL4-8:FHA	DCL4-70B1:FHA
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT5G20320.1 /DCL4		184	783	452	965	1052
AT3G62800.1 /DRB4		168	333	168	245	547
AT1G24450.1/RTL4		45	114	52	100	122
AT2G28380.1/DRB2		11	4	4	5	3
AT4G00420./DRB7.2		8	3			2

Expériences #2						
Fleurs						
AGI/Nom	CFP	DCL4:FHA	DCL4-3:FHA	DCL4-6:FHA	DCL4-8:FHA	DCL4-70B1:FHA
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT5G20320.1 /DCL4		2370	2308	3106	ND	1174
AT3G62800.1 /DRB4		283	360	345	ND	367
AT1G24450.1/RTL4		128	128	79	ND	165
AT2G28380.1/DRB2		5	2		ND	2
AT4G00420./DRB7.2			14	1	ND	26

Expériences #3						
Fleurs						
AGI/Nom	Col 0	DCL4:FHA	DCL4-3:FHA	DCL4-6:FHA	DCL4-8:FHA	DCL4-70B1:FHA
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT5G20320.1 /DCL4		382	485	761	959	868
AT3G62800.1 /DRB4		43	43	53	45	49
AT1G24450.1/RTL4		12	15	10	12	12
AT4G00420./DRB7.2		2	1			

Expériences #4						
Fleurs						
AGI/Nom	CFP	DCL4:FHA	DCL4-3:FHA	DCL4-6:FHA	DCL4-8:FHA	DCL4-70B1:FHA
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT5G20320.1 /DCL4		1936	1328	953	ND	1774
AT3G62800.1 /DRB4		260	268	205	ND	323
AT1G24450.1/RTL4		189	101	45	ND	153
AT2G28380.1/DRB2		6	14	10	ND	18
AT4G00420./DRB7.2		2	17	3	ND	8

Expériences #5						
Plantules						
AGI/Nom	CFP	DCL4:FHA	DCL4-3:FHA	DCL4-6:FHA	DCL4-8:FHA	DCL4-70B1:FHA
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT5G20320.1 /DCL4		53	139	203	398	93
AT3G62800.1 /DRB4		53	94	88	196	36
AT1G24450.1/RTL4				5	12	

Expériences #6						
Plantules						
AGI/Nom	Col 0	DCL4:FHA	DCL4-3:FHA	DCL4-6:FHA	DCL4-8:FHA	DCL4-70B1:FHA
	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres	# Spectres
AT5G20320.1 /DCL4		55	84	99	240	183
AT3G62800.1 /DRB4		5	12	12	11	16

Figure 2.6: Analyse de l'interactome des différentes protéines DCL4

Analyses IP-MS des fractions immunoprécipitées à partir des échantillons présentés en figure 2.5 C. Chacun des différents tableaux représentent un résumé des différentes protéines d'intérêts retrouvées spécifiquement associées à la protéine DCL4 et aux différentes protéines mutantes pour les six répliques techniques réalisés. # **Spectres**: Nombre de spectre de fragmentation MS/MS : Un même peptide peut être fragmenté six fois au maximum.

Dans le but de définir de manière exhaustive l'interactome des différents allèles de DCL4, plusieurs répliques biologiques ont été réalisés à partir de deux types de tissus (fleurs et plantules) (**Figure 2.5C**) provenant des lignées transgéniques décrites précédemment. Comme attendu, l'analyse MS des protéines immunoprécipitées a permis de mettre à nouveau en évidence que DRB4 est la principale protéine interagissant avec DCL4, confirmant son rôle majeur de cofacteur (Hiraguri *et al*, 2005; Nakazawa *et al*, 2007; Fukudome *et al*, 2011). En effet, dans chacune de nos expériences, cette protéine occupe la première place des protéines spécifiquement associées à DCL4, et ce, quel que soit l'allèle testé (**Figure 2.6**).

En dehors de DRB4, aucune autre protéine n'ayant pu être identifiée dans l'intégralité de nos répliques biologiques et techniques, nous avons donc, dans un deuxième temps, recherché des protéines présentes dans, au moins, trois des six expériences réalisées. L'utilisation de ce filtre nous a ainsi permis d'identifier d'autres facteurs parmi lesquels RTL4/NFD2 (AT1G24450) et, de manière beaucoup plus marginale, DRB2 (AT2G28380) et DRB7.2 (AT4G00420) (**Figure 2.6**). Cependant, ces protéines étant également détectées dans les IP de chacun des allèles de DCL4 testés, elles ne représentent pas de candidat potentiel pouvant être responsable du découplage de la biogénèse des sRNA DCL4-dépendant observé.

RTL4/NFD2 (RNASE THREE-LIKE 4 / NUCLEAR FUSION DEFECTIVE 2) est une protéine possédant un domaine RNase III (Portereiko *et al*, 2006; Elvira-Matlot *et al*, 2016). Cette protéine semble impliquée dans la maturation du gamétophyte femelle (Portereiko *et al*, 2006), bien que son rôle dans ce processus biologique ne soit pas clairement défini. De part la présence de ce domaine RNaseIII et de son interaction reproductible avec DCL4, RTL4 semble donc être un candidat intéressant comme nouveau cofacteur de DCL4. L'étude de cette protéine pourrait permettre de révéler une nouvelle fonction de DCL4 au cours de la caryogamie.

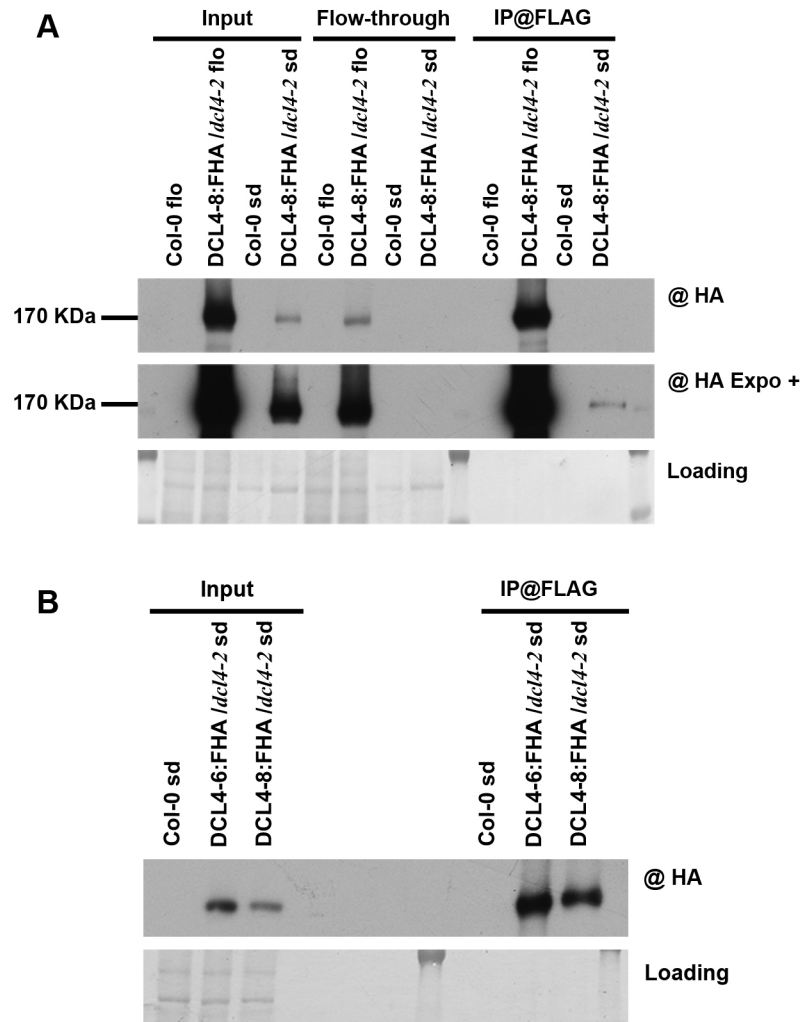
DRB2 est une des DRBs codées par le génome d'*Arabidopsis* et joue un rôle dans la biogénèse des miRNAs DCL1-dépendant au niveau de tissus



spécifiques (Eamens *et al*, 2012; Reis *et al*, 2015). De manière intéressante, la perte de DRB2 conduit également à une diminution de l'accumulation des miRNAs DCL4-dépendant tels que miR822 et miR839, mais n'affecte pas la production des autres sRNAs endogènes produits par ce DCL (Pélissier *et al*, 2011; Eamens *et al*, 2012). Ainsi, DRB2 semble être spécifiquement requise, bien que de manière plus modeste que DRB4, dans la production de miRNAs par DCL4. Le rôle de DRB2 au cours de cette biogénèse est, pour le moment, inconnu. La détection de DRB2 dans nos résultats d'IP-MS de DCL4 pourrait donc apporter un premier élément de réponse quant à son mode d'action. Cependant, ces résultats sont à prendre avec précaution étant donné le faible nombre de spectre de fragmentation MS/MS retrouvé (**Figure 2.6**) et le fait qu'ils soient en contradiction avec une étude préalable de l'interactome de DRB2 qui n'a pas permis de mettre en évidence d'interaction entre cette protéine et DCL4, ou son cofacteur DRB4 (Clavel *et al*, 2015). La question quant à la fonction de DRB2 dans la biogénèse des miRNAs DCL4-dépendant reste donc ouverte, et nécessitera de plus amples investigations.

Enfin, DRB7.2 a également été détectée dans certaines de nos analyses MS (**Figure 2.6**). Cependant, sur la base des résultats présentés dans le chapitre 1 de cette thèse, où aucune interaction entre DRB7.2 et DCL4 n'a pu être mise en évidence par BiFC (Clavel *et al*, 2016), co-IP ou IP-MS (Montavon *et al*, soumis), la présence de quelques peptides correspondant à DRB7.2 dans les IP de DCL4 constitue très probablement un exemple de faux positif qui découle de son interaction avec DRB4 (Clavel *et al*, 2016; Montavon *et al*, soumis).

En résumé, les résultats obtenus au cours de notre approche d'IP-MS décrits ci-dessus suggèrent fortement que le découplage observé dans la biogénèse des différents sRNAs DCL4-dépendant ne provient pas de la perte d'une interaction entre un cofacteur spécifiquement impliqué dans l'une de ces voies et nos allèles mutants. Nous avons alors décidé de tester si les défauts de biogénèse observés chez ces allèles pouvaient provenir d'un défaut de reconnaissance par ces mutants, d'un (ou plusieurs) type(s) de précurseur(s) spécifique(s) des ces sRNAs.



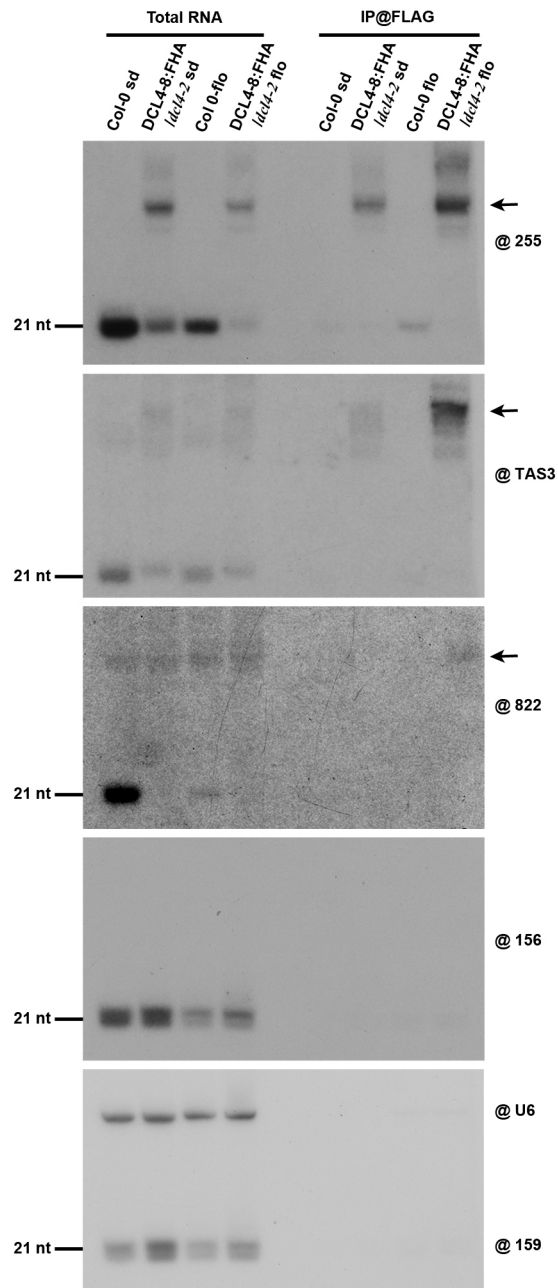
**Figure 2.7: Analyses des efficacités d'immunoprécipitation des expériences de RIP**

**A.** Analyse par western blot des niveaux de la protéines DCL4-8:FHA (@HA) dans la fraction totale (Input), le surnageant après immunoprécipitation (Flow-through) et la fraction immunoprécipitée (Ip@FLAG) à partir de fleurs (flo) ou de plantules (sd). Les ARN associés aux protéines immunoprécipitées ont été analysés par northern blot (cf. figure 2.8). **B.** L'efficacité de l'immunoprécipitation réalisée pour les expériences de RIP-seq à partir de plantes sauvages (Col-0) ou des lignées transgéniques exprimant les protéines DCL4-6:FHA ou DCL4-8:FHA a été confirmée par western blot à l'aide de l'anticorps @HA. Les ARN associées aux protéines immunoprécipitées ont été analysé par séquençage à haut débit (cf. figure 2.10). Loading: La coloration au bleu de coomassie des protéines sert ici de témoin de charge.

### 2.4.3 ÉTUDE DE LA FIXATION DES PRECURSSEURS DE sRNA DCL4-DEPENDANT PAR DCL4-6 ET DCL4-8

En vue de déterminer si l'absence de production de certains sRNAs DCL4-dépendant, observée dans notre série allélique de *DCL4*, provenait d'un défaut de reconnaissance de leur précurseur par ces mutants, nous avons réalisé des expériences de "RNA immunoprecipitation" (RIP) à partir de nos plantes transgéniques exprimant les protéines DCL4-6:FHA ou DCL4-8:FHA. Le choix de ces deux allèles s'est fait d'une part car ceux sont les mutants les plus fortement affectés dans la diversité des sRNA DCL4-dépendant produits, et d'autre part car ils portent leur mutation au niveau de deux domaines distincts (DEAD-Helicase pour DCL4-8, et PAZ pour DCL4-6). Après immunoprécipitation, les ARN associés à ces deux protéines ont été extraits et analysés, soit par northern blot à l'aide de sondes spécifiques des sRNAs DCL4-dépendant (DCL4-8), soit par séquençage à haut débit (RIP-seq; DCL4-6 et DCL4-8). En parallèle, des plantes sauvages Col-0, servant de contrôle négatif, ont été traitées de façon similaire.

Après vérification par western blot de l'efficacité d'immunoprécipitation de ces deux protéines étiquetées (**Figure 2.7A et 2.7B**), l'analyse par northern blot des acides nucléiques associés à DCL4-8 a permis de mettre en évidence la présence d'ARN de haut poids moléculaire sur ces membranes (**Figure 2.8**). En effet, un signal correspondant aux précurseurs du miRNA DCL4-dépendant (@miR822) et des tasiRNA TAS1 et TAS3 (@255, @TAS3) a pu être détecté spécifiquement dans la fraction IP obtenue avec la protéine DCL4-8:FHA et n'a jamais pu être observé dans notre contrôle négatif (**Figure 2.8**). De même, aucun signal correspondant aux miRNAs DCL1-dépendant (@156, @159) ou au contrôle de charge (@snRNA U6) n'a pu être mis en évidence dans la fraction immunoprécipitée par DCL4-8. Ainsi, la corrélation observée entre la détection spécifique d'ARN de haut poids moléculaire en association avec cette protéine, et l'absence de production par celle-ci des sRNAs DCL4-dépendant découlant de ces précurseurs, indiquent que les défauts de biogénèse de ces sRNAs ne

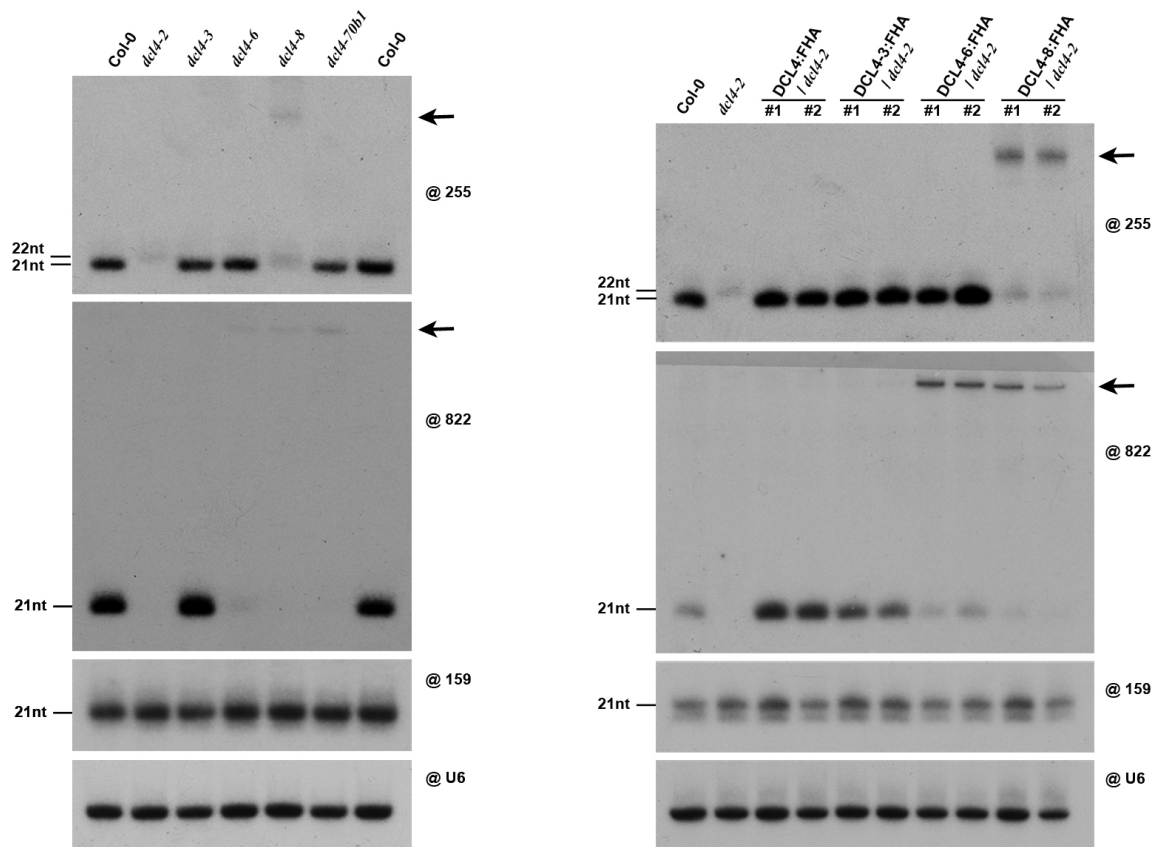


**Figure 2.8: Etude de la fixation des différents précurseurs de sRNAs DCL4-dépendant par DCL4-8**

Les ARN associés à la protéine DCL4-8 immunoprécipitée (RIP, cf. figure 2.7A) ont été analysés par northern blot à l'aide de sondes spécifiques des tasiRNAs (@255, @TAS3), du miRNA DCL4-dépendant miR822 (@822) et des miRNAs DCL1-dépendant miR156 et miR159 (@156, @159). Les flèches indiquent l'apparition de bande de haut poids moléculaire correspondant aux précurseurs des sARNs analysés. Le snRNA U6 (@U6) est utilisé comme contrôle de charge pour la fraction totale.

proviennent pas d'un défaut de fixation de ces précurseurs par DCL4-8. D'autre part, ces résultats valident, au moins pour DCL4-8, l'utilisation de cette approche pour les expériences de RIP-seq.

A ce stade, il est intéressant de noter qu'au cours des analyses par northern blot de notre série allélique de *dcl4* (**Figure 2.9A**), et des lignées transgéniques exprimant ces protéines mutantes (**Figure 2.9B**), nous avons également pu observer la stabilisation d'ARN de haut poids moléculaire dans les fractions "ARN total". En effet, en présence de DCL4-8, un signal de haut poids moléculaire est détecté avec les sondes révélant les sRNAs TAS1, et miR822, alors que seule cette dernière donne un signal en présence de DCL4-6 et DCL4-70b1. A l'inverse, aucun signal n'est détecté à ce niveau pour TAS1 et miR822 dans les plantes sauvages Col-0 ou les lignées transgéniques exprimant les versions étiquetées de DCL4 sauvage ou DCL4-3. Aucune visualisation ne peut malheureusement être faite avec la sonde révélant les tasiRNAs TAS3, car le bruit de fond généré par cette sonde masque totalement la détection d'un signal spécifique de haut poids moléculaire dans les fractions "ARN total". Malgré cela, la détection de ces ARN de grande taille, spécifiquement dans les mutants affectés pour la production des sRNAs correspondants, suggère fortement que ce signal découle de la stabilisation de leurs précurseurs, comme mis en évidence lors des expériences de RIP présentées ci-dessus (**Figure 2.8**). Le corollaire de ces observations est que DCL4-6 (et DCL4-70b1), tout comme DCL4-8, doit vraisemblablement être capable de se fixer aux précurseurs dont elle n'est plus capable de générer les sRNAs correspondants, et ainsi de les stabiliser. En accord avec cette hypothèse, il est intéressant de noter qu'aucun signal correspondant à ces précurseurs n'est détecté dans le mutant *dcl4-2*. Ceci s'explique probablement par le fait qu'en l'absence de protéine DCL4, ces précurseurs sont pris en charge et cliver par la protéine DCL2. A l'inverse, en présence d'une protéine DCL4 non fonctionnelle, mais toujours capable de se fixer à ses précurseurs, DCL2 ne peut remplir son rôle de substitution à DCL4.



**Figure 2.9: Stabilisation des précurseurs des sRNAs DCL4-dépendant.**

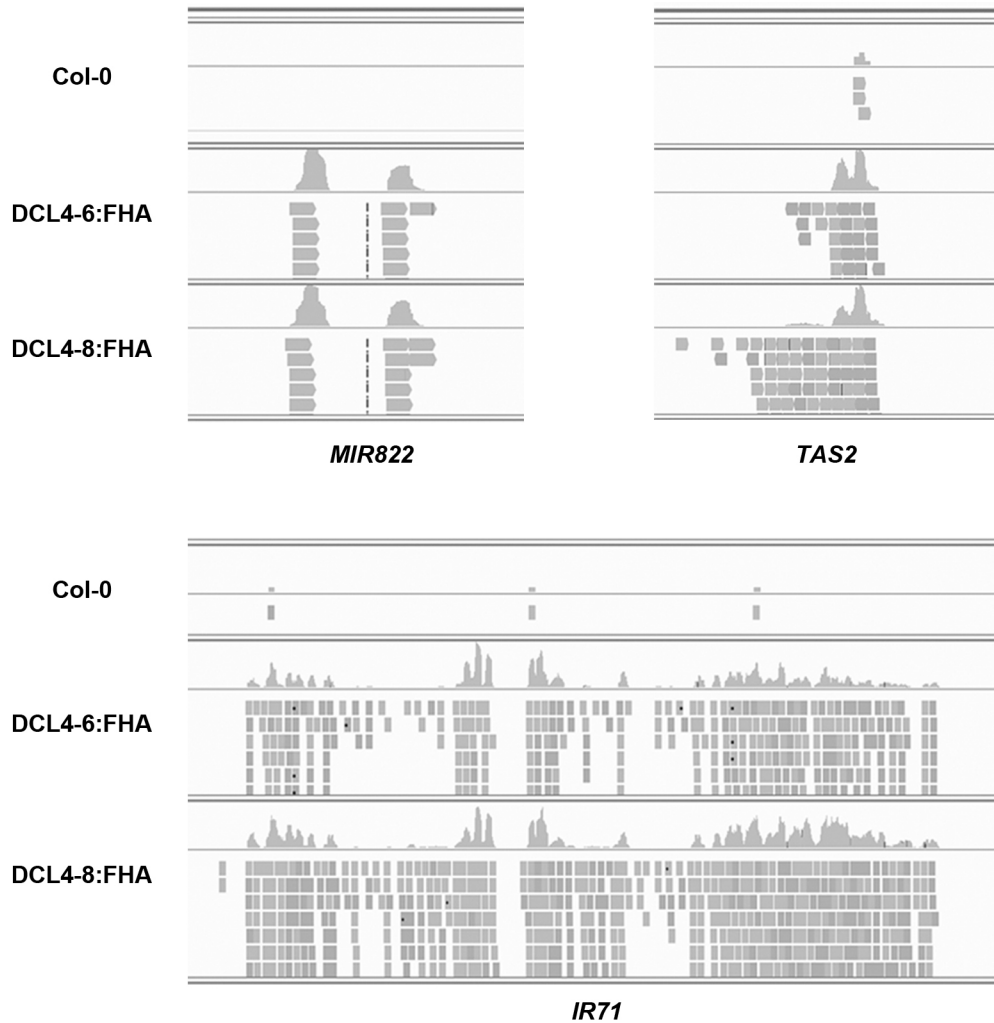
Analyses de l'accumulation du tasiRNA (@255) et du miRNA DCL4-dépendant (@822) dans les plantes présentées en figure 2.1B (gauche) et figure 2.3 (FHA) (droite). Les flèches indiquent l'apparition de bande de haut poids moléculaire correspondant aux précurseurs des sARNs analysés dans la fraction ARN total. Le miRNA DCL1-dépendant miR159 (@159) et le snRNA U6 (@U6) sont utilisé ici comme contrôle de charge.

Cette hypothèse a ensuite été validée par des expériences de RIP-seq (**Figure 2.7B** et **Figure 2.10**). Pour cela, les protéines DCL4-6:FHA et DCL4-8:FHA ont été immunoprécipitées à partir de plantules ou de fleurs, et les ARN associés à ces protéines ont été séquencés à haut débit (RIP-seq). Dans un premier temps, l'analyse bioinformatique des bibliothèques ainsi obtenues s'est focalisée sur les cibles connues de DCL4. Cette analyse a permis de révéler que les précurseurs des tasiRNAs (TAS1, TAS2, TAS3), de divers miRNAs DCL4-dépendant (dont miR822 et miR839; Rajagopalan *et al*, 2006), ainsi que ceux d'endoIR-siRNAs (notamment IR71, IR2039 et IR-AtGP10; Montavon *et al*, soumis), sont spécifiquement détectés dans les IP de DCL4-6 et DCL4-8, mais jamais ou très faiblement dans le contrôle négatif (**Figure 2.10**). Ces résultats sont en adéquation avec le rôle de DCL4 dans la biogénèse de l'ensemble de ces sRNAs. De manière intéressante, si les niveaux de précurseurs de miRNAs DCL4-dépendant observés dans ces bibliothèques sont similaires entre ces deux allèles mutants, la quantité de précurseurs correspondant aux divers tasiRNAs est, quant à elle, fortement enrichie dans les bibliothèques DCL4-8 (**Figure 2.10B**). Cet enrichissement coïncide avec l'incapacité de DCL4-8 à générer les tasiRNAs, là où DCL4-6 demeure compétent pour leur production. A l'inverse, les niveaux similaires de miRNAs DCL4-dépendant entre ces deux allèles illustrent l'incapacité de ces deux protéines à générer ce type de sRNAs. Ces résultats suggèrent fortement (i) que la stabilisation des précurseurs des divers sRNAs DCL4-dépendant observée par northern blot (**Figure 2.8** et **Figure 2.9**) provient de leur fixation par des mutants de DCL4 ayant perdu la capacité de les cliver spécifiquement et (ii) que le découplage des différents allèles dans la biogénèse des sRNAs DCL4-dépendant ne vient pas d'un défaut de fixation à leurs précurseurs.

**A**

	DCL4-6:FHA	DCL4-8:FHA		DCL4-6:FHA	DCL4-8:FHA		DCL4-6:FHA	DCL4-8:FHA
tasiRNA	TAS1C	TAS1A	miRNA	miR822	miR822	endIR-siRNA	IR71	IR71
	TAS2	TAS1B		miR839	miR839		IR2039	IR2039
	TAS3	TAS1C		miR783	miR783		IRAtGP10	IRAtGP10
	TAS3B	TAS2						
		TAS3						
	TAS3B							

**B**



**Figure 2.10: Identification des précurseurs de sRNAs associés à DCL4-6 ou DCL4-8**

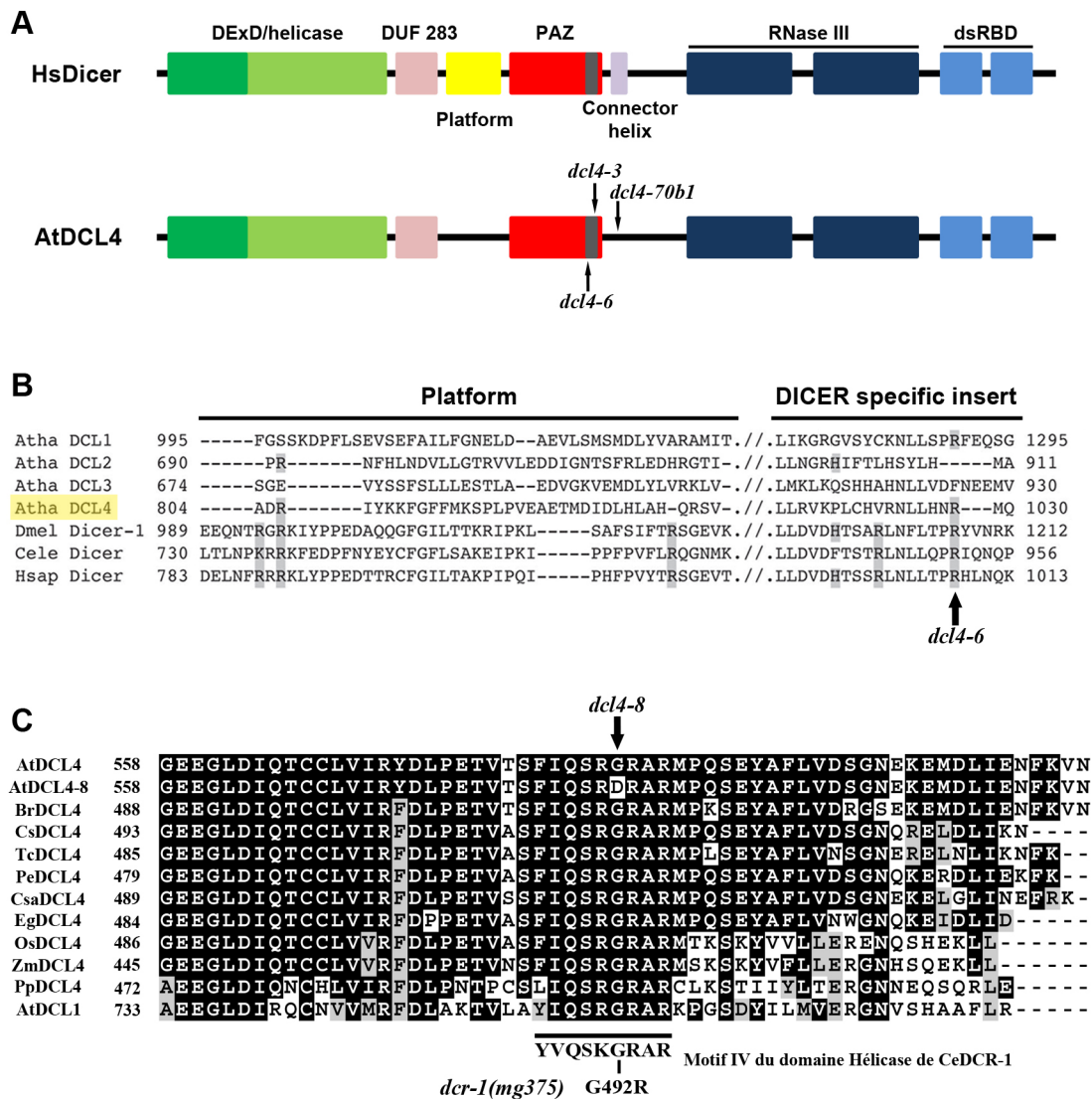
(A) Tableaux résumant les ARN spécifiquement associés ou fortement enrichis dans les fractions immunoprécipitées de DCL4-6 ou DCL4-8. Ces ARN sont triés en fonction des différentes classes de sRNAs qu'ils génèrent (tasiRNA; gauche, miRNA; milieu, endoIR-siRNA; droite). Ces tableaux constituent des listes non exhaustives des différents ARN identifiés à partir de fleurs. (B) Représentation schématique des séquences obtenues au cours des expériences de RIP-seq sur les protéines DCL4-6:FHA ou DCL4-8:FHA, au niveau de loci représentatifs des trois classes de sRNAs endogènes DCL4-dépendant (*MIR822* haut gauche, *TAS2* haut droit, *IR71* bas).



## 2.5 DISCUSSION, PERSPECTIVES

L'utilisation des outils moléculaires développés au cours de cette thèse, dans le but de mieux caractériser le mode d'action de DCL4, a permis de démontrer que le découplage observé chez les différents allèles de cette protéine ne semble pas provenir d'un changement de localisation subcellulaire (**Figure 2.4**), de la perte d'interaction avec des cofacteurs spécifiques (**Figure 2.6**), de la perte de l'activité RNase III à proprement parlée (**Figure 2.1B**), ou d'un défaut de fixation de DCL4 à ses différents substrats (**Figure 2.9**). Ainsi, l'hypothèse restante la plus probable est donc que les mutations portées par ces allèles affectent le positionnement correct du précurseur ARNdb fixé au sein de la protéine, empêchant son clivage efficace par ses domaines RNase III. Cette hypothèse est d'autant plus vraisemblable que plusieurs études, réalisées chez la drosophile, ont non seulement permis de démontrer le rôle essentiel du domaine PAZ des DICERs dans la reconnaissance et la fixation des extrémités 3' et 5' de leurs précurseurs (Zhang *et al*, 2004; MacRae *et al*, 2006, 2007; Welker *et al*, 2011; Park *et al*, 2011), mais aussi que la discrimination des différences d'extrémités, entre les différents précurseurs de sRNAs, contribue à la spécificité des substrats reconnus par DICER-1 (miRNA) et DICER-2 (siRNA) dans cet organisme (Welker *et al*, 2011; Park *et al*, 2011; Cenik *et al*, 2011).

Chez les plantes, si dans le cas de la biogénèse des miRNAs DCL1-dépendant, un certain nombre de déterminants structuraux impliqués dans la reconnaissance et le clivage des précurseurs de ces sRNAs ont pu être identifiés (Bologna *et al*, 2009; Mateos *et al*, 2010; Bologna *et al*, 2013), rien n'est actuellement connu concernant les différents substrats de DCL4. L'utilisation des allèles de cette protéine, caractérisés au cours de cette thèse, pourrait donc représenter un outil de choix afin d'adresser cette question, et ainsi mieux comprendre le mode de fonctionnement du DCL le plus polyvalent chez les plantes.



**Figure 2.11: Conservation des acides aminés mutés dans DCL4-6 et DCL4-8**

(A). Représentation schématique de l'organisation des différents domaines fonctionnels identifiés chez le Dicer humain (HsDicer) ou DCL4 d'*Arabidopsis* (AtDCL4). Le "Dicer specific insert" positionné dans le domaine PAZ est représenté en gris. (B). Alignement de séquence du domaine "platform" et du "Dicer specific insert" présent chez le Dicer humain (Hsap), de *C. elegans* (Cele), DICER-1 de *D. melanogaster* (Dme), et les 4 DCL d'*Arabidopsis* (Atha) (extrait de Fukunaga & Zamore 2014). Les acides aminés conservés, jouant un rôle, chez le Dicer humain, dans la formation de la "5' binding pocket", sont colorés en gris. L'acide aminé muté dans *dcl4-6* est indiqué par une flèche. (C). Alignement de séquence du domaine DEAD-helicase de divers DCL de plantes. La séquence de DICER-1 de *C. elegans* comprenant l'acide aminé muté dans le mutant *dcr-1(mg373)* est indiquée en dessous. Noir: conservé, Gris: mutation conservative, Blanc: non conservé. L'acide aminé muté dans *dcl4-8* est indiqué par une flèche. At; *Arabidopsis thaliana*, Br; *Brassica rappa*, Cs; *Citrus sinensis*, Tc; *Theobroma cacao*, Pe; *Populus euphratica*, Csa; *Cucumis sativus*, Eg; *Eucalyptus grandis*, Os; *Oriza sativa*, Zm; *Zea mays*, Pp; *Physcomitrella patens*

Quels déterminants structuraux (au niveau protéique ou acide nucléique) pourraient donc expliquer le découplage dans la production des divers sRNAs DCL4-dépendant chez les allèles DCL4-3, DCL4-6, DCL4-70b1? En l'absence d'informations précises concernant la structure tridimensionnelle de DCL4, ou des extrémités 5' et 3' de la plupart de ses différents substrats, nous ne pouvons, à l'heure actuelle, qu'émettre de simples suppositions.

A l'exception du mutant *dcl4-8*, les mutations présentes dans les autres allèles sont toutes localisées dans, ou à proximité du domaine PAZ (**Figure 2.1A**, **Figure 2.11A**). Chez les animaux, plusieurs études ont permises d'identifier, dans cette région de DICER, certains acides aminés pouvant contribuer à la spécificité de fixation des ARNdb. Ces études ont montré notamment que certaines tyrosines, positionnées dans le domaine PAZ, forment une poche (3' binding pocket), permettant la fixation des extrémités 3' sortantes (Zhang *et al*, 2004; MacRae *et al*, 2006, 2007). De même, l'existence d'une seconde poche (5' binding pocket), permettant la fixation des extrémités 5' phosphate des ARNdb a pu être mise en évidence. Cette "5' binding pocket" est formée par des acides aminés présents à la fois entre le domaine DUF283 et domaine PAZ (aussi appelé "platform"), dans le domaine PAZ et dans une structure en hélice positionnée entre les domaines PAZ et RNase III ("connector helix") (**Figure 2.11A**) (Park *et al*, 2011; Tian *et al*, 2014; Fukunaga *et al*, 2014). De plus, la présence d'un segment très conservé, et spécifique du PAZ des DICERs ("Dicer specific insert"), participe également à la fixation et au clivage efficace par ces protéines des précurseurs ARNdb (MacRae *et al*, 2006, 2007) (**Figure 2.11A**).

En émettant l'hypothèse que la structure de DCL4 suive un arrangement tridimensionnel similaire à celui décrit pour les DICERs animaux, tout du moins au niveau de ces domaines, il est intéressant de noter que les acides aminés mutés dans les allèles *dcl4-3*, *dcl4-6* et *dcl4-70b1* sont positionnés, soit au niveau du "Dicer specific insert", soit au niveau de la "connector helix" (**Figure 2.11A**).

Ainsi, l'arginine mutée dans le mutant *dcl4-6* est conservée chez de nombreux DICERs animaux mais également dans les protéines DCL1 et DCL4 chez les plantes (**Figure 2.11B**) (MacRae *et al*, 2006; Mukherjee *et al*, 2013). Dans le DICER humain, cet acide aminé, et d'autres arginines conservées et adjacentes, participent à la formation de la "5' binding pocket", nécessaire à la reconnaissance des extrémités 5' phosphates des pré-miRNAs (**Figure 2.11B**) (Park *et al*, 2011; Tian *et al*, 2014). De part la conservation de cette arginine, il est donc tentant de supposer que DCL4 (de même que DCL1) possède également une "5' binding pocket" nécessaire à la reconnaissance des pri-miRNA DCL4-dépendant (**Figure 2.11 B**). Ceci serait en accord avec les défauts de biogénèse observés dans le mutant *dcl4-6*, puisque celui-ci est déficient pour la production des miRNAs (**Figure 2.1 B, 2.3**). Le corollaire de cette hypothèse est que la "5' binding pocket" ne serait pas requise pour la fixation, le positionnement correct, et le clivage des précurseurs de tasiRNAs, puisque leur biogénèse n'est pas affectée par cette mutation (**Figure 2.1 B, 2.3**).

Dans le cas du mutant *dcl4-70b1*, l'acide aminé muté semble entraîner un découplage spécifique dans la production des deux grandes catégories de tasiRNAs. En effet, si ce mutant est toujours capable de produire les tasiRNAs TAS1 (AGO1-dépendant), il ne semble plus en mesure de produire les tasiRNAs TAS3 (AGO7-dépendant) (Cuperus *et al*, 2010) (**Figure 2.1B, 2.3**). Qu'est-ce qui différencie ces deux types de précurseurs ARNdb ? Dans le cas des transcrits *TAS1* (et *TAS2*), leur conversion en ARNdb par RDR6 (Vazquez *et al*, 2004b; Peragine *et al*, 2004), est précédé par la reconnaissance et le clivage de ces transcrits dans leur partie 5' par le complexe miR173/AGO1 (Montgomery *et al*, 2008b; Rajeswaran *et al*, 2012). Ceci va générer un ARNdb dont l'extrémité 5' sera, *a priori*, franche et c'est cette extrémité qui sera ensuite reconnue et clivée par DCL4 en siRNA de 21 nt (Gascioli *et al*, 2005; Xie *et al*, 2005b; Yoshikawa *et al*, 2005; Allen *et al*, 2005). Dans le cas des transcrits *TAS3*, ceux-ci vont être reconnus par le complexe miR390/AGO7 au niveau de deux sites de fixation (un en 5' et un en 3'), mais clivés par ce complexe uniquement au niveau du site situé en 3'

(Axtell *et al*, 2006; Montgomery *et al*, 2008a). RDR6 va alors convertir ce transcrit clivé en ARNdb jusqu'au site de fixation du complexe miR390/AGO7 situé en 5'. Ceci va donc générer un ARNdb dont l'extrémité 5' sera sortante sur environ 220 nt (Rajeswaran & Pooggin, 2012) rendant ainsi impossible l'attaque de cette extrémité par DCL4. C'est donc l'extrémité 2 nt 3' sortante qui sera reconnue et clivée par DCL4 en siRNA de 21 nt. L'acide aminé muté dans *dcl4-70b1* semble donc être requis pour la reconnaissance et le clivage efficace de précurseurs arborants ce type d'extrémité 3' sortante de 2 nt. Le corollaire de cette hypothèse est que les précurseurs de miRNAs DCL4-dépendant doivent également posséder ce genre d'extrémité, puisque leur biogénèse est également affectée dans ce mutant. Cette hypothèse est relativement probable étant donné la nature 2 nt 3' sortant des pré-miRNAs DCL1-dépendant, générée suite au premier clivage de leurs précurseurs (pri-miRNAs) par DCL1.

Les extrémités des précurseurs d'exoIR-siRNAs (et d'endoIR-siRNAs) n'étant pas connues, il est très difficile d'émettre une hypothèse permettant d'expliquer le découplage observé dans le mutant *dcl4-3*, qui n'est affecté que dans la biogénèse de ce type de sRNAs

Enfin, le mutant *dcl4-8* est, quant à lui, impacté dans la biogénèse de tous les sRNAs DCL4-dépendant (**Figure 2.1B, 2.3**) L'acide aminé muté dans cet allèle est localisé au niveau du domaine DEAD-hélicase (**Figure 2.1A, 2.11C**), ayant une fonction de translocase ATP-dépendant, permettant la translocation de la protéine le long de l'ARNdb chez les animaux (Cenik *et al*, 2011; Welker *et al*, 2011). Chez la drosophile un mutant ponctuel de DICER-2 touchant un acide aminé intervenant dans la fixation de l'ATP réduit, *in vitro*, la production des siRNAs provenant de longs ARNdb, mais pas celle des miRNAs indiquant que, chez les insectes, la production de ces derniers ne nécessite pas d'hydrolyse d'ATP (Welker *et al*, 2011). En parallèle, un mutant de DICER-1 (*dcr-1(mg375)*) chez *C. elegans*, exhibant exactement le même acide aminé muté que dans DCL4-8 (**Figure 2.11C**), est toujours capable de cliver les ARNdb ayant des extrémités 3' sortantes, caractéristiques des pré-miRNAs animaux, mais ne semble plus capable cliver les ARNdb avec des

extrémités franches ou 5' sortantes (Welker *et al*, 2011). Chez les plantes, l'activité de DCL1 et DCL4 est également ATP-dépendante, qui dépend de ce domaine DEAD-hélicase (Dong *et al*, 2008; Fukudome *et al*, 2011; Liu *et al*, 2012; Nagano *et al*, 2014). L'observation que le mutant *dcl4-8* soit déficient dans la biogenèse de tous les sRNAs DCL4-dépendant indique, qu'à l'inverse de ce qui a été décrit chez les animaux, leur production, *in vivo*, est strictement dépendante de l'hydrolyse d'ATP, suggérant que DCL4 est une enzyme exclusivement processive.

Afin de mieux comprendre les déterminants structuraux qui dictent la reconnaissance et le clivage des différents précurseurs par DCL4, les outils moléculaires générés durant ces travaux peuvent s'avérer précieux. En effet, l'immunoprécipitation des différentes versions étiquetées de DCL4 et leur utilisation dans des tests de "Dicing" *in vitro* (Fukudome *et al*, 2011) en présence de substrats ARNdb présentant des extrémités et/ou des structures distinctes devrait nous permettre de confirmer ou d'infirmer certaines des hypothèses exposées ci-dessus. De même, la mise au point des expériences de RIP à partir d'allèles de DCL4 incapables de cliver ses substrats ARNdb ouvre la porte à l'identification de nouveaux précurseurs reconnus par cette protéine au niveau de types cellulaires spécifiques, et qui aurait échappé à des analyses réalisées à partir de tissus complexes. De même ces expériences de RIP-seq couplées à des expériences de protection à différents types de RNase, pourraient permettre d'identifier des motifs structuraux particuliers reconnus par DCL4.

Enfin, de part le rôle crucial de DCL4 au cours de la défense antivirale, l'utilisation de ces allèles pour des expériences de RIP en contexte d'infection virale, pourrait s'avérer extrêmement intéressante pour identifier précisément les ARNdb viraux ciblés par cette protéine (intermédiaire de réplication ou structure secondaire intrinsèque à l'ARN viral). Cependant, afin de s'affranchir de l'étape d'amplification de la réponse antivirale, reposant sur la production de nouveaux ARNdb viraux par la machinerie cellulaire du RNA silencing, ces expériences devront être réalisées dans des mutants nuls de RDR6 (Moissiard *et al*, 2007; Wang *et al*, 2010).

***CONCLUSION***  
***GENERALE***

Le RNA silencing constitue à l'heure actuelle un mécanisme central dans la régulation de l'expression des gènes, le maintien et la plasticité du génome, ainsi que dans l'adaptation des organismes aux stress biotiques ou abiotiques. Chez les plantes, au cours de ces vingt dernières années, les différentes voies de ce mécanisme, ainsi que leurs principales fonctions et les protéines majeures qui les composent, ont pu être identifiées. De plus, au cours de ces travaux, il a également pu être constaté que ces différentes voies sont parfois interconnectées (Morel *et al*, 2000; Eamens *et al*, 2008a; Wang *et al*, 2011; Pontier *et al*, 2012) et qu'une même protéine peut agir dans différentes voies (Gascioli *et al*, 2005; Dunoyer *et al*, 2005; Henderson *et al*, 2006; Nakazawa *et al*, 2007; Qu *et al*, 2008; Péliissier *et al*, 2011; Montavon *et al*, soumis).

Ainsi, les travaux réalisés durant cette thèse ont non seulement permis de caractériser une nouvelle fonction pour la protéine DRB4, mais ont aussi révélé l'importance d'étudier la fonction d'une protéine donnée en fonction de son interactome et de sa localisation subcellulaire. En effet, l'étude simultanée de ces deux dernières caractéristiques pour la protéine DRB7.2 a permis de mettre en évidence un nouveau rôle de DRB4, considéré jusqu'à présent comme un simple cofacteur de DCL4, basé sur l'interaction spécifique observée entre ces deux DRBs et la localisation intracellulaire distincte de chacun des interactants de DRB4. L'existence de différents pools subcellulaires de DRB4 n'est pas sans rappeler celle mise en évidence pour la protéine AGO1 chargé en siRNAs ou en miRNAs (Schott *et al*, 2012). Ainsi, il est probable que cette caractéristique s'applique également à d'autres facteurs du RNA silencing. La localisation de DCL4 observée lors de mes travaux semble également aller dans ce sens. En effet, bien que majoritairement cytoplasmique, une certaine proportion de DCL4 est également observée sous forme de foyers discrets au niveau du noyau (Pontes *et al*, 2013; Montavon *et al*, soumis). Si la fonction de la fraction cytoplasmique de DCL4 est en accord avec la biogénèse des siRNAs antiviraux de cette protéine contre des virus se répliquant exclusivement dans ce compartiment subcellulaire, ainsi qu'avec la localisation



cytoplasmique d'autres facteurs clés impliqués dans la production des tasiRNAs (Kumakura *et al*, 2009), la fonction de la fraction nucléaire reste à déterminer. Elle pourrait jouer un rôle dans la production des easiRNAs (McCue *et al*, 2012; Creasey *et al*, 2014), impliqués dans la protection du génome lors de la réactivation d'éléments transposables. Les précurseurs de ces easiRNAs DCL4-dépendant étant produits par RDR6, ceci impliquerait qu'une fraction de RDR6 serait également nucléaire, comme cela a pu déjà être observé (Pontes *et al*, 2013). Cependant, la production de ces easiRNAs n'ayant lieu qu'au niveau de tissus particuliers (Slotkin *et al*, 2009; McCue *et al*, 2012; Creasey *et al*, 2014; Tanurdzic *et al*, 2008), la présence de DCL4 dans le noyau d'autres assises cellulaires suggère l'existence d'autres fonctions. Parmi celles-ci, on peut supposer que la production des miRNAs DCL4-dépendant se fasse exclusivement dans ce compartiment intracellulaire, comme cela a été décrit pour les miRNAs DCL1-dépendant (Fang & Spector, 2007; Song *et al*, 2007; Bologna *et al*, 2013).

En lien avec ce dernier point, les premiers résultats obtenus sur l'étude des différents allèles de DCL4 laissent entrevoir la possibilité de caractériser les éléments structuraux portés par les divers précurseurs ARNdb reconnus par cette protéine. L'identification de ces éléments permettrait de mieux comprendre, d'une part, le mode de reconnaissance de ces substrats par DCL4 et, d'autre part, d'appréhender les différences potentielles entre les substrats des quatre DCLs d'*Arabidopsis* qui semblent participer à la spécificité de la production des divers sRNAs de plantes par chacune de ces protéines. Même si cette spécificité est très probablement multifactorielle (localisation, cofacteurs spécifiques, taille des précurseurs), l'aspect structural des précurseurs joue probablement un rôle clé dans la reconnaissance du substrat par le DCL idoine, assurant ainsi le fonctionnement des différentes voies du RNA silencing et conditionnant *in fine* le rôle biologique correcte des sRNAs produits (Nagano *et al*, 2014; Blevins *et al*, 2015).

L'impact des infections virales sur des plantes à intérêt économique étant considérable, la caractérisation des déterminants structuraux reconnus

par DCL4 sur les précurseurs des vsiRNAs ouvre la possibilité de pouvoir modifier la séquence de DCL4 afin d'optimiser la reconnaissance et le clivage spécifique de ces substrats. Ceci pourrait potentiellement conduire à des applications biotechnologiques intéressantes en permettant, à terme, de générer un "super" DCL4 antivirale.

De même, la transposition de certaines des mutations de DCL4 décrites ci-dessus, sur les autres DCLs de plantes, pourrait permettre d'implémenter des approches similaires pour ces protéines. Ainsi, on peut noter que DCL1 et DCL2 possèdent, toutes les deux, l'arginine conservée au niveau du domaine DEAD-helicase (mutée dans *dcl4-8*). La mutation de cet acide aminé dans ces deux protéines pourrait donc, en théorie, permettre de stabiliser leurs substrats en association avec ces DCLs et leur identification par RIP-seq. En particulier, la caractérisation des substrats de DCL2 pourrait s'avérer intéressante afin de déterminer si cette protéine joue simplement une fonction redondante avec les autres DCLs d'*Arabidopsis*, ou si elle possède des fonctions discrètes qui lui sont propres et qui auraient, jusqu'à présent, échappées à nos analyses.

# ***MATERIEL***

## **SOUCHES BACTERIENNES**

### ***ESCHERICHIA COLI***

La souche d'*Escherichia coli* (*E. coli*), utilisée pour l'amplification des plasmides, est DH5 $\alpha$ <sup>™</sup> dont le génotype est : F<sup>-</sup>  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (rK<sup>-</sup>, mK<sup>+</sup>) *phoA supE44*  $\lambda$ <sup>-</sup> *thi-1 gyrA96 relA1*

### ***AGROBACTERIUM TUMEFACIENS***

La souche d'*Agrobacterium tumefaciens*, utilisée pour la transformation stable de plantes d'*Arabidopsis thaliana* et pour l'expression transitoire de protéines recombinantes dans des feuilles de *Nicotiana benthamiana*, est GV3101 (pMP90). Cette souche dont le génotype n'est pas disponible, contient le gène de résistance à la rifampicine sur son chromosome ainsi que le gène de résistance à la gentamycine sur son plasmide assistant (Ti) portant les gènes nécessaires à la transformation des plantes.

## PLANTES

### ARABIDOPSIS THALIANA

Les plantes utilisés lors de cette étude sont toutes issues de l'écotype Col-0 (sauvage ou mutant) (cf Chapitre 1 §1.1 et 1.2). Elles ont été cultivées, soit en condition de jour court (8h jour / 16h nuit, température 22°C / 18°C), soit en jour long (16h jour / 8h nuit, température 22°C / 18°C).

### NICOTIANA BENTHAMIANA

Les plantes de *N. benthamiana* ont été utilisées pour les expériences de BiFC. Ces plantes ont été cultivées en condition de jour longs (16h jour / 8h nuit, température 22°C / 18°C). Les mêmes conditions de cultures ont été appliquées après agroinfiltration.

## VECTEURS PLASMIDIQUES

### PGEM<sup>®</sup>-T EASY (PROMEGA<sup>®</sup>)

Ce vecteur permet le clonage direct de produits de PCR par la "Technologie TA". Ce vecteur linéarisé contient un résidu thymidyl sortant à chaque extrémité 3' ce qui permet la ligation de produits de PCR contenant un résidu adénosyl sortant en 3'. Ces résidus sont ajoutés de façon matricielle indépendante à l'extrémité 3' des produits de PCR par certaines ADN

polymérase ADN-dépendante du type Taq polymérase. La sélection des bactéries contenant l'insert se fait par le principe de l' $\alpha$  complémentation. Ce vecteur porte le gène de la  $\beta$ -lactamase qui confère aux bactéries transformées la résistance à l'ampiciline permettant de sélectionner les transformants. La cassette de clonage est insérée dans le gène codant pour le peptide  $\alpha$  de la  $\beta$ -galactosidase. Ce gène est sous le contrôle du promoteur lac inductible par l'isopropyl- $\beta$ -D-galactosidase (IPTG). En absence de l'insert, la complémentation de l'activité la  $\beta$ -galactosidase permet, après induction avec de l'IPTG, l'hydrolyse du X-gal (analogue su substrat naturel de l'enzyme) présent dans le milieu de culture. Cette hydrolyse conduit à la coloration du milieu des bactéries transformées par un vecteur sans insert. En revanche, la présence d'un insert dans la cassette de clonage modifie le cadre de lecture du peptide  $\alpha$ , supprimant la complémentation de l'activité  $\beta$ -galactosidase. Les colonies obtenues seront blanches.

### **PDONR™/ZEO (INVITROGEN®)**

Ce vecteur est utilisé comme vecteur donneur lors de clonage par la technologie Gateway®. Ce vecteur possède les sites de recombinaison attP1 et attP2 ce qui permet le clonage des fragments d'intérêt par recombinaison homologue. Le gène *ccdB*, codant pour un inhibiteur de gyrase, est présent entre les sites de recombinaison attP1 et attP2. Les bactéries sensibles à ce gène (comme les DH5 $\alpha$ ) ayant intégrées le vecteur non recombiné ne pourront pas croître. Ce vecteur possède un gène de résistance à la zéocyne permettant de sélectionner les bactéries ayant intégrées le vecteur.

### **PCAMBIA1300**

Ce vecteur est un vecteur binaire utilisé pour l'expression *in panta* de protéines d'intérêt fusionnées à une étiquette (ici la GFP) en partie C-

terminale. Ce vecteur contient le gène de résistance à la kanamycine (pour la sélection en bactérie) et le gène de résistance à l'hygromycine (pour la sélection en plante) et est utilisé lors de clonage par enzymes de restriction.

## **ANTICORPS, AMORCES OLIGONUCLEOTIDIQUE ET SONDÉS**

Les anticorps utilisés lors de ces travaux sont listés en annexe 2. Les oligonucléotides (clonage, génotypage et sonde) ont été commandés chez Eurofin MWG<sup>®</sup> et IDT<sup>®</sup> et les températures de fusion ont été calculées avec les différents programmes de commande en ligne des fournisseurs. Suivant les amorces une qualité basique ("Salt free") ou HPLC a été demandée. Les amorces oligonucléotiques utilisées lors de cette thèse qui ne sont pas présentées dans le chapitre 1 (cf. § 1.2) sont listées en annexe 2.

# ***METHODES***

## **PCR**

Lors de ce travail plusieurs types de PCR ont été réalisées. Pour ces différentes PCR l'appareil utilisé a été le PTC-225 Tetrad Thermal Cycler (MJ Research).

### **PCR POUR L'AMPLIFICATION DE GENES D'INTERET**

Pour ce faire, l'ADN polymérase ADN-dépendante PHUSION (Thermo Fisher Scientific) a été utilisée. Ces réactions ont été effectuées dans un volume final de 20 µl en présence du tampon PCR 1x (La composition de ce tampon est la propriété de Thermo Fisher Scientific et contient notamment l'ADN polymérase ADN-dépendante PHUSION™), et de 0,1 mM de chaque amorce. La première étape de la PCR est une étape de dénaturation de 30 s à 98°C, puis suivie de 35 cycles composés d'une étape de dénaturation de 15 s à 98°C, une étape de 30 s à la température d'hybridation la plus basse des amorces utilisées et une étape d'élongation à 72°C d'une durée variable suivant la taille du fragment à amplifier (environ 30 s/kb). Après les 35 cycles une dernière étape d'amplification de 3 min est programmée

## PCR POUR LA VERIFICATION DE LA PRESENCE D'INSERT DANS LES PLASMIDES OU POUR LES GENOTYPAGES

A cette fin la GoTaq polymérase (Promega®) a été utilisée. Le volume réactionnel est de 20 µl et contient : Du tampon de PCR dilué 1X, 2.5 mM de MgCl<sub>2</sub>, 0.1 µM de chaque amorce, 0.2 µM de chaque dNTPs et 0,075 U/µl d'ADN polymérase ADN-dépendante GoTaq. La PCR commence par une étape de dénaturation de 3 min à 94°C, suivie de 35 cycles composés d'une étape à 94°C pendant 30 s de dénaturation, de 30 s à la température d'hybridation la plus basse des amorces utilisées et une étape d'élongation à 72°C d'une durée variable suivant la taille du fragment à amplifier (1 min/kb). Après les 35 cycles une dernière étape d'amplification de 5 min est programmée.

### CLONAGE

Les ADN<sub>g</sub> et ADN<sub>c</sub> correspondants aux protéines nécessaires à ce travail, ont été clonés, soit à l'aide d'enzymes de restriction, soit grâce à la technologie Gateway®. Les ADN<sub>g</sub> et ADN<sub>c</sub> ont été amplifiés par PCR à partir respectivement, d'ADN ou d'ARN totaux provenant de bourgeons floraux d'*A. thaliana*. Pour les clonages par enzymes de restriction, les ADN<sub>g</sub> correspondants aux séquences codantes des protéines utilisées dans cette études ont été amplifiés, avec ou sans leurs séquences promotrices. Pour les clonages par recombinaison seul les ADN<sub>c</sub> correspondants à la séquence codante ont été amplifiés.



## CLONAGE GATEWAY®

*Réaction BP pour l'obtention de clones d'entrée:* Les réactions de recombinaison sont réalisées dans un volume final de 10 µl selon les instructions du fournisseur. Le volume réactionnel contient 50 ng d'ADNc flanqué des séquences attB, 150 ng de vecteur pDONOR™ZEO et 2 µl de mix BP Clonase™ II (La composition du tampon est la propriété du fournisseur). La réaction de BP recombinaison se fait à 25°C pendant 1h. Afin de stopper la réaction, un traitement avec 1 µl de protéinase K à 20mg/ml pendant 30 min à 37°C est réalisé. Deux µl de ce milieu réactionnel sont utilisés pour la transformation de bactéries compétentes.

*Réaction LR pour l'obtention de clones d'expression:* Pour le deuxième événement de recombinaison le volume réactionnel contient 150 ng de vecteur pDONOR™ZEO contenant l'insert, 150 ng de vecteur de destination et 2µl de LR Clonase™ II Mix (La composition du tampon est la propriété du fournisseur). La réaction de LR recombinaison se fait à 25°C pendant 1h. Afin de stopper la réaction, un traitement avec 1 µl de protéinase K à 20mg/ml pendant 30 min à 37°C est réalisé. Deux µl de ce milieu réactionnel sont utilisés pour la transformation de bactéries compétentes.

## CLONAGE PAR ENZYMES DE RESTRICTION

Les gènes sont amplifiés par PCR à partir d'ADNg en utilisant des amorces spécifiques (Cf. Chapitre 1§1.2) comportant à leurs extrémités les sites de restrictions Sall (amorce sens) et XmaI (amorce antisens). Le vecteur pGEM®-T Easy (Promega®) est linéarisé par l'enzyme EcoR1 et les fragments amplifiés sont introduits dans le plasmide par ligation avec la T4 DNA ligase (Promega®). Les plasmides sont alors amplifiés chez *E. Coli* (souche DH5α), puis extrait des bactéries. Après vérification de l'insert par séquençage, le

vecteur est digéré avec les enzymes de restriction Sall et XmaI est le produit de digestion est déposé sur gel d'agarose 1%. Après migration, le fragment correspondant à la taille attendue est découpé et purifié. Dans le même temps le vecteur d'expression en plante pCambia1300 est également digéré avec les enzymes Sall et XmaI. Le fragment purifié est ensuite introduit dans le vecteur d'expression linéarisé par ligation avec la T4 DNA ligase (Promega®).

## TRANSFORMATIONS

### TRANSFORMATION D'*ESCHERECHIA COLI* PAR CHOC

#### THERMIQUE

Des bactéries *E. coli* DH5 $\alpha$ <sup>TM</sup> compétentes (50  $\mu$ l) sont mises en présence de 10  $\mu$ l de milieu réactionnel de recombinaison BP ou LR ou de 10 ng de plasmide d'expression purifié ou 10  $\mu$ l de réaction de ligation *pGEM*<sup>®</sup>-*T Easy*. Le mélange est incubé 10 min sur la glace puis un choc thermique est appliqué en plaçant successivement le mélange pendant 30 s dans un bain marie à 42°C puis 5 min dans la glace. 1ml de milieu LB (10 g/L bactotryptone, 5 g/L d'extrait de levure, 5 g/L NaCl, pH 7,2) est ensuite ajouté au mélange puis celui-ci est incubé à 37°C durant une heure. Le mélange est ensuite étalé sur de boîte de Pétri de LB agar (milieu LB avec 1% d'agar (p/v)) additionné de l'antibiotique approprié pour la sélection des bactéries transformées. Les boîtes sont ensuite incubées 24 h 37°C.

## TRANSFORMATION D'AGROBACTERIUM TUMEFACIENS

### PAR LA METHODE D'ELECTROPORATION

Des bactéries *A. tumefaciens* GV3101 (pMP90) (50 µl) sont misent en présence d'environ 50 ng de plasmide à transformer dans une cuve d'électroporation (2 mm). Le mélange est incubé 10 min dans la glace puis placé dans un électroporateur et un courant électrique est ensuite appliqué (2500 V, 25 µF, 400 Ω). 1 ml de milieu LB est ensuite ajouté et le tout est incubé pendant 1 h à 2 h à 28°C. 25 et 50 µl du mélange sont ensuite étalés sur deux boîtes de LB agar additionné de rifampicine, de gentamycine et de l'antibiotique adapté à la sélection de la construction. Les boîtes sont incubées 48 h à 28°C.

## TRANSFORMATION D'ARABIDOPSIS THALIANA

Afin d'exprimer les protéines étudiées lors de ce travail *in planta*, des plantes d'*A. thaliana* sauvages ou mutantes ont été transformées suivant la méthode dite du "Floral Dip" (Clough & Bent, 1998). Des bactéries *A. tumefaciens* transformées avec la construction d'intérêt sont mises en culture à 28°C pendant 24 h à 200 rpm dans 250 ml de LB additionné de rifampicine, de gentamycine et de l'antibiotique adapté à la sélection de la construction. Les bactéries sont ensuite culotées par centrifugation à 5000 g pendant 15 min et sont ensuite homogénéisées dans 250 ml de milieu de transformation (5 % (p/v) Sucrose, 2,4 g/L ½ MS, 250 mg/L MES, 0,05% Sylvett, 200 µM Acetosyringone). Les fleurs d'*A. thaliana* sont ensuite plongées pendant 90 s dans le milieu de transformation puis le surplus de liquide resté sur les fleurs est ensuite épongé. Les plantes sont mises en maxi-serre à l'obscurité pendant 24 h puis en condition normale de culture.

## **EXTRACTION D'ADN PLASMIDIQUE ET PURIFICATION DE PRODUIT DE PCR**

Les préparations d'ADN plasmidique ont été réalisées à l'aide du kit "NucleoSpin® Plasmid Quickpure" de Macherey-Nagel suivant les instructions du fournisseur. Les purifications de produit de PCR ont été réalisées à l'aide du kit "NucleoSpin® Gel and PCR Clean-up" de Macherey-Nagel suivant les instructions fournies par le fournisseur.

## **EXTRACTION D'ADN D'*ARABIDOPSIS THALIANA***

Lors de cette étude, plusieurs techniques d'extraction d'ADNg ont été réalisées, suivant le type de tissu et suivant le type d'expérience.

### **EXTRACTION A PARTIR DE BOURGEONS FLORAUX**

0,5 mg de bourgeons floraux sont broyés dans un pilon et un mortier en présence de 700 µl de tampon d'extraction (250 mM Tris-HCl pH 7.5, 25 mM EDTA pH 8.0, 250 mM NaCl, 0,5 % SDS). Un volume de Phénol (pH 8.0)/Chloroforme iso amylique (25/24/1) est ajouté et les échantillons sont centrifugés pendant 10 min à 13000 rpm à la température de la pièce. La phase aqueuse est ensuite transférée dans un nouveau tube et un volume d'isopropanol 100 % est ajouté. Les échantillons sont mélangés 7 à 8 fois puis incubés à la température de la pièce pendant 10 min. Après une nouvelle

centrifugation le culot est lavé avec 350 µl d'éthanol 80 % puis repris dans 50 µl de tampon TE/RNase A (1 ml de 10 mM Tris pH 8.5, 1 mM EDTA, 10 µg/ml RNase A). Les échantillons sont ensuite incubés à 37°C pendant 30 min. Une nouvelle étape d'extraction au phénol/chloroform est réalisée afin d'enlever la RNase A. Les échantillons une fois traités sont repris dans 50 µl de tampon TE (Tris-EDTA). Les ADN ont été utilisés pour clonage.

## **EXTRACTION A PARTIR DE FEUILLES**

Deux jeunes feuilles sont placées dans un tube contenant 1 ml de tampon d'extraction (100 mM Tris-HCl pH 8.0, 50 mM EDTA pH 8.0, 500 mM NaCl, 0,07 % (v/v) 2-β-mercaptoéthano) ainsi que des billes d'acier (2 mm). Les échantillons sont broyés pendant 3 min à 30 Hz dans un broyeur (TissueLyser II Qiagen®). Une fois broyés, 130 µl de SDS 10 % sont ajoutés et les échantillons sont incubés 15 min à 60°C. Afin de précipiter les protéines, 300 µl d'acétate de potassium 5 M sont ajoutés aux échantillons et ceux-ci sont incubés 30 min dans la glace. Les tubes sont ensuite centrifugés 10 min à 10 000 rpm. Huit cent µl du surnageant sont prélevés et un volume d'isopropanol 100 % est additionnés. Après une nouvelle étape de centrifugation à 8 000 rpm pendant 10 min, Les culots sont séchés puis repris dans 50 µl d'eau distillée. Les ADN extrait sont ensuite utilisés pour génotypage.

# PREPARATION DES ADNc ET RT-QPCR

## RETROTRANSCRIPTION DES ARN

Pour toutes les expériences de rétrotranscription présentées, 2 µg d'ARN totaux ont été utilisés.

Traitement DNase: Afin d'éliminer les ADN contaminants dans les échantillons d'ARN, un traitement à la DNase est effectué. Le traitement est réalisé dans un volume de 10 µl en présence de 2 unités de DNase RQ1 (Promega®) et d'un dixième de volume de tampon DNase RQ1 10X (400 mM Tris-HCl pH 8.0, 100 mM MgSO<sub>4</sub>, 10 mM CaCl<sub>2</sub>). Les échantillons sont incubés pendant 30 min à 37°C. La réaction est ensuite stoppée par ajout de 1 µl de la solution de blocage RQ1 DNase Stop Solution (20 mM EGTA pH 8.0). 8 µl de ces ARN traités sont ensuite utilisés pour les réactions de rétrotranscription. Les 2 µl restants seront utilisés comme contrôle d'absence d'ADNg.

Rétrotranscription: Le milieu de rétrotranscription de 20 µl, renferme 8 µl des ARN traités, 25 µM d'oligodT, 5 ng/µl d'hexamères aléatoires (Fermentas), 0,5 mM de chaque dNTP (dATP, dTTP, dGTP, dCTP), 5 mM de DTT, 1X de tampon First-Strand (Invitrogen) (Tampon First Strand5X: 250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM MgCl<sub>2</sub>), 0,4 U/ml de RNase OUT™ (Invitrogen), 10 U/µl de Reverse Transcriptase SuperScript™ IV (Invitrogen). Dans un premier temps, l'ARN, les dNTPs, les hexamères aléatoires ainsi que les oligodT sont mélangés. Les ARN sont alors dénaturés à 65°C 5 min puis maintenues à 4°C pendant 5 min. Le tampon First Strand, le DTT et la RNase OUT™ sont ensuite ajoutés. La Reverse Transcriptase SuperScript™ IV est le dernier élément ajoutés, puis le mélange est incubé 10 min à 25°C puis 50 min à 50°C. La réaction est stoppée par une incubation de 15 min à 70°C. Les échantillons sont enfin dilués dans 20 µl d'eau "RNase Free". Afin de vérifier la qualité des ADNc et l'absence d'ADNg une PCR pour amplifier l'Actine 2 est réalisée.

## QRT-PCR

Chaque réaction de qRT-PCR a été réalisée en *triplicat* sur la même plaque. Chaque réaction se fait dans un volume de 10 µl et déposée sur une plaque 384 puits. Le milieu réactionnel comprend 1 µl d'ADNc, 0,5 µM de chacune des amorces, 1X de mélange LigthCycler480® SYBR Green I Master (Roche) renfermant le tampon, les dNTPs et la FastStart Taq DNA Polymerase. La réaction de PCR ainsi que le suivi de la fluorescence à 530 nm au cours des cycles d'amplification est réalisée dans le LigthCycler480® (Roche). Après une dénaturation de 10 min à 95°C, les échantillons sont soumis à 45 cycles de dénaturation de 15 s à 95°C, d'hybridation à 60°C pendant 20 s et d'élongation de 40 s à 72°C. A la fin de l'amplification, une courbe de dénaturation des produits est réalisée. Le logiciel le LigthCycler480® Software Version 1.5 (Roche) permet ensuite, a partir de ces courbes d'amplification, de déterminer les cycles seuil (Ct) pour chaque échantillon par la méthode du maximum des dérivées secondes. Le niveau de transcrit est normalisé par rapport à l'Actine 2.

## INFECTION VIRALE

### AGROINFILTRATION DE FEUILLES D'ARABIDOPSIS

#### THALIANA

Des bactéries *A. tumefaciens* transformées avec des vecteurs binaires contenant le *Turnip crinkle virus* (TCV) ou le *Tobacco rattle virus-PDS* (TRV-PDS) (cf Chapitre 1 § 1.2) sont misent en culture dans 10 ml de milieu LB additionné de rifampicine, de gentamycine et de l'antibiotique approprié à la construction pendant 24 h à 28°C sous agitation à 200 rpm. Les cultures sont

ensuite centrifugées 5 min à 5000 g et les culots sont repris dans 10 ml de milieu d'induction (10.5 g/l de  $K_2HPO_4$ , 4.5 g/l de  $KH_2PO_4$ , 1 g/l de  $(NH_4)_2SO_4$ , 0.5 g/l de Na-Citrate, 0.1 g/l MES, 0,4% de glycerol, 200  $\mu M$  Acetosyringone). Les bactéries sont misent ensuite à incubée à l'obscurité sous agitation pendant 4 à 5 h. Après une nouvelle étape de centrifugation les culots sont repris dans du tampon d'infiltration (10 mM MES, pH 5.86, 10 Mm  $MgCl_2$ ). L'absorbance à 600 nm de ces solutions est ensuite calculée et les cultures sont diluées afin que l'absorbance à 600 nm soit égale à 0,4 dans le cas du TCV, et égale à 0,8 dans le cas du TRV-PDS. Pour ce dernier, un volume égal de suspensions TRV-1 et de suspension TRV-2-AtPDS sont mélangés avant infiltration. Trois à quatre feuilles d'*A. thaliana*, âgées de 4 à 5 semaines, sont ensuite agroinfiltrées côté abaxial avec une seringue de 1 ml. Dans le cas des infections avec le TRV-PDS le VIGS est observé 14 jours après agroinfiltration et les feuilles systémiques sont récoltées. Dans le cas des infections avec le TCV les feuilles systémiques sont récoltées 14 jours après agroinfiltration.

## **BI-MOLECULAR FLUORESCENCE COMPLEMENTATION (BiFC)**

### **CONSTRUCTIONS PLASMIDIQUES**

Les ADNc de DRB4 et de DRB7.2 ont été rétrotranscrit à partir d'ARN totaux d'*Arabidopsis thaliana* (Col0). Les différentes constructions de ces protéines ont été clonées en utilisant la technologie Gateway (Invitrogen) dans les vecteurs binaires pBIFP1/2/3/4 (cf Chapitre 1 §1.1) permettant d'exprimer des protéines fusionnées soit en C-terminal, à la partie N ou C



terminal de la YFP (pBIFP1 et pBIFP4 respectivement) soit en N-terminal, à la partie N ou C terminal de la YFP (pBIFP2 et pBIFP3 respectivement). DRB4, DRB7.2 ont été exprimés sous le contrôle du promoteur 35S du CaMV. Les vecteurs contenant DCL4 ont été donnés par Marie-Aude TSCHOPP. Cette protéine a été exprimée sous le contrôle de son propre promoteur et cloné suivant la technique GATEWAY®. Le terminateur de ces constructions est le terminateur de la nopaline synthase (terminateur NOS). Tous les vecteurs binaires ont été transformés dans la souche GV3101 (pMP90) d'*Agrobacterium tumefaciens*. Les constructions GST utilisées en contrôle négatif ont été données par le Dr Lukhovitskaya. Afin d'augmenter le signal de BiFC le suppresseur viral de RNA silencing P19 a été co-exprimé (Voinnet *at al*, 2003) avec les différentes combinaisons testées.

## **AGROINFILTRATION DE FEUILLES DE *NICOTIANA***

### ***BENTHAMIANA***

Des bactéries *A. tumefaciens* transformées avec des vecteurs binaires contenant les différentes constructions sont mises en culture dans 10 ml de milieu LB additionné de rifampicine, de gentamycine et de l'antibiotique approprié à la construction pendant 24 h à 28°C sous agitation à 200 rpm. Les cultures sont ensuite centrifugées 5 min à 5000 g et les culots sont repris dans 10 ml de tampon d'infiltration (10 mM MgCl<sub>2</sub>) additionné d'Acetosyringone (200 µM final). Les bactéries sont mises ensuite à incubées sous agitation pendant 2 à 3 h. Après une nouvelle étape de centrifugation les culots sont repris dans du tampon d'infiltration (10 mM MgCl<sub>2</sub>). L'absorbance à 600 nm de ces solutions est ensuite calculée et les cultures sont diluées afin que l'absorbance à 600 nm soit égale à 1,5. Pour chaque combinaison testée un volume égal des différentes cultures est mélangé avant infiltration pour obtenir une OD finale à 0,5. Une à deux feuilles de *N. benthamiana*, âgées de 3 à 4 semaines, sont ensuite agroinfiltrées côté

abaxial avec une seringue de 1 ml. La fluorescence est observée 3 et 4 jours après agroinfiltration.

## OBSERVATION MICROSCOPIQUE

Les observations microscopiques ont été réalisées à l'aide d'un microscope confocal LSM780 ou LSM700 avec l'objectif x40 (Carl Zeiss). Les deux protéines fluorescentes utilisées ont été la GFP et la YFP. La GFP est excitée à 488 nm et son pic d'émission est de 508nm. La YFP est excitée à 512 et son pic d'émission est de 530 nm. Pour les observations des protéines recombinantes DRB7.2:GFP, DCL4:GFP ainsi que les différents allèles de DCL4, la fluorescence a été observée dans des racines d'*Arabidopsis thaliana* transgéniques âgées de 10 à 15 jours. Pour les observations de la YFP lors des expériences de BiFC, la fluorescence a été observée sur des disques foliaires issues de *Nicotiana benthamiana* âgée de 4 à 5 semaines. Les images provenant des lasers GFP, YFP et en lumière blanche (DIC, Differential Interference Contrast) ont été acquises simultanément. Toutes les images obtenues pour une expérience donnée ont été réalisées avec les mêmes paramètres d'intensité. Les images brutes en fichier lsm, ont ensuite été converties en fichier tif et une barre d'échelle a été ajoutée avec le logiciel Image J (NIH).

# EXTRACTION D'ARN ET DES PROTEINES TOTALES

## EXTRACTION D'ARN TOTAUX

Des plantules âgées de 2 à 3 semaines, des rosettes ou des bourgeons floraux d'*Arabidopsis thaliana* sont récoltées et congelés dans l'azote liquide. Les tissus sont ensuite broyés avec un pilon dans un mortier et les ARN totaux sont extraits avec du TRI-Reagent® (Sigma-Aldrich) selon les instructions du fabricant. Les ARN sont solubilisés dans du formamide 50% ou de l'eau stérile et leur concentration est calculé en utilisant un NanoDrop 2000 (Thermo Fisher Scientific). Avant analyse par northern blot, la qualité des ARN est vérifiée par électrophorèse en chargeant 1 à 2µg sur un gel d'agarose (1%) dans du tampon Tris-borate-EDTA 0,5x (45mM Tris-base, 45mM d'acide borique, 1mM EDTA)

## EXTRACTION DES PROTEINES

0,15 à 0.2g de tissu végétal sont broyés dans de l'azote liquide. La poudre ainsi obtenue est transférée dans un tube eppendorf puis homogénéisée dans 600µl du tampon d'extraction (0,7 M Saccharose, 0,5 M Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 0,1 M NaCl, 2 % (v/v) 2-β-mercaptoéthanol, inhibiteurs de protéases (Complete Mini, Roche)). Les protéines sont extraites par vortex en présence d'un volume de Biophénol pH 8.0 (Biosolve). Après centrifugation, les protéines présentes dans la phase phénolique sont précipitées pendant 16h à -20°C en présence de méthanol, acétate d'ammonium 0,1M. Après centrifugation et lavages des culots, les protéines sont reprises dans du tampon de resuspension (3 % SDS, 62.3 mM Tris-HCl pH 8.0, 10% (v/v) Glycerol), puis chauffées 2 min à 95°C. Leur concentration est déterminée par dosage colorimétrique (Bio-rad DC™ Protein Assay) par la méthode de Lowry, selon les instructions du

fournisseur. Le tampon de charge est ajouté (concentrations finales : 10 % glycérol, 4 % SDS, 62.3 mM Tris-HCl pH 6.8, 20 % (v/v) 2-β-mercaptoéthanol, Bleu de Bromophénol 0,003 %) à l'extrait protéique. Les échantillons sont incubés 2 min à 95°C puis stockés à -20°C avant d'être analysés par western blot.

## NORTHERN BLOT

### ELECTROPHORÈSE SUR GEL D'AGAROSE/FORMALDEHYDE

2,5 à 5 µg d'ARN totaux sont séparés par électrophorèse sur un gel d'agarose 1%, formaldéhyde 6% dans du tampon HEPES 1 X pendant 3 h à 130 V. Les ARN sont ensuite transférés par capillarité sur une membrane de nylon (Amersham Hybon-NX ; GE Healthcare), pendant 16 h dans du tampon SSC 20 X. Une fois transférés les ARN sont fixés par irradiation aux UVs (2x 1200mJ dans un Stratalinker 2400, Stratagene). Les ARN fixés sur la membrane sont ensuite mis à pré-hybrider dans du tampon d'hybridation (PerfectHyb™ Plus Sigma-Aldrich) pendant 30 minutes à 65°C avant l'addition de la sonde radioactive préalablement dénaturé 5 min à 95°C. L'hybridation est effectuée pendant 16 h sous agitation douce à 65°C. Afin d'éliminer les hybridations non spécifiques, la membrane est lavée, deux fois 20 min à 65°C avec du tampon de lavage à faible stringence (SSC 0,5X, SDS 0,1%) puis une fois avec du tampon de lavage à forte stringence (SSC 2X, SDS 0,1%) à 65°C pendant 20 min. L'accumulation des ARN viraux étudiés dans ce travail, est ensuite visualisée par autoradiographie.

## ELECTROPHORESE SUR GEL DE POLY-ACRYLAMIDE/UREE

15 à 40 µg d'ARN totaux sont séparés par électrophorèse sur un gel de 17,5% de polyacrylamide (ratio d'acrylamide et de bis-acrylamide 19 :1), urée (7,5M) dans du tampon Tris-borate-EDTA 0,5x pendant 4 heures à 80 V. Les ARN sont ensuite électro-transférés sur une membrane de nylon (Amersham Hybon-NX ; GE Healthcare), pendant 75 min à 300mA dans un tampon Tris-borate-EDTA 0,5x à 4°C, puis fixés soit, par irradiation aux UVs (2x 1200mJ dans un Stratalinker 2400, Stratagene), soit chimiquement 90 min à 60°C avec de l'EDC (*N*-(3-Diméthylaminopropyl)-*N'*-éthylcarbodiimide hydrochloride, Sigma-Aldrich). Les ARN fixés sur la membrane sont ensuite mis à pré-hybrider dans du tampon d'hybridation (PerfectHyb™ Plus Sigma-Aldrich) pendant 30 minutes à 42°C avant l'addition de la sonde radioactive préalablement dénaturé 5 min à 95°C. L'hybridation est effectuée pendant 16 h sous agitation douce à 42°C. Afin d'éliminer les hybridations non spécifiques, la membrane est lavée, trois fois 10 min à 50°C avec du tampon de lavage (SSC 2X, SDS 2%). L'accumulation des petits ARN étudiés dans ce travail, est ensuite visualisée par autoradiographie.

### **SYNTHESE ET PURIFICATION DES SONDES**

#### **RADIOACTIVES**

*Marquage à la PNK:* La réaction de marquage des sondes est réalisée pendant 1 h à 37°C en présence de  $\gamma$ -<sup>32</sup>P-ATP et de l'enzyme PolyNucléotide Kinase (PNK thermo scientific). Les sondes radioactives sont purifiées par centrifugation au travers d'une colonne Sephadex G-25 (GE Healtcare). Les oligonucléotides ainsi marqués correspondent aux séquences

complémentaires des siARN et miARN analysés dans cette étude et sont listés soit dans le chapitre 1 (cf. §1.2) soit en annexe 2.

Marquage par Random Priming: Les sondes radioactives sont synthétisées par Random Priming (Prime-a-gene, Promega®) selon les instructions du fournisseur, à partir d'une matrice d'ADN correspondant à la séquence nucléotidique que l'on souhaite détecter. La réaction est réalisée pendant 1h30 à 37°C en présence de [ $\alpha$ -<sup>32</sup>P]-dCTP et du fragment de Klenow de l'ADN polymérase ADN-dépendante I. Les sondes radioactives sont ensuite purifiées par centrifugation au travers d'une colonne Sephadex G-50 (GE Healthcare). Les oligonucléotides ainsi marqués correspondent aux séquences complémentaires des siARN analysés dans cette étude dans le chapitre 1 (cf. §1.2).

## WESTERN BLOT

### ELECTROPHORESE SUR GEL DE POLY-ACRYLAMIDE

Les protéines sont séparées par la technique du SDS-PAGE (ratio d'acrylamide et de bis-acrylamide 35,5 :1) dans du tampon de migration (Tris-glycine 1X, SDS 0,1 %). Les protéines sont ensuite électro-transférées sur une membrane de PVDF (Immobilon™-P, Merck Millipore) à 4°C à 300 mA pendant 1h30 dans du tampon de transfert (Tris-glycine 1X, Ethanol 20%). La membrane est ensuite lavée sous agitation pendant 2 min dans du tampon PBS-tween (PBS 1X, Tween 20 0,1%) puis saturée dans du tampon PBS-tween contenant 5% de lait. Après saturation la membrane est incubée à 4°C pendant 16h avec l'anticorps primaire, lavée 3 fois 10 min dans du tampon PBS-tween puis incubé 2h avec l'anticorps secondaire à température ambiante

sauf si l'anticorps primaire est couplée à une peroxydase. Enfin après trois lavages de 10 min dans du tampon PBS-tween, les protéines sont révélées par chimiluminescence avec le kit Lumi Light<sup>plus</sup> (Roche).Après révélation, les protéines transférés sont visualisé avec une solution de coloration (0,3% (p/v) bleu de Coomassie R250, 40 % ethanol, 10 % acide acétique) à base de bleu de Coomassie puis décoloré dans une solution de décoloration (40 % ethanol, 10 % acide acétique). La membrane est enfin rincée avec de l'eau déminéralisée.

## **IMMUNOPRECIPITATION DES ARN**

Lors de ce travail plusieurs types d'immunoprécipitations ont été utilisés suivant le type d'anticorps utilisé et suivant le type d'analyse réalisée.

### **IMMUNOPRECIPITATION AVEC BILLES D'AGAROSE**

0.3 g à 0,6 g de tissu végétal sont broyés dans de l'azote liquide et sont homogénéisés dans 1 ml de tampon de lyse (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 % (v/v) glycérol, 0,1 % NP-40, 1 µM MG132, 60 U de RNase OUT<sup>TM</sup> (Invitrogen), inhibiteurs de protéases (Complete Mini, Roche)) puis incubés 20 min à 4°C sur une roue. Les échantillons sont centrifugés deux fois à 13 200 rpm pendant 10 min à 4°C, et le surnageant est transféré dans un nouveau tube. Après la seconde centrifugation, un aliquot est prélevé pour constituer la fraction "Input" et 50 µl de billes d'agarose couplées à l'anticorps Anti-FLAG (EZview<sup>TM</sup> Red ANTI-FLAG<sup>®</sup> M2 Affinity Gel, Sigma-Aldrich-Aldrich) sont ajoutés au reste du surnageant. Les échantillons sont ensuite incubés pendant 4h à 4°C sur une roue, puis centrifuger 1 min à 12 000 g et un aliquot du surnageant est prélevé pour constituer la fraction "Flow-through". Les billes sont ensuite lavées 3 fois pendant 10 min sur une roue à 4°C avec 1 ml de tampon de lavage (50 mM Tris-HCl pH 7.5, 150 mM

NaCl, 10 % (v/v) glycérol, 0,1 % NP-40) puis repris dans 1 ml de TRI-Reagent® (Sigma-Aldrich). Les ARN associés aux protéines immunoprécipitées, sont extraits selon les instructions du fabricant. Les ARN sont ensuite repris soit dans 10 µl de formamide 50 % pour être analysé par northern blot soit dans 10 µl d'eau stérile pour être analysé par séquençage à haut débit.

Les protéines immunoprécipitées sont extraites en prelevant 350 µl de la phase phénolique provenant de l'extraction des ARN et sont précipitées avec de l'acétone 100 %. Les protéines sont reprises dans 35 µl de tampon de resuspension (3 % SDS, 62.3 mM Tris-HCl pH 8.0, 10% (v/v) Glycerol) et du tampon de charge est ajouté (concentrations finales: 10 % glycérol, 4 % SDS, 62.3 mM Tris-HCl pH 6.8, 20 % (v/v) 2-β-mercaptoéthanol, Bleu de Bromophénol 0,003 %). Les protéines sont enfin analysées par western blot afin de vérifier l'efficacité d'immunoprécipitation.

## **IMMUNOPRECIPIATAION AVEC BILLES MAGNETIQUES**

0.1 g à 0,3 g de tissu végétal sont broyés dans de l'azote liquide et sont homogénéisés dans 1 ml de tampon de lyse (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, 100 U de RNase OUT™ (Invitrogen), inhibiteurs de protéases (Complete Mini, Roche)) puis incubés 30 min à 4°C sur une roue. Les échantillons sont centrifugés deux fois à 13 200 rpm pendant 10 min à 4°C, et le surnageant est transféré dans un nouveau tube. Après la seconde centrifugation, un aliquot est prélevé pour constituer la fraction "Input" et 50 µl de billes magnétiques couplées à l'anticorps GFP (Miltenyi Biotec) sont ajouté au reste du surnageant. Les échantillons sont ensuite incubés pendant 30 min à 4°C sur une roue, puis passés sur une colonne (M column Miltenyi Biotec) préalablement préparée avec du tampon de lyse. Un aliquot des échantillons passés est prélevé pour constituer la fraction "Flow-trough". Les billes sont lavées deux fois avec 500 µl de tampon de lyse puis une fois avec 100 µl de tampon de lavage (20 mM Tris-Hcl pH



7.5) et sont éluées avec 2 fois 500 µl de TRI-Reagent® (Sigma-Aldrich) préchauffé à 50°C. Les ARN associés aux protéines immunoprécipitées, sont extraits selon les instructions du fabricant. Les ARN sont ensuite repris dans 10 µl de formamide 50 % pour être analysé par northern blot.

Les protéines immunoprécipitées sont extraite en prelevant 350 µl de la phase phénolique provenant de l'extraction des ARN et sont précipitées avec de l'acétone 100 %. Les protéines sont reprises dans 35 µl de tampon de resuspension (3 % SDS, 62.3 mM Tris-HCl pH 8.0, 10% (v/v) Glycerol) et du tampon de charge est ajouté (concentrations finales : 10 % glycérol, 4 % SDS, 62.3 mM Tris-HCl pH 6.8, 20 % (v/v) 2-β-mercaptoéthanol, Bleu de Bromophénol 0,003 %). Les protéines sont enfin analysées par western blot afin de vérifier l'efficacité d'immunoprécipitation.

## **IMMUNOPRECIPITATION DES PROTEINES**

Lors de ce travail plusieurs types d'immunoprécipitation ont été utilisés suivant le type d'anticorps utilisé et suivant d'analyse réalisée.

### **IMMUNOPRECIPITATION AVEC BILLES D'AGAROSE**

0.3 g à 0,6 g de tissu végétal sont broyés dans de l'azote liquide et sont homogénéisés dans 1 ml de tampon de lyse (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 % (v/v) glycérol, 0,1 % NP-40, 1 µM MG132, inhibiteurs de protéases (Complete Mini, Roche)) puis incubés 20 min à 4°C sur une roue. Les échantillons sont centrifugés deux fois à 13 200 rpm pendant 10 min à 4°C, et le surnageant est transféré dans un nouveau tube. Après la seconde

centrifugation, un aliquot est prélevé pour constituer la fraction “Input” et 50 µl de billes d’agarose couplées à l’anticorps Anti-FLAG (EZview™ Red ANTI-FLAG® M2 Affinity Gel, Sigma-Aldrich) sont ajoutés au reste du surnageant. Les échantillons sont ensuite incubés pendant 4h à 4°C sur une roue, puis centrifugés 1 min à 12 000 g et un aliquot du surnageant est prélevé pour constituer la fraction “Flow-through”. Les billes sont ensuite lavées 3 fois pendant 10 min sur une roue à 4°C avec 1 ml de tampon de lavage (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 % (v/v) glycérol, 0,1 % NP-40) et les protéines sont éluées par compétition avec du peptide FLAG (3X FLAG® Peptide Sigma-Aldrich) dans du tampon d’éluion (50 mM Tris-HCl pH 7.5, 150 mM NaCl) pendant 1 h sous agitation (1050 rpm) à température ambiante. L’éluât est séparé des billes par une centrifugation de 1 min à 12 000g et transféré dans un nouveau tube. Un aliquot est prélevé pour constituer la fraction “immunoprécipité” et le reste est analysé par spectrométrie de masse. Dans le but de vérifier l’efficacité de l’immunoprécipitation, les fractions collectées sont analysées sur gel de polyacrylamide en condition dénaturante.

## **IMMUNOPRECIPIATION AVEC BILLES MAGNETIQUES**

0.1 g à 0,3 g de tissu végétal sont broyés dans de l’azote liquide et sont homogénéisés dans 1 ml de tampon de lyse (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, inhibiteur de protéases (Complete Mini, Roche)) puis incubés 30 min à 4°C sur une roue. Les échantillons sont centrifugés deux fois à 13 200 rpm pendant 10 min à 4°C, et le surnageant est transféré dans un nouveau tube. Après la seconde centrifugation, un aliquot est prélevé pour constituer la fraction “Input” et 50 µl de billes magnétiques couplées soit à l’anticorps HA soit à l’anticorps GFP (Miltenyi Biotec) sont ajoutés au reste du surnageant. Les échantillons sont ensuite incubés pendant 30 min à 1h à 4°C sur une roue, puis passés sur une colonne (M column Miltenyi Biotec) préalablement préparée avec du tampon de lyse. Un aliquot des échantillons passés est prélevé pour constituer la fraction “Flow-through”. Les billes sont lavées deux fois avec 500 µl de tampon de

lyse puis une fois avec 100 µl de tampon de lavage (20 mM Tris-HCl pH 7.5) et sont éluées avec 2 fois 50 µl de tampon de charge (10 % glycérol, 4 % SDS, 62.3 mM Tris-HCl pH 6.8, 20 % (v/v) 2-β-mercaptoéthanol, Bleu de Bromophénol 0,003 %) préchauffé à 95°C.

Pour les analyses de spectrométrie de masse, un aliquot est prélevé pour constituer la fraction “immunoprécipité” et les fractions collectées sont analysés sur gel de polyacrylamide en condition dénaturante afin de vérifier l’efficacité de l’immunoprécipitation.

Pour les expériences de co-immunoprécipitation les fractions collectées sont analysés sur gel de polyacrylamide en condition dénaturante.

## **ANALYSES DES PROTEINES PAR SPECTROMETRIE DE MASSE**

Les analyses de spectrométrie de mass (MS) ont été entièrement réalisées par Philippe Hammann, Lauriane Kuhn et Johana Chicher à la plateforme de Protéomique de l’IBMP (Institut de Biologie Moléculaire et Cellulaire) CNRS Strasbourg-Esplanade.

### **DIGESTION DES PROTEINES**

Les protéines immunoprécipitées sont tout d’abord précipitées avec deux volumes de 100 % de Méthanol dans 0,1 M d’acétate d’ammonium. Le culot contenant les protéines est ensuite lavé deux fois et séché sous vide avant d’être repris dans 50 mM de bicarbonate d’ammonium. Après une

étape de réduction-alkylation (5 mM DTT, 10 mM Iodoacetamide) les protéines sont digérées avec 200 ng de trypsine “sequencing-grade” (Promega®) durant la nuit. Après une centrifugation à 12000g, le surnageant est collecté puis séché sous vide.

## **ANALYSES PAR NANO-LIQUID CHROMATOGRAPHY- ELECTROSPRAY IONIZATION TRIPLETOF MS/MS**

Avant injection, les peptides séchés sont repris dans 15 µl d'acide formique (FA) 0.1% (solvant A). Un tiers de chaque échantillon est injecté dans un NanoLC-2DPlus (module nanoFlexChiP; Eksigent, ABSciex, Concord, Ontario, Canada) couplé à un spectromètre de masse TripleTOF 5600 (ABSciex) fonctionnant en mode positif. Les peptides sont chargés sur une colonne en phase-inverse C18 (Pré-colonne ChiP C-18 300 µM ID x 5mm ChromXP et colonne analytique ChiP C-18 75 µM ID x 15 cm ChromXP; Eksigent). Les peptides sont élués en utilisant un gradient de 5% à 40 % de solvant B (0,1% d'acide formique dans de l'acétonitrile) pendant 1 h à un débit de 300 nL/min. Le TripleTOF 5600 fonctionne en mode «high-sensitivity data-dependant acquisition» dans une gamme de 350 à 1250 m/z avec le logiciel Analyst (v1.6, ABSciex). Une calibration externe du système a été effectuée avant l'injection de chaque échantillon en injectant 10 peptides de «beta-galactosidase trypsinic digest». La méthode de détection dite du Top20 a été utilisée: Jusqu'à 20 des ions les plus intenses chargés de façon multiple (2+ à 5+) ont été sélectionnés pour une fragmentation CID pendant 3,3s.

## **ANALYSES INFORMATIQUES**

Les données brutes sont tout d'abord converties au format “Mascot Generic File” (mgf) puis utilisées pour une recherche dans la base de données

de TAIR (The Arabidopsis Information Resource) additionnée de la base de donnée «decoy» (séquences en sens inverse). Le programme Mascot (version 2.2, Matrix Science, London, UK) utilisant ProteinScape 3.1 (Bruker Daltonics, Leipzig, Germany) a été utilisé pour générer l'algorithme de recherche. Durant cette analyse certaines modifications présentes sur les peptides ont été autorisées. À savoir, N-actylation (protéine), carbaminométhylation (cystéine), oxydation (Méthionine). L'absence d'au maximum deux sites de clivage par la trypsine a été autorisée. La tolérance de masse pour la MS et la MS/MS a été réglée respectivement à 20 ppm et à 0,5 Da. Les peptides identifiés par Mascot ont ensuite été validés en utilisant l'outil Protein Assessment de ProteinScape avec un FDR (False Discovery Rate) de 1%. Les protéines identifiées ont été comptées en fonction du nombre total de spectre MS/MS fragmentés par protéine permettant une stratégie de comptage de spectre quantitative.

Les interactions aspécifiques ont été écartées dans un premier temps en enlevant toutes les protéines présentes dans les contrôles négatifs en incluant les protéines «same set» et «sub set». Une protéine a été considérée comme interagissant que si elle est retrouvée dans tous les réplicas techniques pour un même échantillon. Dans un deuxième temps un second filtre a été appliqué. Les protéines identifiées ont été recherchées dans une base de données élaborée à la plateforme, provenant d'autres contrôles négatifs d'expériences de co-IP @GFP réalisées par d'autre équipe de recherche utilisant cette plateforme de protéomique.

*Note:* Lors des expériences d'immunoprécipitations concernant la partie DCL4, cette base de données n'existait pas encore.

# SEQUENÇAGE A HAUT DEBIT ET ANALYSES BIO- INFORMATIQUES DES ARN ASSOCIES A DCL4-6 ET DCL4-8

## SEQUENÇAGE A HAUT DEBIT

Les ARN associés aux protéines immunoprécipité et extraits, ont été fragmentés puis séquencé par un Illumina Genome Analyser par une compagnie externe. (Fasteris Suisse). Ce protocole de séquençage a donné lieu à l'obtention de banques de séquences de 50 nt. Le séquençage n'ayant pas donné un nombre de séquence suffisant pour certaine banque, un second à été réalisé. Les deux banques obtenues pour chaque échantillon (allant de 4 à 21 million de séquences) ont été ainsi utilisé comme répliques techniques pour les analyses différentielles.

## ALIGNEMENT ET ANALYSE DIFFERENTIELLE

La qualité des banques obtenues a ensuite été vérifiée à l'aide de l'outil FASTQC, puis les séquences ont été alignées sur le génome d'*Arabidopsis thaliana* à l'aide des outils Tophat (v2.0.11; -I 4000), Bowtie (v2.2.1.0) et Samtools (v0.1.19) (97.à 99 % de taux d'alignement).

Pour chaque échantillon le nombre de séquence à été compté en utilisant la méthode du "100 nt Sliding window" avec un incrément de 50 nt par fenêtre (correspondant à la taille des séquences à la sortie du séquenceur) conduisant à l'obtention de 2393360 fenêtres couvrant le génome nucléaire et plastidial d'*Arabidopsis*. Afin de rechercher les séquences spécifiquement retrouvée ou enrichie dans les banques DCL4-6 et DCL4-8 une analyses différentielle à été effectuée. L'analyse différentielle a

été réalisé sur les fenêtres obtenues à l'aide de l'outil DESeq2 (v1.3.32; avec mesure de dispersion, normalisation, "outlierSD" = 8, padj < 0,05 ou <0,1).

## ANNOTATIONS

Les fenêtres ont été fusionnées si elles étaient chevauchantes ou contiguës. Les annotations ont été faites sur ces fenêtres fusionnées par comparaison de leurs positions génomiques avec celle des gènes, snoRNAs, tRNAs, rRNAs, miRNAs, tasiRNAs et les éléments transposables.

*Note:* Une région donnée peut avoir plusieurs annotations et une annotation peut correspondre à plusieurs fenêtres.

## ALIGNMENT MULTIPLE DE SEQUENCES

Les séquences de DCL4 de dix espèces plus ou moins représentatives des plantes vasculaires ont été obtenues sur GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). Une espèce de bryophyte (*Physcomitrella patens*) et deux espèces de monocotylédones (*Oryza sativa*, *Zea mays*,) représentative des poales ont été sélectionnées. Sept espèces d'eudicotylédones représentatives des rosidées ont été retenues. Deux espèces des fabidées (*Cucumis sativus*, *Populus euphratica*) et cinq espèces représentant les malvidées (*Arabidopsis thaliana*, *Brassica rappa*, *Citrus sinensi*, *Eucalyptus grandis*, *Theobroma cacao*) ont été choisies.

L'alignement des séquences des différentes plantes choisies a été réalisé avec le logiciel Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) en utilisant les paramètres par

défaut. La coloration de l'alignement à été obtenue grâce à l'outil BOXSHADE v 3.21).



# ***ANNEXES***

**Annexe 1:** Annexe contenant une partie des données de la Supplemental Table S2 Montavon *et al*, présenté dans le chapitre 1 (cf. §1.2). Cette annexe contient une partie des résultats de séquençage à haut-débit (tasiRNA et 12 premiers endoIR-siRNA) ainsi que la couverture peptidique de DRB4 obtenue lors des expériences d'IP-MS.

ANNEXE 1

tasiRNA count

chr	name	start	end	21 nt				22 nt			
				Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2	Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2
Chr2	TAS1a	11721539	11722468	1214	1115	1505	935	421	420	625	401
Chr1	TAS1b	18549204	18550042	2548	2557	2183	2039	583	757	550	521
Chr2	TAS1c	16537288	16538277	38640	40767	37179	31066	8592	9090	7429	6910
Chr2	TAS2	16539384	16540417	21356	24647	21008	19745	3207	3421	2763	2773
Chr3	TAS3a	5861491	5862437	7887	8774	8952	6871	2000	2113	2151	1754
Chr5	TAS3b	20134200	20134786	3456	3429	3927	2618	832	741	912	602
Chr5	TAS3c	23394005	23394500	180	272	251	183	38	58	54	30
Chr3	TAS4	9417551	9418705	7	2	6	11	0	0	0	0

tasiRNA count

chr	name	start	end	23 nt				24 nt			
				Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2	Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2
Chr2	TAS1a	11721539	11722468	98	143	131	92	321	295	413	260
Chr1	TAS1b	18549204	18550042	82	83	64	50	264	281	254	239
Chr2	TAS1c	16537288	16538277	821	968	810	693	6279	7273	5878	5195
Chr2	TAS2	16539384	16540417	826	1029	760	731	1179	1292	983	1082
Chr3	TAS3a	5861491	5862437	97	143	132	91	333	406	402	304
Chr5	TAS3b	20134200	20134786	44	39	58	49	52	61	75	47
Chr5	TAS3c	23394005	23394500	3	7	3	2	6	6	8	3
Chr3	TAS4	9417551	9418705	0	0	0	0	0	1	0	1

endoIR-siRNA count

chr	name	start	end	Size	21 nt				22 nt			
					Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2	Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2
Chr3	IR71	1963600	1970599	6999	3472	7261	3646	5497	16128	7733	16132	6022
Chr4	AT4G21366	11383049	11388924	5875	3306	3010	2710	2512	825	664	714	534
Chr2	IRAtGP10	818470	824309	5839	310	3519	221	2758	438	1687	374	1394
Chr5	AT5G22960	7682904	7686252	3348	8252	9832	6758	7823	1785	1716	1533	1299
Chr3	IR2039	16158978	16161020	2042	2939	7352	2932	5107	4701	3762	4036	2807
Chr5	x	8313108	8314258	1150	12	65	10	52	4	15	13	11
Chr4	x	11383433	11384546	1113	4	4	0	2	5	1	9	0
Chr4	x	2185400	2186350	950	7	26	6	24	10	6	7	12
Chr3	x	18047400	18048099	699	3	19	4	9	7	4	6	5
Chr3	x	14421301	14421688	387	0	0	0	2	1	0	0	0
Chr4	x	3183663	3184038	375	0	0	0	0	0	0	1	1
Chr1	x	14441302	14441630	328	0	0	1	1	6	0	1	1

endoIR-siRNA count

chr	name	start	end	Size	23 nt				24 nt			
					Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2	Col-0#1	<i>drb7.2</i> #1	Col-0#2	<i>drb7.2</i> #2
Chr3	IR71	1963600	1970599	6999	2944	3858	2728	2855	6377	43435	5899	31896
Chr4	AT4G21366	11383049	11388924	5875	63	97	77	51	1091	5252	1019	4151
Chr2	IRAtGP10	818470	824309	5839	262	989	200	741	1277	11913	1098	8900
Chr5	AT5G22960	7682904	7686252	3348	123	276	119	220	692	2574	568	2133
Chr3	IR2039	16158978	16161020	2042	1943	3283	1770	2222	7847	31296	6901	21084
Chr5	x	8313108	8314258	1150	29	42	23	26	132	211	132	156
Chr4	x	11383433	11384546	1113	1	1	4	1	32	16	33	23
Chr4	x	2185400	2186350	950	18	25	23	26	189	204	208	189
Chr3	x	18047400	18048099	699	89	78	82	64	214	269	220	196
Chr3	x	14421301	14421688	387	1	2	1	0	6	11	5	8
Chr4	x	3183663	3184038	375	0	0	0	0	2	3	2	3
Chr1	x	14441302	14441630	328	5	1	0	2	16	21	12	9

Sequence coverage obtained by Trypsin digestion,  
Searching parameters : peptide mascot score > 30 (Matrix Science) and no missed cleavages

DRB7 (AT4G00420) Sequence coverage : 68 %

1 MDLPPPESSS NPI SSFSVAQ KRMLLQRCEK KFKI RKLNDT VEEDNVTQME  
51 PNFTLENTLV SEPEADLRPH TTEPTTEEET QRSSAKSPLY NLCSVRHWKA  
101 PLYEYI AEGP CHMKI FTGKV TVEMKEDSRI TVLECFGNPQ YKKKI AAEQA  
151 AEAALWYLNK VGYTLQTEKA SGRKGRSKPI SKMMVTGEPV

DRB4 (AT3G62800) Sequence coverage : 59 %

1 MDHVKYKQLQ AYALQHLEL PVYANEREGP PHAPRFRCNV TFCGQTFQSS  
51 EFFPTLKSAE HAAAKI AVAS LTPQSPEGI D VAYKNLLQEI AQKESLLPF  
101 YATATSGPSH APTFTSTVEF AGKVFSGEEA KTKKLAEMSA AKVAFMSI KN  
151 GNSNQTGSPT LPSERQEDVN SNVKSSPQEI HSQPSSKVVV TPDTPSKGI K  
201 VNEDEFPDLH DAPASNAKEI NVALNEPENP TNDGTL SALT TDGVMKNI AS  
251 SSLPI PHNPT NVI TLNAPAA NGI KRNI AAC SSWMPQNPTN DGSETSSCVV  
301 DESEKKKLI M GTGHLSI PTG QHVVCRPWNP EI TLPQDAEM LFRDDKFI AY  
351 RLVKP



**Annexe 2:** Annexe contenant, un complément de liste (cf chapitre 1 §1.2) des amorces oligonucléotidiques utilisées ainsi que la liste des anticopr.

## ANNEXE 2

### Amorces oligonucléotidiques

Nom	Séquence	utilisation
<i>d7ip1</i> LP	AAACTTTACCCGGATCACCAG	géotypage
<i>d7ip1</i> RP	AAACAAACAACCAGGCATTTG	géotypage
@miR390	GGCGCTATCCCTCCTGAGCTT	sonde radioactive

### Anticorps

Nom	provenance	dilution
@DCL1	Lapin	1/1000 èmme
@DCL3	Lapin	1/1000 èmme
@DCL4	Lapin	1/4000 èmme
@AGO1	Lapin	1/8000 èmme
@AGO4	Lapin	1/2000 èmme
@DRB4	Cochon d'Inde	1/5000 èmme
@GFP	Lapin	1/30 000 èmme
@HA	Lapin	1/15 000 èmme

# BIBLIOGRAPHIE

Allen E & Howell MD (2008) MiRNAs in the biogenesis of trans-acting siRNAs in higher plants. *Semin. Cell Dev. Biol.* **21**: 798–804

Allen E, Xie Z, Gustafson AM & Carrington JC (2005) MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**: 207–221

Allen E, Xie Z, Gustafson AM, Sung G-H, Spatafora JW & Carrington JC (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36**: 1282–1290

Angell SM & Baulcombe DC (1997) Consistent gene silencing in transgenic plants expressing a replicating potato virus X RNA. *EMBO J.* **16**: 3675–3684

Aufsatz W, Mette MF, Matzke AJM & Matzke M (2004) The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. *Plant Mol. Biol.* **54**: 793–804

Axtell MJ (2013) Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol* **64**: 137–159

Axtell MJ, Jan C, Rajagopalan R & Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**: 565–577

Baranauskė S, Mickutė M, Plotnikova A, Finke A, Venclovas Č, Klimašauskas S & Vilkaitis G (2015) Functional mapping of the plant small RNA methyltransferase: HEN1 physically interacts with HYL1 and DICER-LIKE 1 proteins. *Nucleic Acids Res.* **43**: 2802–2812

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297

Baulcombe DC (1996) RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol. Biol.* **32**: 79–88

Baulcombe DC (2004) RNA silencing in plants. *Nature* **431**: 356–363

Baumberger N & Baulcombe DC (2005) *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 11928–11933

Bernatavichute Y V, Zhang X, Cokus S, Pellegrini M & Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS One* **3**: e3156

Bernstein E, Caudy AA, Hammond SM & Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366

Bies-Etheve N, Pontier D, Lahmy S, Picart C, Vega D, Cooke R & Lagrange T (2009) RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10**: 649–654



Blevins T, Podicheti R, Mishra V, Marasco M, Tang H & Pikaard CS (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **4**: e09591

Blevins T, Rajeswaran R, Shivaprasad P V, Beknazariants D, Si-Ammour A, Park H-S, Vazquez F, Robertson D, Meins F, Hohn T & Pooggin MM (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* **34**: 6233–6246

Boccaro M, Sarazin A, Thiébeauld O, Jay F, Voinnet O, Navarro L & Colot V (2014) The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathog.* **10**: e1003883

Bollman KM, Aukerman MJ, Park M-Y, Hunter C, Berardini TZ & Poethig RS (2003) HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* **130**: 1493–1504

Bologna NG, Mateos JL, Bresso EG & Palatnik JF (2009) A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J.* **28**: 3646–3656

Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC & Palatnik JF (2013) Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res.* **23**: 1675–1689

Bologna NG & Voinnet O (2014) The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annu. Rev. Plant Biol.* **65**: 473–503

Borges F & Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* **16**: 727–741

Borges F, Pereira PA, Slotkin RK, Martienssen RA & Becker JD (2011) MicroRNA activity in the Arabidopsis male germline. *J. Exp. Bot.* **62**: 1611–1620

Bortolamiol D, Pazhouhandeh M & Ziegler-Graff V (2008) Viral suppression of RNA silencing by destabilisation of ARGONAUTE 1. *Plant Signal. Behav.* **3**: 657–659

Bouché N, Lauressergues D, Gascioli V & Vaucheret H (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* **25**: 3347–3356

Boutet S, Vazquez F, Liu J, Béclin C, Fagard M, Gratias A, Morel J, Créte P, Chen X & Vaucheret H (2003) Arabidopsis HEN1: A genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.* **13**: 843–848

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth LE & Voinnet O (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**: 1185–1190

Brodersen P, Sakvarelidze-Achard L, Schaller H, Khafif M, Schott G, Bendahmane A & Voinnet O (2012) Isoprenoid biosynthesis is required for miRNA function and affects membrane association of ARGONAUTE 1 in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 1778–1783

Brosseau C & Moffett P (2015) Functional and genetic analysis identify a role for Arabidopsis ARGONAUTE5 in antiviral RNA silencing. *Plant Cell* **27**: 1742–1754

Buchmann RC, Asad S, Wolf JN, Mohannath G & Bisaro DM (2009) Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol.* **83**: 5005–5013

- Cao M, Du P, Wang X, Yu Y-Q, Qiu Y-H, Li W, Gal-On A, Zhou C, Li Y & Ding S-W (2014) Virus infection triggers widespread silencing of host genes by a distinct class of endogenous siRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 14613–14618
- Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M & Jacobsen SE (2003) Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. *Curr. Biol.* **13**: 2212–2217
- Cao X & Jacobsen SE (2002) Role of the Arabidopsis DRM Methyltransferases in De Novo DNA methylation and gene silencing. *Curr. Biol.* **13**: 1138–1144
- Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, Cuperus JT & Carrington JC (2012) Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell* **24**: 3613–3629
- Carthew RW & Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642–655
- Castel SE & Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* **14**: 100–112
- Cenik ES, Fukunaga R, Lu G, Dutcher R, Wang Y, Tanaka Hall TM & Zamore PD (2011) Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell* **42**: 172–184
- Cerutti L, Mian N & Bateman A (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem. Sci.* **25**: 481–482
- Chapman EJ & Carrington JC (2007) Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* **8**: 884–896
- Chen X, Liu J, Cheng Y & Jia D (2002) HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development* **129**: 1085–1094
- Chilton M-D, Drummond MH, Merlo DJ, Sciaky D, Montoya AL, Gordon MP & Nester EW (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell* **11**: 263–271
- Clavel M, Pélissier T, Descombin J, Jean V, Picart C, Charbonel C, Saez-Vásquez J, Bousquet-Antonelli C & Deragon J-M (2015) Parallel action of AtDRB2 and RdDM in the control of transposable element expression. *BMC Plant Biol.* **15**: 70
- Clavel M, Pélissier T, Montavon T, Tschopp MA, Pouch-Pélissier MN, Descombin J, Jean V, Dunoyer P, Bousquet-Antonelli C & Deragon JM (2016) Evolutionary history of double-stranded RNA binding proteins in plants: identification of new cofactors involved in easiRNA biogenesis. *Plant Mol. Biol.* **91**: 131–147
- Clough S & Bent A (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **116**: 735–743
- Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC & Martienssen RA (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* **508**: 411–415
- Cuperus JT, Montgomery TA, Fahlgren N, Burke RT, Townsend T, Sullivan CM & Carrington JC (2010) Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 466–471

- Curtin SJ, Watson JM, Smith NA, Eamens AL, Blanchard CL & Waterhouse PM (2008) The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS Lett.* **582**: 2753–2760
- Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC & Voinnet O (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* **313**: 68–71
- Diaz-pendon JA, Li F, Li W & Ding S (2007) Suppression of antiviral silencing by Cucumber Mosaic Virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**: 2053–2063
- Donaire L, Barajas D, Martínez-García B, Martínez-Priego L, Pagán I & Llave C (2008) Structural and genetic requirements for the biogenesis of Tobacco Rattle Virus-derived small interfering RNAs. *J. Virol.* **82**: 5167–5177
- Dong Z, Han M-H & Fedoroff N (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci USA* **105**: 9970–9975
- Dougherty WG & Parks TD (1995) Transgenes and gene suppression: telling us something new. *Curr. Opin. Cell Biol.* **7**: 399–405
- Dunoyer P, Himber C, Ruiz-Ferrer V, Alioua A & Voinnet O (2007) Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.* **39**: 848–856
- Dunoyer P, Himber C & Voinnet O (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat Genet* **37**: 1356– 1360
- Eamens AL, Kim KW, Curtin SJ & Waterhouse PM (2012) DRB2 is required for microRNA biogenesis in Arabidopsis thaliana. *PLoS One* **7**: e35933
- Eamens AL, Smith NA, Curtin SJ, Wang M-B & Waterhouse PM (2009) The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* **15**: 2219–2235.
- Eamens AL, Vaistij FEFE & Jones L (2008a) NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. *Plant J.* **55**: 596–606
- Eamens AL, Wang M-B, Smith NA & Waterhouse PM (2008b) RNA Silencing in plants: Yesterday, Today, and Tomorrow. *Plant Physiol.* **147**: 456–468
- El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi MA, Jacobsen SE, Cooke R & Lagrange T (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**: 2539–2544
- Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K & Tuschl T (2001a) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494–498
- Elbashir SM, Lendeckel W & Tuschl T (2001b) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**: 188–200
- Elvira-Matelot E, Hachet M, Shamandi N, Comella P, Saez-Vasquez J, Zytnicki M & Vaucheret H (2016) Arabidopsis RNASE THREE LIKE2 modulates the expression of protein-coding genes via 24-nt siRNA-directed DNA methylation. *Plant Cell* **28**: 406–425
- Endo Y, Iwakawa H, & Tomari Y (2013) Arabidopsis ARGONAUTE7 selects miR390 through multiple checkpoints during RISC assembly. *EMBO Rep.* **14**: 652–658

Eun C, Lorkovic ZJ, Naumann U, Long Q, Havecker ER, Simon SA, Meyers BC, Matzke AJM & Matzke M (2011) AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in *Arabidopsis thaliana*. *PLoS One* **6**: e25730

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL & Carrington JC (2007) High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One* **2**: e219

Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL & Carrington JC (2006) Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Curr. Biol.* **16**: 939-944

Fang Y & Spector DL (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr. Biol.* **17**: 818-823

Fei Q, Xia R & Meyers BC (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* **25**: 2400-2415

Finnegan EJ (2002) Epialleles a source of random variation in times of stress. *Curr. Opin. Plant Biol.* **5**: 101-106

Finnegan EJ, Margis R & Waterhouse PM (2003) Posttranscriptional gene silencing is not compromised in the *Arabidopsis* *CARPEL FACTORY (DICER-LIKE1)* mutant, a homolog of Dicer-1 from *Drosophila*. *Curr. Biol.* **13**: 236-240

Fire A, Albertson D, Harrison SW & Moerman DG (1991) Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development* **113**: 503-514

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE & Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806-811

Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA & Coupland G (2009) *Arabidopsis* DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev. Cell* **17**: 75-86

Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL & Woo SC (1983) Expression of bacterial genes in plant cells. *Genetics* **80**: 4803-4807

Fukudome A, Kanaya A, Egami M, Nakazawa Y, Hiraguri A, Moriyama H & Fukuhara T (2011) Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of *Arabidopsis* Dicer-like 4. *RNA* **17**: 750-760

Fukunaga R & Zamore PD (2014) A universal small molecule, inorganic phosphate, restricts the substrate specificity of Dicer-2 in small RNA biogenesis. *Cell cycle* **13**: 1671-1676

Garcia-Ruiz H, Carbonell A, Hoyer JS, Fahlgren N, Gilbert KB, Takeda A, Giampetruzzi A, Garcia Ruiz MT, McGinn MG, Lowery N, Martinez Baladejo MT & Carrington JC (2015) Roles and programming of *Arabidopsis* ARGONAUTE proteins during Turnip Mosaic Virus infection. *PLoS Pathog.* **11**: 1-27

Garcia D, Collier SA, Byrne ME & Martienssen RA (2006) Specification of leaf polarity in *Arabidopsis* via the trans-acting siRNA pathway. *Curr. Biol.* **16**: 933-938

Gascioli V, Mallory AC, Bartel DP & Vaucheret H (2005) Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.* **15**: 1494-1500

German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, de Paoli E, Lu C, Schroth GP, Meyers BC & Green PJ (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotech* **26**: 941-946

Ghildiyal M & Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**: 94-108

Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L & Pikaard CS (2012) In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol. Cell* **48**: 811-818

Hamilton AJ & Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950-952

Hammond SM, Bernstein E, Beach D & Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**: 293-296

Hammond SM, Boettcher S, Caudy AA, Kobayashi R & Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**: 1146-1150

Han M-H, Goud S, Song L & Fedoroff N (2004) The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 1093-1098

Harsman A, Schock A, Hemmis B, Wahl V, Jeshen I, Bartsch P, Schlereth A, Pertl-Obermeyer H, Goetze TA, Soll J, Philippar K & Wagner R (2016) OEP40, a regulated glucose-permeable  $\beta$ -barrel solute channel in the chloroplast outer envelope membrane. *J. Biol. Chem.* **291**: 17848-17860

Harvey JJW, Lewsey MG, Patel K, Westwood JH, Heimstädt S, Carr JP & Baulcombe DC (2011) An antiviral defense role of AGO2 in plants. *PLoS One* **6**: e14639

Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH & Baulcombe DC (2010) The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**: 321-334

He X-J, Hsu Y-F, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu H-L, Wang C-S, Jin H & Zhu J-K (2009) An effector of RNA-Directed DNA Methylation in Arabidopsis is an ARGONAUTE 4 and RNA-binding protein. *Cell* **137**: 498-508

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ & Jacobsen SE (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**: 721-725

Herr AJ, Jensen MB, Dalmay T & Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118-120

Himber C, Dunoyer P, Moissiard G, Ritzenthaler C & Voinnet O (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J.* **22**: 4523-4533

Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, Koiwa H, Seki M, Shinozaki K & Fukuhara T (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. *Plant Mol. Biol.* **57**: 173-188

Hooykaas-Van Slogteren GMS, Hooykaas PJJ & Schilperoort RA (1984) Expression of Ti plasmid genes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. *Nature* **311**: 763-764

Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD & Carrington JC (2007) Genome-Wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**: 926-942

Huang Y, Ji L, Huang Q, Vassilyev DG, Chen X & Ma J-B (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* **461**: 823-827

Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T & Ishikawa M (2010) In vitro assembly of plant RNA-Induced Silencing Complexes facilitated by molecular chaperone HSP90. *Mol. Cell* **39**: 282-291

Incarbone M & Dunoyer P (2013) RNA silencing and its suppression: novel insights from in planta analyses. *Trends Plant Sci.* **18**: 382-392

Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D & Schnable PS (2009) Loss of RNA-Dependent RNA Polymerase 2 (RDR2) Function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt Small RNAs. *PLoS Genet.* **5**: e1000737

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J & Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17**: 379-384

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, Patel DJ & Jacobsen SE (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**: 124-128

Johnson LM, Law JA, Khattar A, Henderson IR & Jacobsen SE (2008) SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLoS Genet.* **4**: e1000280

Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M & Matzke AJM (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**: 761-765

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA & Carrington JC (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol.* **5**: e57

Kennerdell JR & Carthew RW (1998) Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 Act in the wingless pathway. *Cell* **95**: 1017-1026

Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* **6**(5): 376-385

Kim YJ, Zheng B, Yu Y, Won SY, Mo B & Chen X (2011) The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. *EMBO J.* **30**: 814-822

Kozomara A & Griffiths-Jones S (2011) MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**: 152-157

Kumakura N, Takeda A, Fujioka Y, Motose H, Takano R & Watanabe Y (2009) SGS3 and RDR6 interact and colocalize in cytoplasmic SGS3/RDR6-bodies. *FEBS Lett.* **583**: 1261-1266

Kurihara Y, Takashi Y & Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**: 206-212

Kurihara Y & Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci USA* **101**: 12753-12758

- Lanet E, Delannoy E, Sormani R, Floris M, Brodersen P, Crete P, Voinnet O & Robaglia C (2009) Biochemical evidence for translational repression by Arabidopsis microRNA. *Plant Cell* **21**: 1762-1768
- Lau P-W, Guiley KZ, De N, Potter CS, Carragher B & MacRae IJ (2012) The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.* **19**: 436-440
- Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ & Jacobsen SE (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**: 385-389
- Law JA, Vashisht AA, Wohlschlegel JA & Jacobsen SE (2011) SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet.* **7**: e1002195
- Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ & Carthew RW (2004) Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**: 69-81
- Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute Y V, Chan SWL, Lagrange T, Pikaard CS & Jacobsen SE (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell* **126**: 93-106
- Li J, Yang Z, Yu B, Liu J & Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.* **15**: 1501-1507
- Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD & Chen X (2015) Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. *Genome Res.* **25**: 235-245
- Lindbo JA, Silva-Rosales L, Proebsting WM & Dougherty WG (1993) Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance. *Plant Cell* **5**: 1749-1759
- Lindow M, Jacobsen A, Nygaard S, Mang Y & Krogh A (2007) Intragenomic matching reveals a huge potential for miRNA-mediated regulation in plants. *PLoS Comput. Biol.* **3**: e238
- Lisch DR (2009) Epigenetic Regulation of Transposable Elements in Plants. *Annual Review of Plant Biology* **60**: 43-66
- Liu C, Axtell MJ & Fedoroff N V (2012) The helicase and RNaseIIIa domains of Arabidopsis Dicer-Like1 modulate catalytic parameters during microRNA biogenesis. *Plant Physiol.* **159**: 748-758
- Liu Q, Yao X, Pi L, Wang H, Cui X & Huang H (2009) The ARGONAUTE 10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in Arabidopsis. *Plant J.* **58**: 27-40
- Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, Chen S, Huang H-W, Cai T & He X-J (2014) The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet.* **10**: e1003948
- Llave C, Xie Z, Kasschau KD & Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**: 2053-2056
- Lobbes D, Rallapalli G, Schmidt DD, Martin C & Clarke J (2006) SERRATE: a new player on the plant microRNA scene. *EMBO Rep.* **7**: 1052-1058
- Lund E, Guttinger S, Calado A, Dahlberg JE & Kutay U (2004) Nuclear export of microRNA precursors. *Science* **303**: 95-98

- Luo Q-J, Mittal A, Jia F & Rock CD (2012) An autoregulatory feedback loop involving PAPI and TAS4 in response to sugars in Arabidopsis. *Plant Mol. Biol.* **80**: 117–129
- Ma J-B, Ye K & Patel DJ (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**: 318–322
- MacRae IJ, Zhou K & Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* **14**: 934–940
- MacRae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA & Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. *Science* **311**: 195–198
- Maillard P V, Ciaudo C, Marchais A, Li Y, Jay F, Ding SW & Voinnet O (2013) Antiviral RNA interference in mammalian cells. *Science* **342**: 235–238
- Mallory A & Vaucheret H (2010) Form, function, and regulation of ARGONAUTE proteins. *Plant Cell* **22**: 3879–3889
- Mallory AC, Hinze A, Tucker MR, Bouché N, Gascioli V, Elmayer T, Laressergues D, Jauvion V, Vaucheret H & Laux T (2009) Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. *PLoS Genet.* **5**: e1000646
- Mallory AC & Vaucheret H (2009) ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. *EMBO Rep.* **10**: 521–526
- Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ & Waterhouse PM (2006) The evolution and diversification of Dicers in plants. *FEBS Lett.* **580**: 2442–2450
- Mateos JL, Bologna NG, Chorostecki U & Palatnik JF (2010) Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Curr. Biol.* **20**: 49–54
- Matzke MA & Birchler JA (2005) RNAi-mediated pathways in the nucleus. *Nat Rev Genet* **6**: 24–35
- Matzke MA & Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**: 394–408
- Matzke MA, Primig M, Trnovsky J & Matzke AJ (1989) Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**: 643–649
- Maunoury N & Vaucheret H (2011) AGO1 and AGO2 act redundantly in miR408-mediated Plantacyanin regulation. *PLoS One* **6**: e28729
- McCue AD, Nuthikattu S, Reeder SH & Slotkin RK (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet.* **8**: e1002474
- Meister G (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* **14**: 447–459
- Metzlaff M, O'Dell M, Cluster P. & Flavell R. (1997) RNA-mediated RNA degradation and Chalcone Synthase A silencing in Petunia. *Cell* **88**: 845–854
- Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ & Qi Y (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**: 116–127



Moissiard G, Parizotto EA, Himber C & Voinnet O (2007) Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA* **13**: 1268-1278

Moissiard G & Voinnet O (2006) RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. *Proc Natl Acad Sci USA* **103**: 19593-19598

Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E & Carrington JC (2008a) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* **133**: 128-141

Montgomery TA, Yoo SJ, Fahlgren N, Gilbert SD, Howell MD, Sullivan CM, Alexander AL, Nguyen G, Allen E, Ahn JH & Carrington JC (2008b) AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc Natl Acad Sci USA* **105**: 20055-20062

Morel J-B, Godon C, Mourrain P, Béclin C, Boutet S, Feuerbach F, Proux F & Vaucheret H (2002) Fertile Hypomorphic ARGONAUTE (ago1) Mutants impaired in Post-Transcriptional Gene Silencing and virus resistance. *Plant Cell* **14**: 629-639

Morel J-B, Mourrain P, Béclin C & Vaucheret H (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr. Biol.* **10**: 1591-1594

Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J & Carbonero P (2007) The family of DOF transcription factors: from green unicellular algae to vascular plants. *Mol. Genet. Genomics* **277**: 379-390

Mourrain P, Béclin C, Elmayan T, Feuerbach F, Godon C, Morel J-B, Jouette D, Lacombe A-M, Nikic S, Picault N, Ré moué K, Sanial M, Vo T-A & Vaucheret H (2000) Arabidopsis SGS2 and SGS3 genes are required for Posttranscriptional Gene Silencing and natural virus resistance. *Cell* **101**: 533-542

Mukherjee K, Campos H & Kolaczowski B (2013) Evolution of animal and plant dicers: Early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Mol. Biol. Evol.* **30**: 627-641

Nagano H, Fukudome A, Hiraguri A, Moriyama H & Fukuhara T (2014) Distinct substrate specificities of Arabidopsis DCL3 and DCL4. *Nucleic Acids Res.* **42**: 1845-1856

Nakazawa Y, Hiraguri A, Moriyama H & Fukuhara T (2007) The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. *Plant Mol. Biol.* **63**: 777-785

Napoli C, Lemieux C & Jorgensen R (1990) Introduction of a chimeric Chalcone Synthase gene into Petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2**: 279-289

Ngô H, Tschudi C, Gull K & Ullu E (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 14687-14692

Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA & Vielle-Calzada J-P (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature* **464**: 628-632

Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O & Pikaard CS (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613-622

- Parent JS, Bouteiller N, Elmayan T & Vaucheret H (2015) Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. *Plant J.* **81**: 223-232
- Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ & Kim VN (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**: 201-205
- Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H & Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 3691-3696
- Park W, Li J, Song R, Messing J & Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr. Biol.* **12**: 1484-1495
- Pélissier T, Clavel M, Chaparro C, Pouch-Pélissier M-N, Vaucheret H & Deragon J-M (2011) Double-stranded RNA binding proteins DRB2 and DRB4 have an antagonistic impact on Polymerase IV-dependent siRNA levels in Arabidopsis. *RNA* **17**: 1502-1510
- Peragine A, Yoshikawa M, Wu G, Albrecht HL & Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev.* **18**: 2368-2379
- Pikaard CS, Haag JR, Ream TS & Wierzbicki AT (2008) Roles of RNA Polymerase IV in gene silencing. *Trends Plant Sci.* **13**: 390-397
- Pontes O, Vitins A, Ream TS, Hong E, Pikaard CS & Costa-Nunes P (2013) Intersection of small RNA pathways in Arabidopsis thaliana sub-nuclear domains. *PLoS One* **8**: e65652
- Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi M-A, Lerbs-Mache S, Colot V & Lagrange T (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19**: 2030-2040
- Pontier D, Picart C, Roudier F, Garcia D, Lahmy S, Azevedo J, Alart E, Laudie M, Karlowski WM, Cooke R, Colot V, Voinnet O & Lagrange T (2012) NERD, a Plant-Specific GW Protein, Defines an Additional RNAi-Dependent Chromatin-Based Pathway in Arabidopsis. *Mol. Cell* **48**: 121-132
- Portereiko MF, Sandaklie-Nikolova L, Lloyd A, Dever CA, Otsuga D & Drews GN (2006) NUCLEAR FUSION DEFECTIVE1 encodes the Arabidopsis RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Physiol.* **141**: 957-965
- Qi X, Bao FS & Xie Z (2009) Small RNA deep sequencing reveals role for Arabidopsis thaliana RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS One* **4**: e4971
- Qin H, Chen F, Huan X, Machida S, Song J & Yuan YA (2010) Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA* **16**: 474-481
- Qu F, Ye X & Morris TJ (2008) Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *Proc Natl Acad Sci USA* **105**: 14732-14737
- Raja P, Sanville BC, Buchmann RC & Bisaro DM (2008) Viral genome methylation as an epigenetic defense against Geminiviruses. *J. Virol.* **82**: 8997-9007
- Rajagopalan R, Bartel DP, Vaucheret H, Trejo J & Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Cell* **20**: 3407-3425

Rajeswaran R, Aregger M, Zvereva AS, Borah BK, Gubaeva EG & Pooggin MM (2012) Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. *Nucleic Acids Res* **40**: 6241-6254

Rajeswaran R & Pooggin MM (2012) RDR6-mediated synthesis of complementary RNA is terminated by miRNA stably bound to template RNA. *Nucleic Acids Res.* **40**: 594-599

Ramachandran V & Chen X (2008) Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science* **321**: 1490-1492

Ratcliff F, Harrison BD & Baulcombe DC (1997) A similarity between viral defense and gene silencing in plants. *Science* **276**: 1558-1560

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B & Bartel DP (2002) MicroRNAs in plants. *Genes Dev.* **16**: 1616-1626

Reis RS, Hart-Smith G, Eamens AL, Wilkins MR & Waterhouse PM (2015) Gene regulation by translational inhibition is determined by Dicer partnering proteins. *Nat. plants* **1**: 14027

Ren G, Chen X & Yu B (2012) Uridylation of miRNAs by HEN1 SUPPRESSOR1 in Arabidopsis. *Curr. Biol.* **22**: 695-700

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B & Bartel DP (2002) Prediction of plant microRNA targets. *Cell* **110**: 513-520

Romano N & Macino G (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* **6**: 3343-3353

Ronemus M, Vaughn MW & Martienssen RA (2006) MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-dependent RNA polymerase in Arabidopsis. *Plant Cell* **18**: 1559-1574

Schauer SE, Jacobsen SE, Meinke DW & Ray A (2002) DICER-LIKE1: blind men and elephants in Arabidopsis development. *Trends Plant Sci.* **7**: 487-491

Shamandi N, Zytnicki M, Charbonnel C, Elvira-Matelot E, Bochnakian A, Comella P, Mallory AC, Lepère G, Sáez-Vásquez J & Vaucheret H (2015) Plants encode a general siRNA suppressor that is induced and suppressed by viruses. *PLoS Biol.* **13**: 1-28

Shivaprasad P V, Chen H-M, Patel K, Bond DM, Santos BACM & Baulcombe DC (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* **24**: 859-874

Slotkin RK, Freeling M & Lisch D (2003) Mu killer causes the heritable inactivation of the Mutator family of transposable elements in *Zea mays*. *Genetics* **165**: 781-797

Slotkin RK, Freeling M & Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat. Genet.* **37**: 641-644

Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA & Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461-472

Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS & Baulcombe DC (2007) An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19**: 1507-1521

Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG & Waterhouse PM (2000) Total silencing by intron-spliced hairpin RNAs. *Nature* **407**: 319–320

Song J-J, Smith SK, Hannon GJ & Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434–1437

Song L, Han M-H, Lesicka J & Fedoroff N (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci USA* **104**: 5437–5442

Sun Z, Guo T, Liu Y, Liu Q & Fang Y (2015) The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. *PLoS Genet.* **11**: e1005598

Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V, Patel DJ & van der Oost J (2014) The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**: 743–53

Takeda A, Iwasaki S, Watanabe T, Utsumi M & Watanabe Y (2008) The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins. *Plant Cell Physiol.* **49**: 493–500

Tanurdzic M, Vaughn MW, Jiang H, Lee T-J, Slotkin RK, Sosinski B, Thompson WF, Doerge RW & Martienssen RA (2008) Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol.* **6**: 2880–2895

Tian Y, Simanshu DK, Ma JB, Park JE, Heo I, Kim VN & Patel DJ (2014) A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. *Mol Cell* **53**: 606–616

Tkaczuk KL, Obarska A & Bujnicki JM (2006) Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. *BMC Evol. Biol.* **6**: 6

Tomari Y, Matranga C, Haley B, Martinez N & Zamore PD (2004) A Protein Sensor for siRNA Asymmetry. *Science* **306**: 1377–1380

Tu B, Liu L, Xu C, Zhai J, Li S, Lopez MA, Zhao Y, Yu Y, Ramachandran V, Ren G, Yu B, Li S, Meyers BC, Mo B & Chen X (2015) Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in Arabidopsis. *PLoS Genet.* **11**: e1005119

Tucker MR, Okada T, Hu Y, Scholefield A, Taylor JM & Koltunow AMG (2012) Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in Arabidopsis. *Development* **139**: 1399–404

Vaillant I & Paszkowski J (2007) Role of histone and DNA methylation in gene regulation. *Curr. Opin. Plant Biol.* **10**: 528–533

van der Krol AR, Mur LA, Beld M, Mol JN & Stuitje AR (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**: 291–299

Vaucheret H (2008) Plant ARGONAUTES. *Trends in Plant Science* **13**: 350–358

Vaucheret H, Vazquez F, Crété P & Bartel DP (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* **18**: 1187–1197

Vazquez F, Blevins T, Ailhas J, Boller T & Meins F (2008) Evolution of Arabidopsis MIR genes generates novel microRNA classes. *Nucleic Acids Res.* **36**: 6429–6438

Vazquez F, Gascioli V, Crété P & Vaucheret H (2004a) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.* **14**: 346-351

Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gascioli V, Mallory AC, Hilbert J-L, Bartel DP & Crété P (2004b) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol. Cell* **16**: 69-79

Voinnet O (2008) Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci.* **13**: 317-328

Voinnet O, Rivas S, Mestre P, & Baulcombe DC (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**: 949-956

Wang F & Perry SE (2013) Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant Physiol.* **161**: 1251-1264

Wang H, Zhang X, Liu J, Kiba T, Woo J, Ojo T, Hafner M, Tuschl T, Chua N-H & Wang X-J (2011) Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. *Plant J.* **67**: 292-304

Wang L, Song X, Gu L, Li X, Cao S, Chu C, Cui X, Chen X & Cao X (2013) NOT2 proteins promote Polymerase II-dependent transcription and interact with multiple microRNA biogenesis factors in Arabidopsis. *Plant Cell* **25**: 715-727

Wang X-B, Wu Q, Ito T, Cillo F, Li W-X, Chen X, Yu J-L & Ding S-W (2010) RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 484-489

Wang X, Zhang S, Dou Y, Zhang C, Chen X, Yu B & Ren G (2015) Synergistic and Independent Actions of Multiple Terminal Nucleotidyl Transferases in the 3' Tailing of Small RNAs in Arabidopsis. *PLoS Genet.* **11**: 1-17

Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T & Patel DJ (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**: 754-761

Wang Y, Sheng G, Juranek SA, Tuschl T & Patel DJ (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**: 209-213

Wassenegger M & Krczal G (2006) Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci.* **11**: 142-151

Waterhouse PM, Graham MW & Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 13959-13964

Welker NC, Maity TS, Ye X, Aruscavage PJ, Krauchuk AA, Liu Q & Bass BL (2011) Dicer's Helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol. Cell* **41**: 589-599

Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA & Carrington JC (2005a) Expression of Arabidopsis MIRNA genes. *Plant Physiol.* **138**: 2145-2154

Xie Z, Allen E, Wilken A & Carrington JC (2005b) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc Natl Acad Sci USA* **102**: 12984-12989

- Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE & Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**: 642-652
- Xu J & Chua N-H (2011) Processing bodies and plant development. *Curr. Opin. Plant Biol.* **14**: 88-93
- Yang L, Wu G & Poethig RS (2012) Mutations in the GW-repeat protein SUO reveal a developmental function for microRNA-mediated translational repression in Arabidopsis. *Proc. Natl. Acad. Sci.* **109**: 315-320
- Yang Z, Ebright YW, Yu B & Chen X (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.* **34**: 667-675
- Yelina NE, Smith LM, Jones AME, Patel K, Kelly KA & Baulcombe DC (2010) Putative Arabidopsis THO/TREX mRNA export complex is involved in transgene and endogenous siRNA biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 13948-13953
- Yi R, Qin Y, Macara IG & Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**: 3011-3016
- Yoshikawa M, Iki T, Tsutsui Y, Miyashita K, Poethig RS, Habu Y & Ishikawa M (2013) 3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 4117-4122
- Yoshikawa M, Peragine A, Park MY & Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.* **19**: 2164-2175
- Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R & Chen X (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**: 932-935
- Yu D, Fan B, MacFarlane S a & Chen Z (2003) Analysis of the involvement of an inducible Arabidopsis RNA-dependent RNA polymerase in antiviral defense. *Mol. Plant. Microbe. Interact.* **16**: 206-216
- Zambryski P, Joos H, Genetello C, Leemans J, Montagu M V & Schell J (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J.* **2**: 2143-2150
- Zamore PD, Tuschl T, Sharp PA & Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33
- Zhai J, Bischof S, Wang H, Feng S, Lee T, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, Liu W, Henderson IR, Meyers BC, Ausin I & Jacobsen SE (2015) A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. *Cell* **163**: 445-455
- Zhai J, Jeong D-H, De Emanuele P, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ & Meyers BC (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* **25**: 2540-2553
- Zhang H, Kolb FA, Jaskiewicz L, Westhof E & Filipowicz W (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**: 57-68
- Zhang H, Ma Z-Y, Zeng L, Tanaka K, Zhang C-J, Ma J, Bai G, Wang P, Zhang S-W, Liu Z-W, Cai T, Tang K, Liu R, Shi X, He X-J & Zhu J-K (2013) DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 8290-8295

Zhang X, Henderson IR, Lu C, Green PJ & Jacobsen SE (2007) Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 4536-4541

Zhang Z, Liu X, Guo X, Wang X-J & Zhang X (2016) Arabidopsis AGO3 predominantly recruits 24-nt small RNAs to regulate epigenetic silencing. *Nat. plants* **2**: 16049

Zhao X, Zhang H & Li L (2013) Identification and analysis of the proximal promoters of microRNA genes in Arabidopsis. *Genomics* **101**: 187-194

Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, Meyers BC, Mo B & Chen X (2012) The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.* **22**: 689-694

Zheng X, Zhu J, Kapoor A & Zhu J-K (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* **26**: 1691-1701

Zhu H, Hu F, Wang R, Zhou X, Sze S-H, Liou L, Barefoot A, Dickman M & Zhang X (2011) Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell* **145**: 242-256

Zilberman D, Cao D & Jacobsen SE (2003) ARGONAUTE4 Control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**: 716-719





## Biogénèse des siRNAs endogènes chez *Arabidopsis thaliana*

Etude fonctionnelle de DRB7.2, une nouvelle protéine de fixation à l'ARN double brin et développement d'outils moléculaires pour la caractérisation du mode d'action de DCL4

### Résumé

Les ARN double brin (ARNdb) sont les molécules clés initiatrices du RNA silencing, à partir desquelles les différentes classes de petits ARN (sRNAs), conférant la séquence spécificité de ce mécanisme, vont être produit. Chez la plante modèle *Arabidopsis thaliana*, le clivage des divers ARNdb en sRNAs est opéré par quatre enzymes de type RNase III, nommées DCL1 à DCL4, dont l'activité peut être assistée par des protéines fixant l'ARNdb (DRBs).

Au cours de cette thèse, j'ai pu caractériser la fonction d'une nouvelle DRB, DRB7.2. Les résultats obtenus m'ont permis de démontrer que cette protéine régule la production d'une classe particulière de sRNAs endogènes, les endoIR-siRNAs, en séquestrant spécifiquement leurs précurseurs ARNdb, inhibant ainsi leur clivage par les différents DCLs. En parallèle, j'ai également développé des outils moléculaires afin d'étudier le mode d'action du DCL le plus polyvalent chez les plantes, DCL4. La caractérisation détaillée de ces outils a permis de révéler le rôle clé de déterminants structuraux distincts (protéiques ou nucléiques) impliqués dans la spécificité de reconnaissance et de clivage des divers substrats ARNdb par cette enzyme.

Mot-clés : DRBs, sRNAs, endoIR, séquestration, DCL4, outils moléculaire, spécificité de reconnaissance

### Abstract

RNA silencing is initiated by double-stranded RNA (dsRNA) molecules that will be processed into various classes of small RNAs (sRNAs), which confer the sequence-specificity of this mechanism. In the model plant *Arabidopsis thaliana*, dsRNA processing is mediated by four distinct RNaseIII-like enzymes, named DCL1 to DCL4, which can be assisted by dsRNA-binding proteins (DRBs).

During my PhD, I was able to characterize in details the function of a new DRB protein, DRB7.2. Our results revealed that this protein regulates the accumulation of a specific class of endogenous sRNAs, the endoIR-siRNAs, by selectively sequestering their dsRNA precursors and inhibiting their cleavage by the DCLs. In parallel, I also developed molecular tools to study the mode of action of the most versatile DCL in plants, DCL4. Detailed characterization of these tools revealed key roles of distinct structural determinants (at the protein or RNA level), implicated in the specificity and cleavage efficiency of the various dsRNA substrates by DCL4.

Keywords DRBs, sRNAs, endoIR, sequestration, DCLs, molecular tools, structural determinants