

Mécanismes et fonctions de l'ARN interférence induit par l'ARN double brin chez *Paramecium tetraurelia*

INTRODUCTION	1
I. PREAMBULE	1
A. Découverte de l'ARN interférence	1
B. Rôle essentiel de l'ARNi dans de nombreux processus biologiques	2
C. De l'outil génétique au traitement thérapeutique	3
II. MECANISMES D'ARNi INDUITS PAR ARNDB CHEZ LES EUKARYOTES.....	4
A. Origine des ARNdb.....	4
1) ARNdb exogènes.....	4
2) ARNdb endogènes	5
B. Clivage par une endonucléase Dicer.....	8
1) Domaines de fonction et structure des protéines Dicer.....	8
2) Clivage précis et phasé des ARNdb par Dicer.	10
C. Mécanismes du chargement des ARNdb sur les complexes RISC.....	11
1) Diversité et structure des protéines Ago et Piwi	11
2) Chargement du duplex d'ARN sur la protéine Ago	12
3) Implication de Dicer dans le complexe RISC	13
4) Les autres protéines du complexe RISC.....	13
D. Sélection et élimination du brin passager.....	16
E. Répression de l'ARNm.....	17
1) Clivage du messager (siRNA) ou inhibition de la traduction (miRNA)	17
2) Autres fonctions des protéines Ago dans l'ARN interférence.....	18
F. Fonctions essentielles des RdRPs dans l'ARNi	20
1) Conservation des RdRPs au sein des eucaryotes	20
2) Rôle dans l'amplification du signal de répression.....	21
3) Les autres fonctions des RdRPs	28
G. Modifications des sRNAs.....	29

1)	Méthylation en 3'	29
2)	Uridylation ou adénylation en 3'	30
III.	LES VOIES D'ARNi CHEZ <i>PARAMECIUM TETRAURELIA</i>	33
A.	Les ciliés comme organismes modèles.	33
1)	Le dimorphisme nucléaire	34
2)	Les évènements sexuels	35
3)	Les réarrangements génomiques au cours du développement du MAC.....	37
4)	Le génome de <i>P. tetraurelia</i>	37
B.	Les voies d'ARN interférence chez la paramécie	38
1)	Les gènes candidats	38
2)	Les voies de siRNAs.....	43
3)	La voie des scnRNAs	51
IV.	PRESENTATION DU SUJET	54
	RESULTATS	57
I.	PREMIERE PARTIE : IDENTIFICATION ET CARACTERISATION DES GENES IMPLIQUES DANS L'ARNi INDUITE PAR ARNDb	56
A.	A forward genetic screen reveals essential and non-essential RNAi factors in <i>P. tetraurelia</i>	56
B.	Localisation intracellulaire de Pds1	72
C.	Analyse du gène PDS2.....	73
1)	Croisement du mutant 1.28.....	73
2)	La protéine membranaire putative Pds2.	74
3)	Conservation de PDS2.	75
4)	PDS2 n'est pas impliqué la voie d'ARNi induite par transgène.	77
D.	Compléments sur le rôle de Rdr2 au cours des cycles sexuels	78
5)	L'allèle 5.32 du gène RDR2	78
6)	Effet de la mutation 5.32 à l'autogamie	79
7)	Croisement du mutant 5.32.....	80
E.	Identification des mutations induites par les UV.....	83
II.	DEUXIEME PARTIE : ETUDE DU ROLE DES GENES DE L'ARNi DANS LA PRODUCTION DES siRNAs	84
A.	Primary and secondary siRNA biosynthesis involving distinct RdRP activities in <i>Paramecium tetraurelia</i>	84

B. Deux classes de siRNAs uridylés.....	118
RESUME	120
CONCLUSION DISCUSSION.....	128
A. Sous quelle forme l'ARN est-il importé dans le cytoplasme ?	130
B. Le rôle de l'uridylation des siRNAs.....	132
C. Pas d'amplification de l'ARNi induite par ARNdb chez P. tetraurelia	133
D. La fonction endogène des siRNAs induits par feeding d'ARNdb.....	135
REFERENCES BIBLIOGRAPHIQUES	136
ANNEXES	145
I. ANNEXE 1 : DOCUMENTS SUPPLEMENTAIRES DE L'ARTICLE 1.....	147
II. ANNEXE 2: AUTRES DOCUMENTS DE LA PARTIE I DES RESULTATS	168
III. ANNEXE 3: DOCUMENTS SUPPLEMENTAIRES DE L'ARTICLE 2.....	173

Introduction

I. Préambule

A. Découverte de l'ARN interférence

L'effet de la répression homologie dépendante a été observé par le laboratoire de Jorgensen en 1990. Dans l'optique de renforcer la couleur violette des fleurs de pétunia, les auteurs ont injecté un transgène codant une enzyme (l'anthocyanine) de la voie de biosynthèse du pigment qui donne la couleur violette à la fleur : la chalcone synthétase (CHS). Contrairement à l'effet souhaité, l'expression du gène endogène a été abolie et les fleurs sont devenues totalement ou partiellement blanches (Figure 1). Cet effet étant totalement réversible, les auteurs ont suggéré que l'inhibition du gène était dû à une interaction directe ou indirecte entre le gène endogène et le transgène (Napoli et al. 1990^[1]). Des résultats similaires ont été observés au même moment dans le laboratoire de Stuitje (van der Krol et al. 1990^[2]).



Figure 1: Décoloration des fleurs de pétunia. L'injection d'un transgène homologue à un gène de la voie de biosynthèse du pigment donnant la couleur violette à la fleur induit sa décoloration partielle ou totale. a) fleur sauvage, b-d) fleurs après différentes injections (Napoli et al. 1990^[1]).

En 1995, Su Guo et Kenneth J. Kemphues ont découvert qu'une molécule d'ARN pouvait éteindre spécifiquement un gène chez *C. elegans* (Guo & Kemphues 1995^[3]). L'injection de l'ARN anti-sens au gène *par1* dans les gonades de *C. elegans* induit la répression du gène endogène. Etonnement les auteurs ont observé le même effet avec une

molécule d'ARN sens. Chez les plantes, ce mécanisme fut observé à plusieurs reprises (dans les pétunias, chez *Arabidopsis* ainsi que dans les plants de tabac, tomate et pommes de terre) et fut appelé *post-transcriptionnal gene silencing* (PTGS). De plus, il a été montré que des ARN viraux sont également capables d'éteindre spécifiquement des gènes (Baulcombe 1996^[4]). C'est seulement en 1998 que le mécanisme d'ARN interférence induit par ARN double-brin (ARNdb) fut décrit. Andrew Fire et Craig C. Mello ont étudié l'effet de l'injection d'ARN sens, d'ARN anti-sens ou d'un mélange des deux sur l'expression de plusieurs gènes de *C. elegans*. Ils ont remarqué que l'extinction de l'expression du gène est bien plus efficace avec le mélange des deux ARNs et en ont déduit que c'est l'injection d'ARNdb qui induit l'ARN interférence (Fire et al. 1998^[5]). Pour expliquer l'effet de l'injection d'ARN-sens ou anti-sens seul, les auteurs suggèrent que la transcription *in vitro* induit la contamination des préparations d'ARN-sens par quelques molécules d'ARN anti-sens (et inversement) conduisant à la production de quelques molécules d'ARNdb suffisantes pour induire l'ARNi. Le fait que la répression soit inductible par une quantité faible d'ARNdb et qu'elle puisse être maintenue dans la descendance des organismes injectés suggère un mécanisme d'amplification de l'ARN interférence (Fire et al. 1998^[5]). Cette découverte leur a valu le prix Nobel de physiologie et de médecine en 2006.

B. Rôle essentiel de l'ARNi dans de nombreux processus biologiques

L'ARN interférence est un mécanisme biologique essentiel conservé dans l'immense majorité des eucaryotes à l'exception de certaines espèces tel que *Saccharomyces cerevisiae*, *Trypanosoma cruzi*, *Leishmania major* (Excavata), *Cyanidioschyzon merolae* (Archaeplastida) et *Plasmodium falciparum* (Chromalveolata) (Shabalina & Koonin 2008^[6]). On pense généralement que la fonction première de l'ARN interférence, c'est-à-dire la raison pour laquelle ce mécanisme a été conservé depuis le dernier ancêtre commun aux eucaryotes, est la défense contre les parasites moléculaires tel que les virus ou les transposons. Ce mécanisme essentiel de défense antiviral a maintenant été démontré dans un grand nombre d'espèces (plantes et métazoaires notamment). Cependant ces mécanismes d'ARNi ont aussi des rôles essentiels dans de nombreuses fonctions cellulaires. Ainsi, des milliers de petits ARN (sRNAs) sont maintenant référencés dans les bases de données et sont classés en 3 grandes catégories : les microRNAs (miRNAs), les *short interfering RNA* (siRNAs) et les piwi-interacting RNA (les piRNAs). Il est probable que les fonctions de ces sRNAs soient apparues dans un second temps au cours de l'évolution des eucaryotes, suite au détournement de la fonction de défense pour servir d'autres fonctions cellulaires. Par exemple, les miRNAs sont absents dans un grand nombre d'eucaryotes et sont certainement apparus indépendamment chez les plantes et les métazoaires, ils sont donc postérieurs au dernier ancêtre commun des eucaryotes. Au contraire, on peut imaginer que les piRNAs et scnRNAs, spécifiquement impliqués dans la reproduction sexuelle des organismes, étaient présents chez

le dernier ancêtre commun (bien qu'ils soient absents chez les plantes par exemple). Toutes ces fonctions nécessitent des mécanismes variés et plusieurs voies d'ARNi différentes qui seront décrites dans la deuxième partie de cette introduction.

C. De l'outil génétique au traitement thérapeutique

Depuis la découverte du fonctionnement de l'ARN interférence, de très nombreuses études scientifiques ont cherché à utiliser ce mécanisme comme outil d'analyse génétique. En effet, théoriquement tous les gènes d'un organisme peuvent être éteints par ARN interférence à l'unique condition que cette voie cellulaire existe chez l'organisme étudié. Pour utiliser cette méthode il suffit de connaître la séquence du gène que l'on souhaite éteindre, d'utiliser cette séquence pour synthétiser l'ARNdb (*in vitro* ou en transformant une bactérie par un plasmide d'expression) et d'avoir une méthode pour l'introduire dans les cellules de l'organisme (injection, alimentation,...). Cette méthode d'analyse fonctionnelle des gènes est donc utilisée très régulièrement dans la plupart des organismes modèles. Chez les mammifères, les voies d'ARNi existent, cependant l'injection de longs ARNdb induit immédiatement la réponse interféron qui par l'activation de la protéine kinase R bloque globalement la synthèse des protéines et provoque la mort cellulaire (Gantier & Williams 2007^[7]). Pour remédier à ce problème, des double-brin d'ARN plus courts, entre 21 et 27 paires de base (pb) ont été utilisés. Ces duplex de petits ARN, qui peuvent être des duplex de siRNAs ou de shRNAs (*short hairpin RNAs*), induisent l'ARNi mais pas la réponse interféron. Ces découvertes ouvrent alors des perspectives très importantes sur le traitement de toutes les pathologies qui entraînent la surproduction de protéines, en particulier les oncogènes mais aussi potentiellement toutes les maladies infectieuses virales !

II. Mécanismes d'ARNi induits par ARNdb chez les eucaryotes

Dans la deuxième partie de cette introduction, les mécanismes impliqués dans le fonctionnement des voies d'ARN interférences induites par ARNdb chez les eucaryotes seront détaillés. Au cours de ma thèse, mon attention s'est portée plus particulièrement sur deux points: 1) les mécanismes de biogénèse des siRNAs et les protéines impliquées dans leur synthèse 2) l'amplification de l'ARN interférence induite par les ARN polymérase ARN dépendantes (RdRPs). Ces mécanismes seront plus longuement détaillés. Enfin, dans la troisième partie de cette introduction (page 32), je détaillerai les voies d'ARNi végétatives chez l'organisme modèle que j'ai étudié au cours de ma thèse : le cilié *Paramecium tetraurelia*.

A. Origine des ARNdb

Les voies d'ARNi induites par l'ARNdb peuvent être classées en fonction de l'origine des ARNdb, endogènes ou exogènes. Les ARNdb d'origine endogène sont souvent associés à la régulation des gènes de l'organisme et au contrôle de l'expression des transposons alors que ceux d'origine exogène sont associés à la défense de l'organisme contre les virus et autres parasites moléculaires pouvant provenir de l'environnement.

1) ARNdb exogènes

Le nématode *Caenorhabditis elegans* a été le premier animal pour lequel on a observé l'activation des voies d'ARNi en réponse à des amorces exogènes et où les voies d'ARNi activées ont été caractérisées (Fire et al. 1998^[5], Sarkies & Miska 2013^[8]). Depuis, deux sources d'ARNdb exogènes ont été mises en évidence chez plusieurs organismes. Les virus à ARN d'une part et les ARNdb présents dans l'environnement (notamment produits par les bactéries) d'autre part. Si les virus sont une source naturelle connue d'ARNdb, la production d'ARNdb chez les bactéries est au contraire artificielle. Il n'y a pour le moment aucune évidence d'une prise en charge par la machinerie d'ARNi des ARN double-brin ou simple brin produits naturellement par les bactéries.

a. ARN des particules virales

La capacité à utiliser l'ARN d'origine virale pour activer les voies d'ARNi a été montré chez de nombreux eucaryotes, les plantes (Hohn & Vazquez 2011^[9]), les nématodes (Sarkies

& Miska 2013^[8]) et même jusqu'aux mammifères récemment (notamment sur des lignées cellulaires de souris et humaines) (Maillard et al. 2013^[10]).

b. ARNdb importés de l'environnement

La capacité de certains organismes à activer les voies d'ARNi à partir d'ARNdb présents dans l'environnement suggère que des mécanismes de transport sont capables d'importer l'ARNdb de l'environnement vers le cytoplasme. Il est possible de nourrir les *Caenorhabditis elegans* de bactéries *E.coli* transformées pour qu'elles expriment à partir d'un plasmide un ARNdb (Tabara et al. 1998^[11]). Cette méthode appelée "feeding d'ARNdb" active l'ARN interférence chez *C. elegans*. Afin de découvrir des gènes impliqués dans l'ARNi, les auteurs ont réalisé un crible de mutagenèse au méthanesulfonate d'éthyle (EMS). Le principe est de faire du *feeding* d'un ARNdb homologue à un gène essentiel du nématode. Ainsi, les nématodes capables de faire de l'ARNi vont éteindre leur propre gène essentiel et vont mourir. Au contraire, les nématodes déficients pour l'ARNi vont résister au *feeding* d'ARNdb et vont survivre (Tabara et al. 1999^[12]). Ce crible a notamment permis d'identifier un transporteur de type ABC (ATP binding cassette), haf-6, comme impliqué dans l'ARNi induite par ARNdb (Sundaram et al. 2006^[13]). Par la suite, parmi les 61 transporteurs ABC du génome de *C. elegans*, 10 ont montré une certaine implication dans des voies d'ARNi et dans la répression de transposons (Sundaram et al. 2008^[14]). Un autre transporteur, SID-2, a également été décrit comme impliqué dans le traitement des ARNdb de l'environnement (Winston et al. 2007^[15]). La construction d'une protéine de fusion GFP a révélé que SID-2 permet l'import d'ARNdb de la lumière intestinale vers le cytoplasme. Étonnement, la sensibilité aux ARNdb de l'environnement est rare au sein des *Caenorhabditis*, seul 2 espèces sur 8 testées ont montré cette sensibilité (Winston et al. 2007^[15]). De plus, l'injection de SID-2 dans *C.remanei* qui en est dépourvu suffit à lui conférer la sensibilité aux ARNdb de l'environnement ce qui suggère que cette fonction a été perdue chez certaines espèces au cours de l'évolution des nématodes (Nuez & Felix 2012^[16]). Cette capacité à importer l'ARNdb présent dans les bactéries a pu être utilisée comme outil d'analyse génétique chez les nématodes mais aussi chez les ciliés (Galvani & Sperling 2002^[17]).

D'autres part l'import direct d'ARNdb présents dans l'environnement est couramment utilisé dans les lignées cellulaires de drosophiles (Zhou et al. 2014^[18]).

2) ARNdb endogènes

Il existe de nombreuses sources endogènes d'ARNdb. La nomenclature des sRNAs peut être associée à leur origine, leur fonction ou encore les protéines impliquées dans leur biogénèse. Par exemple, les noms des siRNAs endogènes chez *A. thaliana* sont donnés en fonction de leurs origines : hc-RNA (hétérochromatique), NAT-siRNA (transcrits anti-sens

naturels), rasiRNA (associé aux répétitions), ... (Figure 2). Ces sRNAs ont souvent des similitudes dans leurs caractéristiques et au niveau des protéines impliquées dans leur biogénèse.

Il semble alors plus simple de distinguer les sRNAs endogènes en deux classes. 1) les sRNAs qui ont pour origine une molécule d'ARN qui a la capacité intramoléculaire de former une structure de type tige-boucle et 2) les sRNAs qui ont pour origine deux ARNs complémentaires qui peuvent former un double brin (Figure 2). La première catégorie de précurseur forme principalement des miRNAs alors que la seconde forme les siRNAs. Il existe cependant des structures de type "épingle à cheveux" plus grandes que la normale et qui permettent la production de siRNAs (Dunoyer et al. 2010^[19]). Dans la suite de cette introduction, l'accent sera mis sur les siRNAs (endogènes ou exogènes).

De nombreux ARNdb endogènes sont dépendants d'une RdRP (ARN polymérase ARN dépendante) pour être synthétisés. Les mécanismes de production d'ARNdb par les RdRPs seront décrits dans la section correspondante (page 19).

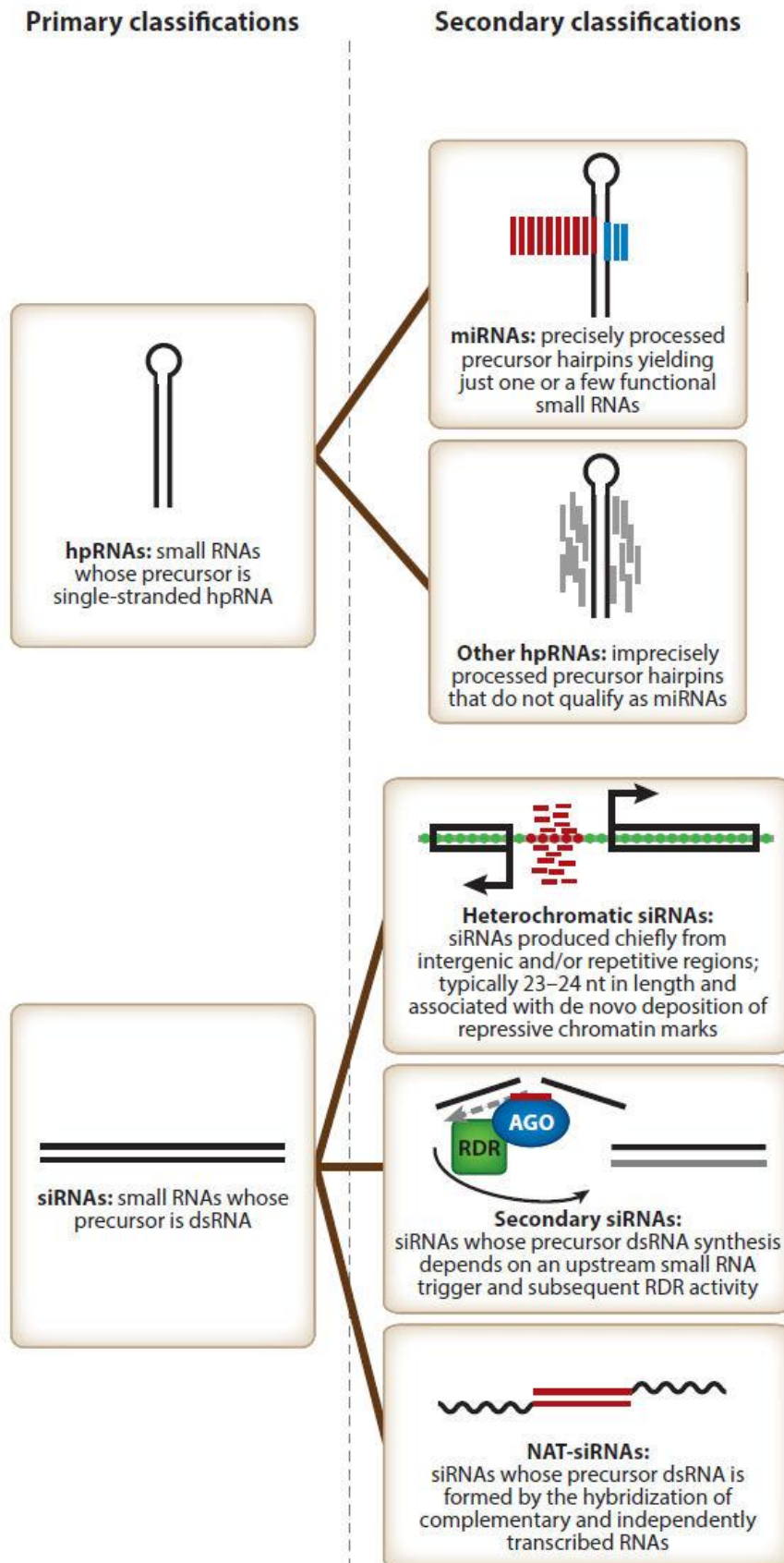


Figure 2: Classification des ARNdb endogènes chez *A. thaliana* (Axtell 2013^[20]). La nomenclature des sRNAs dépend de leur origine, un double-brin mono-moléculaire (structure en épingle à cheveux) ou bi-moléculaire, ainsi que des protéines impliquées dans leur biogénèse.

B. Clivage par une endonucléase Dicer.

1) Domaines de fonction et structure des protéines Dicer.

Les protéines de la famille Dicer ou Dicer-like sont des endonucléases de type RNase-III capables de reconnaître et cliver des ARNdb. Ces protéines produisent des duplex d'ARN courts (entre 21 et 28 nt) avec 2 nucléotides sortants en 3'. Les Dicer sont conservés chez tous les eucaryotes pourvus de voies d'ARNi et possèdent en général deux domaines endonucléases (endoND1 et endoND2), un domaine de liaison à l'ARNdb et une grande extension N-terminale qui contient un domaine hélicase et un domaine PAZ (Piwi Argonaute Zwiille) (Figure 3).

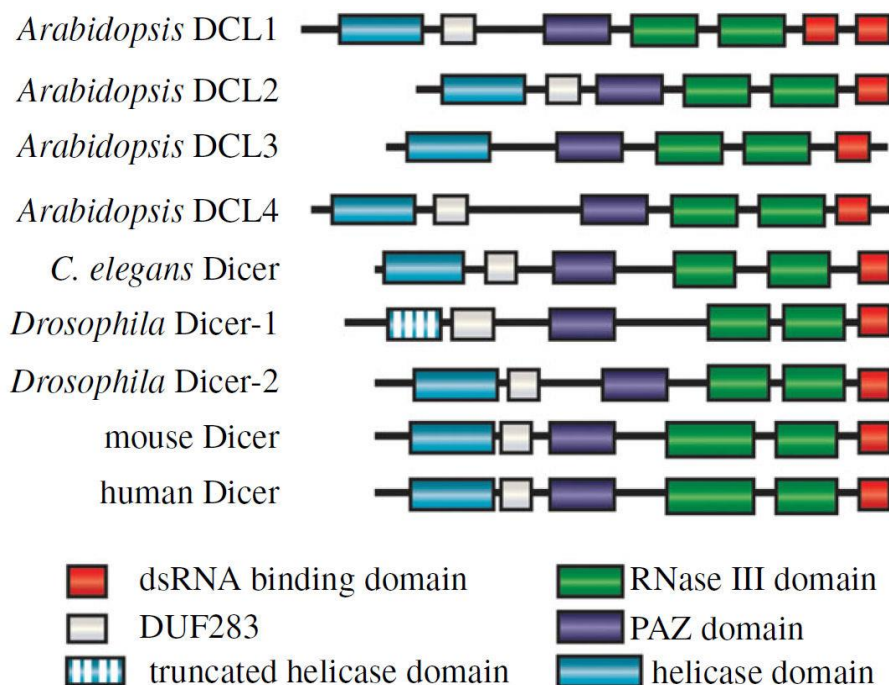


Figure 3: Domaines protéiques de l'endonucléase Dicer conservés chez les eucaryotes. (Johanson et al. 2013^[21])

Le domaine hélicase est présent dans les Dicer de la plupart des eucaryotes (mammifères, diptères, nématodes, plantes et levures), cependant certaines protéines Dicer ne possèdent pas ce domaine et restent fonctionnelles ce qui suggère que cette activité est dispensable dans certains cas. La protéine Dicer du protozoaire *Giardia intestinalis* (Macrae et al. 2006^[22]) ou encore Dicer-1 de *Drosophila melanogaster* requise pour la synthèse des miRNAs (Kim et al. 2006^[23]) n'ont pas de domaine hélicase. Il a été montré que le domaine hélicase peut être requis pour la production de certains siRNAs mais pas pour d'autres. Certains siRNAs endogènes de *C. elegans* requièrent le domaine hélicase contrairement aux

siRNAs exogènes et aux miRNAs (Welker et al. 2010^[24]). Les auteurs suggèrent que l'hélicase aiderait l'enzyme à se déplacer sur certains longs ARNdb permettant plusieurs clivages successifs sans se dissocier de l'ARNdb. Cette processivité de Dicer serait particulièrement importante lorsque le nombre d'ARNdb est faible pour augmenter l'efficacité de production des sRNAs.

Le domaine PAZ est indispensable à la fonction de Dicer, il détermine la position des 2 nt sortants en 3' du duplex de siRNA. Ce serait donc la position de ce domaine dans la protéine qui détermine la taille du siRNA (Cook & Conti 2006^[25]). Ce domaine est lié aux deux domaines endonucléase par un connecteur en hélice qui structure la protéine mais lui donne aussi une certaine flexibilité (Figure 4).

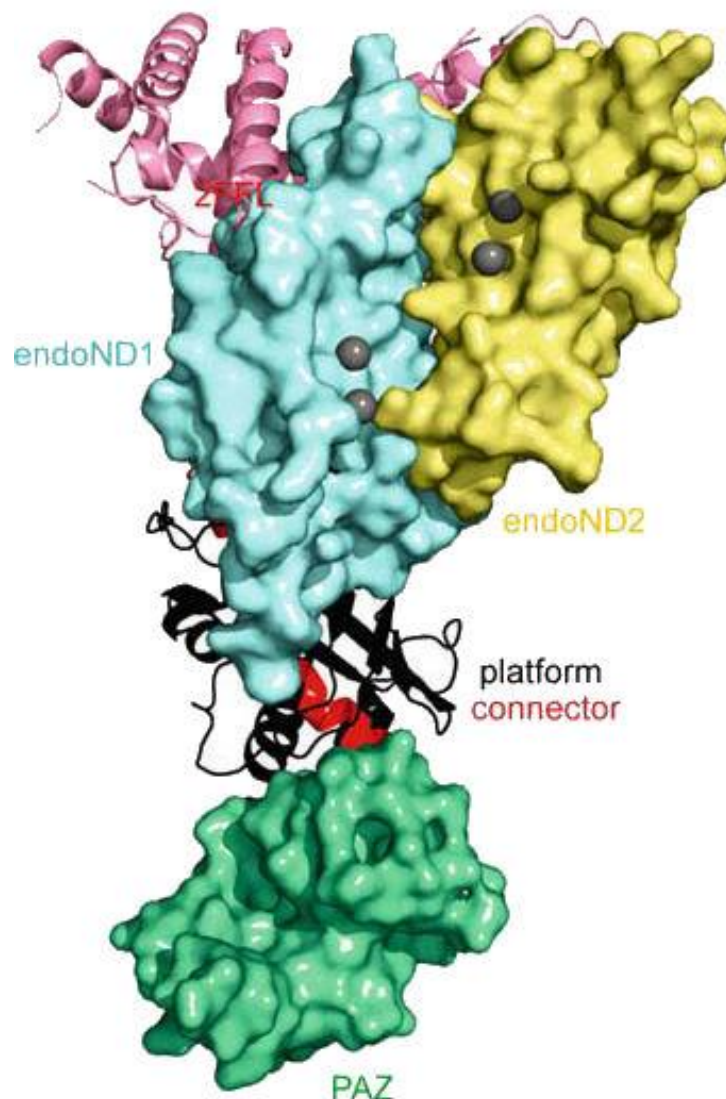


Figure 4: Structure cristallographique de Dicer chez *Giardia intestinalis* (Ji 2008^[26]). Les deux domaines endonucléases (endoN1 et endoN2) forment une cavité qui peut recevoir un ARNdb. Le domaine PAZ, lié aux endonucléases par un connecteur va fixer les deux nucléotides 3' sortants et ainsi définir la taille du duplex de sRNAs.

Les deux domaines endonucléases portent l'activité RNase-III de Dicer et sont les plus conservés. Ils forment un dimère intramoléculaire avec une cavité en son centre qui permet le passage du double-brin d'ARN (Figure 4). La signature du domaine RNase-III est formée par 8 résidus très conservés (ERLEMLGD pour le domaine endoND1 de l'Homme). 4 acides aminés (E D D E) sont responsables de l'activité catalytique (Figure 5). Les 8 acides aminés responsables de l'activité catalytique sont situés aux deux extrémités de la cavité formée par les 2 endonucléases.

Dcl2Ci	1018	VLTATSSSAAREASNYQR	-L	E	F	L	G	D	S	L	K	L	H	T	S	I	O	L	1121	M	S	T	K	V	I	A	D	V	V	E	A	L	I	G	--A		
Dcr2Dm	1194	FIAAITASSAADVFDMER	-L	E	F	L	G	D	S	F	L	K	L	S	A	T	L	Y	L	1361	I	P	N	K	V	I	A	D	T	L	E	A	L	I	G	--V	
Dcr1Sp	914	ACQALTSAESQLNFDYDR	-L	E	F	Y	G	D	C	F	L	K	L	G	A	S	I	T	V	1017	L	S	V	K	R	I	A	D	M	V	E	A	S	I	G	--A	
Dcl1Ci	1046	RAQQIHFORGM-GKNYER	-L	E	F	L	G	D	C	F	L	K	M	A	T	S	I	S	L	1115	L	A	D	K	T	I	A	D	V	C	E	A	L	I	G	A	C
Dcr1Ce	1326	ILTALTTSNAAAGMSLER	-E	E	T	I	G	D	S	F	L	K	F	A	T	T	D	Y	L	1503	I	S	D	K	S	I	A	D	V	E	A	L	I	G	--V		
Dcr1Hs	1287	ILQALTLNASDGFNER	-L	E	N	L	G	D	S	F	L	K	H	A	I	T	T	Y	L	1544	I	A	D	K	S	I	A	D	C	V	E	A	L	I	G	--C	
Dcr1Dm	1682	ILQALTMSNANDGINLER	-L	E	T	I	G	D	S	E	L	K	Y	A	I	T	T	Y	L	1898	I	P	D	K	S	I	A	D	C	V	E	A	L	I	G	--A	
DcrOs	1050	ILEAITTLRCSEDFSMER	-L	E	L	L	G	D	S	V	L	K	Y	A	V	S	C	H	L	1179	M	C	S	K	T	I	A	D	C	V	E	A	L	I	G	--A	
Dcl3At	957	ILEAVTTLTCPEFSMER	-L	E	L	L	G	D	S	V	L	K	Y	V	A	S	C	H	L	1087	V	V	S	K	S	V	S	D	C	A	E	A	L	I	G	--A	
Dcl4At	1103	VLEALTTEKCHERLSLER	-L	E	V	L	G	D	A	F	L	K	F	A	V	S	R	H	L	1230	L	Y	K	K	T	I	A	D	V	V	E	A	L	I	G	--A	
Dcl1At	1359	ILEALTAASCOETFCYER	-A	E	L	L	G	D	A	V	L	K	V	V	S	R	F	L	1497	L	S	S	K	T	I	A	D	V	V	E	A	L	I	G	--V		
Dcl2At	979	VLEAITTKKCEDQFHLES	-L	E	T	L	G	D	S	F	L	K	Y	A	V	C	Q	L	1092	L	K	R	K	S	V	A	D	V	V	E	S	L	I	G	--A		
Dcr1Tt	958	TITERDLIQIFQKKQLEKN	V	E	L	G	K	Q	Y	F	R	F	I	T	S	I	N	V	1081	I	R	N	K	Q	F	T	I	I	Y	I	I	G	--I	T			
Dcr2Tt	1312	SLTSLEYAVNNKESLER	-Y	E	F	L	G	D	I	V	L	K	C	L	S	T	Q	I	1466	I	K	G	K	T	I	A	D	T	L	E	S	I	G	--V			
Dcr1Pt	1310	CFQSQQFNQ-DQQTNYQV	-L	E	F	L	G	D	A	N	L	K	L	L	S	S	I	E	V	1438	I	P	E	K	V	H	S	D	V	V	E	A	L	I	G	--A	
Dcr2Pt	1101	RLDEILFKSLDLSSHKYN	-E	L	L	G	H	Q	Y	K	F	L	I	A	V	V	L	1225	I	S	N	Q	E	F	F	Q	F	I	Q	N	S	---	L				
Dcr3Pt	1082	ILNSLNVS----AYKLED	-L	-	M	L	G	S	S	Y	K	F	L	I	T	I	Q	L	1195	I	S	D	I	D	F	K	Q	F	L	Q	S	L	Q	---			

Figure 5: Alignement des deux régions les plus conservées du premier domaine RNase-III de Dicer chez les eucaryotes. Les acides aminés catalytiques sont indiqués en rouge. Pt = *Paramecium tetraurelia*, Tt = *Tetrahymena thermophila*, At = *Arabidopsis thaliana*, Os = *Oryza sativa*, Dm = *Drosophila melanogaster*, Hs = *Homo sapiens*, Ce = *Caenorhabditis elegans*, Ci = *Coccidioides immitis*, Sp = *Saccharomyces pombe*. (Lepere et al. 2009^[27])

Enfin, le domaine de liaison à l'ARNdb comme le domaine hélicase semblent ne pas être requis pour l'activité de Dicer chez un certain nombre d'eucaryote. La fonction de liaison à l'ARNdb est porté entre autre par le domaine DUF283 (Dlatic 2006^[28]). La structure minimale de Dicer est donc composée d'un dimère d'endonucléase et du domaine PAZ reliés par un connecteur central.

2) Clivage précis et phasé des ARNdb par Dicer.

Dicer est une protéine capable de couper très régulièrement les ARNdb. Chez *A. thaliana*, les 4 protéines Dicer-like sont impliquées dans des voies différentes et produisent des sRNAs dont la taille est variable. Dcl1 est associée uniquement aux miRNAs de 20-22 nt, Dcl2 aux nat-siRNAs de 22 nt, Dcl3 aux siRNAs hétérochromatiques (casiRNA ou hsiRNA) de 24 nt et enfin Dcl4 aux tasiRNAs (trans-acting) de 21 nt (Liu et al. 2009^[29], Xie & Qi 2008^[30]). De la même manière, chez *D. melanogaster*, les protéines Dicer-1 et Dicer-2 sont respectivement requises pour la synthèse des miRNAs (22 nt) et des siRNAs (21 nt) (Lee et al. 2004^[31]). La

taille du connecteur entre les domaines PAZ et RNase-III est à l'origine de cette spécialisation des protéines Dicer qui semble donc essentielle à la diversité des voies d'ARN interférence identifiables notamment par la taille des sRNAs. Contrairement à la plupart des eucaryotes, il n'y a qu'une seule protéine Dicer chez les mammifères.

Chez certains organismes, la procéssivité de Dicer couplée à cette capacité de couper des duplex de siRNAs de taille fixe génère des siRNAs phasés. Ce terme signifie qu'ils sont produits régulièrement sans se chevaucher et sans espace entre eux. C'est le cas par exemple des tasiRNAs de plantes (voir figure 13).

C. Mécanismes du chargement des ARNdb sur les complexes RISC

Malgré la diversité des voies de sRNAs chez les eucaryotes, tous les petits ARNs sont finalement chargés sur un complexe RISC (RNA-induced silencing complex). La structure et les protéines impliquées dans le complexe varient selon les voies et selon les espèces mais il contient au minimum une protéine de type Argonaute (Kuhn & Joshua-Tor 2013^[32]). Le sRNA va diriger Argonaute vers sa cible (un long ARN simple brin) ce qui va induire sa dégradation ou le blocage de sa traduction. Les argonautes sont donc des protéines essentielles à la fonction de l'ARN interférence.

1) Diversité et structure des protéines Ago et Piwi

Les protéines de la famille Argonaute sont souvent en grand nombre dans les génomes eucaryotes. On n'en dénombre pas moins de 26 chez *C. elegans*, 10 chez *A. thaliana*, 8 chez l'homme (1 seule catalytique), 5 chez *D. melanogaster* et 1 seule chez *S. pombe*. Cette diversité des protéines argonautes reflète la complexité des voies d'ARNi dans chaque organisme.

Les argonautes sont caractérisées par 3 domaines principaux: un domaine PAZ, le domaine MID (au milieu) et le domaine PIWI. Le domaine PAZ, de même que dans la protéine Dicer, a pour fonction de fixer l'extrémité 3' de l'ARN, le domaine MID sert d'ancrage au 5' phosphate du brin guide et le domaine PIWI a une structure proche de la RNase H et porte la fonction de clivage de l'ARN ciblé, appelée *slicing*. 3 acides aminés catalytiques formant le motif DDX (X étant D ou H) sont conservés dans la plupart des domaines PIWI (Kuhn & Joshua-Tor 2013^[32]) (Figure 6).

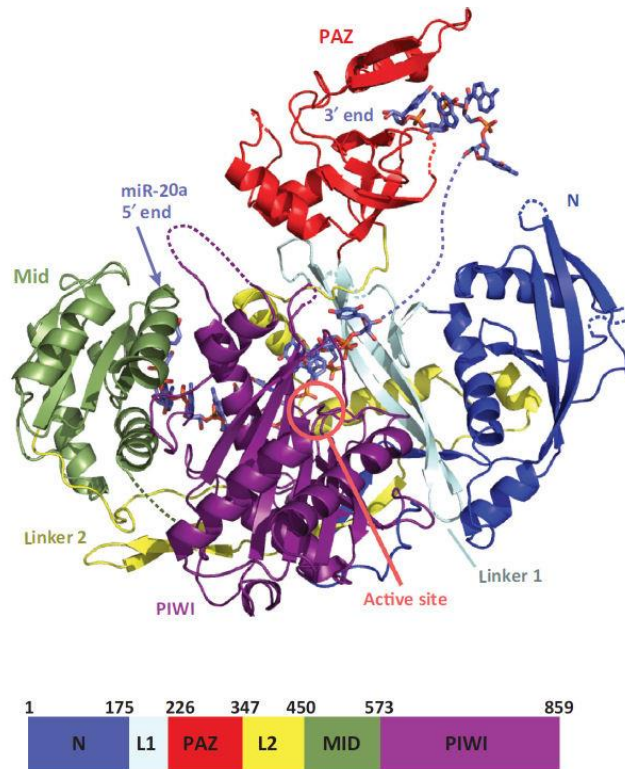


Figure 6: Structure de la protéine Ago2 humaine.
(Kuhn & Joshua-Tor 2013^[32]).

Les protéines argonautes peuvent être divisées en deux sous-familles : les protéines de type AGO (similaires en séquence à la protéine AGO1 chez *A. thaliana*) et les protéines de type PIWI (similaires aux protéines Piwi de *D. melanogaster*). Les mécanismes de chargement des piRNAs dans les PIWI étant peu connus, seuls les mécanismes impliquant les argonautes de la sous-famille AGO seront décrits plus précisément dans les paragraphes suivants. Chez *C. elegans*, il existe une troisième famille de protéines argonautes, aussi distante des PIWI que des AGO et spécifique aux nématodes : les WAGO (worms argonaute) (Meister 2013^[33]). Les WAGO forment une famille de 12 protéines et sont impliquées dans la prise en charge des siRNAs secondaires (Yigit et al 2006, Gu et al 2009).

2) Chargement du duplex d'ARN sur la protéine Ago

Dicer produit un ARNdb de 21 à 28 nt et un des deux brins du duplexe va être reconnu par AGO. Le brin chargé dans la protéine sera alors le brin guide et l'autre brin, maintenu dans un premier temps par l'appariement des bases sera le brin passager (les facteurs influant sur le choix du brin guide seront discutés dans les paragraphes suivants). Les mismatches dans les précurseurs de miRNAs vont induire leur chargement sur des argonautes spécialisées dans les voies de miRNAs alors que les duplex sans mismatch seront chargés dans les argonautes des voies de siRNAs. C'est par exemple le cas chez *D. melanogaster* où Ago1 charge les miRNAs

et Ago2 les siRNAs. De même chez *C. elegans*, les duplex de miRNAs et de siRNAs sont clivés par la même protéine Dicer mais sont chargés dans des Ago différentes (RDE-1 pour les siRNAs et ALG-1 / 2 pour les miRNAs). Cependant, ce critère de sélection n'est pas strict chez *C. elegans*. Plus le nombre de mismatch sur le duplex de sRNA diminue plus l'interaction avec RDE-1 augmente. Certains miRNAs endogènes avec peu de mismatch (let-7, miR-71) peuvent être chargés sur les deux Ago (Steiner et al. 2007^[34]). Cette observation met en évidence une interaction possible entre les voies de siRNAs et de miRNAs. Au sein d'une même classe de sRNAs, les AGO peuvent être très spécialisés. Chez *A. thaliana*, des siRNAs de 21 nt sont chargés dans AGO1 (les siRNAs viraux, siRNAs induits par transgène et les ta-siRNAs) et des siRNAs de 24nt sont chargés dans AGO4 (siRNAs impliqués dans la maintenance de l'hétérochromatine) (Vaucheret 2008^[35]). Ces deux Ago vont ensuite induire des voies d'ARNi différentes. Certaines Ago sont aussi spécifiques du nucléotide en 5' du brin guide. Les nucléotides favorisés peuvent varier d'une espèce à l'autre. Chez *Neurospora crassa*, l'argonaute QDE-2 charge préférentiellement les sRNAs qui portent une uridine en 5' (Lee et al. 2009^[36]) alors que les WAGO des nématodes sont spécifiques des siRNAs commençants par une guanine (22G siRNAs) (Grishok 2013^[37]). Enfin, chez les plantes, AGO2 et AGO4 recrutent majoritairement les siRNAs avec un A en 5', AGO1 les miRNAs avec un U et AGO5 les siRNAs avec un C (Mi et al. 2008^[38]).

3) Implication de Dicer dans le complexe RISC

Chez plusieurs organismes Dicer est associé au complexe RISC, qui ne le contient pas dans la définition couramment admise. *In vitro*, chez l'homme, l'interaction forte entre les Ago et les protéines Dicer a pu être montrée par purification d'affinité (Meister et al. 2005^[39]). Cette interaction entre Dicer et Ago chez certains organismes pourrait augmenter l'efficacité de l'ARNi en réduisant le nombre d'étapes nécessaires à la production d'un sRNA. En effet, le couplage du clivage du long ARNdb et du chargement sur la protéine argonaute permettrait d'éviter que des duplex de sRNAs ne soient perdus ou dégradés avant leur chargement sur le complexe RISC.

4) Les autres protéines du complexe RISC

a. Les HSP

Chez les plantes, les HSP90 sont nécessaires *in vitro* au fonctionnement des complexes RISC. Un dimère d'HSP90 pourrait se lier à Ago1 (étape 1) pour induire un changement de conformation et permettre la fixation de l'ARNdb (étape 2). Une fois Ago1 chargé, l'hydrolyse de l'ATP induirait le relargage des HSP90 (étape 3) et le retour à la configuration initiale de Ago1 pourrait déclencher le clivage du brin passager (étape 4) (Iki et al. 2010^[40]) (Figure 7).

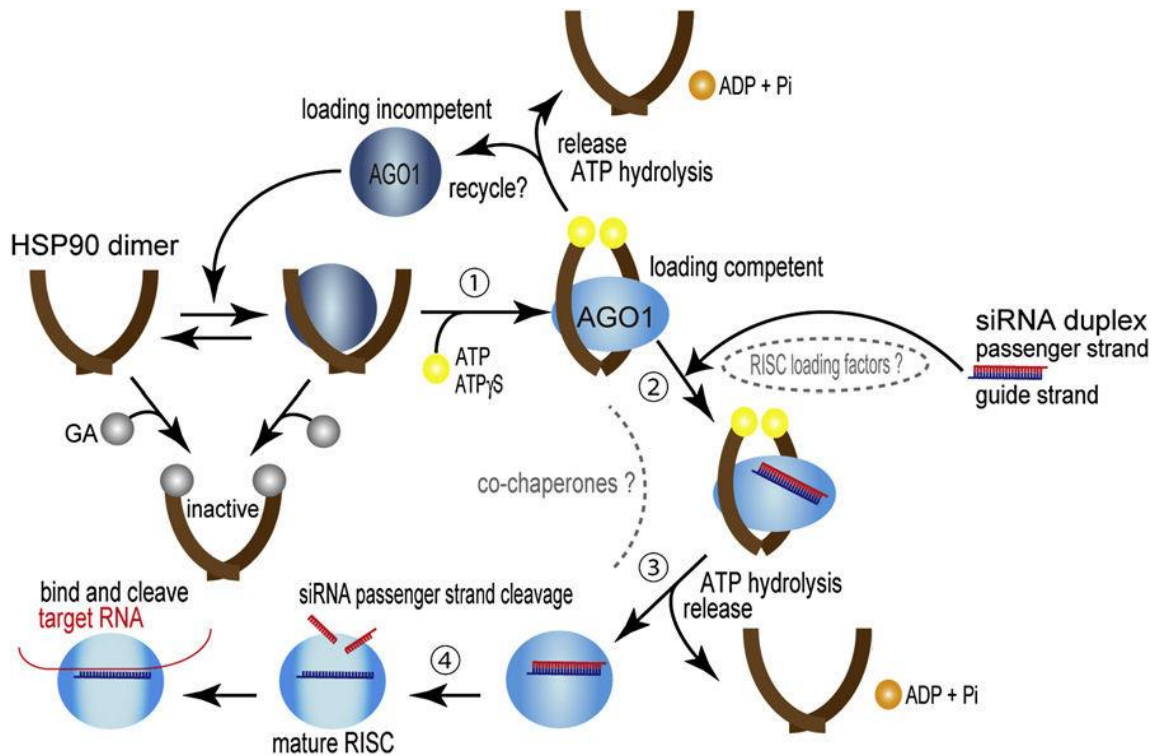


Figure 7: Mécanisme possible pour la formation d'un complexe RISC actif chez *A. thaliana* (Iki et al. 2010^[40]).

Un mécanisme tout à fait similaire a été décrit chez *D. melanogaster* (Iwasaki et al. 2010^[41]). Enfin, dans des lignées cellulaires humaines, les protéines chaperonnes HSP90 sont aussi essentielles à la stabilité du complexe RISC avant le chargement d'un siRNA. Après traitement à la geldanamycine, inhibiteur de HSP90, l'abondance des miRNAs est maintenue mais le chargement des siRNAs dans Ago2 est fortement affectée (Johnston et al. 2010^[42]).

b. Les protéines de liaison aux ARNdb

Les interactions entre Dicer et Ago nécessitent très souvent la présence d'autres facteurs. Dans des lignées cellulaires humaines, TRBP qui contient 3 domaines de liaison à l'ARNdb est capable de se lier à Dicer lorsque qu'un duplex de sRNAs vient d'être clivé, puis de recruter la protéine Ago (Chendrimada et al. 2005^[43]). Cette protéine a donc la capacité de coupler les deux premières étapes de l'ARNi : le clivage par Dicer et le chargement sur le complexe RISC (Figure 8).

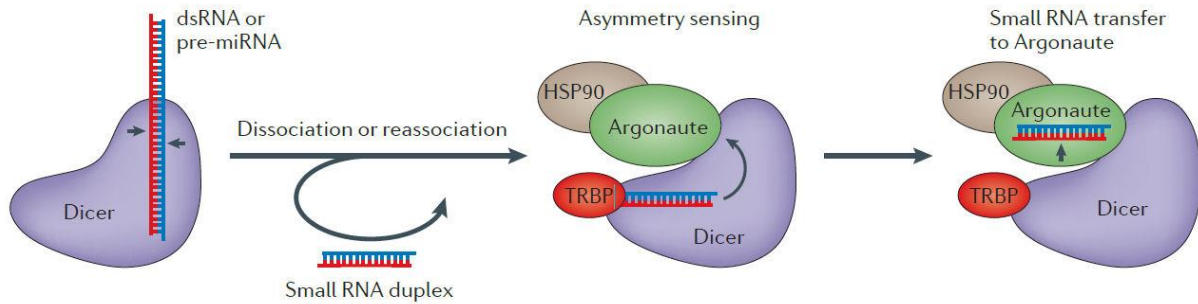


Figure 8: Modèle pour le chargement des ARNdb dans les protéines Ago. La protéine TRBP de liaison à l'ARNdb oriente la molécule d'ARN et participe à son chargement dans Ago (Meister 2013^[33]).

Un rôle similaire a été prédit pour la protéine R2D2 chez *D. melanogaster* (Liu et al. 2003^[44]). Deux autres fonctions essentielles ont été décrites pour cette protéine : 1) Ce facteur permet la localisation de Dcr2 dans des *loci* cytoplasmiques qui contiennent les protéines Ago2 et limiterait ainsi le chargement des siRNAs dans Ago1 (normalement utilisé pour les miRNAs) (Nishida et al. 2013^[45]). 2) La fixation de R2D2 d'un côté ou de l'autre de l'ARNdb serait déterminante pour le choix du brin passager et du brin guide (Tomari et al. 2004^[46]).

D. Sélection et élimination du brin passager

Le principal critère de sélection du brin guide (qu'il s'agisse des miRNAs ou des siRNAs) est thermodynamique. Le brin du duplex avec l'extrémité 5' la moins stable est préférentiellement chargé dans les protéines Ago. Une simple liaison hydrogène supplémentaire sur un brin peut suffire à en faire le brin guide (Schwarz et al. 2003^[47]). De plus, le chargement dans Ago requiert que le brin guide soit phosphorylé (Nykanen et al. 2001^[48]). Dans les voies de siRNAs, la séparation des deux brins du duplex est dépendante du clivage endonucléolytique du brin passager. Les domaines Piwi sont capables de couper la liaison phosphodiester du brin passager des duplex de siRNAs ce qui réduit considérablement la stabilité thermodynamique du duplex et facilite sa séparation (Kwak & Tomari 2012^[49]). Au contraire, dans les voies de synthèse des miRNAs, les mismatches entre les deux brins empêchent la coupure et le brin passager est séparé du brin guide sans être clivé. Dans les deux cas, le domaine N terminal d'Ago est particulièrement important dans la séparation des deux brins (Kwak & Tomari 2012^[49]) (Figure 9).

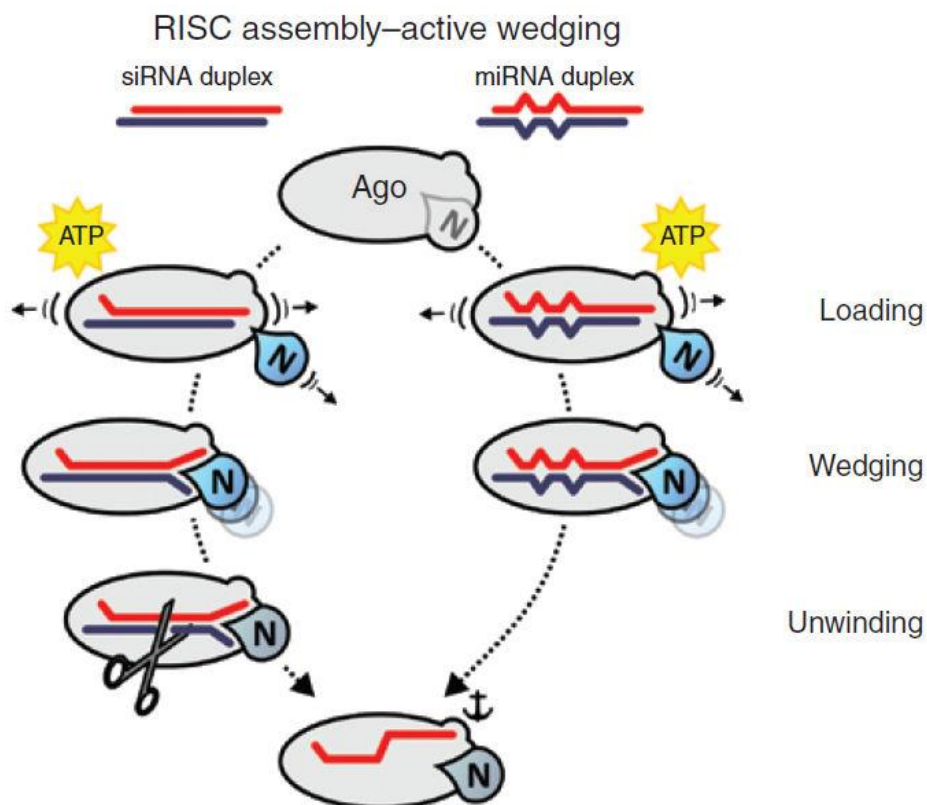


Figure 9 : Modèle pour l'assemblage de RISC chez l'homme (Kwak & Tomari 2012^[49]). Après chargement du duplex de siRNAs ou de miRNAs dans Ago, le domaine N de la protéine initie l'ouverture du duplex. Dans le cas d'un duplex de siRNAs, l'ouverture nécessite le clivage du brin passager contrairement aux duplex de miRNAs qui peuvent être séparés sans clivage.

Chez *D. melanogaster*, Ago2 est capable de cliver le brin passager en fragments de 9 et 12 nt qui sont ensuite retirés du complexe RISC par une protéine de liaison à l'ARN simple brin: C3PO (Liu et al. 2009^[50]). Les hélicases (RHA, MOV10 ou Armitage) faisant partie du complexe RISC sont aussi impliquées dans l'ouverture des deux brins du duplex (Meister 2013^[33]). De manière très similaire, chez *Neurospora*, l'ago QDE-2 et son activité de clivage sont requises pour produire des siRNAs ce qui indique que QDE-2 est capable de cliver le brin passager. QDE-2 agit avec la protéine QIP qui retire le brin passager clivé (Maiti et al. 2007^[51]).

Enfin, chez *C. elegans*, l'ago RDE-1 est nécessaire au chargement de l'ARNdb et au clivage du brin passager mais n'a pas de rôle majeur dans le clivage de l'ARNm ciblé. En effet, RDE-1 va induire le recrutement des RdRPs pour la production de siRNAs secondaires mais sera très peu efficace dans la dégradation directe des ARNm. L'activité de clivage de RDE-1 est donc spécifiquement impliquée dans l'élimination du brin passager (Steiner et al. 2009^[52], Pak et al. 2012^[53]).

E. Répression de l'ARNm.

1) Clivage du messenger (siRNA) ou inhibition de la traduction (miRNA)

Lorsque le complexe RISC et notamment l'activité slicer de l'argonaute sont fonctionnels, l'interaction entre un siRNA et sa cible induit généralement le clivage de l'ARNm alors qu'un miRNA va inhiber la traduction en séquestrant les ARNm dans des *foci* cytoplasmiques (Bartel 2009^[54]). La complémentarité entre le sRNA et sa cible pourrait être à l'origine de la distinction entre ces deux mécanismes. Dans le cas d'une interaction parfaite (au moins pour la moitié 5' du sRNA), l'ARNm serait clivé par l'argonaute alors qu'une interaction imparfaite inhiberait la traduction. Dans les cas où le miRNA s'aligne parfaitement sur sa cible, l'argonaute est alors capable de cliver l'ARNm.

2) Autres fonctions des protéines Ago dans l'ARN interférence.

En plus de la répression de la traduction des ARNm dans le cytoplasme, les protéines Ago peuvent être nucléaires et être impliquées dans différents processus (Figure 10).

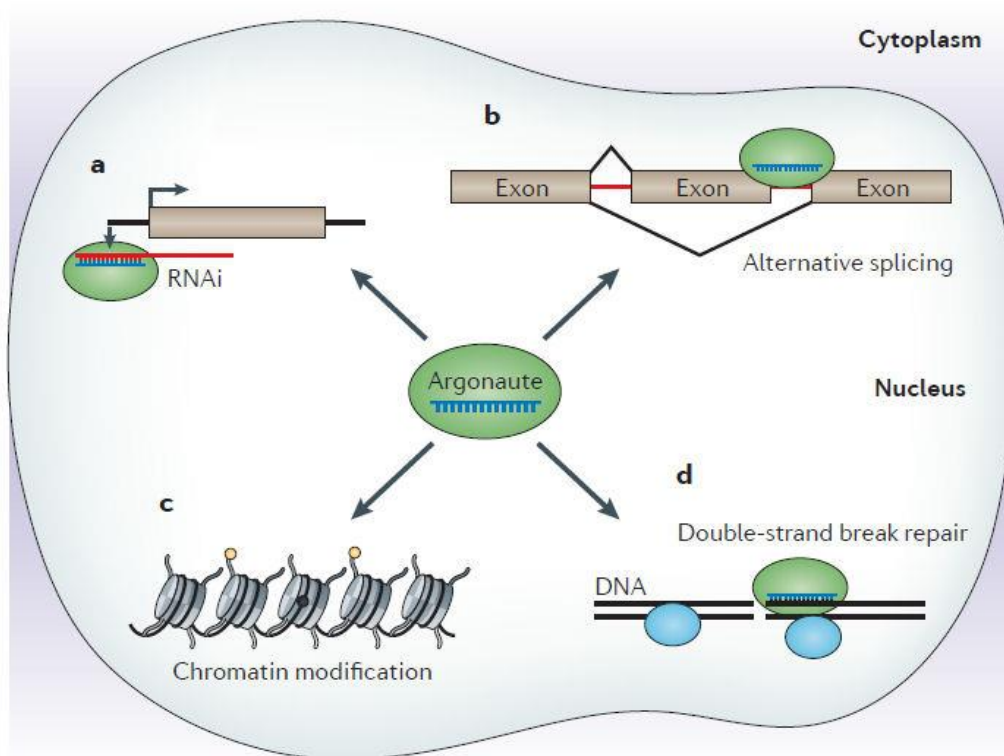


Figure 10: Les fonctions nucléaires d'argonaute (Meister 2013^[33]). Les protéines argonautes peuvent être localisées dans le noyau. En plus de la répression classique des ARNm, les Ago sont impliquées dans l'épissage alternatif, la réparation des cassures double-brin et les modifications de la chromatine

a. Répression transcriptionnelle

Chez *S. pombe*, Ago1 cible les transcrits naissants des régions répétées centromériques pour induire la répression transcriptionnelle. Le complexe RITS (transcriptional gene silencing complex) recrute alors les méthyl-transférases qui vont méthyler la lysine 9 de l'histone H3 ce qui construit un site de liaison pour les protéines à chromodomaines de la famille HP1 (Chp1, Chp2 et Swi6) qui vont empêcher toute transcription de l'hétérochromatine (Grewal 2010^[55]).

Des mécanismes similaires ont été montrés chez les plantes (voir figure 13 page 22) et chez *C. elegans* où l'ARNi protège l'organisme contre la multiplication des transposons en les réprimant au niveau transcriptionnel. Ces marques déposées sur la chromatine peuvent être

conservées sur plusieurs générations. En particulier, l'argonaute HRDE-1 (heritable RNAi-defective 1), qui induit la tri-méthylation de H3K9, est nécessaire à la transmission épigénétique de l'information (Meister 2013^[33]).

b. Epissage alternatif

L'ARN interférence a aussi un rôle dans l'épissage alternatif des introns chez les mammifères. Certains siRNAs ciblent spécifiquement les séquences introniques et influent sur leur épissage. Chez l'homme, cet effet est dépendant d'AGO1 qui est capable d'interagir avec les composants de la machinerie d'épissage. Récemment, ce même mécanisme a été découvert chez *D. melanogaster* (Taliaferro et al. 2013^[56]).

c. Réparation des cassures double-brin

Chez *A. thaliana*, des sRNAs sont produits spécifiquement autour des cassures double-brin du génome (diRNAs). Ces diRNAs s'associent à Ago2 et sont nécessaires pour la réparation (Wei et al. 2012^[57]) (Figure 11).

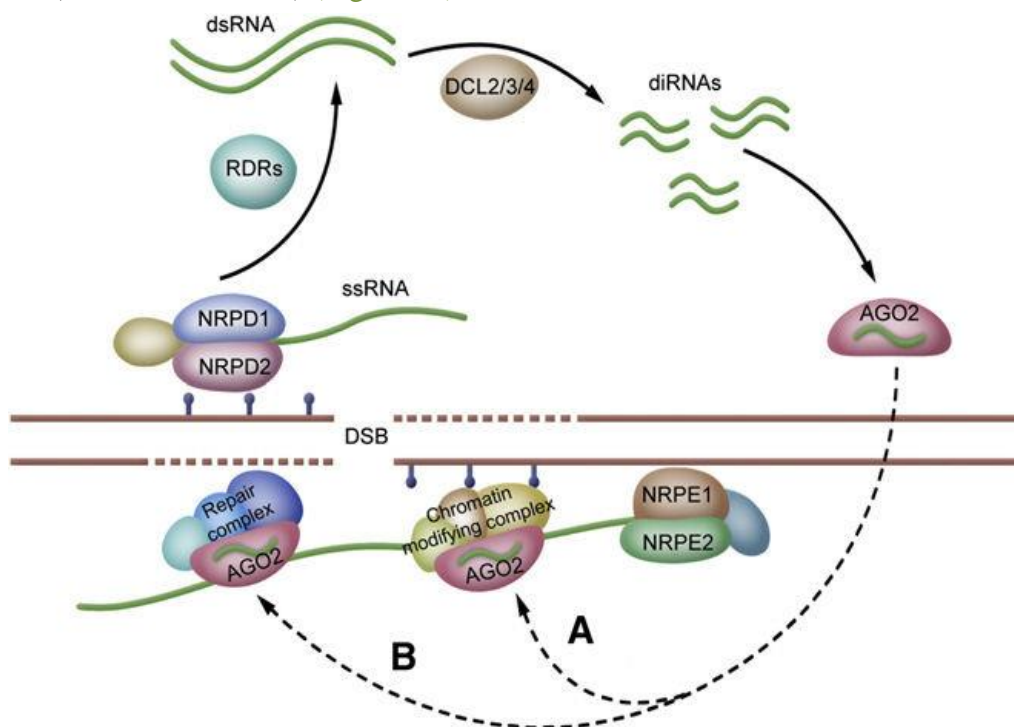


Figure 11: modèle pour la réparation des cassures double-brin chez *A. thaliana* (Wei et al. 2012^[57]). Des transcrits sont générés par la polymérase IV (NRPD1 et NRDP2) à proximité des cassures double-brin. Des RdRPs convertissent les transcrits en double-brin qui vont ensuite être clivés en diRNAs par l'action des 3 Dicer DCL2/3/4. AGO2 charge ces diRNAs et cible les transcrits générés par la polymérase V (NRPE1 et NRPE2). AGO2 serait alors capable de modifier la chromatine localement (A) ou directement d'interagir avec la machinerie de réparation de l'ADN(B).

Ce mécanisme a aussi été identifié chez *Neurospora Crassa* où les cassures double-brin de l'ADN provoquent la production de transcrits aberrants (entre 200 pb et 2 kb) qui sont probablement générés puis convertis en ARNdb par QDE-1. En effet, QDE-1 peut avoir à la fois une activité RdRP et DdRP (voir paragraphe page 27). Ces ARNdb endogènes induisent la transcription de l'argonaute QDE-2. Les sRNAs associés à QDE-2 (qiRNAs), quelques nucléotides plus courts que les siRNAs de 25 nt de *N. crassa*, vont être impliqués dans la réparation des cassures double-brin en recrutant les complexes de réparation de l'ADN au niveau des cassures (Lee et al. 2009^[36]).

F. Fonctions essentielles des RdRPs dans l'ARNi

Les ARN polymérase ARN dépendantes (RdRPs) sont conservées dans la plupart des eucaryotes mais absentes de nombreux métazoaires (les insectes et les vertébrés notamment). L'implication de ces protéines dans l'ARNi a été découvert chez *N. crassa* avec QDE-1 (Cogoni & Macino 1999^[58]) puis chez *C. elegans* avec la protéine EGO-1 (Smardon et al. 2000^[59]) et enfin chez *A. thaliana* avec la RdRP sde1 (Dalmay et al. 2000^[60]). Depuis, de nombreuses RdRPs ont été étudiées avec des fonctions et des mécanismes d'action divers.

1) Conservation des RdRPs au sein des eucaryotes

Des analyses phylogénétiques ont permis de retracer l'histoire évolutive des RdRPs. 161 RdRPs putatives dans 51 génomes eucaryotes, des protistes aux organismes pluricellulaires, ont été classés en 3 groupes: les RDR α , β et γ (Zong et al. 2009^[61]). Ces trois groupes distincts de RdRPs sont conservés chez les chromalvéolés et les champignons. Deux de ces trois groupes sont conservés chez les animaux et les plantes ce qui suggère qu'il existe des fonctions spécifiques à ces 3 types de RdRPs (Figure 12). Il est intéressant de remarquer que les groupes perdus chez les plantes et les animaux ne sont pas les mêmes. Le groupe RDR β a été perdu chez les plantes et groupe RDR γ a été perdu chez les animaux. Si les RdRPs du groupe α sont très bien caractérisés, les fonctions des RdRPs des deux autres groupes sont moins connues.

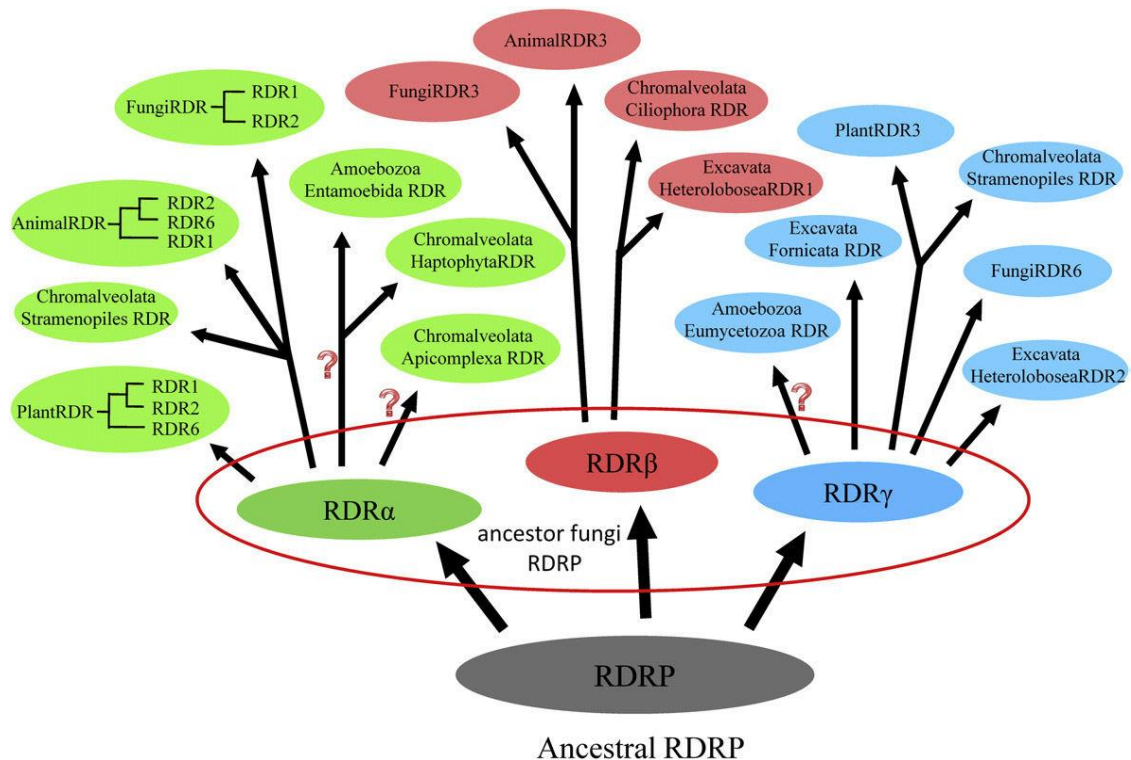


Figure 12 : Modèle de l'évolution des RdRPs chez les eucaryotes.(Zong et al. 2009^[61]). L'eucaryote ancestral aurait possédé 3 RdRPs avant la séparation des grands groupes d'eucaryotes. Les points d'interrogations indiquent les incertitudes sur les positions des gènes chez les protistes

2) Rôle dans l'amplification du signal de répression

Il a été montré que le signal de répression peut être très fort malgré la faible quantité d'ARNdb introduit (Fire et al. 1998^[5]). Ce résultat suggère que le nombre de sRNA est bien supérieur à la quantité d'ARNm dans l'organisme. A l'exception du mécanisme d'amplification des piRNAs (décrit ci-dessous), l'amplification du signal d'ARNi passe par la production de siRNAs secondaires. Les siRNAs secondaires sont des siRNAs produits à partir d'un ARNm ciblé par un siRNA primaire à l'aide notamment des RdRPs. Les modèles expliquant le fonctionnement des RdRP dans la synthèse de siRNAs secondaires diffèrent selon les organismes. Les modèles plantes, animaux et champignons sont présentés ci-dessous.

a. Le modèle plante

Il existe 6 protéines RdRPs chez *A. thaliana*. 3 d'entre elles sont du groupe α (RDR1, 2 et 6), elles possèdent les 6 résidus catalytiques DLDGD et sont très bien caractérisées. Les 3 autres RdRPs (RDR3, 4 et 5), de type γ , ont un motif catalytique atypique : DFDGD et n'ont pour le moment pas de fonction connue (Willmann et al. 2011^[62]). RDR2, nécessaire à la

production des hc-siRNAs (hétérochromatiques siRNA), est impliqué dans la répression transcriptionnelle. RDR1 et RDR6 sont les deux RdRPs impliquées dans la synthèse de siRNA secondaires. Si RDR6 a de nombreuses fonctions dans plusieurs voies, RDR1 est plus spécialisée dans la production et l'amplification des siRNAs d'origine virale (Figure 13).

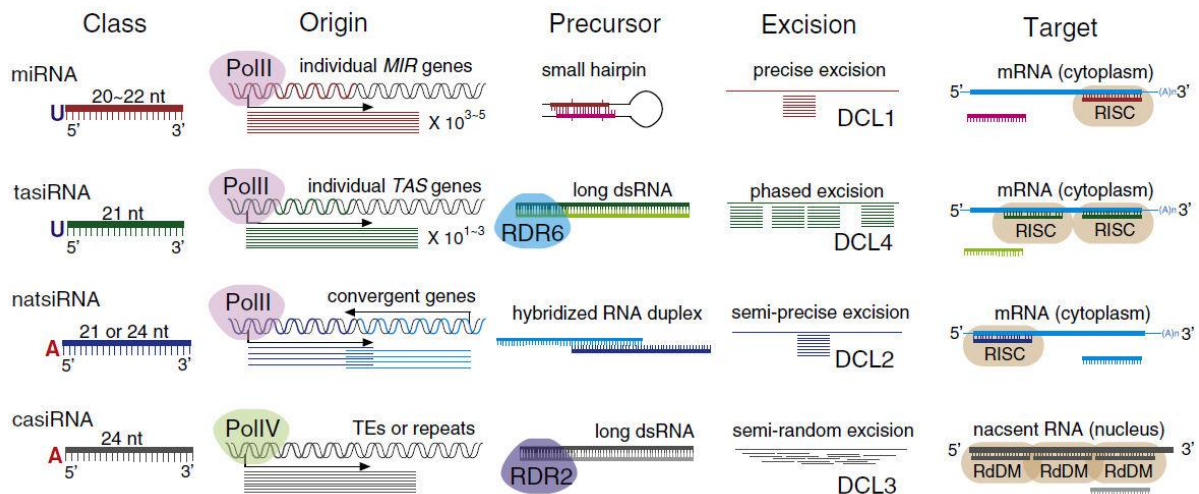


Figure 13 : Les voies de sRNA chez *A. thaliana* (Wang et al. 2011^[63]). Les 4 classes de sRNAs chez *A. thaliana* (miRNA, tasiRNA, natsiRNA, casiRNA/hc-siRNA) sont prises en charge par 4 Dicer différentes (DCL1-4). Les voies tasiRNA et casiRNA nécessitent une RdRP pour synthétiser l'ARNdb avant le clivage par Dicer. RDR6 prend en charge les tasiRNAs et induit la production de siRNAs secondaires phasés qui vont cibler des ARNm cytoplasmiques alors que RDR2 induit la production de siRNAs primaires et non phasés qui vont agir sur les ARNs naissants dans le noyau.

Les siRNA secondaires des plantes, appelés tasiRNAs (trans-acting siRNAs), sont de 21 nt et ont un U en 5'. Ils sont caractérisés par une disposition particulière sur leur cible: ils sont phasés. Cette disposition est typique d'un clivage processif par Dicer à condition que les bornes du double-brin initial soient fixes. Les tasiRNAs sont induits par le clivage d'un ARN, provenant de la transcription d'un gène TAS (ta-siRNA-generating loci), par un miRNA. Il y a deux mécanismes possibles pour la production des tasiRNAs (Figure 14) (Fei et al. 2013^[64]). Dans le modèle "one hit", un miRNA de 22 nt est chargé dans AGO1 et clive le transcrite ce qui induit le recrutement de RDR6 et SGS3 (protéine dont la fonction n'est pas caractérisée). La RdRP est ensuite capable de synthétiser un ARNdb à partir du fragment 3' (en aval). Le second modèle "two hit" démarre d'un transcrite ciblé par deux miRNAs de 21 nt chargés dans AGO7. Le clivage ne se fait que sur le miRNA le plus en 3' ce qui va permettre le recrutement de RDR6 et SGS3 sur le fragment en 5'. Ce modèle où un seul site est clivé sur les deux n'a pu être montré que sur un transcrite chez *A. thaliana* (Axtell et al. 2006^[65]). Il existe d'autres *loci* dans d'autres génomes de plantes où les deux miRNAs sont chargés dans Ago1 et clivent le transcrite aux deux positions. Les longs ARNdb synthétisés par les RdRPs sont ensuite clivés par DCL4 pour former les siRNAs secondaires. Ils sont ensuite capables d'éteindre le transcrite d'où ils proviennent (cis-acting siRNAs) ou d'autres cibles (trans-acting siRNAs).

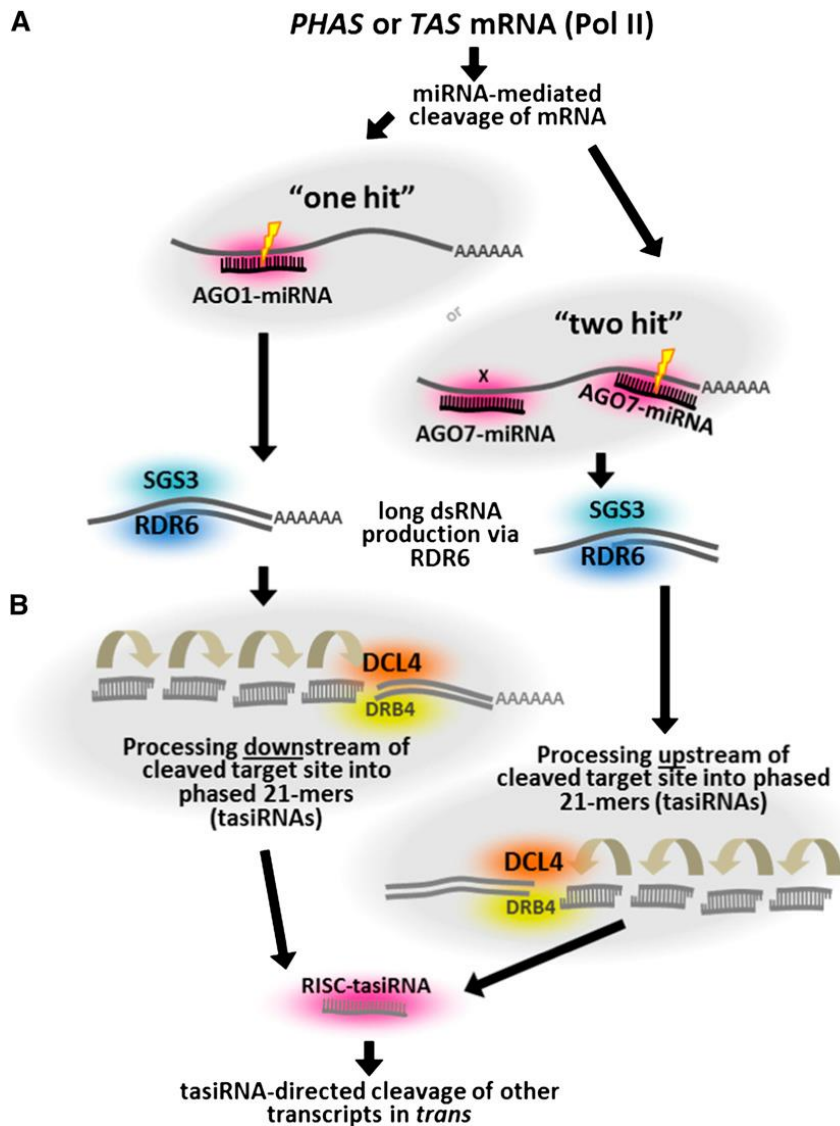


Figure 14: Modèle de production des siRNA secondaires chez *A. thaliana*. (Fei et al. 2013^[64]).

Les siRNAs de 21 nt phasés ont été recherchés globalement dans le génome d'*Arabidopsis* afin d'identifier les ARNm capables d'agir comme substrat pour la production de siRNAs secondaires. Des prédictions bio-informatiques suivies par des validations expérimentales ont montré que des tasiRNAs peuvent eux-mêmes induire la production de tasiRNAs (Chen et al. 2007^[66]). La production de siRNAs secondaires peut donc se faire en cascade chez *A. thaliana* au sein de plusieurs familles de gènes avec des séquences nucléotidiques proches (Figure 15). Ce mécanisme permet la répression globale d'une famille de gènes.

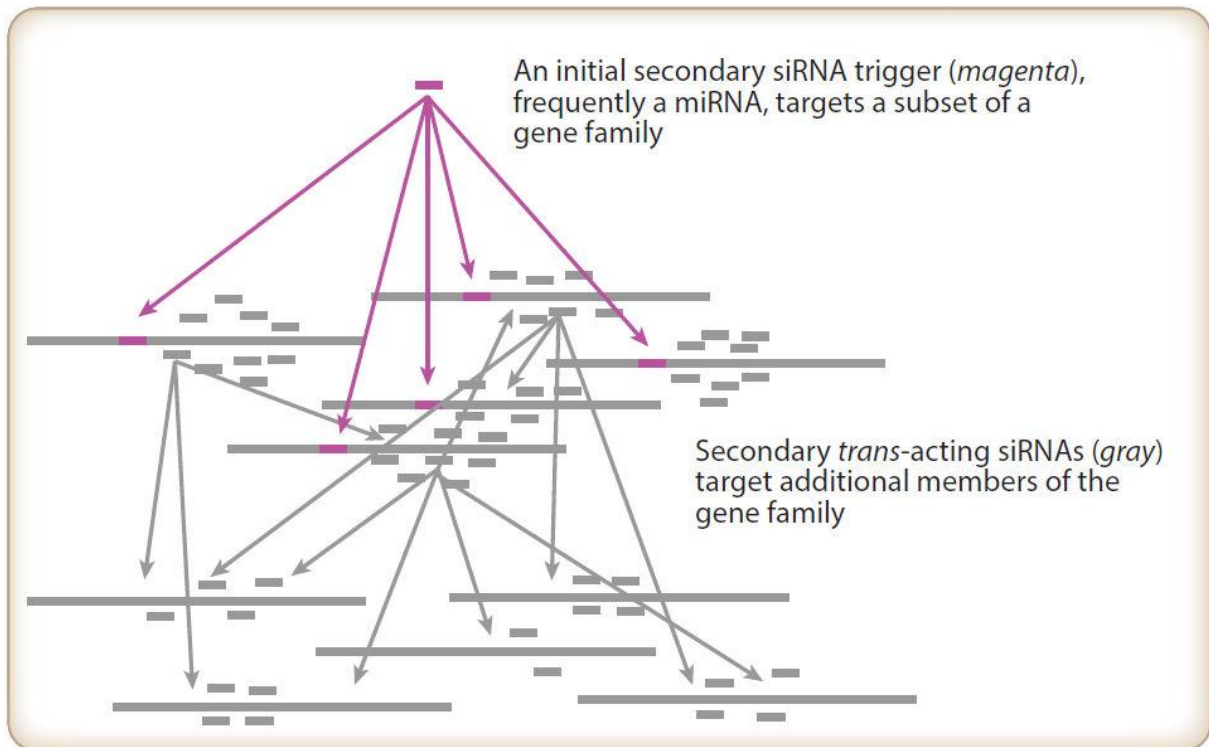


Figure 15: Production de siRNAs secondaires en cascade. Le siRNA secondaire initial (violet) peut cibler 5 transcrits différents qui vont produire des siRNAs tertiaires. Ces siRNAs tertiaires vont à leur tour induire la production de siRNAs quaternaires,...(Axtell 2013^[20]).

b. Le modèle animal

Le modèle de production des siRNAs secondaires des animaux est très différent de celui des plantes. Le génome de *C. elegans* possède 4 RdRPs putatives du groupe RDR α (RRF-1, 2 et 3 et EGO-1). RRF-1 et EGO-1 sont impliquées dans la production de siRNAs secondaires respectivement dans le soma et dans la lignée germinale. RRF-3 est impliquée dans la répression transcriptionnelle et la modification de la chromatine et la fonction de RRF-2 n'a pas encore été découverte (Smardon et al. 2000^[59]). Les secondaires sont de 22nt et ont un fort biais 5'G (22G-siRNAs). Ils peuvent être synthétisés indépendamment de Dicer *in vitro* (Aoki et al. 2007^[67]) et *in vivo* (Pak & Fire 2007^[68]) (Figure 16 A). Les RdRPs de *C. elegans* sont capables de synthétiser directement des siRNAs fonctionnels à partir d'un transcrit sans qu'ils aient besoin d'être recoupés par Dicer. Trois conséquences majeures sur les caractéristiques des siRNAs secondaires en découlent. 1) Chaque siRNA secondaire synthétisé étant un démarrage de transcription, ils portent nécessairement un 5' triphosphate. Ils sont ainsi distinguables des siRNAs primaires qui sont 5' mono-phosphate en conséquence du clivage par Dicer. 2) Ces siRNAs sont tous anti-sens à l'ARNm dont ils proviennent. En effet, pour produire un siRNA sens à partir d'un transcrit sens, il faut nécessairement un double-brin d'ARN clivé par Dicer ou bien un transcrit anti-sens. 3) Ils ne sont pas phasés contrairement aux siRNAs secondaires des plantes. Une étude biochimique récente a montré que RRF-1 et EGO-1 sont en complexe avec l'hélicase DRH-3 et la protéine tudor EKL-1

(Thivierge et al. 2012^[69]) dans la production de certains siRNAs endogènes. Les siRNAs secondaires sont ensuite pris en charge par les argonautes de la famille WAGO pour réprimer leur cible.

Dans l'étude de (Pak et al. 2012^[53]) , une méthode est développée pour distinguer les siRNAs primaires des siRNAs secondaires. Des nématodes ont été soumis au *feeding* d'un ARNdb homologue à un gène endogène mais possédant une substitution tous les 25 nucléotides. Ainsi, après séquençage haut-débit, les siRNAs qui portent les substitutions sont identifiés comme des siRNAs primaires car synthétisés à partir de l'ARNdb initiateur et les siRNAs qui ne les portent pas sont donc des secondaires synthétisés à partir de l'ARNm. Etonnement, les siRNAs primaires sont en faible quantité (5%) en comparaison aux siRNAs secondaires 22G (95%) (Figure 16 B et C). Avec des RdRPs aussi efficaces chez *C. elegans*, le système d'amplification des siRNAs pourrait théoriquement fonctionner indéfiniment et provoquer une dissémination du signal d'ARNi à d'autres ARNm par simple homologie de séquence. Cependant, les auteurs observent que deux facteurs limitent l'amplification du signal d'ARNi. 1) L'ARNdb et les produits de clivage par Dicer ne peuvent pas servir de matrice aux RdRPs 2) Seul les siRNAs primaires sont capables de recruter les RdRPs pour amplifier le signal d'ARNi. En effet, les siRNAs secondaires 22G sont très efficaces et suffisants pour éteindre l'expression d'un ARNm mais très peu efficaces pour produire des siRNAs tertiaires. Ces deux limitations seraient suffisantes pour avoir une réponse ARNi à la fois efficace et très sélective.

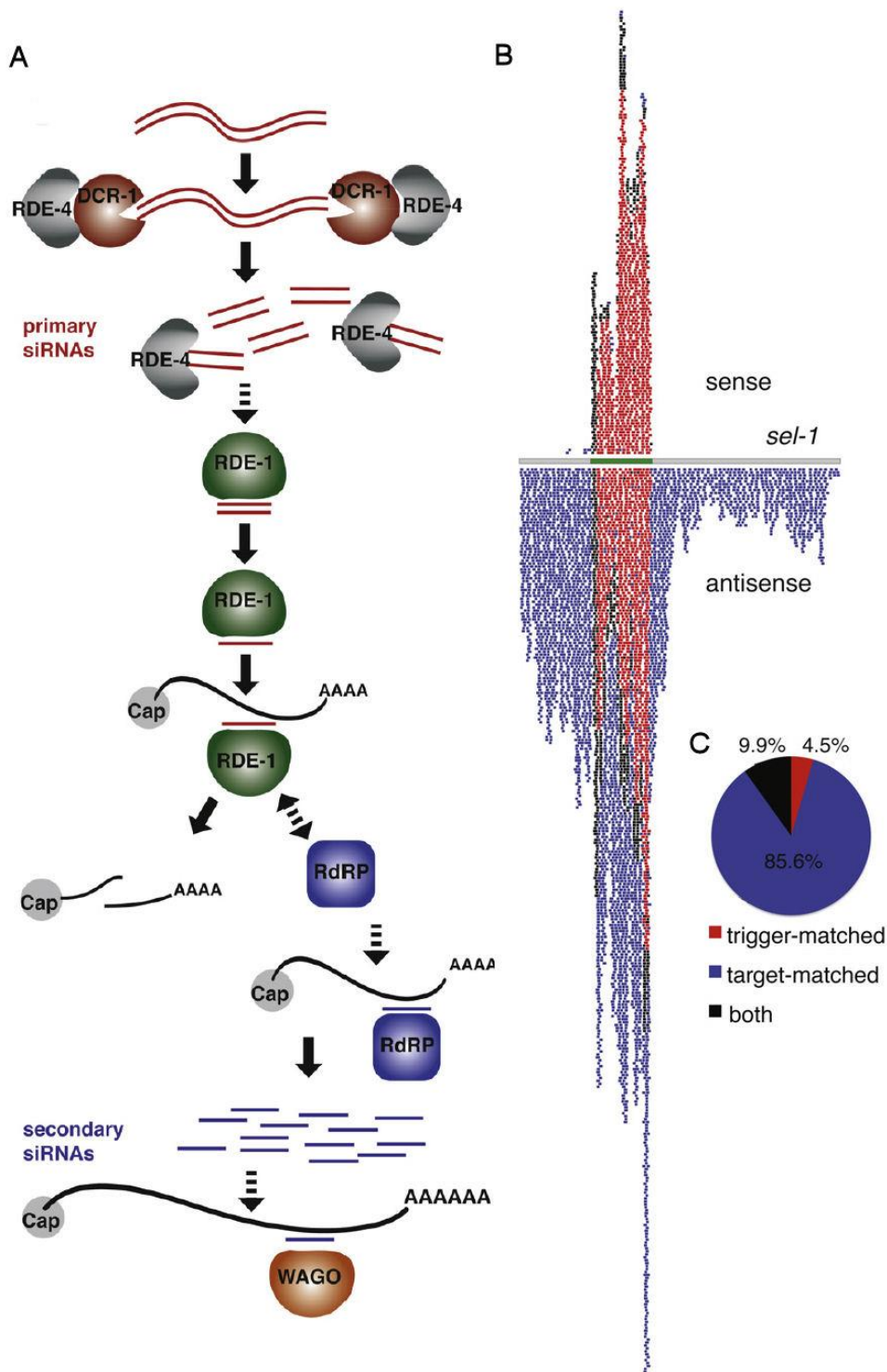


Figure 16: Production des siRNAs secondaires chez *C. elegans* (Pak et al. 2012^[53]). A) Les longs ARNdb sont clivés par Dcr-1 puis la protéine de liaison à l'ARNdb RDE-4 aide au chargement du duplex de siRNAs sur l'argonaute RDE-1. Après élimination du brin passager, RDE-1 est dirigé vers l'ARNm complémentaire au siRNA et recrute une RdRP (RRF-1 ou Ego-1) sur l'ARN ciblé. La RdRP induit la production de siRNAs secondaires 22G qui vont être chargés sur les argonautes de type WAGO pour induire le clivage de la cible. B) Représentation schématique des siRNAs primaires (rouge) ou secondaires (bleu) alignés sur le gène *sel-1* après induction de l'ARNi avec un ARNdb homologue à la partie centrale du gène (en vert). C) Pourcentage de siRNAs primaires (trigger-matched) ou secondaires (target-matched) ou non définis (both).

c. Le modèle champignon

Les fonctions des RdRPs de champignons dans l'amplification de l'ARNi sont moins bien décrites. *In vitro* la RdRP QDE-1 (du groupe RDR γ) de *Neurospora crassa* est capable de produire de longs ARN double-brin mais aussi de synthétiser des ARN courts de 9 à 21nt ce qui pourrait indiquer que cette RdRP est capable de produire plusieurs types de siRNAs (Makeyev & Bamford 2002^[70]). Cependant *in vivo*, QDE-1 n'est pas requis dans l'ARNi induit par ARNdb ce qui suggère que QDE-1 n'est pas impliqué dans la production de siRNAs secondaires mais seulement dans la synthèse d'ARNdb à partir de transcrits aberrants (Lee et al. 2009^[36]). L'implication d'une RdRP dans la production de siRNAs secondaires a été montrée chez *Mucor circinelloides*, distant phylogénétiquement des autres organismes modèles champignons (Calo et al. 2012^[71]). Chez *M. circinelloides*, RDR-1 (du groupe RDR γ) est impliqué dans la production de siRNAs induits par injections de transgènes mais n'a pas de rôle évident sur la production des siRNAs secondaires. Au contraire, RDR-2 (du groupe RDR α) est impliqué à la fois dans la répression induite par des transgènes et par ARNdb et serait à l'origine de la synthèse de siRNAs secondaires de 21nt. RDR-2 serait donc impliqué dans l'accumulation des siRNAs primaires de 25nt en synthétisant de longs ARNdb et dans la production directe de siRNA secondaires de 21nt (Calo et al. 2012^[71]). La différence majeure entre les séquences RDDR-1 et RDR-2 est l'interruption du domaine RdRP de RDR-1 par 53 acides aminés (Calo et al. 2012^[71]).

d. Amplification indépendante des RdRPs : le cycle ping-pong

Le cycle ping-pong est le seul mécanisme d'amplification de sRNAs indépendant des RdRPs. Cette voie de production essentielle au fonctionnement des piRNAs a d'abord été caractérisée dans les lignées germinales de drosophiles puis chez les poissons et les mammifères. Le mécanisme d'amplification commence par la prise en charge des piRNAs matures par le complexe RISC contenant les argonautes de type PIWI (Ago3 et Aubergine chez *D. melanogaster*). L'activité de coupure de PIWI permet le clivage du transcrit ciblé (en général un transposon) et génère ainsi une extrémité 5'. Cette extrémité est chargée dans une autre PIWI et l'extrémité 3' du transcrit est coupée par un mécanisme inconnu jusqu'à générer un piRNA sens mature. Le piRNA sens cible alors un transcrit anti-sens qui va produire par le même mécanisme de nouveaux piRNAs anti-sens (Figure 17) (Meister 2013^[33]). Contrairement à ce que le schéma suggère, la production de piRNAs sens et anti-sens par le cycle ping-pong n'est pas symétrique. Chez *D. melanogaster*, un mécanisme permet d'assurer que la quantité de piRNAs anti-sens, donc capables de cliver les transcrits sens, va toujours être plus importante que la quantité de piRNAs sens. Ce cycle ping-pong hétérotypique

nécessite la protéine Qin (qui contient 5 domaines Tudor et une E3 ligase) mais le mécanisme induisant ce biais reste à être élucidé (Zhang et al. 2011^[72]).

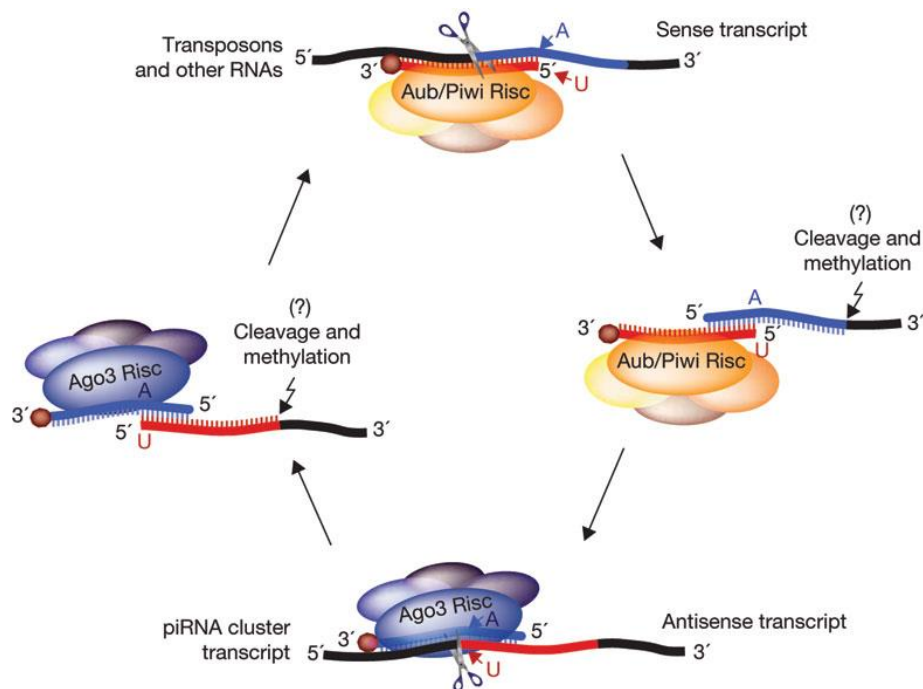


Figure 17: Le cycle ping-pong. Représentation schématique du cycle d'amplification des piRNAs chez *D. melanogaster*. (SEILA & SHARP 2008^[73])

3) Les autres fonctions des RdRPs

Les RdRPs ne sont pas seulement impliquées dans l'amplification de l'ARNi par production de siRNAs secondaires mais ont aussi des fonctions essentielles dans la production d'ARNdb qui vont être pris en charge par Dicer pour former des siRNAs primaires. Il existe de nombreuses voies d'ARNi qui sont dépendantes des RdRPs pour les siRNAs primaires. La production des siRNAs hétérochromatiques chez *A. thaliana* en est un exemple. Les siRNAs hétérochromatiques sont dérivés des régions répétées intergéniques du génome des plantes. Ils sont déterminants pour certaines modifications de la chromatine notamment la méthylation de l'histone H3K9. Chez les plantes, les transcrits naissants générés par la polymérase IV dans ces régions vont être pris en charge par la RDR2 pour synthétiser de long ARNdb qui sont ensuite clivés par Dcl3 pour former des siRNAs de 23-24nt (Figure 18). Ils sont ensuite chargés sur AGO4 et ciblent les transcrits naissants générés par la polymérase V. Cette fixation des hc-siRNA sur les transcrits naissants induit le recrutement des protéines qui vont modifier la chromatine.

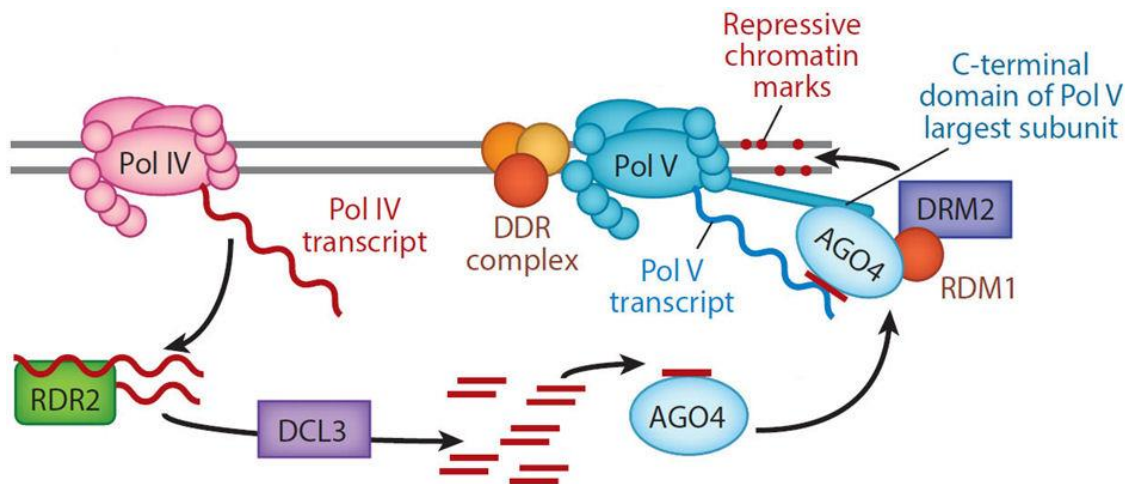


Figure 18 : Modèle pour la biogénèse et la fonction des siRNAs hétérochromatiques chez les plantes (Axtell 2013^[20]).

Plus généralement, les RdRPs ont un rôle essentiel dans la synthèse de siRNAs primaires à chaque fois que l'ARN initiateur est sous forme simple-brin et non double-brin.

G. Modifications des sRNAs

Plusieurs exemples dans la littérature montrent que les modifications des petits ARNs sont essentielles au fonctionnement des voies de d'ARNi. Cependant les fonctions moléculaires de ces modifications restent méconnues dans beaucoup d'organismes (Ji & Chen 2012^[74], Kim et al. 2010^[75]).

1) Méthylation en 3'

Les 3 types de sRNAs (mi- pi- si-ARNs) requièrent parfois la méthylation 2'-O sur le ribose du nucléotide 3'. Le rôle de la méthylation dans l'ARNi a été étudié pour la première fois dans le mutant *hen1* (HUA Enhancer 1) chez *A. thaliana* qui est incapable d'accumuler les miRNAs et les siRNAs (Park et al. 2002^[76]). Hen1 possède un domaine N-terminal de liaison à l'ARNdb et un domaine C-terminal conservé des bactéries aux métazoaires contenant le domaine de liaison au donneur de méthyl SAM (S-adénosylmethionine). Hen1 est capable de métyler les duplex de miRNAs (Yu et al. 2005^[77]) et de siRNAs (Yang et al. 2006^[78]). Cette méthylation est dépendante de la taille des duplex (21-24 nt) mais aussi de la présence des 2 nucléotides sortants issus du clivage par Dicer (Yang et al. 2006^[78]).

Chez les animaux, les homologues de Hen1 ne possèdent pas le domaine N terminal de liaison à l'ARNdb. Ces Hen1 agissent sur les piRNAs simple brin déjà chargés dans les Ago et non sur les duplex. Leur absence entraîne l'activation de certains transposons (Horwich et al. 2007^[79]). Chez *C. elegans*, Hen-1 agit sur les ARNs 26G de classe ERGO et stabilise les siRNAs (Billi et al. 2012^[80]). En absence de HEN1, l'uridylation des sRNAs est fréquemment

observée aussi bien chez les plantes que chez les animaux (Kamminga et al. 2010^[81], Kamminga et al. 2012^[82]). La 2'-O méthylation serait donc une modification essentielle à la stabilité des sRNAs en empêchant leur uridylation qui entrainerait leur dégradation par une exonucléase 3'-5' (Figure 19).

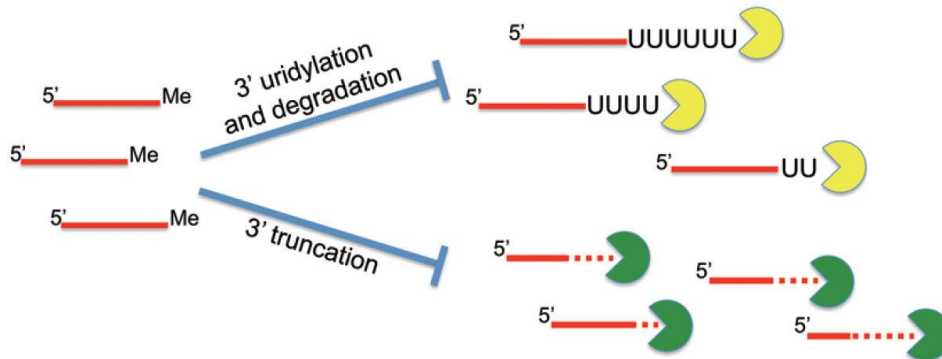


Figure 19 : La méthylation protège les sRNAs de l'uridylation et de la dégradation. (Ji & Chen 2012^[74])

Par exemple, chez le poisson zèbre, le séquençage des sRNAs a révélé que le pourcentage de piRNAs uridylés augmente dans un mutant Hen1 (Kamminga et al. 2010^[81]). De plus, le pourcentage de piRNAs uridylés est plus important sur les transposons actifs que sur les transposons inactifs. Plus généralement, la méthylation des sRNAs des plantes, des piRNAs des animaux et des siRNAs des drosophiles protègent contre l'addition de nucléotides en 3'.

2) Uridylation ou adénylation en 3'

L'uridylation, ou dans une moindre mesure l'adénylation des sRNAs, a été observée dans la plupart des espèces. Elle concerne tous les types de sRNAs (si- mi- et piRNAs) et principalement en absence de hen1. Cette uridylation a d'abord été interprétée comme un signal de dégradation des sRNAs cependant plusieurs études ont aussi montré que l'uridylation pouvait aussi être essentielle à leur stabilité et avoir des rôles biologiques essentiels.

e. Modifications des miRNAs

De nombreuses études traitent des modifications des miRNAs appelés "isomiRs". Ces modifications peuvent être des raccourcissements en 5' ou 3' ou bien l'addition de nucléotides en 3' (Wyman et al. 2011^[83]). Ces isomiRs peuvent avoir une activité et une stabilité différente des miRNAs non modifiés. Pour certains miRNAs, la fréquence des modifications est variable selon le tissu ou le stade de développement indiquant qu'il s'agit d'un processus régulé biologiquement.

L'identification des protéines responsables de l'addition des nucléotides en 3' des miRNAs a permis de mieux comprendre leur rôle. Par exemple, chez l'homme et la souris, miR-122 est adénylé par la nucléotidyl-transférase PAPD4 (GLD-2) ce qui augmente sa stabilité et miR-26a est uridylé par ZCCHC11 ce qui diminue son efficacité à éteindre sa cible (Katoh et al. 2009^[84], Jones et al. 2009^[85]).

Chez les métazoaires, les additions 3' sur les miRNAs sont principalement des adénylations ou des uridylations.

f. Modifications 3' des siRNAs et piRNAs

Chez l'algue verte *Chlamydomonas reinhardtii*, l'absence de la nucléotidyl-transférase Mut-68 induit une augmentation de la quantité de siRNAs et une diminution de la quantité de siRNAs uridylés (Ibrahim et al. 2010^[86]). La sous-unité RRP6 de l'exosome est impliquée dans la dégradation des siRNAs uridylés. Chez *C. elegans*, l'abondance des endo-siRNAs (22G-RNA associés à l'argonaute CSR-1) est dépendante de la nucléotidyl-transférase CDE-1. Dans un mutant nul *cde-1*, la quantité de siRNAs de la lignée germinale de la voie CSR-1 augmente, entraînant des problèmes de ségrégation de chromosomes à la mitose et à la méiose (Wedeles et al. 2013^[87]). Contrairement à *S. pombe* où les complexes RITS-RDRC ciblent l'hétérochromatine, les siRNAs de la voie CSR-1 de *C. elegans* se fixent majoritairement sur l'euchromatine notamment au niveau des variants d'histones impliqués dans la formation des nucléosomes centromériques chez les eucaryotes (HCP-3/CENP-A). Les chromosomes de *C. elegans* étant holocentriques (kinétochores répartis sur la totalité du chromosome), il est possible que le ciblage de la chromatine par les siRNAs chargés dans CSR-1 influence directement la formation des holocentromères (Wedeles et al. 2013^[87]). L'uridylation dans ce cas permet de déstabiliser les siRNAs ce qui est indispensable à leur fonction sur la chromatine. Une hypothèse est que le chargement de siRNAs sur l'argonaute CSR-1 aurait besoin d'être très temporaire pour qu'il y ait toujours des argonautes libres capable de charger les nouveaux siRNAs. Si c'est le cas, la nucléotidyl transférase permettrait de réduire le temps de demi-vie du complexe siRNA/CSR-1. Au contraire, MUT-2(RDE-3) est une uridyl-transférase nécessaire pour l'accumulation des siRNAs. En absence de MUT-2, Dicer est toujours capable de cliver les longs ARNdb ingérés mais l'argonaute RDE-1 n'est plus capable de dissocier les duplex de siRNAs (Jose et al. 2011^[88]). Cette nucléotidyl transférase pourrait donc avoir un rôle essentiel de modification des duplex de siRNAs.

Enfin, chez le cilié *T. thermophila* les siRNAs de 23-24nt sont pour 50% d'entre eux modifiés avec une uridine non génomique (Lee & Collins 2006^[89]). La purification par affinité des protéines associées à Rdr1 et Dcr1 a permis d'identifier deux nucléotidyl-transférase (RDN1 et RDN2). RDN1 (essentielle et exprimée à un niveau constant au cours du cycle de vie) et RDN2 (non essentielle et particulièrement exprimée à la conjugaison) semblent agir dans deux complexes RDRC différents (Lee et al. 2009^[90]). *In vitro*, ces deux nucléotidyl-

transférases ajoutent spécifiquement des uridines sur les siRNAs. Une hypothèse sur le rôle de ces uridyl-transférases dans la stabilité des siRNAs a été proposé dans l'étude de (Talsky & Collins 2010^[91]). Des expériences biochimiques montrent qu'en présence d'un complexe RDRC muté pour RDN2, un ARN spécifique induira la production de siRNAs avec une diversité moins importante qu'avec un complexe RDRC sauvage. Les auteurs suggèrent que les uridines ajoutés en 3' permettent la formation d'une boucle qui servirait de matrice à la RdRP pour synthétiser le brin complémentaire. En absence de ces uridines, c'est l'extrémité de l'ARN lui-même qui formerait une boucle pour permettre la synthèse d'ARNdb ce qui la rendrait inaccessible au clivage par Dicer. (Figure 20) (Talsky & Collins 2010^[91]).

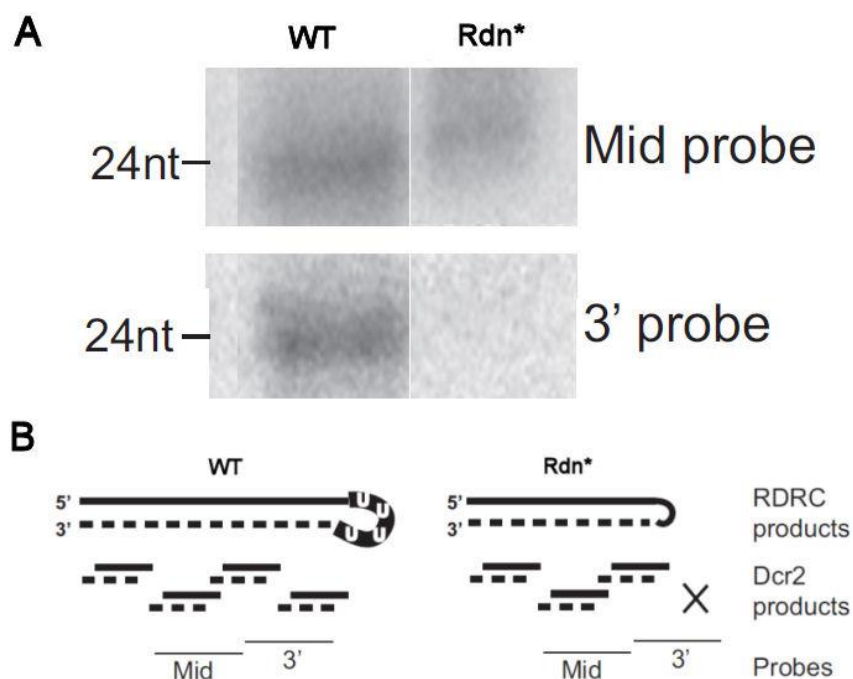


Figure 20: Modèle du rôle de l'uridylation pour la synthèse d'un ARNdb. A. Northern blot des siRNAs de 24nt avec deux sondes différentes : au milieu du transcrit ou à l'extrémité 3' dans un contexte sauvage (gauche) ou muté pour la nucléotidyl-transférase Rdn (droite) **B** Dans le cas sauvage (à gauche), l'ARN est uridylé par Rdn puis le complexe RDRC synthétise le brin complémentaire (ligne pointillée) en utilisant comme matrice la queue polyU repliée pour former une boucle. Dans le mutant nul pour Rdn (à droite), le complexe RDRC utilise l'extrémité 3' de l'ARN pour former la boucle et synthétiser le brin complémentaire. Après clivage par Dicer, les siRNAs à l'extrémité 3' sont présents dans le sauvage mais pas dans le mutant. (Talsky & Collins 2010^[91])

III. Les voies d'ARNi chez *Paramecium tetraurelia*.

A. Les ciliés comme organismes modèles.

Les protozoaires ciliés constituent un groupe monophylétique d'eucaryotes unicellulaires appartenant à la division des *Alveolata* (Figure 21) (Adl et al. 2012^[92]). Ce groupe formé par plus de 8000 espèces possède deux caractéristiques communes: la présence de cils à la surface qui leur servent à nager et à capter leur nourriture ainsi qu'un dimorphisme nucléaire. Ces organismes sont phylogénétiquement légèrement plus proches des plantes que des animaux ou des champignons, ils sont cependant susceptibles d'avoir des caractéristiques moléculaires communes avec les uns ou avec les autres.

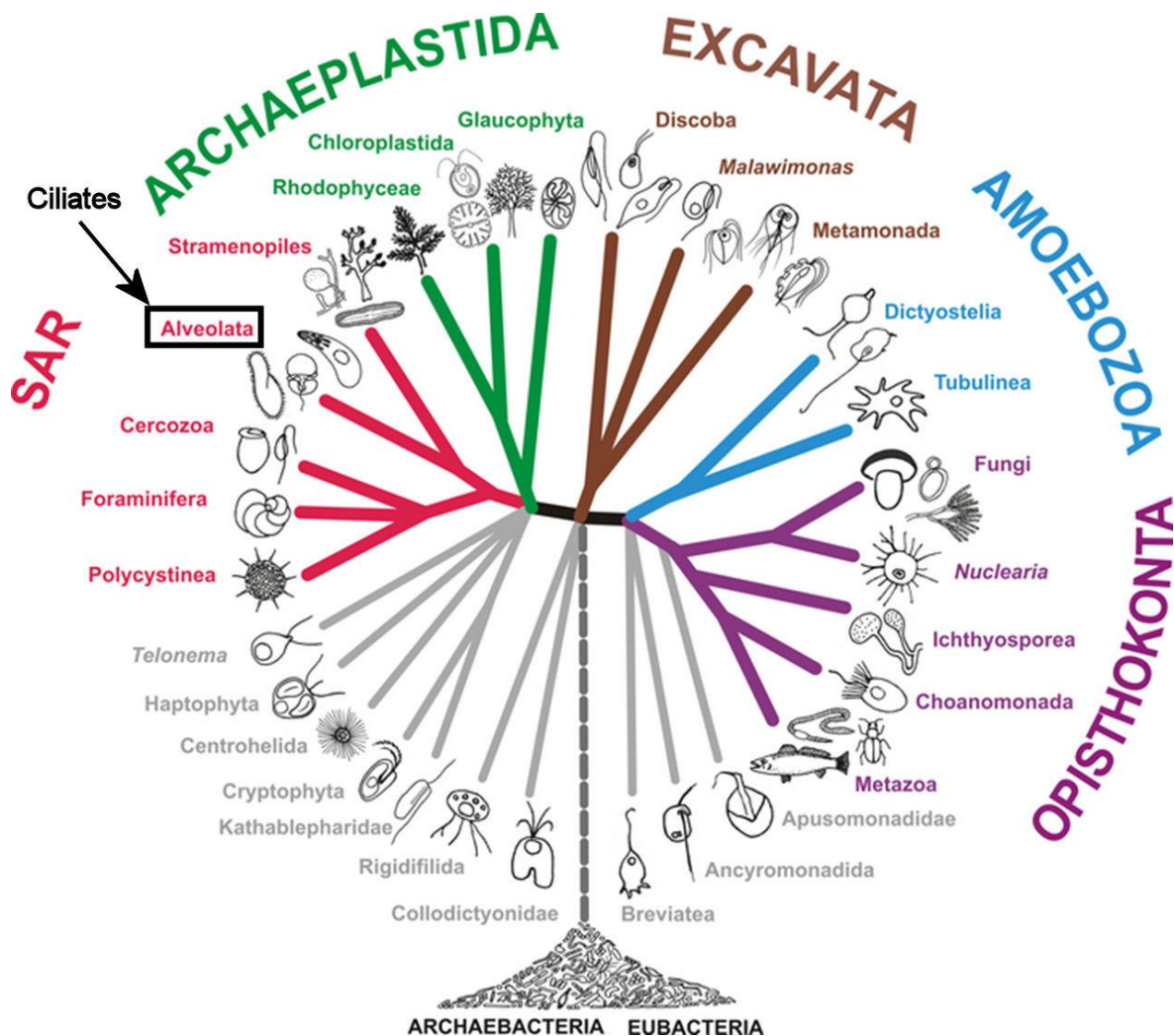


Figure 21 : Arbre phylogénétique des eucaryotes. (Adl et al. 2012^[92])

1) Le dimorphisme nucléaire

Tous les ciliés sont caractérisés par un dimorphisme nucléaire. Les fonctions somatiques et germinales de l'organisme sont séparées dans deux types de noyaux qui coexistent au sein de la même cellule. Le macronoyau (MAC) somatique et le (ou les) micronoyaux diploïdes germinaux (mic) (Figure 22). Le macronoyau des ciliés est hautement polyploïde ; la ploïdie peut varier selon les espèces (environ $45n$ chez *Tetrahymena thermophila* et $800n$ chez *Paramecium tetraurelia*). En croissance végétative il n'y a apparemment pas de mécanisme de ségrégation précise des copies d'ADN dupliquées pendant la réplication, le mac se divise donc par division amitotique. La grande ploïdie permet de retarder la survenue d'hétérogénéité marquée du nombre de copies des chromosomes. Le mac somatique est responsable de la transcription génique au cours de la croissance végétative, il est donc responsable du phénotype de la cellule. A l'inverse, les mic dont le nombre peut varier selon les espèces (1 chez *Tetrahymena*, 2 chez *P.aurelia* par exemple) sont inactifs en croissance végétative. En effet, des cellules générées sans mic sont capables de survivre normalement en croissance végétative mais sont incapables de donner une progéniture viable. Les mic sont diploïdes et réalisent une mitose à chaque division cellulaire. Chez la paramécie, chaque cycle cellulaire dure environ 6h à la température optimale de croissance : 27°C . Leur durée de vie est d'environ 200 divisions. Passé cette limite, les clones végétatifs deviennent sénescents et finissent par mourir s'ils ne passent pas par la reproduction sexuelle pour former un nouveau mac.

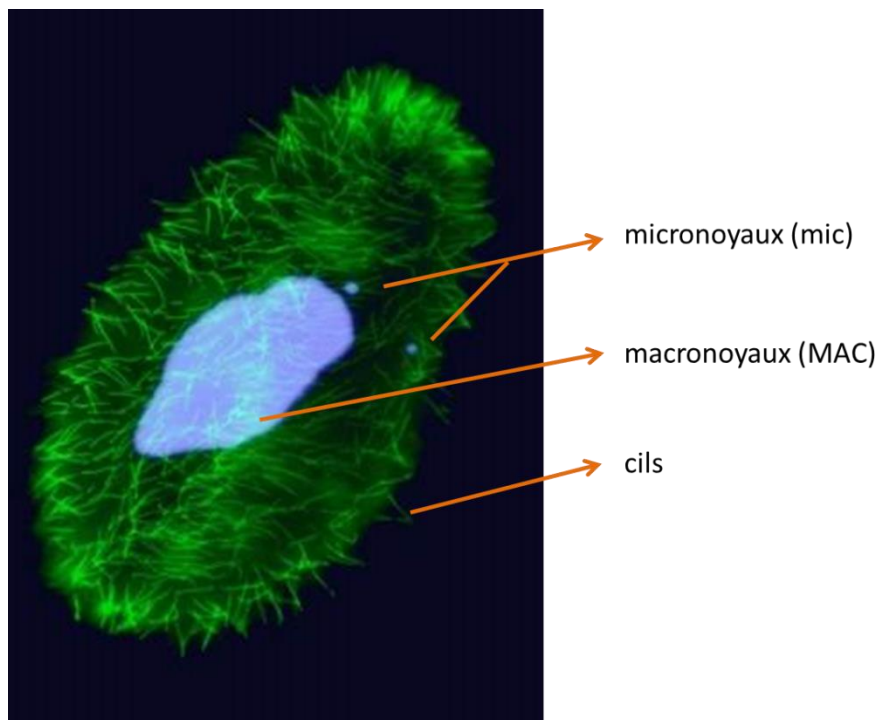


Figure 22: Photo de *Paramecium tetraurelia*. Les deux noyaux germinaux (mic) et le noyau somatique (MAC) sont marqués au DAPI. Les cils sont marqués par un anticorps anti-tubuline et sont colorés en vert.

2) Les évènements sexuels

Chez les ciliés, les évènements sexuels sont généralement induits par la privation de nourriture (*starvation*). Le ou les mic réalisent les deux divisions méiotiques, un seul noyau haploïde est conservé aléatoirement alors que tous les autres sont dégradés, un seul des deux allèles parentaux est donc conservé. Le noyau restant subit alors une mitose conduisant à la production de deux noyaux haploïdes génétiquement identiques (Figure 23).

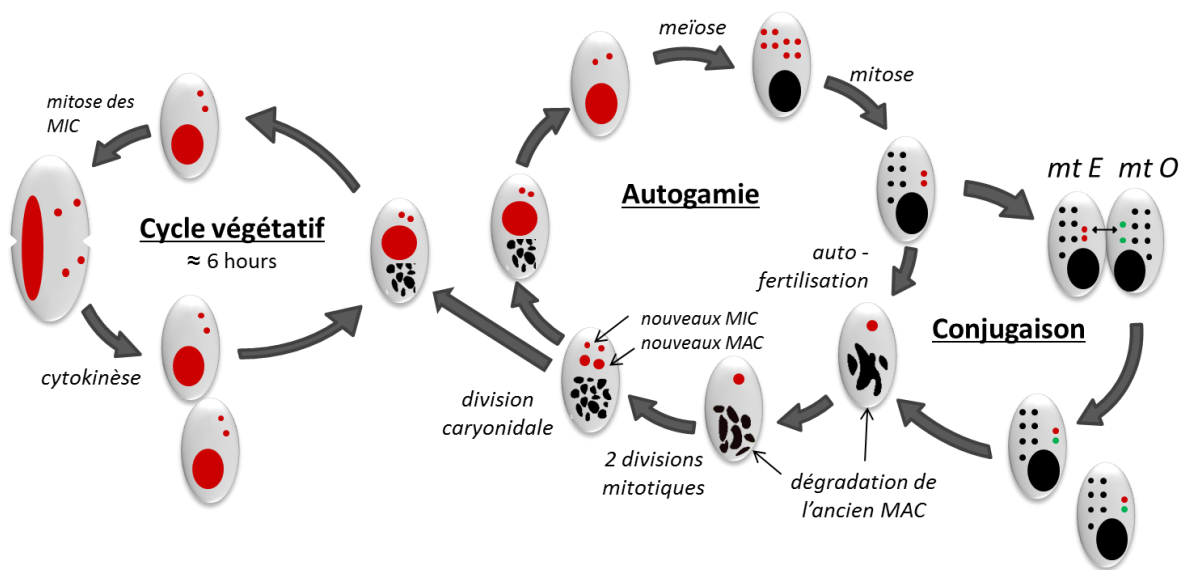


Figure 23 : Les cycles sexuels de *P. tetraurelia*. Les micronoyaux (mic) et macronoyaux (mac) de la paramécie sont symbolisés par des ronds rouge ou noirs lorsqu'ils sont en cours de dégradation.

Au cours de la conjugaison, deux individus de types sexuels différents échangent réciproquement un noyau haploïde ce qui conduit à la génération de deux organismes hétérozygotes et génétiquement identiques (Figure 24). Le macronoyau somatique est quant à lui dégradé et un nouveau macronoyau se développe par endoréplication à partir d'une copie du noyau zygotique (Coyne et al. 2012^[93]). Ces processus généraux sont communs à la plupart des ciliés cependant certaines caractéristiques sont spécifiques de certaines espèces.

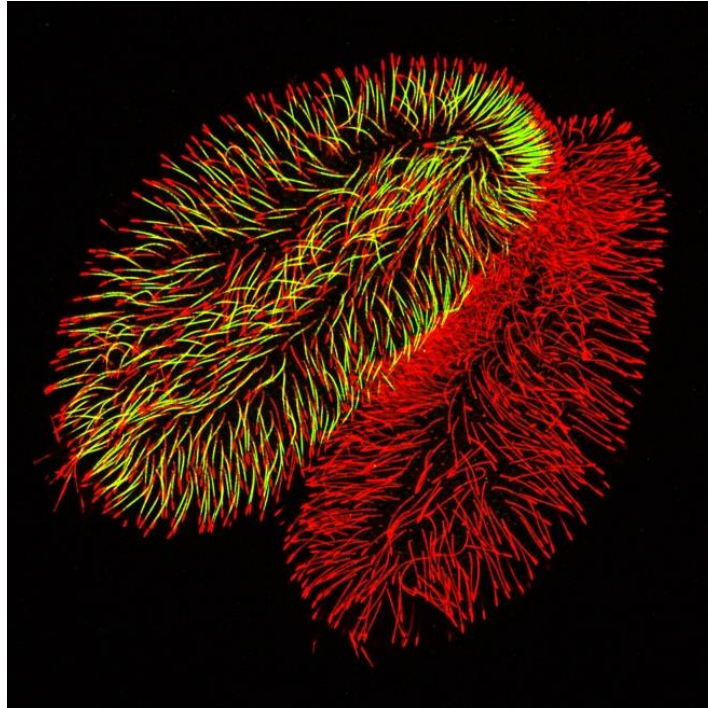


Figure 24: Photo en microscopie confocale de deux paramécies en conjugaison. Les cils des paramécies des deux types sexuels sont marqués avec un anticorps anti tubuline (rouge). Les cils de la paramécie de type sexuel E est également marquée à la GFP (protéine BBS2), ce qui permet de les distinguer. (Singh et al. 2014^[94])

Chez les espèces du complexe *P.aurelia*, le noyau zygotique subit deux divisions mitotiques qui vont former respectivement deux nouveaux micronoyaux et deux nouveaux macronoyaux. Lors de la première division végétative, les deux mic se divisent par mitose et les deux nouveaux macronoyaux sont répartis dans les deux cellules filles appelées caryonides. Les caryonides possèdent donc deux mic et un mac (Figure 23).

De nombreux ciliés dont les paramécies ont la capacité d'entrer dans un processus d'autogamie. L'autogamie est une auto-fertilisation conduisant à la fusion des deux noyaux haploïdes pour former un noyau zygotique diploïde. L'individu devient donc entièrement homozygote. L'autogamie se déclenche sous deux conditions: La cellule doit être âgée d'un minimum de 20 divisions végétatives pour *P. tetraurelia* (5 jours à la vitesse de croissance optimale) et doit être en carence alimentaire (dans un milieu de culture épuisé en bactéries et nutriments).

Ces deux modes de reproduction sexuelle permettent donc d'obtenir soit une cellule entièrement homozygote (autogamie) soit deux cellules hétérozygotes mais génétiquement identiques (conjugaison). *P. tetraurelia* est donc un modèle eucaryote idéal pour réaliser des croisements et des études génétiques.

3) Les réarrangements génomiques au cours du développement du MAC

Lorsque le nouveau mac se développe à partir du noyau zygotique, deux types de réarrangements génomiques essentiels ont lieu. Premièrement, l'excision imprécise de séquences répétées tel que les séquences mini-satellite et les transposons ce qui conduit souvent à la fragmentation du génome. Des télomères sont ajoutés à la plupart de ces fragments pour former les nombreux chromosomes mac (Le Mouel et al. 2003^[95]). Deuxièmement, l'excision précise d'environ 45 000 séquences uniques et courtes appelées IES (internal eliminated sequences) (Duharcourt et al. 1998^[96]). 98% de ces IES sont en copie unique dans le génome et 93% font entre 26 et 150 paires de bases (pb). Ces IES peuvent être aléatoirement intergéniques ou présentes dans les séquences codantes (il y a en moyenne 1 IES par gène) (Arnaiz et al. 2012^[97]). Leur excision précise est donc indispensable à la reconstitution de gènes fonctionnels dans le macronoyau. Ces éléments sont dérivés de la famille de transposons TC1/mariner. Les IES sont toutes bornées par deux di-nucléotides 5'-TA-3' qui recombinent en un seul TA pendant l'excision. De nombreuses protéines dont piggyMac (Pgm), une transposase domestiquée de type piggyBac, sont impliquées dans leur excision. La répression de Pgm par ARN interférence empêche les coupures double-brin de l'ADN ce qui affecte drastiquement l'excision des IES. Pgm pourrait donc porter la fonction endonucléase nécessaire à leur excision (Baudry et al. 2009^[98]).

4) Le génome de *P. tetraurelia*

Un grand nombre de particularités propres aux génomes *Paramecium* ont pu être révélées grâce au séquençage du génome mac de l'espèce *Paramecium tetraurelia* (Aury et al. 2006^[99]). Le génome somatique est relativement petit (environ 72Mb). Il a pu être assemblé en 697 scaffolds dont 188 font plus de 45kb. La présence de séquences télomériques aux extrémités pour plus de la moitié des scaffolds suggère qu'il y a au moins une centaine de chromosomes dans le génome MAC. Le génome mac est très compact (78% de séquences codantes) avec environ 40 000 gènes prédits. En conséquence, les séquences non codantes sont de très petite taille : en moyenne 352pb pour les régions intergéniques et 25pb pour les introns (2,9 introns par gène épissé en moyenne). Enfin, 3 duplications globales du génome (WGD) ont affecté le génome de *P. tetraurelia*. En effet, plus de la moitié des gènes ont au moins un paralogue (Figure 25). Par absence de pression de sélection, un grand nombre de ces homologues (environ 50% de la dernière duplication) ont dégénéré et sont maintenant des pseudogènes mais certains sont encore fonctionnels et source d'une redondance génétique très importante. La dernière duplication du génome pourrait être à l'origine de la spéciation du groupe *Paramecium aurelia* en 15 espèces physiologiquement indistinguables mais incapables de conjuguer entre elles (Coleman 2005^[100]).

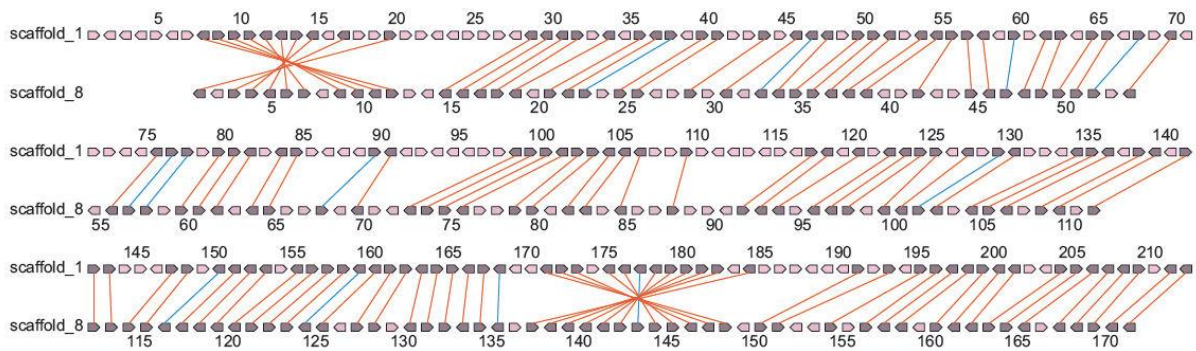


Figure 25: Comparaison de deux scaffolds chez *P. tetraurelia*. Les flèches représentent les gènes qui possèdent un paralogue (en marron) ou qui n'en n'ont pas (en rose). Les relations de paralogie sont indiquées par des barres obliques.

B. Les voies d'ARN interférence chez la paramécie

1) Les gènes candidats

Les gènes appartenant aux 3 grandes familles Dicer, Argonaute et RdRPs ont été recherchés par similarité de séquences dans le génome de *P. tetraurelia*

a. Les gènes de type Dicer ou Dicer-like

Le génome de *Paramecium tetraurelia* contient 8 gènes codants pour des domaines RNaseIII (Lepere et al. 2009^[27], Sandoval et al. 2014^[101]). Trois d'entre eux (Dcr1-3) possèdent un sous-domaine hélicase DExH suivis par le domaine DUF283 et deux domaines RNaseIII à l'extrémité C terminale. Ces trois protéines Dicer d'environ 1600 acides aminés ont donc une structure très proche des Dicer de *S. pombe*, *C. elegans* ou *A. thaliana*. DCR2 et DCR3 sont mutés sur les 4 acides aminés catalytiques des deux domaines RNaseIII ce qui suggère que seul DCR1 est actif. Les 5 autres protéines Dicer-like 1-5 sont plus courtes, entre 600 et 800 acides aminés (aa) et ne possèdent que les deux domaines RNaseIII (Figure 26).

La même analyse chez un autre cilié *Tetrahymena thermophila* a révélé des résultats similaires : deux Dicer possèdent les domaines hélicases et RNaseIII mais les deux domaines catalytiques sont mutés pour une des deux protéines et une seule protéine Dicer-like structurellement similaire aux Dcl1-5 de *P. tetraurelia* (Mochizuki & Gorovsky 2005^[102]).

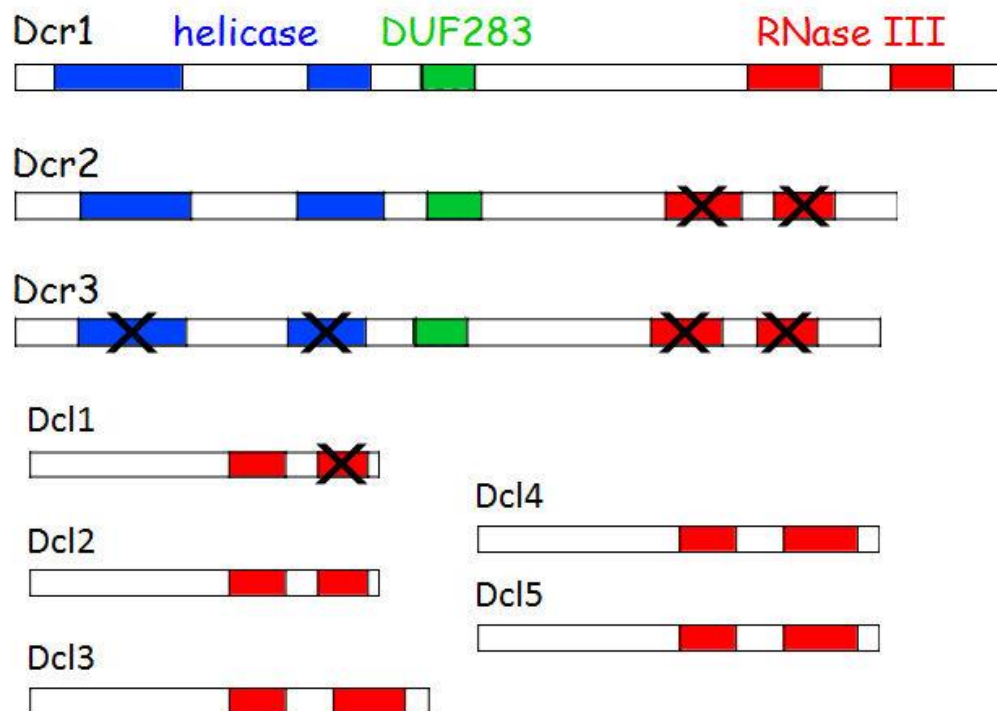


Figure 26: Structure des protéines Dicer et Dicer-like chez *P. tetraurelia* (Lepere et al. 2009^[27], Sandoval et al. 2014^[101]).

b. Les gènes Piwi.

Il y a 17 gènes du sous-clade PIWI de la famille argonaute chez *P. tetraurelia*. Ces gènes ont été nommés PTIWI01-PTIWI15 (PTIWI04 étant un pseudogène) (Bouhouche et al. 2011^[103]). Aucun gène du sous clade Ago n'a été trouvé (Figure 27). Il en est de même chez *T. thermophila* où 12 gènes PIWI ont été trouvés mais aucun appartenant au sous-clade Ago. 11 des 17 protéines Ptiwi portent la triade d'acides aminés DDH essentiels à l'activité de *slicing* (capacité de clivage d'un ARN). Ces résidus sont différents pour Ptiwi07 (IDQ) et pour les deux paralogues Ptiwi12 et Ptiwi15 (EDH). Ces 15 gènes ont souvent un ou deux paralogues issus des 3 dernières duplications du génome et peuvent être regroupés en 6 familles (Figure 28).

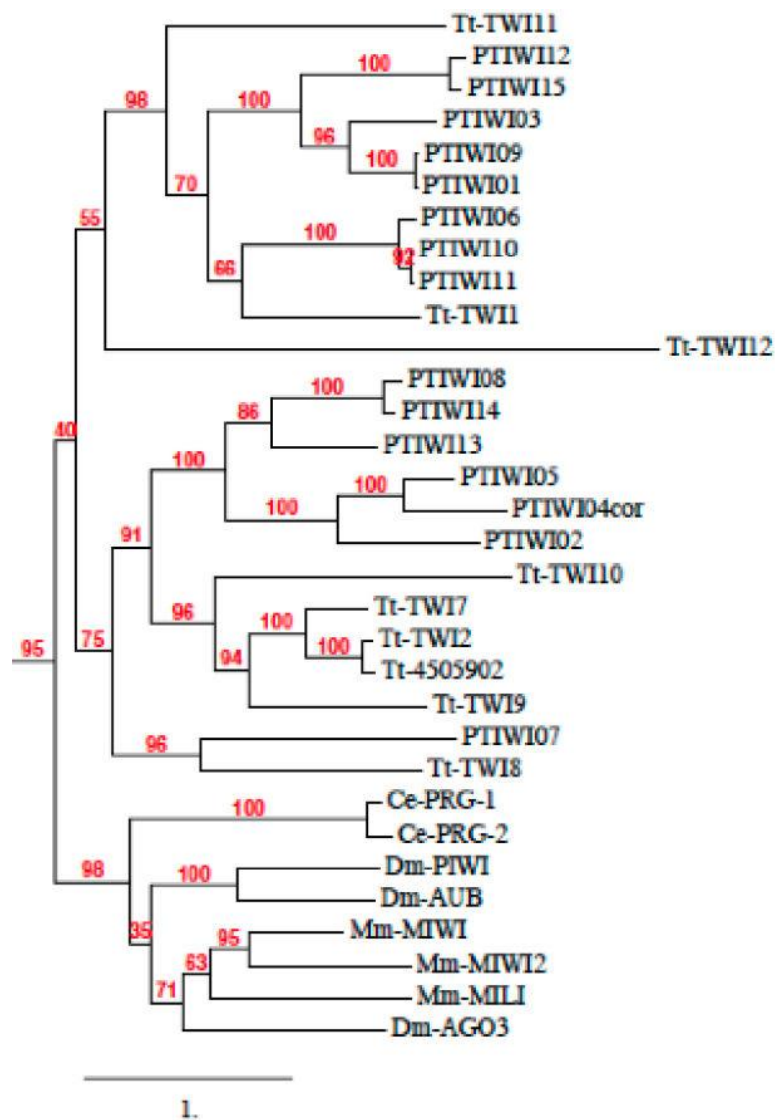


Figure 27: Arbre phylogénétique des argonautes du sous clade Piwi. Les valeurs de bootstrap sont indiquées en rouge. PTIWI = *Paramecium tetraurelia* piwi, Tt = *Tetrahymena thermophila*, Dm = *Drosophila melanogaster*, Mm = *Mus musculus*, Ce = *Caenorhabditis elegans* (Bouhouche et al. 2011^[103])

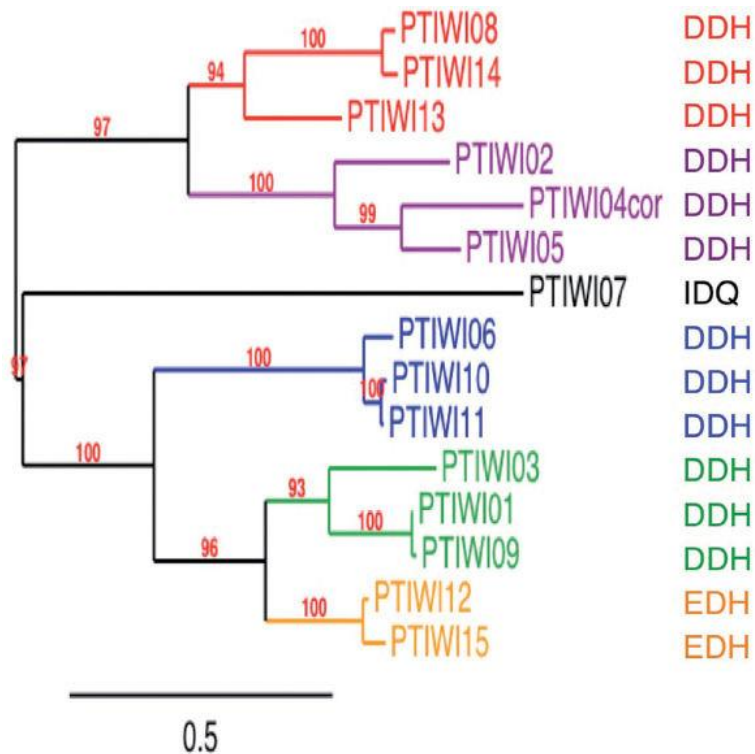


Figure 28 : Arbre phylogénique PTIWI. Les 6 familles de Piwi chez *P. tetraurelia* sont représentées par des couleurs différentes. Les 3 résidus catalytiques de chaque Ptiwi sont indiqués à droite (Bouhouche et al. 2011^[103])

c. Les RdRPs

Il y a 4 gènes codants potentiellement des RdRPs chez la paramécie (RDRP1-4). RdRP4 est le paralogue WGD2 de RdRP1 et porte de nombreuses mutations dans sa séquence (notamment une délétion de 7 acides aminés dans la région catalytique). De plus, alors que RdRP1 2 et 3 sont homologues à la RdRP (Rdr1) de *Tetrahymena*, RdRP4 a largement divergé suggérant que cette RdRP est un pseudogène non fonctionnel. Seules RdRP1 et RdRP2 semblent pouvoir être catalytiquement actives chez *Paramecium tetraurelia* et possèdent les résidus DLDGD du domaine catalytique (résidus très conservés au sein des eucaryotes). Ces mêmes résidus sont mutés dans RdRP3 (DEFDD) (Figure 29) (Marker et al. 2010^[104]).

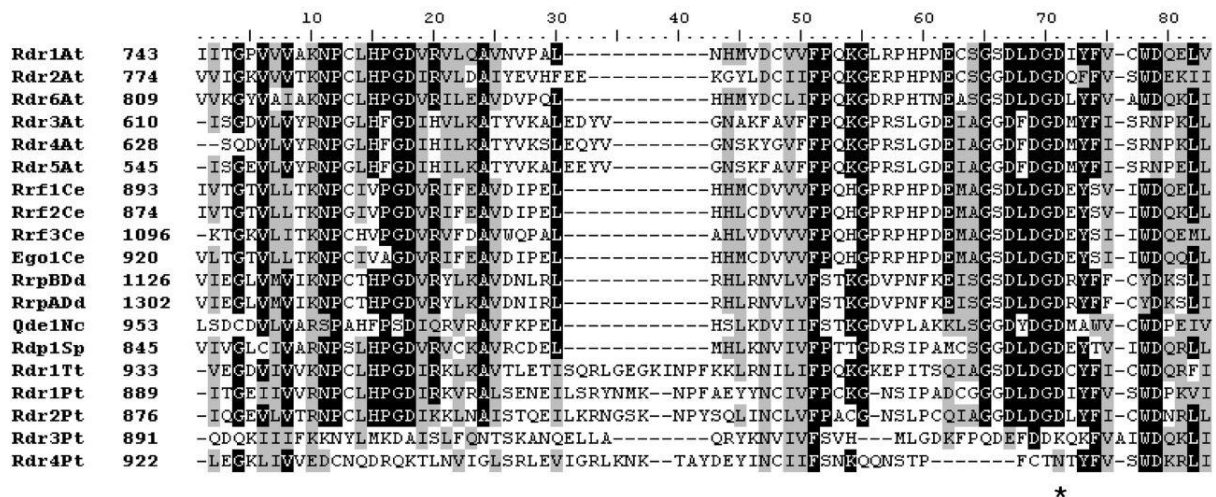


Figure 29 : Alignement MUSCLE du domaine catalytique des RdRPs chez *P. tetraurelia* et d'autres eucaryotes. Les acides aminés conservés sont indiqués en gris et noir, l'astérisque indique l'acide aspartique nécessaire à l'activité catalytique de QDE-1 (*Neurospora crassa*). Pt = *Paramecium tetraurelia*, Tt = *Tetrahymena thermophila*, At = *Arabidopsis thaliana*, Ce = *Caenorhabditis elegans*, Sp = *Saccharomyces pombe*, Dd = *Dictyostelium discoideum*, Nc = *Neurospora crassa*. (Marker et al. 2010^[104])

Les RdRPs 1 et 2 appartiennent au groupe β des RDR d'après l'étude phylogénétique de (Zong et al. 2009^[61]) et seraient donc plus proches de certains métazoaires, comme l'anémone marine *Nematostella vectensis* et le céphalocordé *Branchiostoma florida* que des plantes. Cependant les mécanismes d'action des RdRPs chez ces deux espèces n'ont pas été étudiés. Les RdRPs impliquées dans les voies d'ARN interférence chez *C. elegans* appartiennent au groupe α . Le gène RDR3 n'a pas été analysé dans cette même étude.

d. Expression des gènes d'ARNi

Il est possible de trier ces gènes (Dicer Piwi et RdRP) en deux groupes en fonction de leur profil d'expression au cours du cycle de la paramécie. 1 Dicer-like et 3 Dicer, 7 (PTIWI2, 4, 5, 12, 13, 14, 15) et les 3 RdRPs présentent un niveau d'expression constant au cours du cycle aussi bien en croissance végétative qu'à l'autogamie.

Au contraire, 2 Dicer-like (DCL2, DCL3) et 8 Ptiwi (PTIWI 1, 3, 6, 7, 8, 9, 10, 11) sont fortement exprimés au cours des événements sexuels (à l'autogamie ou à la conjugaison). Le pic d'expression de chacun de ces gènes peut varier, certains sont très exprimés en début de méiose alors que d'autres s'expriment pendant le développement du nouveau MAC. Ces différents profils suggèrent des fonctions différentes pour chacun de ces gènes (Figure 30). (Lepere et al. 2009^[27], Bouhouche et al. 2011^[103], Marker et al. 2010^[104]).

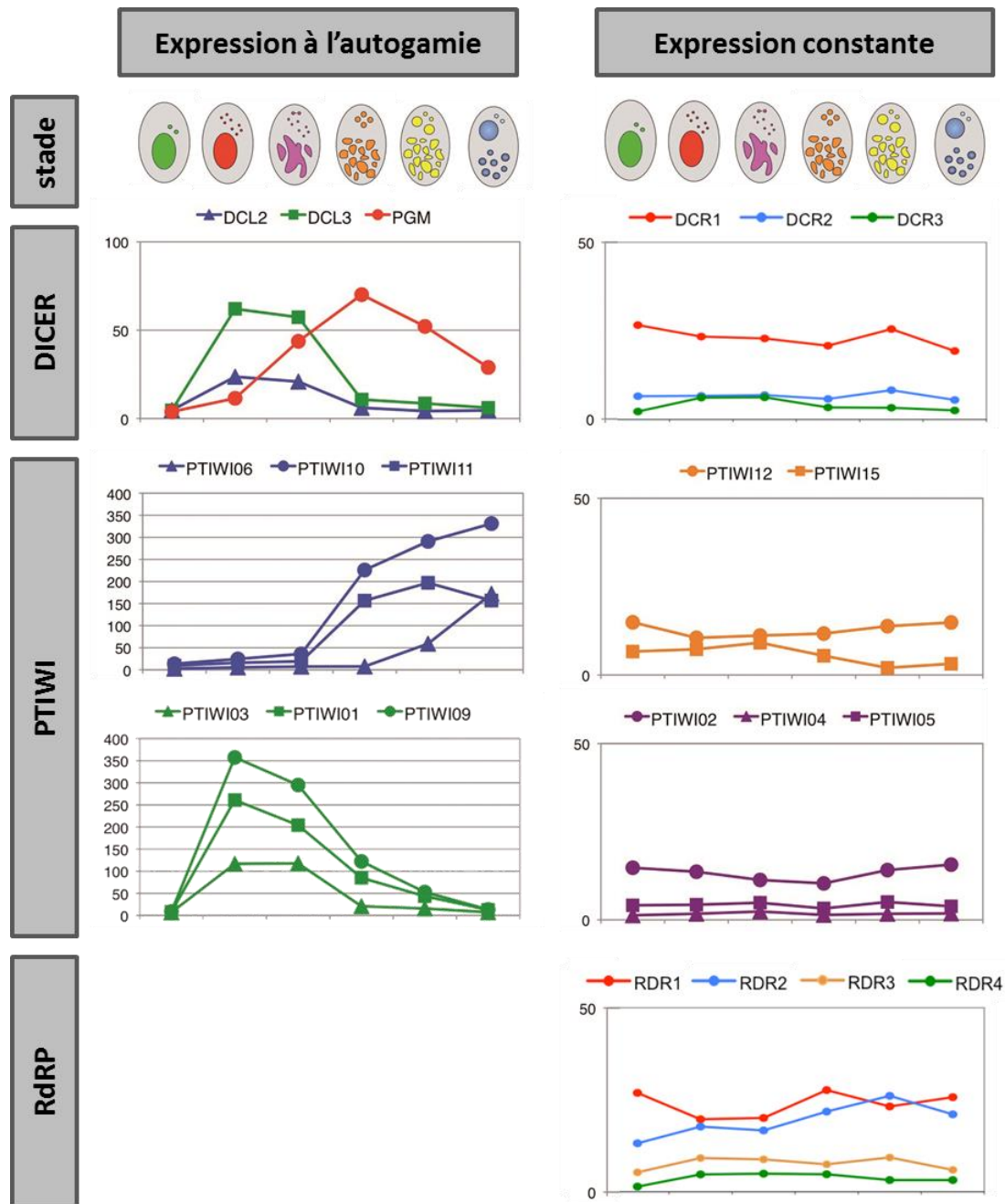


Figure 30: Profil d'expression des gènes impliqués dans l'ARNi au cours de l'autogamie. Les quantités relatives d'ARNm pour chaque gène ont été déterminées à l'aide de microarray (unité arbitraire). Les gènes présentant un pic d'expression à l'autogamie (à gauche) ou constamment exprimés au cours du cycle (à droite) sont représentés.

2) Les voies de siRNAs

Il existe au moins deux voies de siRNAs chez la paramécie qui se distinguent entre autre par leur méthode d'induction. La voie inductible par transgène d'une part et la voie inductible par *feeding* d'ARNdb d'autre part. Ces deux voies sont représentées schématiquement dans la [figure 31](#) et sont détaillées dans les deux paragraphes suivants.

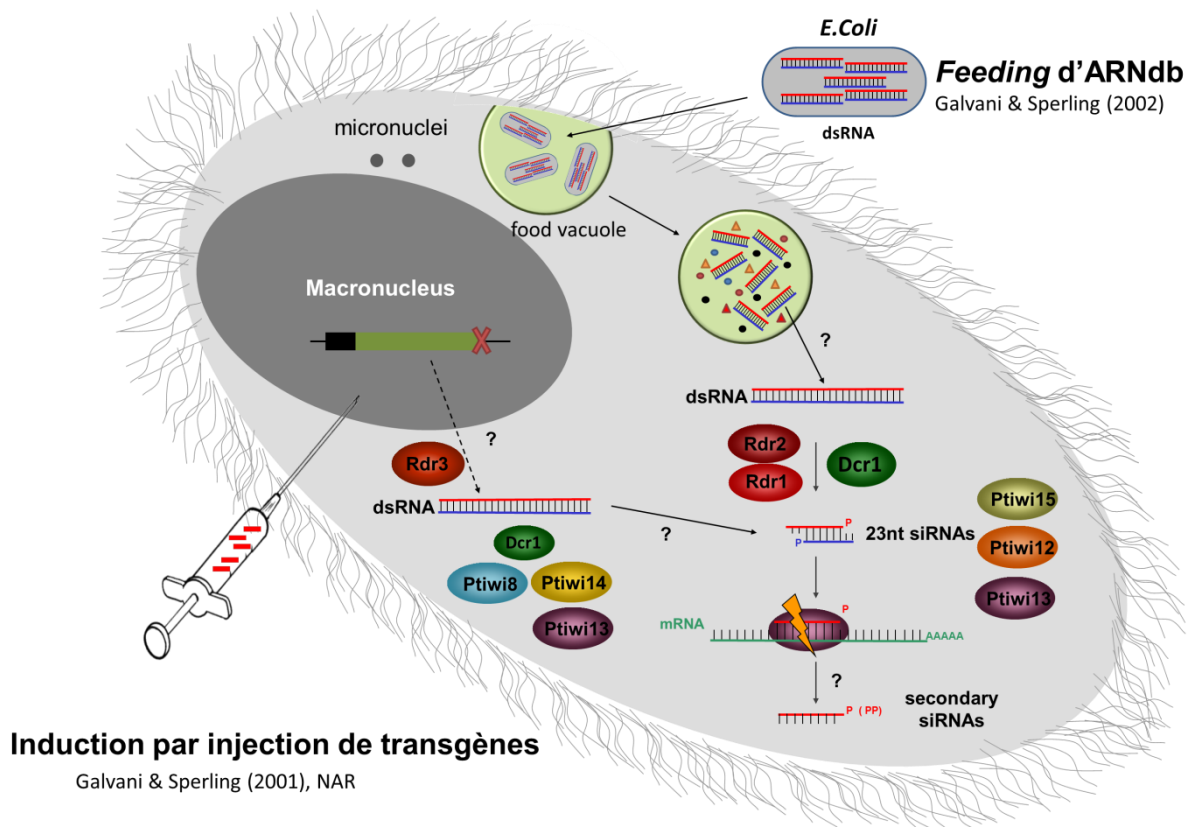


Figure 31: Représentation schématique des deux voies de siRNAs chez la paramécie. La voie d'ARNi induite par *feeding* d'ARNdb utilise les capacités de *P. tetraurelia* à ingérer des bactéries par phagocytose puis à extraire les longs ARNdb des vacuoles digestives (par un mécanisme inconnu). Les longs double-brin sont pris en charge par Dcr1 et nécessitent les deux RdRPs RdR1 et RdR2 pour produire des siRNAs primaires de 23nt. Les études réalisées jusque-là n'apportaient aucune évidence sur la production de siRNAs secondaires. 3 gènes PTIWI sont impliqués dans cette voie (PTIWI12, 13 et 15). La voie ARNi induite par injection de transgènes dans le macronoyau implique des protéines en commun avec la voie par *feeding* : Dicer1 et Ptiwi13 ainsi que des protéines spécifiques : Ptiwi8, Ptiwi14 et RdR3.

a. La voie inductible par transgène

Depuis 1998, on sait que l'injection d'un transgène dans le mac de la paramécie peut induire la répression du gène endogène homologue (Ruiz et al. 1998^[105]). La répression induite par un transgène est maintenue pendant toute la croissance végétative de la paramécie mais est perdue à l'autogamie ou à la conjugaison. En effet, l'ancien mac étant dégradé et remplacé par un nouveau formé à partir du mic, tout l'ADN injecté est perdu à ce moment. Dans cette première étude, les auteurs montrent que l'efficacité de la répression est dépendante d'au moins deux critères: la similarité de séquence et le nombre de copies stablement maintenues dans le mac. L'injection de la séquence codante de la famille multigénique TMP composée d'environ 100 gènes a révélé que seuls les ARNm des

paralogues avec plus de 85% d'identité en séquence nucléotidique étaient réprimés. Les autres gènes de la famille, plus divergents, n'étaient pas affectés. Cette découverte montre que le signal d'extinction est transmis par homologie de séquence et qu'en conséquence il est difficile par cette méthode d'éteindre spécifiquement un gène lorsqu'il possède des paralogues proches (et c'est très souvent le cas chez la paramécie suite à la dernière duplication du génome). Le nombre de copies du transgène injecté dans le mac a aussi une influence sur l'efficacité de la répression. Il n'est pas possible de contrôler le nombre de copies injectées et maintenues dans le mac cependant les auteurs ont montré qu'il faut au minimum 25 copies par génome haploïde donc environ 20 000 copies dans le mac pour induire l'extinction. De plus, alors que l'injection du gène entier (avec le promoteur et le 3'UTR) et quel que soit le nombre de copies n'affecte habituellement pas l'expression du gène, l'injection de la séquence codante peut suffire à le réprimer. Ce résultat montre que l'absence des séquences régulatrices du gène et un grand nombre de copies suffisent à déclencher la répression et suggère donc un mécanisme de répression post-transcriptionnel (Ruiz et al. 1998^[105]).

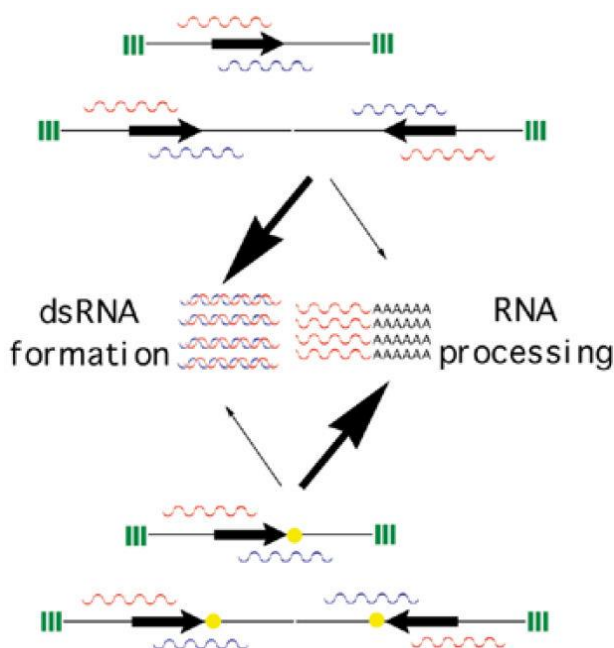


Figure 32: Conséquences de l'injection de transgènes dans le MAC. Les transgènes avec (en bas) ou sans la séquence 3' non codante (en haut) sont représentés. En absence du 3', l'injection du transgène dans le mac induit majoritairement la production d'ARNdb (flèche large). En présence du 3' l'injection induit majoritairement la production d'un ARNm normal et empêche sa prise en charge par l'ARNi.

Cette hypothèse fut confirmée trois ans plus tard par A.Galvani et L.Sperling. Les auteurs ont montré d'une part que seul l'absence du 3'UTR est requise pour induire la répression et d'autre part que l'injection d'un transgène avec ou sans promoteur conduit à la production d'ARN de taille aberrante dans les deux orientations. Cette observation indique que l'injection d'un transgène induit la répression du gène par PTGS et suggère la production d'un long ARN double-brin à l'origine de l'activation de la voie d'ARN interférence (Galvani & Sperling 2001^[106]) (Figure 32).

Enfin, les gènes codants pour des protéines de type Dicer, Piwi et RdRP (voir ci-dessus) ont été testés pour leur implication dans l'ARNi induite par transgène. Les gènes testés sont éteints par *feeding* d'ARNdb (voir paragraphe suivant) dans

des paramécies injectées avec un transgène. Si l'expression du gène homologue au transgène est maintenue alors le gène éteint par *feeding* d'ARNdb est impliqué dans la voie d'ARNi induite par transgène.

Parmi les gènes de type Dicer et Dicer-like, seul DCR1 a pu être impliqué dans l'ARNi induite par transgène (Lepere et al. 2009^[27]). D'autre part, trois gènes Piwi (PTIWI08, 14, 13) semblent nécessaires à cette voie ainsi que la RdRP RDR3 (Bouhouche et al. 2011^[103], Marker et al. 2010^[104]).

b. La voie inductible par *feeding* d'ARNdb

La méthode pour éteindre un gène par "*feeding* d'ARNdb" a été établie en 2002 chez la paramécie. (Galvani & Sperling 2002^[17]). Les paramécies sont nourries avec la souche HT115 d'*E.coli*, déficiente pour la RNaseIII et possédant une ARN polymérase T7 inductible à l'IPTG. Cette souche est transformée avec un plasmide (L4440) contenant une séquence particulière flanquée de deux promoteurs T7 convergents. Après induction à l'IPTG, les bactéries produisent de l'ARNdb complémentaire à cette séquence. Si la séquence entre les deux promoteurs T7 est homologue à une portion de gène de la paramécie ce gène est alors réprimé par ARNi. Cette méthode a d'abord été testée sur des gènes non essentiels impliqués dans la formation ou la décharge des trichocystes (TMP4a et ND7). Les trichocystes sont des organites éjectiles, entourés d'une membrane et placés sous la surface des paramécies. En cas de stress, les trichocystes sont éjectés violemment de la cellule ce qui lui permettrait de se déplacer rapidement et d'échapper à un éventuel prédateur (Figure 33). Il est possible d'induire l'exocytose des trichocystes à l'aide d'acide picrique (phénotype tric+) (Galvani & Sperling 2000^[107]). Après *feeding* d'une portion de gène de TMP4a ou ND7 et dans un délai de 48h, les paramécies deviennent incapables d'éjecter les trichocystes (phénotype tric-). Le phénotype est maintenu tant que les paramécies sont nourries avec les bactéries *E.coli* produisant de l'ARNdb.

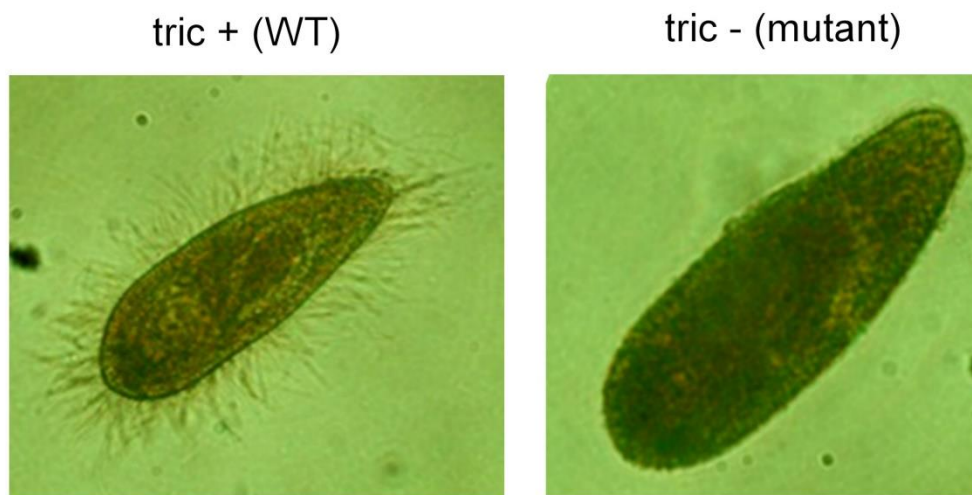


Figure 33 : Phénotypes *tric+* et *tric-* dans de l'acide picrique. A gauche, une paramécie sauvage capable d'éjecter ses trichocystes en présence d'acide picrique. A droite, une cellule mutante ou inactivé pour ND7 et donc incapable d'éjecter ses trichocystes.

Les effets de la répression du gène sont rapidement abolis lorsque les paramécies sont transférées dans un milieu avec des bactéries ne produisant pas d'ARNdb, il n'y a donc *a priori* aucun maintien de la répression. Cette observation suggère qu'il faut un apport continu en ARNdb pour maintenir la répression et donc que les ARNdb dans le cytoplasme deviennent rapidement en quantité insuffisante soit parce qu'ils sont dilués par la croissance végétative soit parce qu'ils sont rapidement dégradés ou encore utilisés pour l'ARN interférence. Pour maintenir la répression en absence d'apport extérieur d'ARNdb, il faudrait qu'ils soient amplifiés ce qui ne semble pas être le cas. Cette voie d'ARNi induite par *feeding* est utilisée très régulièrement comme outil de génétique inverse chez la paramécie. Elle est en effet plus facile à mettre en place et à contrôler que l'injection de transgènes.

Les gènes nécessaires au fonctionnement de cette voie ont été étudiés par ARNi récursif. Des expériences de "double-*feeding*" ont été réalisées avec le gène candidat à tester pour son implication dans l'ARNi et le gène rapporteur (en général ND7). Si le gène rapporteur ne peut plus être réprimé alors le gène testé est certainement impliqué dans le mécanisme de répression par *feeding*. Par cette méthode, DCR1, RDR1, RDR2 et PTIWI13 semblent impliqués dans la voie d'ARNi induite par *feeding* (Lepere et al. 2009^[27], Bouhouche et al. 2011^[103], Marker et al. 2010^[104]). La co-répression PTIWI12-PTIWI15 semble aussi affecter l'ARNi par *feeding* cependant l'effet est plus faible (phénotype partiel) et met plus de temps à apparaître (3 jours) (Bouhouche et al. 2011^[103]).

Il est à noter que les analyses par ARNi récursif ne permettent pas de conclure que sur des résultats positifs aussi bien pour la voie d'ARNi induite par transgène que celle induite par *feeding* d'ARNdb. En effet, le fait que la répression d'un gène par *feeding* d'ARNdb n'affecte pas le fonctionnement de la voie peut être la conséquence de deux phénomènes : le gène testé

n'est pas impliqué dans la voie ou bien le *feeding* n'est pas assez efficace pour réprimer le gène testé (entre autre parce qu'il est lui-même nécessaire au fonctionnement de la voie). Il est donc tout à fait possible que d'autres gènes soient impliqués dans ces voies d'ARN interférence même s'ils n'ont pas été trouvés par ARNi récursif.

Ces deux méthodes d'induction de l'ARNi (par transgène ou par *feeding*) activent donc deux voies de production de siRNAs différentes puisque certains gènes impliqués sont spécifiques d'une voie ou de l'autre. C'est par exemple le cas des RdRPs : Rdr1 et Rdr2 ne semblent impliqués que dans la voie par *feeding* alors qu'au contraire Rdr3 est uniquement nécessaire à la voie par transgène (Marker et al. 2010^[104]).

c. Caractéristiques des siRNAs chez la paramécie

Les siRNAs de *P. tetraurelia* sont majoritairement de 23 nt et sont de ce fait distinguables des scnRNAs de 25 nt (voir paragraphe p 51) (Garnier et al. 2004^[108]).

Un petit nombre de sRNAs issu de l'extraction d'ARN total de paramécies en conjugaison et soumises au *feeding* ND7 ont été clonés puis séquencés dans l'étude de Lepère et al en 2009. Les auteurs ont montré que le *feeding* d'ARNdb entraîne la production de siRNAs primaires sur la totalité de l'ARNdb.

Sur 58 siRNAs qui ont pu être alignés sur l'ARNdb, deux classes ont été identifiées. La première classe de siRNAs (environ la moitié) sont majoritairement de 23 nt, non modifiés et s'alignent en quantité équivalente sur les deux brins de l'ARNdb. Certains de ces siRNAs se situent aux extrémités de l'ARNdb inducteur, sur des séquences qui n'ont aucune similarité avec ND7 (séquences appelées *polylinkers*). Ces siRNAs sont donc synthétisés à partir de l'ARNdb et sont clairement identifiables comme siRNAs primaires. La seconde classe de siRNAs est composée de siRNAs modifiés avec une ou plusieurs adénines non génomiques et sont donc plus longs (24,25 nt). Étonnement, ces siRNAs modifiés sont tous anti-sens au transcrit ND7 et montrent un fort biais 5'U. Enfin, un unique siRNA adénylé a été aligné sur ND7 en dehors de la région complémentaire à l'ARNdb pouvant suggérer une production de siRNAs secondaires par transitivité (Figure 34).

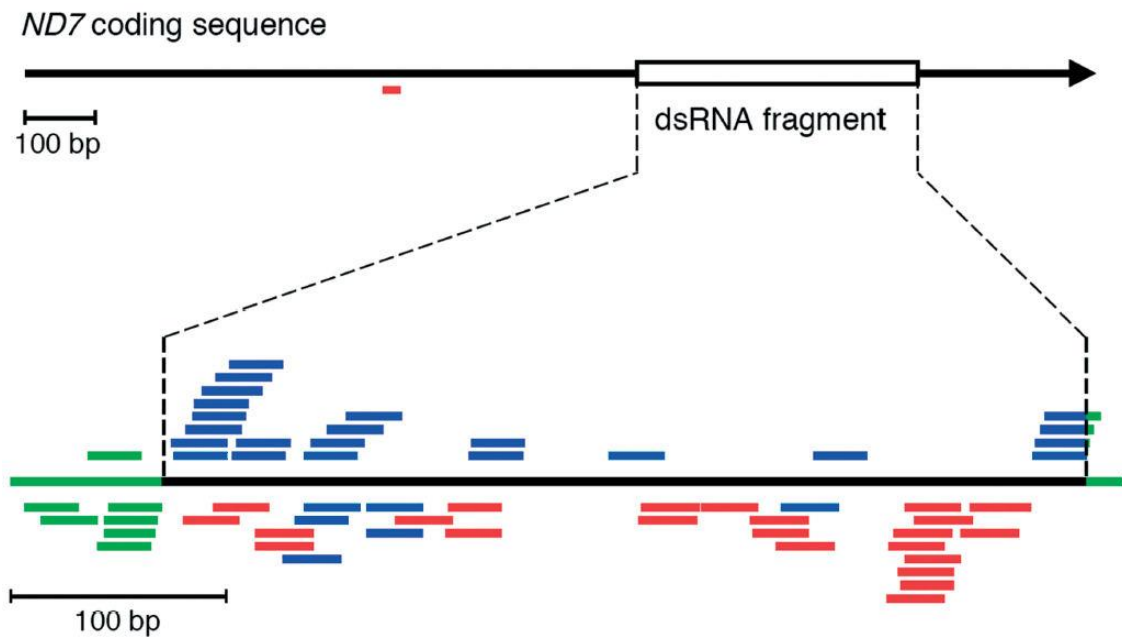


Figure 34: Distribution des siRNAs sur ND7. La flèche noire représente le gène ND7 complet et le rectangle blanc, la position du fragment cloné dans le vecteur *feeding* afin de produire l'ARNdb. La ligne du dessous représente l'ARNdb produite dans *E.coli*. Les extensions en vert de chaque côté de ND7 sont les séquences du vecteur entre les deux promoteurs T7 convergents (régions *polylinker*). Les siRNAs sens et anti-sens sont alignés au-dessus et au-dessous de cette ligne respectivement. Les siRNAs primaires sont en vert, les siRNAs non modifiés sont en bleu et les siRNAs polyadénylés sont en rouge. Un seul siRNA a pu être trouvé sur la séquence ND7 en amont du fragment *feeding* (Garnier et al. 2004^[108]).

Ces résultats doivent être interprétés pour le contexte particulier de la conjugaison où la machinerie d'ARNi est principalement utilisée pour synthétiser des scnRNAs et où le nombre de siRNAs induit par l'ARNdb devient donc très faible en proportion. De plus, l'unique siRNA sur ND7 observé en dehors de la région complémentaire à l'ARNdb ne permet pas de conclure d'un point de vue statistique quant à la production de siRNAs secondaires.

La synthèse de siRNAs secondaires a également été suggérée dans l'étude de (Marker et al. 2010^[104]). Dans cette étude les siRNAs sont étudiés par *Northern Blot* et l'orientation des siRNAs est spécifiquement révélée par une sonde sens (détection des siRNAs anti-sens) ou une sonde anti-sens (détection des siRNAs sens). Deux tailles différentes pour les siRNAs ont été observées dans cette étude. D'une part une grande quantité de siRNAs de 22 nt principalement anti-sens pouvant être des siRNAs secondaires et d'autre part des siRNAs primaires de 23 nt, répartis sur les deux brins (Figure 35).

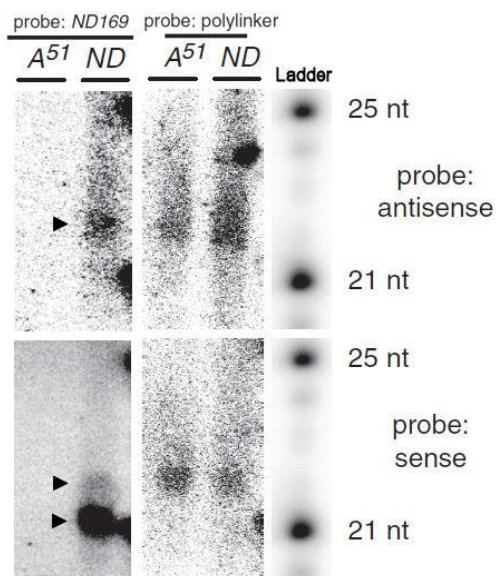


Figure 35 : Caractéristiques des siRNAs induits par ARNdb. Les gènes ND169 (ND) ou A^{51} (gène contrôle) sont éteints par *feeding* d'ARNdb. Les siRNAs produits sont détectés avec une sonde sens (en bas) ou anti-sens (en haut) homologue au gène ND169 (à gauche) ou au polylinker présent dans les deux *feeding* d'ARNdb (à droite). Les flèches indiquent les deux tailles de siRNAs observés. Les siRNAs primaires hybridés au polylinker sont en quantité équivalente sur les deux brins et sont de 23 nt. Deux classes de siRNAs sont hybridés sur ND169 : des siRNAs de 22 nt essentiellement anti-sens et des siRNAs de 23 nt répartis sur les deux brins. (Marker et al. 2010^[104])

Cette faible différence de taille reste à être confirmée car différents facteurs peuvent influencer la migration des siRNAs sur blot tel que le taux de purine, l'état de phosphorylation, la méthylation, ... Des tests biochimiques ont été réalisés pour répondre à ces questions. Après traitement des siRNAs avec une exonucléase spécifique des ARN 5'-monophosphate, tous les siRNAs semblent être dégradés ce qui suggère que les siRNAs sont tous 5' monophosphate. De même, le traitement au periodate suivi par la β -élimination permet de retirer le dernier nucléotide en 3' seulement s'il est 2'-OH et 3'-OH. Après ce traitement les siRNAs migrent 1,5 nt plus loin dans le gel montrant qu'ils ne sont pas résistants au traitement et donc non-méthylés en 3'. Au contraire les siRNAs induits par transgène sont résistants et sont donc méthylés en 2' ou en 3' du ribose 3' terminal (Marker et al. 2010^[104]).

d. Les siRNAs endogènes

Très peu de siRNAs produits par le génome de la paramécie ont été caractérisés jusqu'à présent. Le seul locus de sRNAs endogène identifié (cluster22) est un locus intergénique de 200 pb situé sur le scaffold 22 entre deux gènes convergents (Lepere et al. 2009^[27]). Il a été identifié après clonage et séquençage de 1024 sRNAs au cours d'un *feeding* du gène ND7 à la conjugaison parmi lesquels 9 sRNAs de 23 nt ont pu être alignés sur le cluster22. 5 d'entre eux ont une queue poly A et 7 commencent avec un U. Ces 9 sRNAs sont sur le même brin suggérant que ce sont bien des siRNAs endogènes produits par le même mécanisme que les siRNAs adénylés alignés sur le gène ND7 ciblé par l'ARNdb.

De plus, les antigènes de surface chez *P. tetraurelia* semblent être régulés par des voies d'ARNi. Au moins 11 antigènes de surface peuvent être exprimés de manière mutuellement exclusive. Le gène exprimé définit le sérotype du clone. Malgré de nombreuses études traitant du sérotype des paramécies, la fonction de ces antigènes de surface reste inconnue. Il a été montré que l'expression des antigènes de surface est régulée de manière post-transcriptionnelle (Simon et al. 2006^[109]) et que cette régulation est dépendante de RDR3 (Marker et al. 2010^[104]). L'analyse des siRNAs homologues aux antigènes de surface permettrait la compréhension de ces mécanismes de régulation.

3) La voie des scnRNAs

Une voie d'ARN interférence exprimée uniquement au cours de la reproduction de la paramécie a également été identifiée (Lepere et al. 2009^[27]). Cette voie produit des scanRNAs d'environ 25 nt avec une signature 5'UNG qui sont requis pour l'élimination d'un grand nombre d'IES du génome mac (environ 1/3 des 45000 identifiés) au cours des événements sexuels. Les scnRNAs ont de fortes similitudes avec les piRNAs. Structurellement, ces deux classes de sRNAs produisent des molécules de plus grande taille que les siRNAs et portent un U en 5'. De plus, les piRNAs et les scnRNAs sont nécessaires à la reproduction sexuelle et sont impliqués dans le contrôle des éléments transposables (les IES étant certainement dérivés de transposons). Les scnRNAs sont produits à partir du génome mic pendant la méiose. Le mic est transcriptionnellement inactif en phase végétative mais il est actif en début de méiose et transcrit dans les deux orientations, probablement sur la totalité de sa séquence (Figure 36).

Les double-brin formés requièrent Dcl2 et Dcl3 (Lepere et al. 2009^[27]) pour être clivés en scnRNAs. Ces scnRNAs sont donc homologues aux séquences, exoniques, introniques et intergéniques des deux brins du génome germinale. L'extinction par *feeding* des gènes Dcl2 et 3 induit des défauts majeurs de réarrangements à l'autogamie et à la conjugaison et provoque la mort de la progéniture sexuelle (Lepere et al. 2009^[27]). PTIWI01 et PTIWI09 sont également requis pour le fonctionnement des scnRNAs (Bouhouche et al. 2011^[103]). Un modèle a pu être construit pour expliquer comment les scnRNAs vont aller cibler spécifiquement les IES du macronoyau en développement pour induire leur élimination (Figure 36). Les scnRNAs sont transportés du micronoyau vers le macronoyau parental pour cibler les transcrits des séquences mac par homologie de séquence. Le génome MAC est probablement transcrit sur la totalité de sa séquence et sur les deux brins au cours des événements sexuels (Lepere et al. 2008^[110]). Les scnRNAs qui n'ont pas de cible dans l'ancien macronoyau vont finalement être dirigés vers le nouveau macronoyau en développement et cibler les transcrits naissants qui étaient donc absents dans l'ancien MAC (Coyne et al. 2012^[93]) (Figure 36). L'élimination ARN-dépendante des IES permet à chaque génération de reproduire un génome mac identique au génome mac parental.

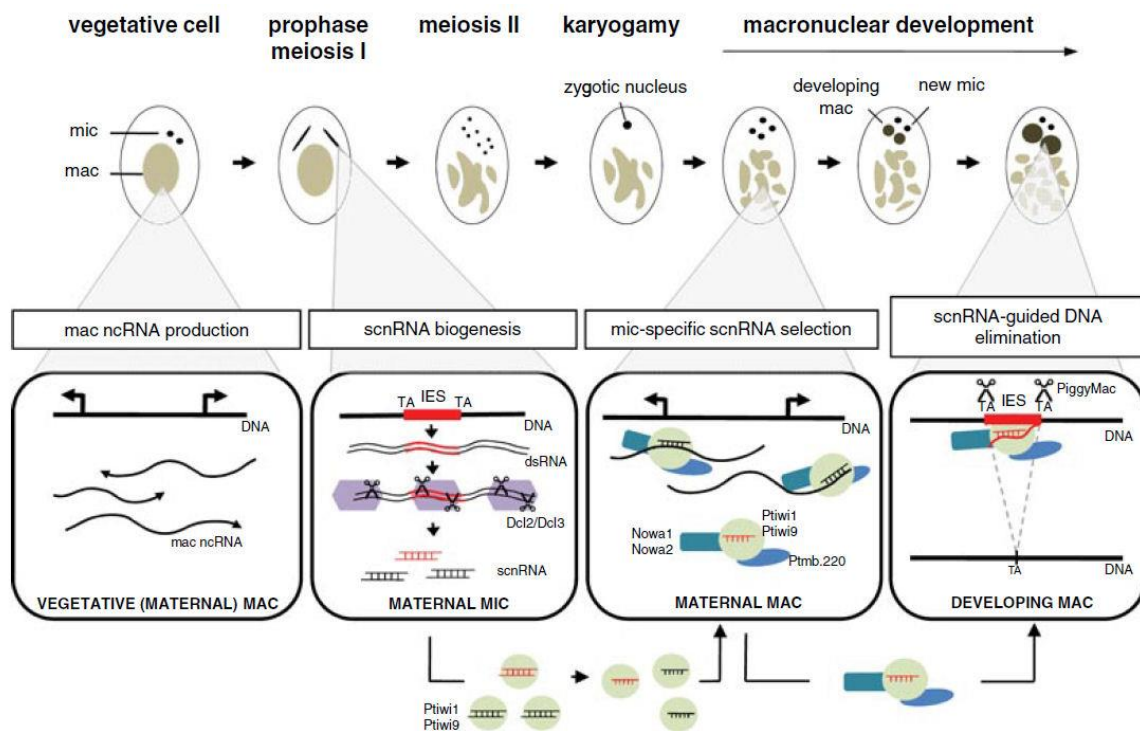


Figure 36 : Modèle des scnRNAs chez *P. tetraurelia*. (Coyne et al. 2012^[93])

Un mécanisme d'amplification des scnRNAs a récemment été caractérisé. A un stade tardif au cours du développement du nouveau mac, des iesRNAs légèrement plus grands que les scnRNAs (25-28 nt) et dépendants de Dcl5 apparaissent. Ces iesRNAs sont présents uniquement sur les séquences des IES ce qui suggère qu'ils sont produits à partir des scnRNAs une fois qu'ils ont été sélectionnés dans l'ancien mac. Cependant les mécanismes contrôlant l'amplification et le rôle de ces scnRNAs secondaires restent à être découverts (Sandoval et al. 2014^[101]).

Cette voie des scnRNAs est également essentielle à la détermination du type sexuel des paramécies. En effet, le type sexuel E est dépendant de l'expression de la protéine transmembranaire mtA et le type sexuel O est exprimé lorsque le promoteur et le début de la séquence codante de mtA est excisé par la voie des scnRNAs de la même manière qu'une IES (Singh et al. 2014^[94]) (Figure 37). Le type sexuel des paramécies est donc hérité épigénétiquement, il est indépendant de la séquence germinale mais dépend de la rétention ou non du promoteur de mtA dans le macronoyau somatique parental. Cette découverte démontre la capacité d'exaptation de cette voie d'ARN interférence médiée par les scnRNA en faveur du contrôle de l'expression des gènes cellulaires (Singh et al. 2014^[94]).

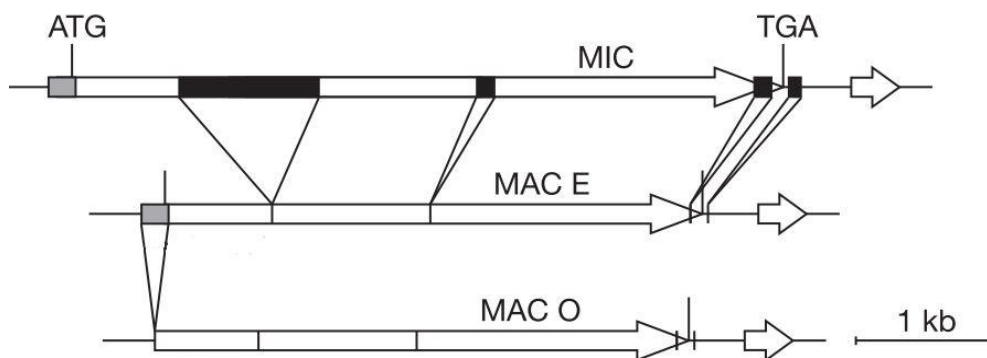


Figure 37: Représentation schématique du gène mtA dans le micronoyau germinale et dans les macronoyaux de type sexuel E ou O. Les rectangles noirs indiquent la position des IES, le rectangle gris la séquence contenant le promoteur de mtA excisé dans les paramécies de type sexuel O.

IV. Présentation du sujet

Les études précédentes ont permis d'identifier clairement deux voies d'ARN interférence végétatives chez la paramécie : la voie inductible par transgène et la voie induite par *feeding* d'ARNdb. Par ARN interférence récursive, un certain nombre de gènes Dicer, Piwi et RdRP ont été impliqués dans ces voies. Mon travail de thèse a consisté à étudier la voie d'ARNi induite par *feeding* d'ARNdb afin d'essayer d'en comprendre les mécanismes moléculaires mais aussi la fonction endogène. 3 grandes questions sur cette voie d'ARNi ont orienté mes recherches.

Les gènes impliqués dans la voie d'ARNi par *feeding* :

L'identification de gènes par ARN interférence récursif est une méthode qui nécessite d'avoir un *a priori* sur les gènes impliqués dans la voie. De plus, les résultats sont parfois partiels et un résultat négatif ne signifie pas que le gène testé n'est pas impliqué. De ce fait, il est très probable que d'autres gènes soient impliqués dans ces voies. Par exemple, la paramécie est capable d'extraire les longs ARN double-brin des vacuoles digestives pour initier l'ARNi. Il pourrait exister un mécanisme actif mettant en jeu une ou plusieurs protéines membranaire pour permettre le transport de ces ARNs. L'identification de ces gènes semble alors essentielle pour comprendre le fonctionnement de cette voie.

La fonction des RdRPs dans la voie d'ARNi :

Le fait que deux RdRPs soient impliquées dans la synthèse de siRNAs primaires pose au moins deux questions. La première concerne le rôle de ces RdRPs. RdR1 et RdR2 sont nécessaires à la synthèse des siRNAs primaires dans une voie d'ARNi qui *a priori* ne nécessite pas la synthèse d'un ARNdb puisqu'il est déjà sous cette forme dans les bactéries ingérées par la paramécie. La seconde question est de comprendre pourquoi les deux RdRPs sont nécessaires. Elles semblent non redondantes puisque l'inhibition d'une seule RdRP suffit à rendre la cellule déficiente pour l'ARNi, cependant, la quantité de protéine restante de l'autre RdRP pourrait ne pas être suffisante pour une répression efficace. Enfin, ces deux RdRPs pourraient aussi être impliquées dans la synthèse d'éventuels siRNAs secondaires.

Mécanismes de production de siRNAs secondaires :

La production de siRNAs secondaires chez la paramécie n'a pas pu être montrée clairement dans les études précédentes, cependant l'existence de 3 RdRPs dans les voies végétatives de siRNAs et le biais anti-sens des siRNAs le suggère fortement. Les mécanismes

conduisant à la synthèse de siRNAs secondaires et plus généralement les caractéristiques moléculaires des siRNAs chez la paramécie restent à être étudiés.

Pour répondre à ces questions, deux approches très différentes mais complémentaires ont été mises en œuvre dans le laboratoire. La première approche a été de réaliser d'un crible génétique direct afin d'obtenir des mutants mendéliens déficients pour l'ARNi et d'identifier *de novo* des gènes impliqués dans la voie. Les résultats obtenus sont présentés dans la première partie de la section suivante et répondent au moins partiellement à la première question. La seconde approche a été de séquencer les petits ARN produits après induction de l'ARNi par *feeding* dans des paramécies sauvages ou dans les mutants obtenus par le crible précédent. Cette analyse a été réalisée pour comprendre le rôle de chacun des gènes dans la voie mais aussi pour analyser les caractéristiques des siRNAs primaires et secondaires et donc pour répondre aux deux autres questions.

Résultats

I. Première partie : Identification et caractérisation des gènes impliqués dans l'ARNi induite par ARNdb

A. *A forward genetic screen reveals essential and non-essential RNAi factors in P. tetraurelia*

Un crible de mutagenèse a permis d'obtenir des mutants mendéliens déficients pour l'ARNi, dont les génomes ont été séquencés afin d'identifier sans *a priori* des gènes impliqués dans la voie d'ARN interférence induite par ARNdb. Au total, 49 allèles ont ainsi été identifiés pour 6 gènes : un Dicer (DCR1), deux RdRPs (RDR1 et RDR2), une nucléotidyl-transférase (CID1) et deux nouveaux gènes (PDS1 et PDS2). Certains de ces gènes (RDR2 et DCR1) sont également impliqués dans une autre voie d'ARNi (inductible par des transgènes). De manière intéressante, l'analyse de la diversité allélique obtenue indique que ces derniers sont essentiels, tandis que les gènes spécifiquement impliqués dans le traitement de l'ARNdb ingéré ne le sont pas. Les résultats de ce crible de génétique direct ont été publiés dans Nucleic Acid Research en janvier 2014. L'article complet est présenté dans les pages qui suivent et les documents supplémentaires se trouvent dans **[l'annexe I](#)**

A forward genetic screen reveals essential and non-essential RNAi factors in *Paramecium tetraurelia*

Simone Marker^{1,†,*}, Quentin Carradec^{1,2}, Véronique Tanty¹, Olivier Arnaiz³ and Eric Meyer¹

¹Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS, Inserm, U1024, CNRS, UMR 8197, Paris F-75005, France, ²Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, F-75252 Paris cedex 05, France, ³CNRS UPR3404 Centre de Génétique Moléculaire, Gif-sur-Yvette F-91198 cedex, France and ⁴Université Paris-Sud, Département de Biologie, Orsay, F-91405, France

Received November 13, 2013; Revised January 31, 2014; Accepted February 10, 2014

ABSTRACT

In most eukaryotes, small RNA-mediated gene silencing pathways form complex interacting networks. In the ciliate *Paramecium tetraurelia*, at least two RNA interference (RNAi) mechanisms coexist, involving distinct but overlapping sets of protein factors and producing different types of short interfering RNAs (siRNAs). One is specifically triggered by high-copy transgenes, and the other by feeding cells with double-stranded RNA (dsRNA)-producing bacteria. In this study, we designed a forward genetic screen for mutants deficient in dsRNA-induced silencing, and a powerful method to identify the relevant mutations by whole-genome sequencing. We present a set of 47 mutant alleles for five genes, revealing two previously unknown RNAi factors: a novel *Paramecium*-specific protein and a Cid1-like nucleotidyl transferase. Analyses of allelic diversity distinguish non-essential and essential genes and suggest that the screen is saturated for non-essential, single-copy genes. We show that non-essential genes are specifically involved in dsRNA-induced RNAi while essential ones are also involved in transgene-induced RNAi. One of the latter, the RNA-dependent RNA polymerase RDR2, is further shown to be required for all known types of siRNAs, as well as for sexual reproduction. These results open the way for the dissection of the genetic complexity, interconnection, mechanisms and natural functions of RNAi pathways in *P. tetraurelia*.

INTRODUCTION

Small RNA pathways have greatly diversified in function during the evolution of eukaryotes (1,2) but generally rely on the production of 21–28nt short RNAs (sRNAs) from single- or double-stranded RNA precursors. Once loaded onto Argonaute-containing effector complexes, sRNAs can target complementary transcripts and silence homologous sequences by diverse mechanisms, such as posttranscriptional or transcriptional inhibition or even DNA elimination (1,3). In RNA interference (RNAi), short interfering RNAs (siRNAs) are produced by endonucleases of the Dicer family from long double-stranded RNA (dsRNA) (4,5). The initial dsRNA triggers may be products of RNA-dependent RNA polymerases (RdRPs), enzymes that also act at downstream steps in some organisms to amplify the silencing response through the synthesis of secondary siRNAs (6,7).

Different small RNA pathways often coexist in a single organism, involving distinct types of sRNAs. In plants, animals and fungi, miRNAs and endogenous siRNAs regulate the expression of somatic and germline-specific genes, and specific endogenous siRNAs control transposons in somatic cells (2,3). Various organisms can also respond to environmental dsRNA, for instance some nematode and insect species that process dsRNA ingested within food bacteria or entering cells by endocytosis (8). Exogenously induced RNAi pathways were shown in several species to mediate antiviral defense (9–12). Despite the diversity of trigger molecules and sRNA responses, the different pathways involved can be interconnected by sharing some protein factors (3).

In ciliates, a phylum characterized by nuclear dimorphism, a meiosis-specific class of small RNAs (scnRNAs) targets the elimination of specific sequences during development of the somatic macronucleus (MAC) from the germline micronucleus (MIC) (13–17). As a result, the

[†]To whom correspondence should be addressed. Tel: +49 681 302 4938; Fax: +49 681 302 2703; Email: simone.marker@uni-saarland.de

*Present affiliations:

Saarland University, Centre for Human and Molecular Biology, Molecular Cell Dynamics, Campus A2 4, 66123 Saarbrücken, Germany.

© The Author(s) 2014. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

MAC genome is cleared of transposable elements and derived single-copy internal eliminated sequences (IESs) (18,19). In addition, siRNA pathways are active throughout the life cycle, as shown in *Tetrahymena thermophila* (20,21) and *Paramecium tetraurelia* (15,22). In *P. tetraurelia*, experimental induction of gene silencing revealed that different trigger molecules recruit different RNAi factors and produce different classes of siRNAs. It was first discovered that transformation of the MAC with high-copy transgenes lacking the 3' UTR results in aberrant transcripts on both strands (22,23) and silencing of homologous endogenous genes (23,24). Transgene-induced silencing requires the Dicer protein Dcr1 (15), the Piwi proteins Ptiwi13 and Ptiwi14 (25), and the RdRP Rdr3. Although Rdr3 appears to be catalytically inactive, it was implicated in the regulation of surface antigen gene expression and in the accumulation of endogenous siRNAs of unknown function (22). Transgene-induced siRNAs are 5' monophosphate and modified at the 3' end, presumably by 2'-O-methylation (22).

A second pathway, which allows *Paramecium* to process exogenous dsRNA, was found to involve a distinct but overlapping set of RNAi factors. Feeding cells with *Escherichia coli* bacteria engineered to produce dsRNA, equivalent to the 'feeding' technique in *Caenorhabditis elegans* (26,27), induces silencing of endogenous genes homologous to the dsRNA (28). In this pathway, Dcr1 (15) and the two RdRPs Rdr1 and Rdr2 (22) are required for accumulation of two types of siRNAs: primary siRNAs, predominantly 23 nt in length and produced from the original dsRNA trigger, and secondary siRNAs, presumably produced from the target mRNA (15,22). In contrast to the primer-independent, 5' triphosphate secondary siRNAs produced by direct RdRP activity in *C. elegans* (29–31), *P. tetraurelia* dsRNA-induced siRNAs are 5' monophosphate. Three Piwi proteins (Ptiwi13, Ptiwi12 and Ptiwi15) appear to be involved in this pathway (25).

The MIC genome of *P. tetraurelia* has undergone at least three whole genome duplications (WGDs), resulting in ~40,000 genes in the MAC genome, which is responsible for all gene expression. 68% of genes still retain their duplicates from the most recent WGD (WGD1), and the conservation of protein sequences suggests that most duplicate pairs are functionally redundant (19). WGDs may explain in part the large number of genes encoding core RNAi factors: eight Dicer- and Dicer-like genes (including one WGD1 duplicate pair), 17 Piwi genes (two unpublished) (six WGD1 pairs) and four RdRP genes (no WGD1 duplicate). Only a few of these genes have been assigned any function, making *P. tetraurelia* a very interesting model to study the genetic and mechanistic complexity of small RNA-mediated pathways. Due to the lack of an established method to transform the

MIC genome, however, reverse genetics approaches have been restricted to RNAi thus far, providing only limited possibilities for the functional analysis of genes involved in RNAi. Targeting a gene involved in its own silencing (recursive RNAi (32)) is a self-defeating process which cannot be completed, and is not conclusive when no effect is observed.

In this study, we describe the outcome of a forward genetic screen for mutants deficient in dsRNA-induced silencing, based on lethal dsRNA 'feeding', which was expected

to yield loss-of-function alleles for genes that do not have functionally redundant paralogs. A combined strategy, including the development of a powerful method to identify mutations by whole-genome resequencing, allowed us to characterize 71 RNAi-deficient mutants and yielded a total of 47 alleles for three known and two previously unknown genes. Analyses of allelic diversity revealed a non-essential dsRNA processing machinery but provided evidence for an essential RNAi function during the life cycle, and suggested that the screen is saturated for non-essential, single-copy genes. The dsRNA- and transgene-induced RNAi pathways are further shown to share essential protein factors.

MATERIALS AND METHODS

Paramecium strains, cultivation and genetic analyses

Mutagenesis and other experiments were carried out with wild-type strain 51 of *P. tetraurelia*, mating type E. RNAi mutants were back-crossed either with strain 51 or with strains carrying the *nd7-1* mutation (33) as a genetic marker (after three rounds of back-cross into strain 51). Unless otherwise specified, cells were grown at 27°C in wheat grass powder (Pines International Co., Lawrence, KS, USA) infusion // // // // medium bacterized with *Klebsiella pneumoniae* the day before use and supplemented with 0.8 mg/ml sitosterol. Genetic analyses of mutants were carried out according to standard procedures (34). F1 phenotypes were recorded for each of the two karyonides resulting from the two exconjugants of at least two conjugating pairs, and at least 30 F2 clones resulting from autogamy of F1 heterozygotes were studied.

Random mutagenesis and screening for RNAi-deficient mutants

UV mutagenesis (254 nm) was carried out on several independent batches of 200,000 pre-autogamous cells, according to Cohen (35), with a modified dose of 650 J/m². Cells were allowed to undergo two divisions before autogamy was triggered by starvation to make MIC mutations homozygous in the MICs and MACs of progeny. Three batches in which autogamy reached >90% of the cells, and showing a post-autogamous survival rate of 75–80%, were selected for screening. To isolate mutants deficient in dsRNA-induced RNAi, post-autogamous cells were first fed with three volumes of *Klebsiella* medium (~1–2 vegetative divisions) and starved again to eliminate fragments of the old MAC (monitored by DAPI staining), which would otherwise complement mutations in the new MAC. Cells were then washed and transferred to three volumes of NSF "feeding medium" (*E. coli* producing NSF dsRNA, see below) and grown until mass lethality (~36 hours, four divisions). Surviving cells were isolated and grown in standard *K. pneumoniae*-bacterized medium. Each clone was replicated for storage and further experiments. RNAi deficiency was verified by feeding each clone with dsRNAs from the unrelated genes *ND7* and *ND169* (33), both required for trichocyst discharge, a quantitative phenotype that allows to distinguish full and partial RNAi deficiencies (see below).

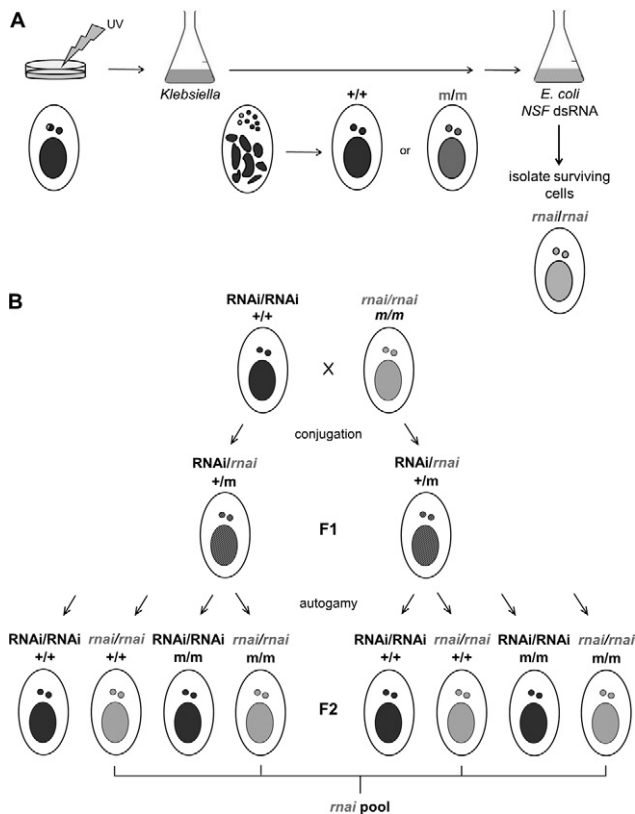


Figure 1. Mutagenesis screen and identification of unknown RNAi factors. (A) UV irradiated cell cultures were starved to induce autogamy, a self-fertilization process that leads to complete homozygosity of the MIC genome. Re-feeding cells with *Klebsiella pneumoniae* medium allowed development of the new MACs from the resulting homozygous zygote within two cell divisions, leading to expression of the mutated genes. Transfer of cells into *NSF* (N-ethylmaleimide-sensitive factor) dsRNA feeding medium allowed isolation of potential *rnaI* mutants, as silencing of the essential *NSF* gene leads to lethality only in wild type cells (28,40,41). (B) To identify mutations in unknown RNAi factors, silencing deficient mutants were crossed with the wild type, resulting in heterozygous F1 progeny. After autogamy, F2 clones become homozygous for the *rnaI*- (grey) or wild type (black) allele. Most other (UV-induced) mutations (grey) are not linked with the target mutation (*rnaI*-) and segregate independently. F2 clones carrying the *rnaI*- allele were pooled for MAC DNA extraction and whole genome-sequencing. In this data set only the *rnaI*- mutation is covered by 100% of SNP-containing reads, whereas all other mutations are ideally covered by 50% SNP-containing and 50% wild type reads.

Induction of RNAi and phenotypic analyses

Production of dsRNA in *E. coli* strain HT115DE3, feeding to *Paramecium* cells, transgene-induced silencing of *ND169*, and monitoring of trichocyst discharge phenotypes were carried out as described previously (22,36). *Nsf*-depleted cells were checked for viability after 36 hours, the time at which typically all cells were dead or dying in a successful silencing. To reveal partial loss-of-function phenotypes, *ND169* dsRNA producing *E. coli* were typically mixed in 1:5 ratio with *ICL7a* dsRNA producing bacteria.

Plasmid constructs and re-sequencing of known genes

To induce silencing of *NSF* and of *ND169* by dsRNA feeding or by transgene (pTI-) plasmid constructs produced pre-

viously (22,28) (Supplementary Table S1). For silencing of *PDS1* and *CID1*-like genes fragments of the coding region were cloned into the plasmid L4440 (Supplementary Table S1) and dsRNA was synthesized in *E. coli* HT115 DE3 (Supplementary Figure S1). Complementation of mutants was achieved by microinjection of the entire open reading frame and flanking sequences, including the endogenous promoter, cloned into the pUC18 plasmid vector (Supplementary Table S1). For re-sequencing of known genes, coding sequences including flanking regions were amplified by PCR using a proof reading polymerase (Phusion[®] High Fidelity, Thermo Fisher Scientific, Waltham, MA, USA). Products were purified by phenole extraction and PEG precipitation (7.5% PEG 8000 final) and sequenced with marginal and internal primers (oligonucleotide sequences on request).

Purification of MAC DNA, whole-genome sequencing and identification of SNPs

RNAi mutants carrying an unknown mutation were back-crossed to wild type and one karyonide of each exconjugant F1 clone was brought to F2 by autogamy. 30 F2 clones were isolated and tested for RNAi deficiency by feeding of *ND169* or *NSF* dsRNA. Mutant clones showing silencing deficiency were raised to large cultures. Immediately prior to DNA extraction, vegetative mutant cultures of both mating types were fused and MAC DNA was extracted from this pool (~ 3 mio cells) according to the protocol used for DNA purification from developing MACs (18). Briefly, cells were concentrated, washed, supplemented with 3 volumes of lysis buffer (0.25 M sucrose, 10 mM MgCl₂, 10 mM Tris pH 6.8, 0.2% Nonidet P-40) and lysed at 4°C with a Potter-Elvehjem homogenizer, allowing cell- but not MAC lysis. MACs were collected and purified from bacteria and cell debris by ultracentrifugation through a 2.1M sucrose cushion. Recovered MACs were lysed and DNA was extracted as described (37). 5 g of RNaseA-treated genomic DNA were used for library construction. DNA was sequenced by single-end (75 bp read length) or paired-end (50 or 100 bp read length) strategies using Illumina GAI and HiSeq next-generation sequencers (overview of sequence output for each mutant F2 pool see Supplementary Table S2). The reads were mapped to the reference *P. tetraurelia* strain 51 genome (18,19) using BWA (38) with default parameters and the mapping indexed with samtools (39). Custom Perl scripts and the SAMtools mpileup function were used to identify positions where at least 80% of the calls differed from the reference genome, at positions with at least 10-X coverage and less than 300-X coverage. The list of SNPs was then curated to identify candidate mutations for experimental validation.

Small RNA analysis by northern blot

Total RNA extraction and small RNA northern blots were carried out as described (22). As previously, two adjacent 50nt sense oligonucleotide probes matching to the dsRNA region were used to detect *ND169* siRNAs, as most of the produced siRNAs are antisense to the target transcript. Similarly, endogenous cluster22 siRNAs were detected with

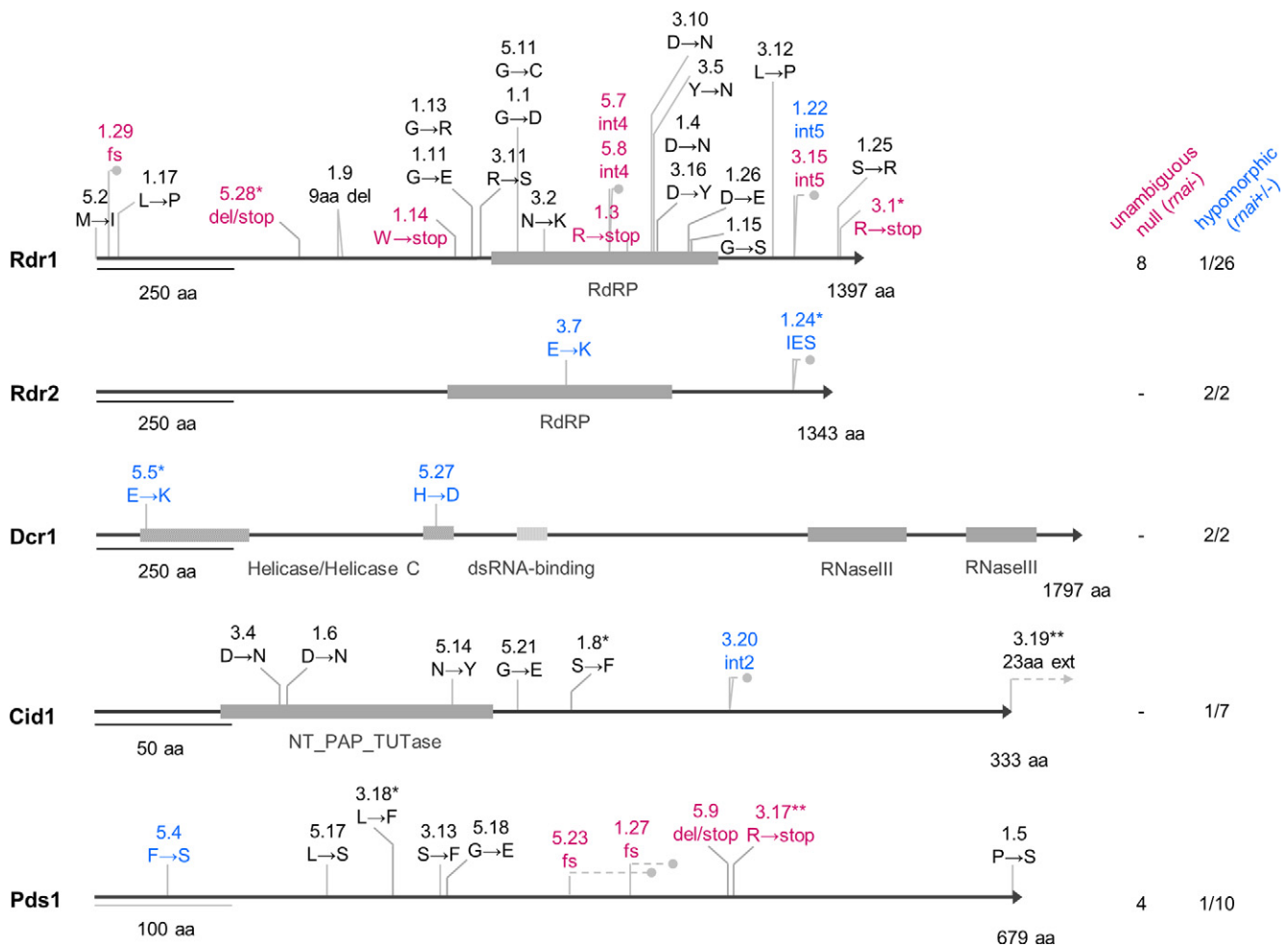


Figure 2. Alleles identified from a screen for mutants deficient in dsRNA-inducible RNAi. Unambiguous null alleles (purple, as defined in the text) were obtained only for *RDR1* and *PDS1*. The location of premature stop codons resulting from frameshifting mutations is shown by grey dots. The 5' TA boundary of the only IES (76 bp) in the *RDR2* gene is mutated to AA in the *rdr2-1.24* allele, leading to complete retention of the IES in the MAC (Supplementary Figure S8B). Hypomorphic alleles are shown in blue. Gene accession numbers are available in supplementary data (Supplementary Table S1). Co-segregation of the phenotype with the mutation in F2 was verified for alleles labelled with an asterisk (*). Alleles complemented by injection of linear DNA encoding the wild-type gene with its natural flanking regions are labelled with two asterisks (**).

two adjacent 50nt oligonucleotide probes, complementary to the top strand (relative to transcription of the 5'-marginal ORF).

RESULTS

Lethal dsRNA feeding allows isolation of Mendelian mutants fully or partially deficient in RNAi

To isolate mutants of the dsRNA-induced silencing pathway, we first treated wild-type cells with UV light to induce random mutations in the MIC genome. Cell populations were then allowed to undergo autogamy, a self-fertilization sexual process in which only one of the parental MIC alleles is retained, and made homozygous in the MICs and MACs of progeny (Figure 1A). Cultures were then screened by feeding them an *E. coli* strain producing dsRNA homologous to *NSF* (N-ethylmaleimide-sensitive factor), an essential gene involved in exocytosis and membrane traffic (28,40,41). *NSF* dsRNA feeding rapidly kills the wild type, so that only mutant cells deficient in dsRNA-induced

RNAi would be able to survive. 150 surviving cells were isolated, and the RNAi-deficient phenotypes were confirmed by dsRNA feeding targeting different genes (*NSF*; *ND7* and *ND169*, two single-copy, unrelated genes involved in the exocytosis of secretory granules called trichocysts (33,42), for details see below, Figure 7B). After sorting out non-viable, false positive or strongly hypomorphic clones (detailed outcome of the screen in Supplementary Table S3), a set of 79 RNAi-deficient cell lines was established.

71 lines were further analyzed. Of these, 64 were fully deficient in RNAi, showing normal growth rates in *NSF* feeding medium and wild-type levels of trichocyst discharge upon *ND7* and *ND169* feeding. 7 lines were only partially compromised in RNAi, suggesting hypomorphic mutations. We reasoned that feeding cells with a smaller amount of dsRNA would be more sensitive to partial defects in the RNAi machinery. Consistently, diluting the silencing trigger (e.g. *ND169* dsRNA bacteria in a 1:5 ratio with bacteria producing dsRNA from *ICL7a*, which encodes a non-essential centrin (24)) revealed stronger phenotypes in hypomorphic

mutants, comparable to the full RNAi deficiency, although this still led to complete silencing in wild-type cells.

Back-crossing selected lines to the wild type confirmed that RNAi phenotypes were due to Mendelian and recessive loss-of-function mutations: F1 heterozygotes showed a wild-type phenotype, and F2 homozygotes obtained by autogamy of the F1s showed the expected 1:1 segregation of wild-type and mutant phenotypes upon dsRNA feeding (Supplementary Table S4A). Although these segregation analyses would not be sufficient to dissociate closely linked mutations, the total number of UV-induced mutations in the genome was later estimated to be ~30 per irradiated cell on average after autogamy (Supplementary Table S5A; mutations see Supplementary Table S5B), indicating that in most cell lines the RNAi deficiency is due to a single mutation.

The mutant screen reveals known and new genes involved in dsRNA-induced RNAi

To identify the mutations in RNAi-deficient lines, we developed a whole-genome sequencing procedure to distinguish the mutation causing the phenotype from other, irrelevant UV-induced mutations, after a single round of back-cross to the wild type (Figure 1B). Briefly, F1 heterozygotes were taken through autogamy, and sets of ~30 homozygous F2 clones were tested for their capacity to respond to dsRNA feeding. Deep sequencing of MAC DNA from pools of ≥15 independent RNAi-deficient F2s will identify the RNAi mutation as being present in 100% of reads, while other, unlinked mutations show up only in 50% of reads on average. This strategy was used in combination with complementation tests and resequencing of genes previously implicated in the pathway to identify candidate mutations in all 71 lines. 45 of them contained mutations in *RDR1*, *RDR2* or *DCR1* (Table 1), confirming the role of these genes in dsRNA-induced RNAi.

The remaining 26 lines were found to carry mutations in either of two new genes, for which one allele was initially identified by the whole-genome sequencing procedure. The missense mutation 1.8 (Figure 2; Supplementary Table S6) is located in the predicted gene PTETG9100013001 (ParameciumDB), encoding a protein of the nucleotidyl transferase family (also referred to as non-canonical polyA/U RNA polymerases (43)) (Figure 3A). It is related to the first described non-canonical polyA polymerase, *Cid1* of *S. pombe*, as well as to *S. pombe*'s *Cid12* (Supplementary Figure S2), the first protein among others of that family shown to be involved in small RNA mediated silencing (44). The gene was therefore named *CID1*. *Cid1* is closely related to *Rdn2* of *T. thermophila* (Supplementary Figure S2; Figure 6A), which physically interacts with the RdRP *Rdr1* in an 'RdRC' complex (45,46). *CID1* is expressed at constant levels throughout the life cycle (Supplementary Figure S3).

Mutation 3.18 lies in a 2-kb single-copy gene on MAC scaffold 6 (PTETG600032001) (Figure 2; Supplementary Table S6), expressed throughout the life cycle (Supplementary Figure S3). Blast and conserved domain searches did not reveal any significant hit, even in other available ciliate genomes (*Tetrahymena thermophila*, *T. borealis*, *T. el-*

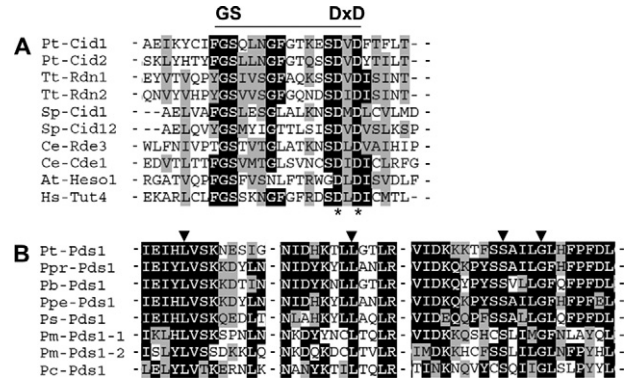


Figure 3. Two newly identified proteins involved in dsRNA-induced RNAi. Protein alignments were performed using the MUSCLE v4 software (77). (A) *Cid1* is a nucleotidyl transferase. Aspartic-acid residues of the predicted catalytic triad (*) required for nucleotidyl transferase activity of *A. thaliana* Heso-1 (51) and *T. thermophila* *Rdn1* and *Rdn2* (46) are mutated in *cid1*-1.6 and *cid1*-3.4. Pt, *Paramecium tetraurelia*; Tt, *Tetrahymena thermophila*; Sp, *Schizosaccharomyces pombe*; Ce, *Caenorhabditis elegans*; At, *Arabidopsis thaliana*; Hs, *Homo sapiens*. (B) *Pds1* is conserved in *Paramecium* species. Examples of missense alleles obtained for Pt-*Pds1* in the RNAi mutant screen are indicated with black arrowheads (substituting amino acids see Supplementary Table S6). Pt, *P. tetraurelia*; Ppr, *P. primaurelia*; Pb, *P. biaurelia*; Ppe, *P. pentaurelia*; Ps, *P. sexaurelia*; Pm, *P. multimicronucleatum*, Pc, *P. caudatum*.

liotti, *T. malaccensis* (Oligohymenophora) and *Oxytricha trifallax* (Hypotricha). In contrast, the gene is present in all sequenced species of the *Paramecium* genus, allowing identification of conserved amino acids (Figure 3A; Supplementary Figure S4). Deduced from its specificity for dsRNA-induced RNAi (see below), we named this gene *PDS1* (*Paramecium* dsRNA-induced RNAi-specific protein 1).

These candidate mutations were shown to co-segregate with the RNAi-deficient phenotype in each of the F2 clones included in the genome-sequencing pools, and microinjection of the cloned wild-type genes into the MAC of mutants (*cid1*-3.19 and *pds1*-3.17) rescued the RNAi deficiency (data not shown). In addition, depletion of these proteins by recursive RNAi experiments (see below, Figure 7D) resulted in an RNAi-deficient phenotype. We conclude that *CID1* and *PDS1* are required for dsRNA-induced silencing.

The RNAi response to dsRNA from food bacteria is not essential for viability

Our screen for RNAi-deficient mutants was unlikely to reveal genes with functionally redundant WGD1 ohnologs, unless disruption of one copy led to a dosage effect. For single-copy genes, we could expect null alleles of non-essential genes and hypomorphic alleles of essential genes. Unambiguous null alleles were here defined as those containing nonsense mutations or frameshifts resulting in premature stop codons, due to indels or mutations in intron splice sites (47), except in cases where the best part of the protein is conserved and where there is experimental evidence for a partial RNAi deficiency. An allele was categorized as hypomorphic if the RNAi deficient phenotype was only partial and/or dsRNA feeding-induced siRNAs were reduced, but still detectable on northern blots. Indeed,

Table 1. Mutants obtained in a screen for dsRNA-induced RNAi deficiency.

	mutants	alleles	function	homologs
<i>RDR1</i>	41	26	RdRP	<i>Ti, At, Ce, Sp</i>
<i>RDR2</i>	2	2	RdRP	<i>Ti, At, Ce, Sp</i>
<i>DCR1</i>	2	2	RNase III Dicer	<i>Ti, At, Ce, Sp, Dm, Mm, Hs</i>
<i>CID1</i> *	16	7	nucleotidyl transferase	<i>Ti, At, Ce, Sp, Dm, Mm, Hs</i>
<i>PDS1</i> *	10	10	unknown	<i>Paramecium</i>

Mutants in six different genes, each represented by several alleles, were obtained. The first alleles to be discovered for the two new genes (*) *CID1* and *PDS1* were identified by whole-genome sequencing of pooled F2 clones (see Figure 1).

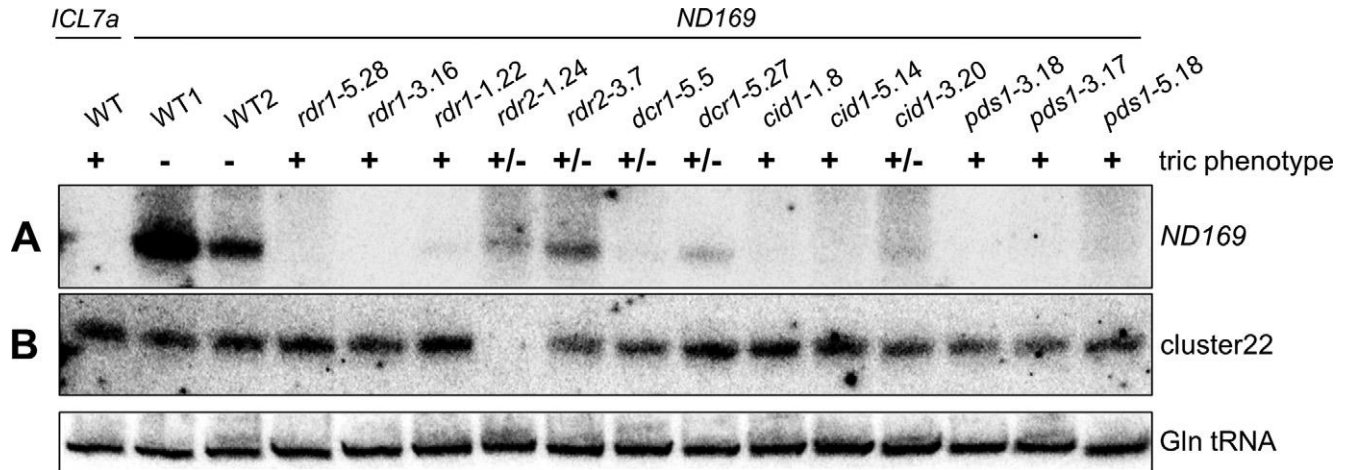


Figure 4. Abolished siRNA accumulation in RNAi mutants. (A) Northern blot analysis of *ND169* siRNAs revealed failure or strong reduction of accumulation of exogenously-induced siRNAs in all mutants, consistent with previous findings from depletion of Dcr1, Rdr1 and Rdr2 by RNAi (15,22,25). This suggests a role of these RNAi factors upstream or within siRNA biosynthesis or stabilization. It is of note that silencing efficiencies can vary from one dsRNA feeding experiment to another, possibly due to contamination by traces of the standard food bacterium *Klebsiella* (showing partial Ampicillin resistance) which overgrows dsRNA-producing *E.coli*. This may explain differences in the total level of dsRNA-induced siRNAs in different wild type cultures. The *ND169* probe corresponds to a 100 bp region of the dsRNA and is sense oriented, as antisense siRNAs represent the predominant fraction of siRNAs in the dsRNA region (22). (B) Detection of endogenous siRNAs from an intergenic region of scaffold 22 (cluster22) revealed that *rdr2* mutants are unable to accumulate these siRNAs. The probe is complementary to the predominant fraction of siRNAs (top strand) of this region. The lower panel shows hybridization to glutamine tRNA as a loading control.

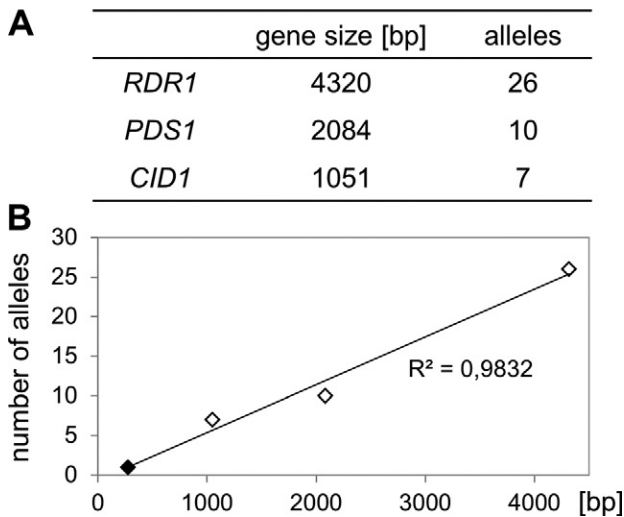


Figure 5. Saturation of the screen for non-essential single copy genes involved in dsRNA-induced RNAi. Gene sizes of non-essential RNAi factors (A) plotted against the number identified mutant alleles obtained (B). Extrapolation by linear regression determined the mutation frequency to 1 in 172 bp of coding sequence.

the diversity of alleles obtained for each of the genes hit appeared to reflect their importance for cellular viability (Figure 2): 41 *rdr1* mutants were found, representing 26 different alleles. Among these were 8 putative null alleles (e.g. *rdr1-5.28* or *rdr1-5.7*). Missense alleles were mostly non-conservative substitutions in conserved residues, two of which in the putative catalytic core region (*rdr1-1.4* D1021N and *rdr1-3.16* D1021Y) (48–50) (Supplementary Table S6). Similarly, the 10 alleles obtained for *PDS1* included null alleles, suggesting that this gene, like *RDR1*, is not required for viability. Although no unambiguous null allele was found among the 7 *cid1* alleles, some of the mutations changed highly conserved amino acids (*cid1-3.4* D68N and *cid1-1.6* D70N) shown to be required for uridylyl transferase activity (46,51). In all putative null mutants tested, a complete loss of dsRNA-induced siRNAs was observed on northern blots (Figure 4A and Supplementary Figure S6). RNAi-deficient *rdr1*, *pds1* and *cid1* null mutants are fully viable throughout the life cycle (vegetative growth and sexual events), and no other phenotypic anomaly was observed. Furthermore, the F2 (Supplementary Table S4B) and F3 generations of an *rdr1-3.1/cid1-1.8* double homozygote showed normal vegetative growth in standard conditions and at high temperature (34°C). We conclude

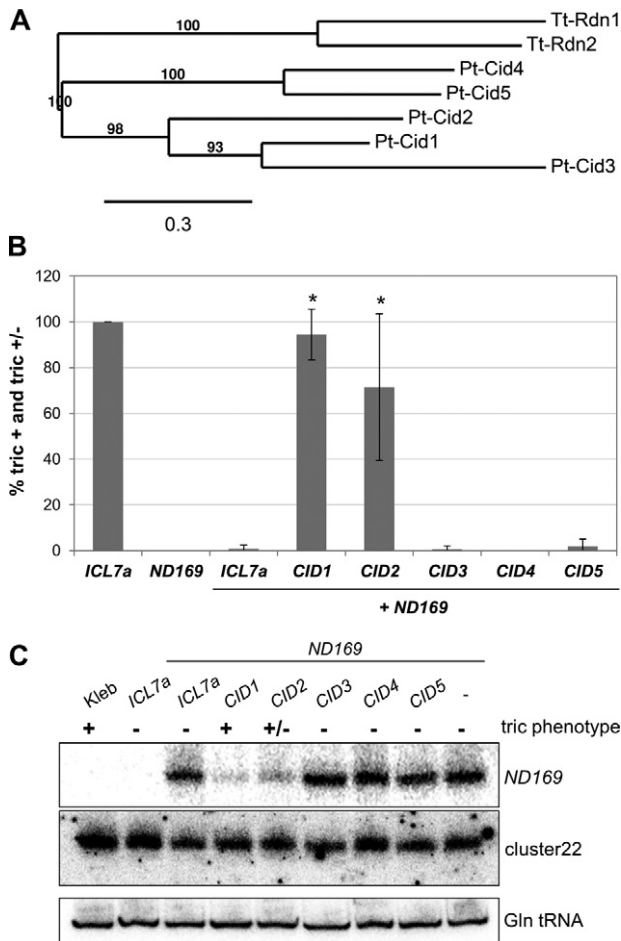


Figure 6. Cid2 is also involved in dsRNA-induced silencing. (A) Phylogenetic relationship of the Cid1-like proteins clustering with Pt-Cid1 (all Cid1-like genes identified in the *P. tetraurelia* MAC genome and generation of phylogenetic trees see Supplementary Figure S2 and Table S7). (B) Double knock down experiment of Pt-Cid1-like genes and *ND169* by dsRNA feeding. *ND169* was used as a reporter for silencing; the gene was considered as silenced when cells showed complete deficiency of trichocyst discharge (tric-) and inhibition of silencing was measured as proportion of cells in the culture showing partial (tric+/-) or complete (tric+) reversion of silencing. Silencing was significantly inhibited upon knockdown of *CID1* and *CID2* (P-values <0.03 (*), Mann-Whitney U test, significance level 0.05; n = 3; standard deviation is shown). Phenotypes were recorded 72 hours after the first feeding. Bacteria were fed in equal amounts. Triple knock down experiments (*CID3*+*CID4*+*ND169*; *CID3*+*CID5*+*ND169*; *CID4*+*CID5*+*ND169*) did not show inhibition of *ND169* reporter silencing (data not shown). Cross-silencing between *CID1* and *CID2* genes is unlikely, as the maximum length of perfect homology is 12nt, and 18nt with one central mismatch. (C) Northern blot analysis of associated *ND169* siRNAs. The *ND169* probe corresponds to a 100 bp region of the dsRNA and is sense oriented. The lower panel shows hybridization to glutamine tRNA as a loading control.

that the capacity to synthesize siRNAs from dsRNA ingested with food is not essential in laboratory conditions.

Nevertheless, the dsRNA-induced RNAi pathway appears to be highly conserved in the *Paramecium* genus, since tblastn searches identified at least one copy of the *RDR1*, *CID1* and *PDS1* genes in all sequenced species of the *P. aurelia* complex, as well as in *P. multimicronucleatum* and *P. caudatum* (Supplementary Figure S5). Essential factors playing a role in this pathway in *P. tetraurelia* (see below)

are also highly conserved (not shown). This suggests that the pathway may fulfill an important function required for long-term survival of these species in their natural environment.

The screen is saturated for non-essential single-copy genes

The large numbers of alleles obtained for 3 non-essential, single-copy genes suggests that the screen is saturated for this category of genes. For *RDR1*, *PDS1* and *CID1*, the number of alleles is roughly proportional to the length of the gene; the relationship indicates that, with the number of mutant lines analyzed here, one mutation was obtained for every 172 bp of coding sequence on average (Figure 5). Thus, there may not be any other non-essential, single-copy gene involved in the dsRNA-induced RNAi pathway.

Essential genes may be involved in dsRNA-induced RNAi

In contrast to *RDR1*, *CID1* and *PDS1*, only two alleles of *RDR2* were identified. *rdr2-3.7* is an amino acid substitution (E859K) within a less conserved region of the RdRP domain, and *rdr2-1.24* retains an intragenic IES (MIC-specific IES) located downstream of the conserved RdRP domain region, due to a T-to-A mutation in one of the TA boundaries that are required for excision during macronuclear development (52) (Supplementary Figure S8B). Both alleles are hypomorphic, as indicated by partial RNAi deficiency and by partial loss of siRNAs upon dsRNA feeding (Figure 4A and Supplementary Figure S6). The greater reduction in siRNA levels in *rdr2-1.24* was paralleled by a weaker silencing of the *ND169* target using dilute dsRNA trigger (not shown). Similarly, for *DCR1* only two hypomorphic missense alleles were identified. Interestingly, the mutations are located within the N-terminal helicase (*dcr1-5.5* E87K) or helicase C domains (*dcr1-5.27* H622D). *Dcr1-5.5* shows a stronger RNAi deficiency than *dcr1-5.27*, judging from both phenotypes and siRNA levels (Figure 4A and Supplementary Figure S6). These findings suggest that the helicase and helicase C domains play a role in the biosynthesis of dsRNA-derived siRNAs, unlike the helicase domain of *C. elegans*' Dcr-1, which is only required for a subset of endo-siRNAs (53). The fact that *RDR2* and *DCR1* are only represented in the mutant collection by a small number of hypomorphic alleles suggests that they are essential genes for which null alleles would lead to lethality. Thus, RNAi may fulfill essential functions in *P. tetraurelia*.

Another, putatively essential nucleotidyl transferase is involved in dsRNA-induced silencing

Whereas no other *PDS1*-related gene could be identified in the *P. tetraurelia* genome, 22 Cid1-like non-canonical poly(A)-polymerases were found (Supplementary Table S7). According to phylogenetic analyses (Supplementary Figure S2), 5 of them, including Cid1, fall in a common branch with Rdn1 and Rdn2 of *T. thermophila* (Figure 6A). None of them is an ohnolog to another. To analyze their function in dsRNA-induced RNAi, we used a recursive RNAi approach (22), i.e. we silenced each of them by dsRNA feeding, simultaneously with *ND169* as a reporter.

Apart from *CID1*, targeting *CID2* also prevented silencing of *ND169* by dsRNA feeding (Figure 6B). This was correlated with a reduced accumulation of *ND169* siRNAs (Figure 6C). It is of note that silencing of a gene involved in its own silencing has only limited efficiency, explaining the partial impairment of *ND169* silencing and siRNA accumulation. The results indicate that, like *CID1*, *CID2* is involved in dsRNA-induced RNAi. Although it is a single-copy gene, our screen did not yield a mutant allele, raising the hypothesis that *CID2* is essential for viability.

Putatively essential RNAi factors are shared with the transgene-induced silencing pathway

Although transgene-induced siRNAs are biochemically different from dsRNA-induced siRNAs, Dcr1 and Ptiwi13 are shared between these two pathways. Like *DCR1*, *PTIW13* is possibly an essential gene, as mutants were not obtained in our screen, probably due to inefficient recovery of hypomorphic mutants. Rdr3, a factor specifically involved in transgene-induced silencing, also seems to be essential, as its silencing results in a reduced vegetative growth rate (22). To test whether the newly identified genes are also involved in transgene-induced silencing, we induced silencing of *ND169* by injecting a transgene construct (pTI-) truncated at the 3' end of the coding sequence (Figure 7A), as previously described (22). Candidate genes were then silenced in injected clones by dsRNA (recursive RNAi, as described previously (15,22,25)), and the *ND169* silencing phenotype was monitored. To increase the sensitivity of the silencing response to depletion of potentially involved factors, we used injected clones that showed only moderate *ND169* silencing efficiency (*tric+/-*) (Figure 7B). Cells showed a strong decrease in silencing efficiency upon *CID2* silencing and, as expected, upon *RDR3* and *PTIW13* silencing (Figure 7C). Surprisingly, using this experimental setup we were also able to detect a significant reduction of transgene-induced silencing efficiency when *RDR2* was co-silenced, which had not been observed previously on strongly silenced clones using the same *RDR2* silencing construct (22). In contrast, targeting *CID1*, *PDS1* or *RDR1* did not result in reduced silencing of the *ND169* transgene, although dsRNA-induced RNAi was inhibited in control experiments (Figure 7D). This suggests that these genes are not involved in the transgene-induced RNAi pathway. The involvement of the putative essential gene *DCR1* in silencing induced by the *ND169* transgene construct (as well as *ND169* dsRNA) could also be confirmed (not shown), as previously shown for a promoter-less transgene (15). We conclude that *RDR1*, *CID1* and *PDS1* are specific for the dsRNA-induced pathway, while genes hypothesized to be essential according to the above criteria (*RDR2*, *CID2*, *DCR1*, *PTIW13*) are involved in both dsRNA- and transgene-induced RNAi pathways.

Rdr2 mutants are compromised in endogenous siRNA accumulation and show reduced viability after sexual reproduction

Factors of the transgene-induced RNAi pathway may have essential functions in any phase of the life cycle, including vegetative growth, sexual events, and development. We

checked whether mutants are impaired in the accumulation of a cluster of endogenous siRNAs of unknown function (cluster22) which are produced throughout the life cycle from an intergenic locus (15). Northern blot analysis revealed that their accumulation was completely suppressed in the *rdr2-1.24* mutant (Figure 4B). Signal quantification (Supplementary Figure S6) showed only a slight, if any, reduction for the *rdr2-3.7* allele, consistent with its strongly hypomorphic nature. Tested mutations in *RDR1*, *CID1* and *PDS1* had little or no effect. Surprisingly, *dcr1* mutants showed only slight (if any) reduction in cluster22 siRNA levels as well, as also observed after targeting *DCR1* by dsRNA-induced RNAi (not shown). Similarly, targeting *CID2* did not result in any detectable reduction (Figure 6C), though this may be due to inefficient depletion by recursive RNAi. Despite the molecular phenotype evidenced here, the *rdr2* mutants did not differ from the wild type in viability or growth rate during vegetative growth.

In contrast, *rdr2* mutants showed a range of defects during sexual reproduction. Survival of homozygotes for the stronger allele *rdr2-1.24* was found to be affected by a long starvation period (≥ 3 days), but only when this immediately followed autogamy, and not when the same starvation was experienced by vegetative clones that had already undergone 7–8 divisions (Supplementary Figure S7). New MAC development nevertheless appeared to be cytologically normal, as judged from DAPI staining, and the reason for this post-autogamous defect remains unknown.

Conjugation was also impaired in both *rdr2* mutants. After a cross to the wild type, the F1 clones derived from one or both parents were unviable for 35% (18/51) of conjugating pairs and the frequency of proper formation of heterozygous new MACs seemed reduced (Supplementary Figure S8A, Table S4A). The most striking effect was that autogamy of the viable, heterozygous F1s most often (9/10; $p = 0.01$) gave rise to F2 progeny in which close to 100% of clones were unviable or showed regeneration of the F1 MAC (Supplementary Figure S8, Table S4A). This was observed for F1s deriving from both parents and suggests a dominant-negative maternal effect of the heterozygous F1 MAC with incomplete penetrance, because the rare cases where the F2 progeny survived gave the expected 1:1 segregation of *rdr2* and wild-type F2 homozygotes, and both types of clones survived equally well. The reason why post-autogamous lethality is much stronger for heterozygotes than for *rdr2* homozygotes, and the nature of the cellular process affected, remain unclear. Testing excision of some IESs in post-autogamous, newly developed MACs (*rdr2-1.24*) did not reveal any DNA rearrangement defect for maternally controlled (scnRNA-dependent) or non-maternally controlled IESs (data not shown). Nevertheless, our observations indicate that *RDR2* has essential functions during sexual reproduction. A similar phenotype was not observed in any other RNAi mutant.

DISCUSSION

A whole-genome sequencing method to identify mutations potentiates a powerful screen

In this study, we have used a powerful and straightforward mutagenesis screen to identify genes required for the RNAi

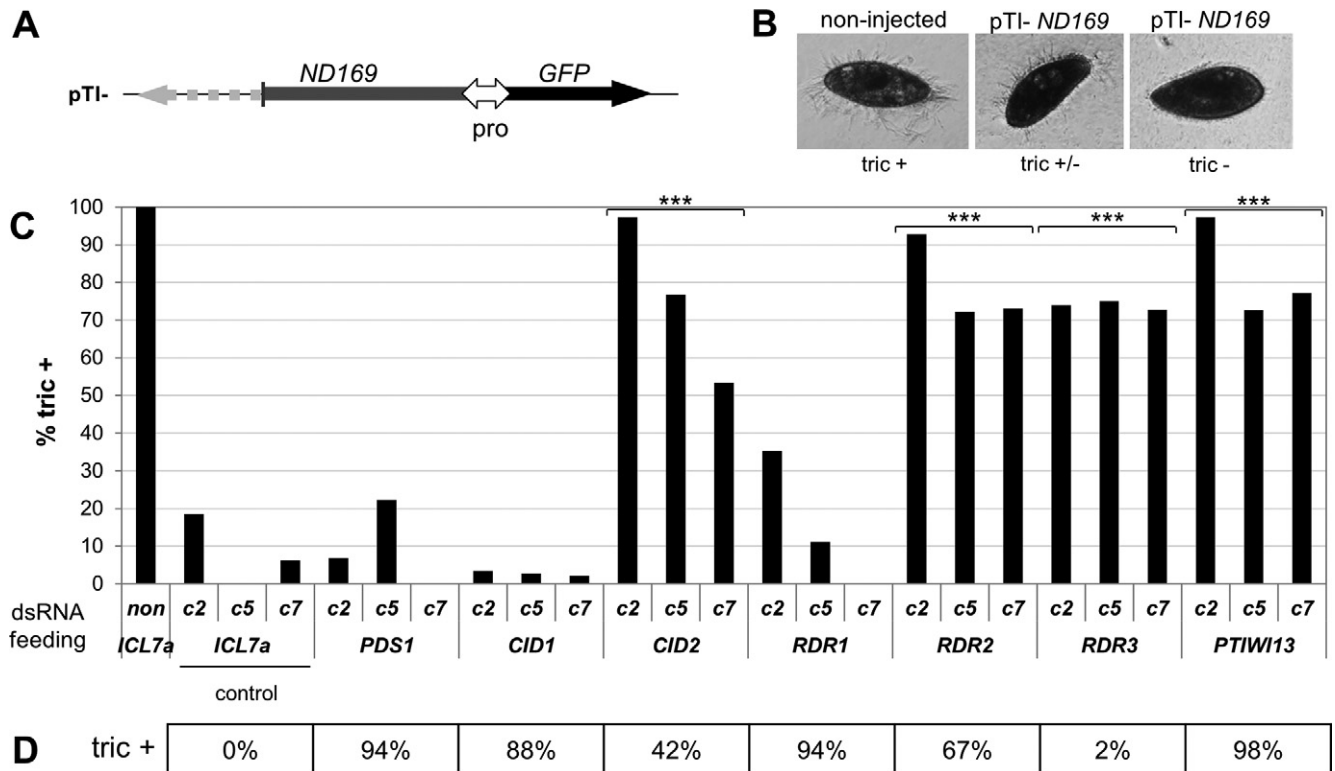


Figure 7. dsRNA-induced RNAi factors sharing functions with the transgene-induced silencing pathway. (A) *ND169* transgene construct pTI- carrying a 3' truncated version of the *ND169* coding sequence driven by a bidirectional constitutive promoter. GFP fluorescence was monitored as control for injection and expression (not shown). (B) *Nd169* is involved in the exocytosis of secretory granules (trichocysts) docked at the plasma membrane. Wild-type cells expell trichocysts upon treatment with picric acid (tric+). By microinjection of pTI- into the MAC strongly silenced clones showing complete trichocyst discharge defects were obtained (tric -) (not shown), as well as three clones with moderate silencing effect (tric +/-), c2, c5, and c7. Note that clone c2 showed up to 18% of cells with tric+ phenotypes in control cultures (*ICL7a* feeding). (C) Knock down of RNAi factor genes by dsRNA feeding in clones with moderate transgene-induced silencing phenotypes. De-repression of *ND169* silencing was measured as the percentage of cells showing a complete wild type phenotype (tric +). A significant difference to the *ICL7a* control was detected upon silencing of *CID2*, *RDR2*, *RDR3* and *PTIW13* (P-values <0.00001 (***), one-way ANOVA, significance level 0.005). Phenotypes recorded 120 hours after the first feeding are shown. Note that pTI- induced silencing efficiencies varied slightly in different cultures of the same injected clone. (D) The effect of RNAi factor depletion on dsRNA-induced silencing was verified in parallel as a positive control by double knock down, mixing RNAi factor and *ND169* dsRNA-producing bacteria in equal amounts. The percentage of tric+ cells, i.e. the degree of inhibition of *ND169* silencing, is shown (mean of two replicates for *CID1* and *PDS1*). *ICL7a* control dsRNA feeding does neither inhibit dsRNA- nor transgene-induced silencing; *PTIW13* is involved in both pathways, whereas *RDR3* is specific for transgene- and *RDR1* for dsRNA-induced silencing, as previously described (22,25).

response to dsRNA feeding in *P. tetraurelia*: dsRNA homologous to the essential gene *NSF* efficiently kills the wild type, and can be fed to large populations of cells. Mutational studies have long benefited from the easy genetics and short generation time of this organism (54–58), but the identification of mutations relied on functional complementation and sib-selection sorting of an indexed library (59,60), a time-consuming method ill-suited for large mutant collections. We have designed an efficient procedure to identify phenotype-causing mutations after a single round of back-cross to the wild type. Other, unlinked mutations segregate independently among F2 homozygotes, so that whole-genome sequencing of pools of phenotypically mutant and wild-type F2s will show the RNAi mutation in 100% and 0% of reads, respectively, while most irrelevant mutations should be present in 50% of reads from both pools. In practice, sequencing of the wild-type pool was found to be unnecessary; 15-X coverage of pools of ~20 mutant clones, with 75-bp single reads, was often sufficient to narrow down the list of candidate mutations to a single

one, although 100-bp paired-end sequencing proved more efficient. Distinguishing closely linked mutations would require more F2s and a higher coverage, but the standard mutagenesis conditions used (~20% lethality in homozygous progeny of irradiated cells) were shown to result in an average of ~30 mutations in the 72-Mb MAC genome of viable progeny, making cases of tight linkage infrequent. Similar approaches have been used in *C. elegans* and *A. thaliana* (61,62), but here the possibility to generate large numbers of entirely homozygous recombinant F2s by autogamy makes the method faster and simpler, dispensing with the need for extensive back-crossing or pre-mapping.

Mutations causing full or partial RNAi deficiency identify known and new genes

Mutations were identified in 71 RNAi-deficient cell lines, yielding a total of 47 different alleles for only 5 genes (Figure 8). Three of them (*RDR1*, *RDR2* and *DCR1*) were previously implicated in dsRNA-induced RNAi on the basis of recursive RNAi experiments. The 2 new genes included the

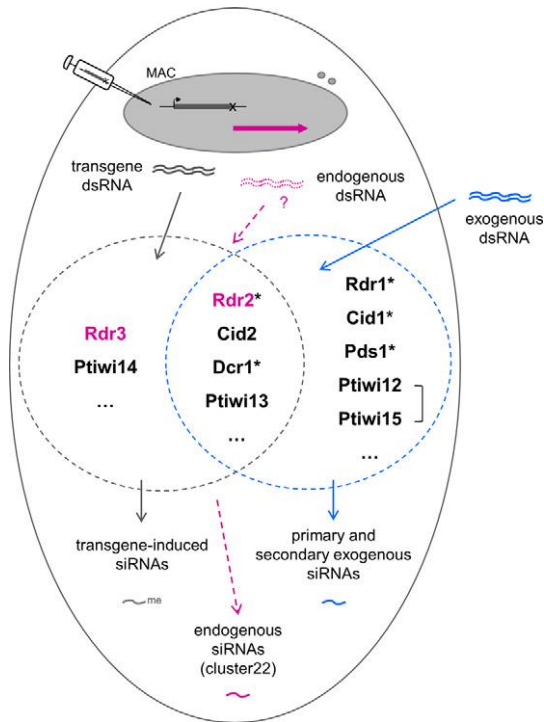


Figure 8. Sharing of RNAi core factors by non-essential exogenously and essential endogenously triggered RNAi pathways in *P. tetraurelia*. In this study, mutants were obtained for genes labelled with an asterisk, and genes shared between the transgene-induced (dark grey) and the dsRNA-induced pathway (blue) were found to be essential. Brackets connecting two genes symbolize WGD1 duplicate genes. *CID2* was identified by recursive RNAi in this study. Other genes were identified previously (15,22,25). The *RDR2* gene is required for the accumulation of endogenous siRNAs (purple) of the cluster22 locus, as previously found for *RDR3* (22). Note that production of dsRNA from this locus is speculative, and that other endogenous siRNAs may require different factors. The figure is only meant to depict genetic requirements, not the subcellular localization or mechanistic roles of the corresponding proteins.

nucleotidyl transferase *CID1* and the *Paramecium*-specific gene *PDS1*. Among the 47 alleles, 14 were isolated more than once; in 3 cases the same alleles were recovered from different mutagenized cultures and thus originated from different mutation events. This suggests that our analysis provides a rather exhaustive view of the genes and alleles that can be identified by the experimental scheme used.

The latter allowed recovery of some hypomorphic alleles (7 out of 47) causing only partial RNAi deficiencies, as confirmed by functional tests using dilute dsRNA trigger. The rather low frequency of such alleles may be due to the ~7 vegetative divisions allowed for fully resistant mutants in *NSF* dsRNA feeding medium, before cell isolation: although hypomorphic mutants were able to survive that step, their growth rate was probably greatly reduced. The experimental scheme would thus have to be modified to yield a higher frequency of hypomorphic alleles, the only ones that can be obtained for essential genes.

The dsRNA- and transgene-induced RNAi pathways share essential genes

Analysis of the diversity of alleles obtained for each of these 5 genes showed that they fall in two groups. For *RDR1*, *CID1*, and *PDS1*, the many alleles recovered included unambiguous null ones such as nonsense mutations and/or non-conservative changes of conserved residues, which all abolished accumulation of siRNAs upon dsRNA feeding. Although these mutants were completely deficient in the silencing response to dsRNA, they did not show any other phenotype at any stage of the life cycle, indicating that the capacity to synthesize siRNAs from dsRNA ingested as food is not required for short-term viability in laboratory conditions. In contrast, only two alleles were obtained for each of *DCR1* and *RDR2* and these were hypomorphic, only partially impairing the silencing response and siRNA accumulation. This strongly suggests that these are essential genes in which null mutations would be lethal at some stage of the life cycle. In particular, these results establish a clear distinction between *RDR1* and *RDR2*, two RdRPs which could not be distinguished in previous recursive RNAi tests applied during the vegetative phase (22).

Strikingly, these two groups of genes also differ in the specificity of their involvement in RNAi pathways (Figure 8). We confirmed that *DCR1* is also required for transgene-induced silencing and showed that the same is true of *RDR2*, although this had not been seen previously. Here, recursive RNAi tests (by dsRNA feeding) were made more sensitive by using a transformed clone showing only moderate silencing of the *NDI69* reporter (as proposed in a model for recursive RNAi dynamics (32)). In the same conditions, recursive silencing of *RDR1*, *CID1*, and *PDS1* did not affect transgene-induced silencing. Thus, the only essential genes uncovered by our screen are those also involved in transgene-induced RNAi. Two more genes likely belong to this group. By testing close paralogs of *CID1*, we found that *CID2* is involved in both pathways, as previously shown for *PTIWI13*. The fact that no mutation was recovered in these genes is consistent with the idea that they are also essential.

A third, presumably non-catalytic RdRP gene previously implicated in transgene-induced silencing, *RDR3*, is further required for normal vegetative growth rate and accumulation of cluster22 endogenous siRNAs. The latter differ from transgene-induced siRNAs in that they lack a modification of the terminal ribose, and interestingly, were also absent or reduced in the hypomorphic *rdr2* mutants. *RDR2* thus appears to be required for all known types of siRNAs in *Paramecium* (Figure 8). However, the wild-type growth rate of the mutants suggests that its putative essential function may not be needed during the vegetative phase.

In contrast to *rdr2*, no effect was observed in *dcr1* mutants or *CID134* knockdown cells on cluster22 siRNAs. Although we cannot completely exclude a false negative result due to the hypomorphic nature of the *dcr1* alleles and the inefficient knockdown of *CID134*, this may indicate that these endogenous siRNAs are not dependent on all components of the transgene-induced pathway and are thus produced by a different pathway (Figure 8). This is supported by the observation that they do not carry a 3' modification.

RDR2 is involved in sexual reproduction

The crosses of *rdr2* mutants to the wild type revealed defects in sexual reproduction. F1s showed a high lethality rate and occasional failure of the formation of heterozygous MACs. However, it is unclear whether this was significant in F1s deriving from the wild-type parent (which would point to defects in pre-zygotic stages such as karyogamy). More striking was the frequent unviability of most or all F2 clones produced by autogamy of some – though not all – viable heterozygotes, regardless of F2 genotype or F1 parental origin. The stochastic nature of the effect raises the possibility that it is due to defects in maintenance of the germline in mutant homozygotes, such as defects in chromosome segregation during MIC mitosis, which would have no impact on vegetative growth of mutants, but would lead to aneuploidy in the F1 and therefore lethality in the F2. Such defects have been observed for instance in RNAi mutants in *S. pombe*, including the RdRP Rdp1 (63,64) and the nucleotidyl transferase Cid12 (44,65), which are both essential for pericentromeric heterochromatin formation (44,66). Stochastic loss of MIC chromosomes during vegetative growth would make each mutant clone potentially unique in its ability to foster viable F2 progeny.

The genetic complexity of the dsRNA-induced RNAi pathway

Previous studies have probed the genetic complexity of RNAi pathways (67–71). In *C. elegans*, a genome-wide RNAi screen for genes involved in silencing induced by RNA hairpin expression identified 90 genes, 54 of which are essential for viability (68). In our study, the screen seems to be saturated only for the sub-category of non-essential, single-copy genes, all 3 of which are specifically involved in dsRNA-induced RNAi. Additional non-essential genes are likely to have been missed because they retain their WGD1 duplicates, which very often have redundant functions. Indeed, 68% of all *P. tetraurelia* genes still have their WGD1 duplicate (19). Such may be the case of *PTIWI12/PTIWI15*, previously shown to be involved in dsRNA-induced silencing but for which co-silencing of both duplicates was required to detect an effect (25). Other genes may have been missed because they are essential, since our scheme yielded only a low frequency of hypomorphic alleles. Essential genes may be shared with the transgene-induced pathway; in addition, processing of dsRNA from food bacteria could be coupled to other cellular processes such as uptake of nutrients, as shown in *D. melanogaster* S2 cells and *C. elegans* (72,73). Thus, the total number of genes contributing to dsRNA-induced RNAi might be much higher than the 9 known so far.

Environmental RNAi in Paramecium

Environmental RNAi, *i.e.* the processing of exogenous dsRNA into functional siRNAs, has been observed in a number of organisms (8), but has not yet been reported to be essential for cellular viability. It is also dispensable for short-term viability of *P. tetraurelia* in laboratory conditions. However, the genes involved are highly conserved in the *Paramecium* genus, suggesting they may have important functions for the long-term survival of these species

in their natural environment. Exogenous RNAi may serve as an antiviral defense, a major function of the pathway in *D. melanogaster*, *C. elegans* and other species (9–12), even though no *Paramecium* virus has been identified so far. If most of the genes in the pathway are conserved for the processing of viral RNA, only a few more may be required for the uptake of dsRNA from phagosomes, and these are not necessarily conserved. Such a situation is seen with some *Caenorhabditis* species that lack the dsRNA uptake mechanism, but respond to exogenous dsRNA when the *C. elegans* dsRNA transporter SID-2 is expressed (74). The identification of novel genes and the availability of mutants will help investigations into the mechanism and natural functions of dsRNA uptake.

Biosynthesis of dsRNA triggers and siRNAs

Although it is still unclear how RdRPs are specifically recruited to their target template during biogenesis or maintenance of silencing triggers *in vivo*, recent studies in *S. pombe*, *C. elegans* and *T. thermophila* point to the formation of RdRP complexes (RdRC) as an important step (44,46,75,76). In the ciliate *T. thermophila*, a single RdRP (Rdr1) interacts in a mutually exclusive manner with either of two nucleotidyl transferases, forming an essential RdRC with Rdn1, or a non-essential one with Rdn2. The latter contains an associated factor, either Rdf1 or Rdf2 (46,76). The role of these specific RdRCs *in vivo* is still unclear (21,46). In *P. tetraurelia*, it is tempting to speculate that Rdr1 and Cid1 form a non-essential RdRC which specifically acts in dsRNA-induced silencing, while Rdr2 and Cid2 form an essential RdRC acting in both pathways (Rdf1 and Rdf2 are not conserved in the *P. tetraurelia* genome). The reason why two RdRCs should be required for dsRNA-induced silencing, however, will remain puzzling until their precise functions are identified. They may have distinct roles in the production of primary and secondary siRNAs (15,22). The present work opens the way for an analysis of siRNAs and their precursors in each mutant, which is likely to provide insight into the processing mechanisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online, including: Supplementary Figures S1–S8 and Supplementary Tables S1 and S7.

ACKNOWLEDGMENT

We thank Jean Cohen for providing plasmid constructs and advice in UV-induced mutagenesis, Michael Lynch, Thomas G. Doak and Casey McGrath for sharing unpublished sequence data of *Paramecium* species (funded by NSF EF-0328516-A006), Linda Sperling, Martin Simon and Vincent Récamier for comments on the manuscript, and all lab members for critical discussions and continuous support. The sequencing of RNAi mutant MAC genomes benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr).

FUNDING

This work was supported by the Agence Nationale de la Recherche [“Investissements d’Avenir” program ANR-10-LABX-54 MEMO LIFE/ANR-11-IDEX-0001-02 Paris Sciences et Lettres* Research University, grant ANR-08-BLAN-0233 “ParaDice” and ANR-12-BSV6-0017 “IN-FERNO” to E.M.]; the Fondation pour la Recherche Médicale [“Equipe FRM” grant to E.M.] and the Fondation Pierre Gilles de Gennes [post-doctoral fellowship to S.M.]. This study was carried out in the context of the CNRS-supported European Research Group “Paramecium Genome Dynamics and Evolution” and the European COST Action BM1102.

REFERENCES

1. Cerutti,H. and Casas-Mollano,J.A. (2006) On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.*, **50**, 81–99.
2. Ghildiyal,M. and Zamore,P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.*, **10**, 94–108.
3. Ketting,R.F. (2011) The many faces of RNAi. *Dev Cell*, **20**, 148–161.
4. Meister,G. and Tuschl,T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*, **431**, 343–349.
5. Tomari,Y. and Zamore,P.D. (2005) Perspective: machines for RNAi. *Genes Dev.*, **19**, 517–529.
6. Baulcombe,D.C. (2007) Molecular biology. Amplified silencing. *Science*, **315**, 199–200.
7. Voinnet,O. (2008) Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci.*, **13**, 317–328.
8. Whangbo,J.S. and Hunter,C.P. (2008) Environmental RNA interference. *Trends Genet.*, **24**, 297–305.
9. Félix,M.A., Ashe,A., Piffaretti,J., Wu,G., Nuez,I., Belicard,T., Jiang,Y., Zhao,G., Franz,C.J., Goldstein,L.D. *et al.* (2011) Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLoS Biol.*, **9**, e1000586.
10. Lu,R., Maduro,M., Li,F., Li,H.W., Broitman-Maduro,G., Li,W.X. and Ding,S.W. (2005) Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*. *Nature*, **436**, 1040–1043.
11. Wilkins,C., Dishongh,R., Moore,S.C., Whitt,M.A., Chow,M. and Machaca,K. (2005) RNA interference is an antiviral defence mechanism in *Caenorhabditis elegans*. *Nature*, **436**, 1044–1047.
12. Wang,X.-H., Aliyari,R., Li,W.-X., Li,H.-W., Kim,K., Carthew,R., Atkinson,P. and Ding,S.-W. (2006) RNA Interference Directs Innate Immunity Against Viruses in Adult *Drosophila*. *Science*, **312**, 452–454.
13. Mochizuki,K., Fine,N.A., Fujisawa,T. and Gorovsky,M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*, **110**, 689–699.
14. Lepère,G., Betermier,M., Meyer,E. and Duharcourt,S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.*, **22**, 1501–1512.
15. Lepère,G., Nowacki,M., Serrano,V., Gout,J.F., Guglielmi,G., Duharcourt,S. and Meyer,E. (2009) Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.*, **37**, 903–915.
16. Mochizuki,K. and Gorovsky,M.A. (2004) Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.*, **18**, 2068–2073.
17. Garnier,O., Serrano,V., Duharcourt,S. and Meyer,E. (2004) RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **24**, 7370–7379.
18. Arnaiz,O., Mathy,N., Baudry,C., Malinsky,S., Aury,J.M., Wilkes,C.D., Garnier,O., Labadie,K., Lauderdale,B.E., Le Mouel,A. *et al.* (2012) The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.*, **8**, e1002984.
19. Aury,J.M., Jaillon,O., Duret,L., Noel,B., Jubin,C., Porcel,B.M., Segurens,B., Daubin,V., Anthonard,V., Aiach,N. *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**, 171–178.
20. Lee,S.R. and Collins,K. (2006) Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev.*, **20**, 28–33.
21. Couvillion,M.T., Lee,S.R., Hogstad,B., Malone,C.D., Tonkin,L.A., Sachidanandam,R., Hannon,G.J. and Collins,K. (2009) Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes Dev.*, **23**, 2016–2032.
22. Marker,S., Le Mouel,A., Meyer,E. and Simon,M. (2010) Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res.*, **38**, 4092–4107.
23. Galvani,A. and Sperling,L. (2001) Transgene-mediated post-transcriptional gene silencing is inhibited by 3’ non-coding sequences in *Paramecium*. *Nucleic Acids Res.*, **29**, 4387–4394.
24. Ruiz,F., Vayssie,L., Klotz,C., Sperling,L. and Madeddu,L. (1998) Homology-dependent gene silencing in *Paramecium*. *Mol. Biol. Cell*, **9**, 931–943.
25. Bouhouche,K., Gout,J.F., Kapusta,A., Betermier,M. and Meyer,E. (2011) Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.*, **39**, 4249–4264.
26. Timmons,L., Court,D.L. and Fire,A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene*, **263**, 103–112.
27. Timmons,L. and Fire,A. (1998) Specific interference by ingested dsRNA. *Nature*, **395**, 854–854.
28. Galvani,A. and Sperling,L. (2002) RNA interference by feeding in *Paramecium*. *Trends Genet.*, **18**, 11–12.
29. Sijen,T., Steiner,F.A., Tijssen,K.L. and Plasterk,R.H. (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, **315**, 244–247.
30. Aoki,K., Moriguchi,H., Yoshioka,T., Okawa,K. and Tabara,H. (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J*, **26**, 5007–5019.
31. Pak,J. and Fire,A. (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science*, **315**, 241–244.
32. Marshall,W.F. (2008) Modeling recursive RNA interference. *PLoS Comput. Biol.*, **4**, e1000183.
33. Skouri,F. and Cohen,J. (1997) Genetic approach to regulated exocytosis using functional complementation in *Paramecium*: identification of the ND7 gene required for membrane fusion. *Mol. Biol. Cell*, **8**, 1063–1071.
34. Sonneborn,T.M. (1970) Methods in *Paramecium* research. *Methods Cell Physiol.*, **4**, 241–339.
35. Cohen,J. (1980) Cytotoxic versus mutagenic effect of ethyl methanesulfonate on *Paramecium tetraurelia*. *Mutat. Res.*, **70**, 251–254.
36. Simon,M.C., Marker,S. and Schmidt,H.J. (2006) Posttranscriptional control is a strong factor enabling exclusive expression of surface antigens in *Paramecium tetraurelia*. *Gene Expr.*, **13**, 167–178.
37. Gratiyas,A. and Betermier,M. (2003) Processing of Double-Strand Breaks Is Involved in the Precise Excision of *Paramecium* Internal Eliminated Sequences. *Mol. Cell. Biol.*, **23**, 7152–7162.
38. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
39. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Subgroup,G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
40. Froissard,M., Kissmehl,R., Dedieu,J.C., Gulik-Krzywicki,T., Plattner,H. and Cohen,J. (2002) N-ethylmaleimide-sensitive factor is required to organize functional exocytotic microdomains in *paramecium*. *Genetics*, **161**, 643–650.
41. Kissmehl,R., Froissard,M., Plattner,H., Momayez,M. and Cohen,J. (2002) NSF regulates membrane traffic along multiple pathways in *Paramecium*. *J. Cell Sci.*, **115**, 3935–3946.
42. Froissard,M., Keller,A.M., Dedieu,J.C. and Cohen,J. (2004) Novel secretory vesicle proteins essential for membrane fusion display extracellular-matrix domains. *Traffic*, **5**, 493–502.

43. Stevenson,A.L. and Norbury,C.J. (2006) The Cid1 family of non-canonical poly(A) polymerases. *Yeast*, **23**, 991–1000.
44. Motamedi,M.R., Verdel,A., Colmenares,S.U., Gerber,S.A., Gygi,S.P. and Moazed,D. (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell*, **119**, 789–802.
45. Talsky,K.B. and Collins,K. (2010) Initiation by a eukaryotic RNA-dependent RNA polymerase requires looping of the template end and is influenced by the template-tailing activity of an associated uridylyltransferase. *J. Biol. Chem.*, **285**, 27614–27623.
46. Lee,S.R., Talsky,K.B. and Collins,K. (2009) A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *RNA*, **15**, 1363–1374.
47. Jaillon,O., Bouhouche,K., Gout,J.F., Aury,J.M., Noel,B., Saudemont,B., Nowacki,M., Serrano,V., Porcel,B.M. and Segurens,B. et al. (2008) Translational control of intron splicing in eukaryotes. *Nature*, **451**, 359–362.
48. Makeyev,E.V. and Bamford,D.H. (2002) Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell*, **10**, 1417–1427.
49. Sugiyama,T., Cam,H., Verdel,A., Moazed,D. and Grewal,S.I. (2005) RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 152–157.
50. Curaba,J. and Chen,X. (2008) Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J. Biol. Chem.*, **283**, 3059–3066.
51. Zhao,Y., Yu,Y., Zhai,J., Ramachandran,V., Dinh,T.T., Meyers,B.C., Mo,B. and Chen,X. (2012) The Arabidopsis nucleotidyl transferase HES01 uridylylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.*, **22**, 689–694.
52. Mayer,K.M. and Forney,J.D. (1999) A mutation in the flanking 5'-TA-3' dinucleotide prevents excision of an internal eliminated sequence from the Paramecium tetraurelia genome. *Genetics*, **151**, 597–604.
53. Welker,N.C., Pavelec,D.M., Nix,D.A., Duchaine,T.F., Kennedy,S. and Bass,B.L. (2010) Dicer's helicase domain is required for accumulation of some, but not all, *C. elegans* endogenous siRNAs. *RNA*, **16**, 893–903.
54. Van Houten,J. and Preston,R.R. (1987) Eukaryotic Unicells: How Useful in Studying Chemoreception? *Ann. N. Y. Acad. Sci.*, **510**, 16–22.
55. Preer,J.R., Preer,L.B., Rudman,B. and Barnett,A. (1987) Molecular Biology of the Genes for Immobilization Antigens in Paramecium1. *J. Protozool.*, **34**, 418–423.
56. Sonneborn,T.M. (1974) Paramecium aurelia, In: King,R. (ed.), *Handbook of Genetics*. Plenum press, New York, pp. 469–594.
57. Vayssie,L., Skouri,F., Sperling,L. and Cohen,J. (2000) Molecular genetics of regulated secretion in paramecium. *Biochimie*, **82**, 269–288.
58. Saimi,Y. and Kung,C. (1987) Behavioral Genetics of Paramecium. *Annu. Rev. Genet.*, **21**, 47–65.
59. Haynes,W.J., Vaillant,B., Preston,R.R., Saimi,Y. and Kung,C. (1998) The cloning by complementation of the pawn-A gene in Paramecium. *Genetics*, **149**, 947–957.
60. Keller,A.M. and Cohen,J. (2000) An indexed genomic library for Paramecium complementation cloning. *J. Eukaryot. Microbiol.*, **47**, 1–6.
61. Greenberg,M.V.C., Ausin,I., Chan,S.W.L., Cokus,S.J., Cuperus,J.T., Feng,S., Law,J.A., Chu,C., Pellegrini,M., Carrington,J.C. et al. (2011) Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics*, **6**, 344–354.
62. Zuryn,S., Le Gras,S., Jamet,K. and Jarriault,S. (2010) A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics*, **186**, 427–430.
63. Hall,I.M., Noma,K. and Grewal,S.I. (2003) RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 193–198.
64. Volpe,T., Schramke,V., Hamilton,G.L., White,S.A., Teng,G., Martienssen,R.A. and Allshire,R.C. (2003) RNA interference is required for normal centromere function in fission yeast. *Chromosome Res.*, **11**, 137–146.
65. Win,T.Z., Stevenson,A.L. and Wang,S.W. (2006) Fission yeast Cid12 has dual functions in chromosome segregation and checkpoint control. *Mol. Cell Biol.*, **26**, 4435–4447.
66. Volpe,T.A., Kidner,C., Hall,I.M., Teng,G., Grewal,S.I. and Martienssen,R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
67. Vastenhouw,N.L., Fischer,S.E.J., Robert,V.J.P., Thijssen,K.L., Fraser,A.G., Kamath,R.S., Ahringer,J. and Plasterk,R.H.A. (2003) A Genome-Wide Screen Identifies 27 Genes Involved in Transposon Silencing in *C. elegans*. *Curr. Biol.*, **13**, 1311–1316.
68. Kim,J.K., Gabel,H.W., Kamath,R.S., Tewari,M., Pasquinelli,A., Rual,J.F., Kennedy,S., Dybbs,M., Bertin,N., Kaplan,J.M. et al. (2005) Functional genomic analysis of RNA interference in *C. elegans*. *Science*, **308**, 1164–1167.
69. Elmayer,T., Balzergue,S., Beon,F., Bourdon,V., Daubremet,J., Guenet,Y., Mourrain,P., Palauqui,J.C., Vernhettes,S., Vialle,T. et al. (1998) Arabidopsis mutants impaired in cosuppression. *Plant Cell*, **10**, 1747–1758.
70. Cogoni,C. and Macino,G. (1997) Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 10233–10238.
71. Matzke,M., Aufsatz,W., Kanno,T., Daxinger,L., Papp,I., Mette,M.F. and Matzke,A.J. (2004) Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim. Biophys. Acta*, **1677**, 129–141.
72. Saleh,M.C., van Rij,R.P., Hekele,A., Gillis,A., Foley,E., O'Farrell,P.H. and Andino,R. (2006) The endocytic pathway mediates cell entry of dsRNA to induce RNAi silencing. *Nat. Cell Biol.*, **8**, 793–802.
73. Ulvila,J., Parikka,M., Kleino,A., Sormunen,R., Ezekowitz,R.A., Kocks,C. and Ramet,M. (2006) Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in *Drosophila* S2 cells. *J. Biol. Chem.*, **281**, 14370–14375.
74. Nuez,I. and Felix,M.A. (2012) Evolution of susceptibility to ingested double-stranded RNAs in *Caenorhabditis* nematodes. *PLoS One*, **7**, e29811.
75. Duchaine,T.F., Wohlschlegel,J.A., Kennedy,S., Bei,Y., Conte,D. Jr, Pang,K., Brownell,D.R., Harding,S., Mitani,S., Ruvkun,G. et al. (2006) Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell*, **124**, 343–354.
76. Lee,S.R. and Collins,K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat. Struct. Mol. Biol.*, **14**, 604–610.
77. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797

B. Localisation intracellulaire de Pds1

Le gène PDS1 identifié dans le crible de la partie précédente est un gène impliqué dans l'ARN interférence induite par *feeding* d'ARNdb, cependant aucun motif protéique et aucune structure particulière n'ont pu être trouvés par homologie de séquence sur cette protéine. Afin d'en apprendre un peu plus sur la fonction de cette protéine dans l'ARNi, la construction du gène PDS1 fusionné à la GFP en C-terminal sous contrôle de son promoteur endogène a été injectée dans le macronoyau du clone 3.17, un mutant nul pour Pds1 (Figure 1 de l'article). Les clones mutants injectés ont été de nouveau sensibles à l'ARNi induite par *feeding* ce qui montre que Pds1 est bien le gène muté impliqué dans l'ARNi et que la construction PDS1-GFP est fonctionnelle. La fluorescence a été observée sur les cellules vivantes immobilisées dans une microgoutte de BSA 0,2% (Figure 38). La protéine Pds1 semble être diffuse dans tout le cytoplasme. La localisation cytoplasmique de Pds1 ne nous renseigne pas plus sur sa fonction, cependant elle rend moins probable l'hypothèse que Pds1 pourrait être impliqué dans l'import des ARNdb des vacuoles vers le cytoplasme. En effet, si c'était le cas, on aurait pu s'attendre à une localisation spécifique au niveau des vacuoles digestives.

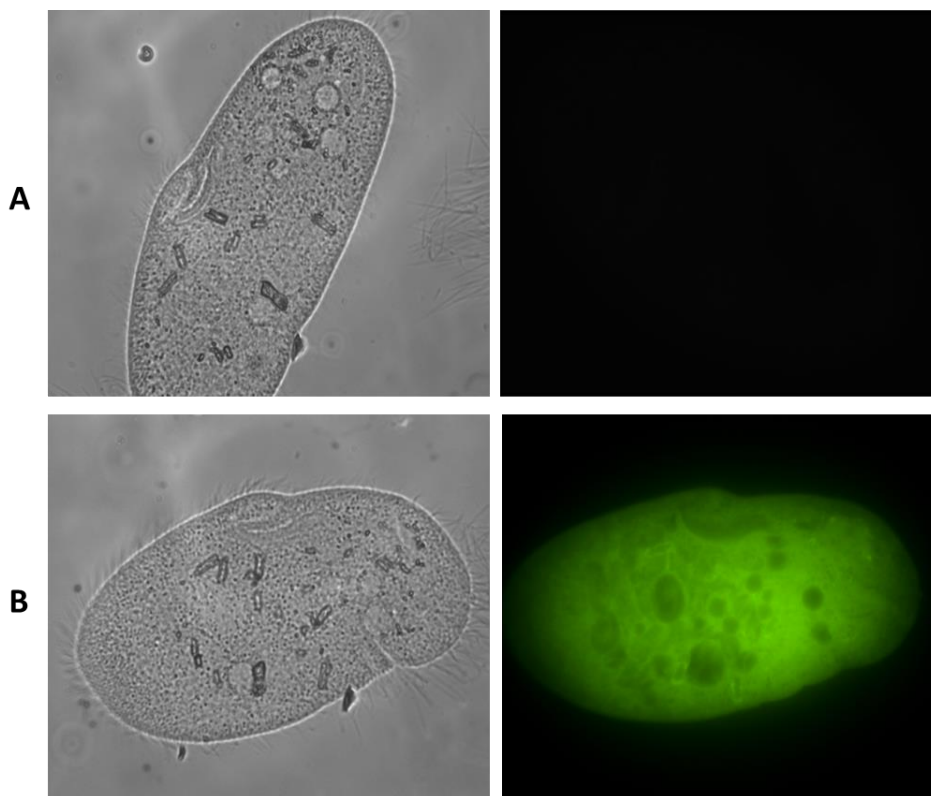


Figure 38: Localisation intracellulaire de Pds1. A) Clone 3.17 non injecté B) Clone 3.17 injecté avec la construction PDS1-GFP.

C. Analyse du gène PDS2

Les résultats présentés dans l'article concernent l'analyse de 71 clones parmi les 79 déficients pour l'ARN interférence obtenus dans le crible. 6 des 8 derniers clones ont été analysés en utilisant la même procédure. Les gènes déjà identifiés comme impliqués dans l'ARNi induite par *feeding* d'ARNdb (DCR1, Ptiwi13, RDR1, RDR2, Cid91 et Pds1) ont été séquencés dans ces 8 derniers mutants.

Pour 5 de ces 6 clones, le gène a ainsi été identifié. 3 mutations dans DCR1, 1 mutation dans RDR1 et 1 dans RDR2 (Figure 39). La mutation dans le clone 1.28 n'a pas pu être trouvée dans les gènes déjà connus.

Mutant	Gène muté	Nucléotide	codon	Acide aminé
5.6	Rdr1	A -> C	AGT -> CGT	Ser -> Arg
5.32	RdR2	C -> T	CGA -> TGA	Arg -> Stop
1.19	Dcr1	C -> T	?	?
1.20	Dcr1	C -> T	TCT -> TTT	Ser -> Phe
1.7	Dcr1	A -> T	AGA -> AGT	Arg -> Ser
1.28	non muté dans les gènes déjà connus			
5.1	non analysé			
5.10	non analysé			

Figure 39 : Gènes mutés dans les derniers clones du crible. La substitution, le codon affecté et l'acide aminé changé sont indiqués

1) Croisement du mutant 1.28

Le mutant 1.28 est donc muté dans un autre gène impliqué dans l'ARN interférence. En plus d'une résistance totale à l'ARNi, ce mutant présente une croissance végétative beaucoup plus lente que les paramécies sauvages. En effet, à 27°C ce mutant fait une à deux divisions par jour contre 4 pour une paramécie mutante. Pour identifier le gène muté dans ce clone et pour déterminer si ce gène est aussi responsable du phénotype de croissance lente, ce mutant a été croisé avec une paramécie sauvage. Les couples qui ont réalisé l'échange des mic ont permis la génération d'une progéniture F1 hétérozygote qui présente le phénotype sauvage de sensibilité à l'ARN interférence (par *feeding* du gène NSF, létal pour les paramécies sensibles) ainsi qu'une croissance normale. La ou les mutations responsables de ces deux phénotypes sont donc récessives. Les F1 d'un des couples ont été amenées à l'autogamie et 60 clones F2 de chaque côté du croisement ont été isolés. Les clones F2 du côté sauvage du croisement ont été testés pour leur vitesse de croissance et leur sensibilité à l'ARNi. Parmi les 46 clones survivants, on observe une ségrégation 2:2 du phénotype de résistance à l'ARNi interférence qui corrèle exactement avec le phénotype de vitesse de croissance (Figure 40).

Phénotype ARNi	Vitesse de croissance	Clones F2
sensible	Rapide	25
résistant	Lente	21
sensible	Lente	0
résistant	Rapide	0
Morts		14
Total		60

Figure 40: Ségrégation F2 du croisement 1.28 par sauvage. Le nombre de clones présentant une des 4 combinaisons avec les deux phénotypes possible: sensibilité à l'ARNi et de vitesse de croissance, a été compté.

Ce résultat indique que le mutant 1.28 est muté dans un seul gène et que cette mutation lui apporte les deux phénotypes de résistance à l'ARN interférence et de croissance végétative lente. Les F2 mutantes ont été rassemblées et leurs génomes ont été séquencés ensemble par la méthode décrite dans l'article.

2) La protéine membranaire putative Pds2.

Le séquençage du mac du mutant 1.28 a révélé une unique mutation dans une région non annotée du scaffold 22. A l'aide des données de RNAseq, les modèles de gènes ont pu être positionnés (Figure 41 A et B). La mutation G vers A dans le codon TGG (tryptophane) entraîne l'apparition d'un codon stop (TAG) anticipé dans un gène de 652 pb contenant un intron de 25 pb. Ce gène a été nommé PDS2 (Paramecium dsRNA-inducedRNAi-specific protein 2). Une particularité de ce gène est qu'il est chevauchant avec le gène en amont dans un autre cadre de lecture : les 4 dernières paires de bases ATGA du gène en amont servent aussi d'ATG au gène PDS2 (Figure 41 A). Les transcrits entre PDS2 et le gène en amont étant bien séparés, il est peu probable que ces deux gènes fassent partie de la même unité de transcription (Figure 41 B). Le promoteur de PDS2 pourrait être contenu dans l'extrémité 3' du gène amont. Les séquences nucléotidiques et protéiques de toute la région sont dans la figure 1 de l'annexe 2.

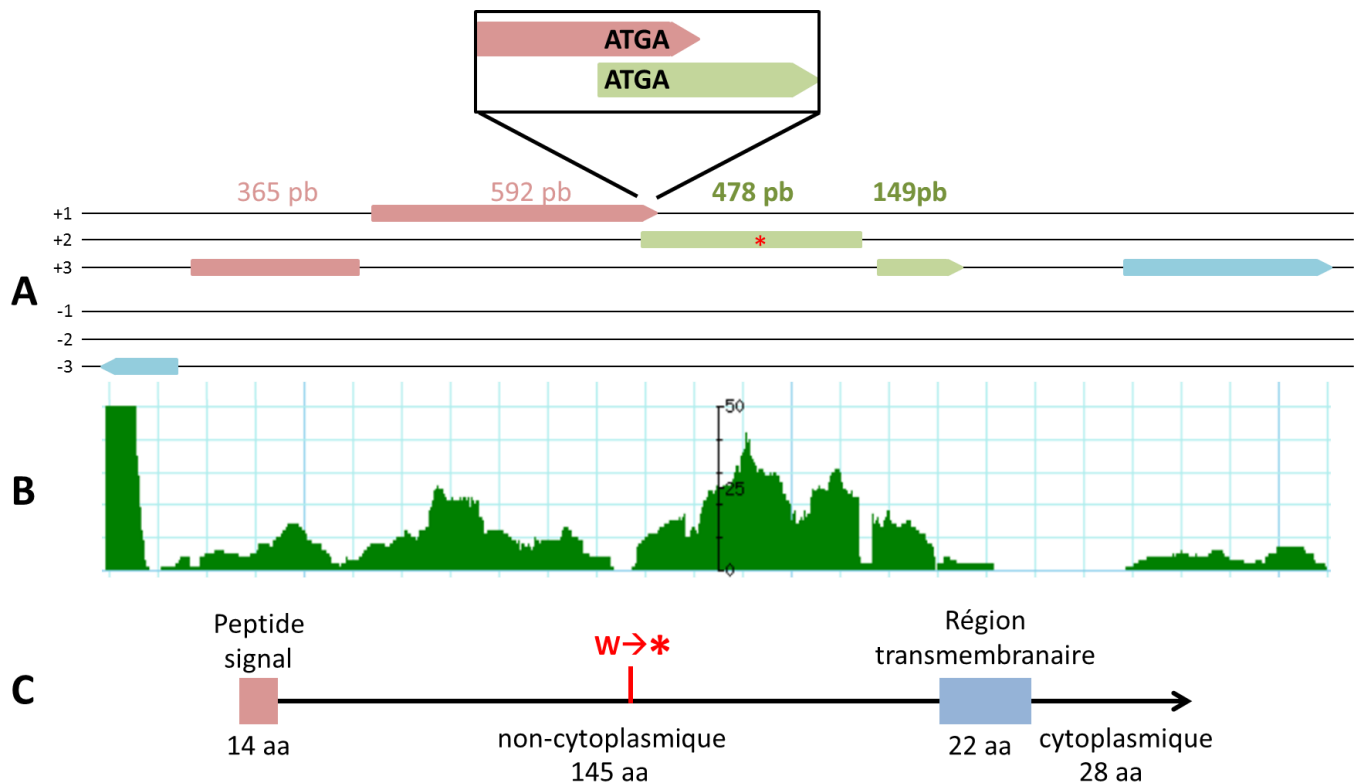


Figure 41: Caractéristiques de PDS2. A) Modèles de gènes sur le scaffold22 de *P. tetraurelia* entre les positions 314,600 et 317,099 dans les 6 cadres de lecture. Les exons sont représentés par des rectangles, verts pour PDS2, rouges pour le gène en amont et bleu pour les autres gènes de la région. L'astérisque représente la position de la mutation. B) Données de RNAseq dans un échantillon sauvage en croissance végétative. C) Représentation de la protéine Pds2 et des domaines identifiés. La région transmembranaire et le peptide signal ont été identifiés respectivement par les programmes TMHMM (v2.0) et SignalP (v4.1) accessible sur le site web (<http://www.cbs.dtu.dk/services/>) avec les paramètres par défauts.

La recherche de motifs structuraux sur cette protéine a permis de mettre en évidence la présence d'une région transmembranaire de 22 acides aminés dans la deuxième moitié de la protéine ainsi qu'un peptide signal de 14 aa (Figure 41 C et Figure 42). Cette protéine membranaire impliquée dans l'ARNi pourrait donc avoir un rôle dans l'import des ARNdb des vacuoles digestives vers le cytoplasme. Nous posons donc l'hypothèse que PDS2 est fixée à la membrane des vacuoles digestives et a un rôle déterminant dans leur fonctionnement. Si c'est le cas, l'absence de PDS2 pourrait avoir un impact fort sur la capacité des paramécies à se nourrir ce qui expliquerait la croissance végétative plus lente observée chez le mutant 1.28.

3) Conservation de PDS2.

Par homologie de séquence protéique, les orthologues de PDS2 ont pu être identifiés dans les autres espèces de paramécies du groupe *aurelia* avec un très fort pourcentage d'identité (>87%) mais aussi chez *P.miltimicronucleatum* et *P.caudatum* (respectivement 68 et 69% d'identité). Ce gène a aussi été retrouvé chez les ciliés du genre *Tetrahymena* et chez *Oxytricha* où la séquence est beaucoup plus divergente (respectivement 29% et 23% d'identité). Chez toutes ces espèces, le peptide signal et le segment transmembranaire sont bien conservés (Figure 42).

4) PDS2 n'est pas impliqué la voie d'ARNi induite par transgène.

L'hypothèse que Pds2 est la protéine nécessaire à l'import des ARNdb des vacuoles digestives vers le cytoplasme pourrait être remise en cause si ce gène était également impliqué dans la voie d'ARN interférence induite par transgène qui ne requiert pas l'import d'ARN. Pour répondre à cette question, un plasmide *feeding* complémentaire à une portion du gène PDS2 (fragment de 339 pb à 140 pb du début de la séquence codante) a été exprimé dans des bactéries *E. coli*. La fonctionnalité de cette construction a été testée en double *feeding* sur des paramécies sauvages (Figure 43 A). Les cellules nourries à la fois avec la construction ND169 conférant et l'ARNdb PDS2 restent de phénotype tric+ et sont donc résistantes à l'ARNi de même que les doubles *feeding* ND169 + RDR2 et ND169 + CID1. Au contraire le contrôle ND169 + ICL est sensible à l'ARNi (phénotype tric -). La construction *feeding* Pds2 est donc fonctionnelle. Le transgène pTI- portant le gène ND169 tronqué à son extrémité 3' (voir le matériel et méthode de l'article) a ensuite été injecté dans des cellules sauvages et différents *feeding* ont été réalisés (Figure 43 B). Parmi les *feeding* CID2, ICL, RDR2, seul le gène RDR2 est impliqué dans la voie d'ARNi induite par transgène et sa répression conduit donc à une résistance à l'ARNi induite par transgène et au maintien du phénotype tric+. CID2 et ICL, qui ne sont pas impliqués dans cette voie restent sensibles à l'ARNi et de phénotype tric -. Il en est de même pour les cellules soumises au *feeding* PDS2 qui reste sensibles à l'ARNi et sont de phénotype tric-. PDS2 n'est donc *a priori* pas impliqué dans la voie d'ARN interférence induite par transgène. L'hypothèse que cette protéine est capable d'importer les longs double-brin d'ARN des vacuoles digestives reste donc vraisemblable.

A		B			
cellules sauvages		cellules PTI-			
Phénotype tric		Phénotype tric			
		+	-		
Double <i>feeding</i> d'ARNdb	ND169 + ICL	-	ICL	0	26
	ND169 + R2	+	RdR2	31	0
	ND169 + Cid1	+	Cid1	0	37
	ND169 + Pds2	+	Pds2	0	42
			Feeding d'ARNdb		

Figure 43: PDS2 n'est pas impliqué dans l'ARNi induite par transgènes. A) Double *feeding* sur des cellules sauvages. +: phénotype tric+ de résistance à l'ARNi, -: phénotype tric- de sensibilité à l'ARNi B) Simple *feeding* sur des cellules injectées avec le transgène ND169 (PTI-). Les chiffres indiquent le nombre de paramécies observées pour chaque phénotype.

La détermination de la localisation intracellulaire de Pds2 à l'aide de la construction d'une protéine fusionnée à la GFP pourrait nous permettre de confirmer cette hypothèse.

D. Compléments sur le rôle de Rdr2 au cours des cycles sexuels

Le rôle de RdR2 au cours des événements sexuels, autogamie et conjugaison, a pu être précisé suite à l'analyse d'un troisième allèle trouvé dans le crible plus tardivement.

5) L'allèle 5.32 du gène RDR2

Le clone 5.32 est muté dans le domaine RdRP du gène RDR2. La substitution C → T dans un codon CGA (arginine) introduit un codon stop TGA anticipé. La mutation se situe au deux tiers de la protéine, 37 acides aminés après le motif catalytique DLDGD. Il est donc très probable que RdR2 dans ce clone soit non fonctionnel. Ce nouvel allèle de RdR2 serait donc en contradiction avec la conclusion de l'article qui suggère que RdR2 est un gène essentiel et que c'est la raison pour laquelle seuls deux allèles partiels ont été trouvés. Pour résoudre ce paradoxe, plusieurs expériences ont été menées.

Même si l'allèle 5.32 suggère que la protéine RdR2 n'est pas fonctionnelle, pour déterminer si ce mutant est phénotypiquement nul, la *feeding* de l'ARNdb homologue au gène non essentiel ND169 (induisant le phénotype tric-) a été réalisé sur des paramécies portant les 3 allèles RdR2 (3.7, 1.24 ou 5.32) et sur des paramécies sauvages. Afin d'augmenter la sensibilité du test, les *E.coli* exprimant l'ARNdb ND169 ont été dilués dans des proportions variables avec des *E.coli* exprimant un ARNdb homologue au gène ICL qui n'a aucune influence sur l'exocytose des trichocystes (Figure 44).

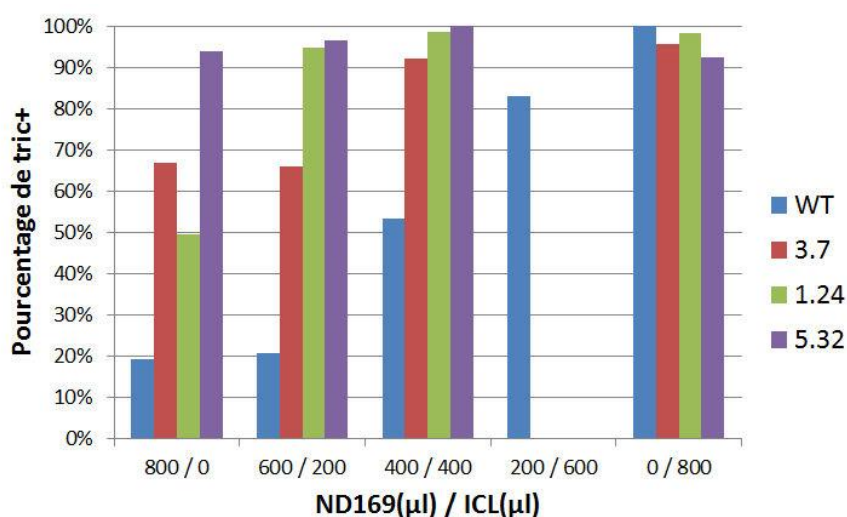


Figure 44: Résistance au *feeding* des mutants RdR2. Les histogrammes représentent le pourcentage de cellules capable d'éjecter les trichocystes (phénotype tric+) après *feeding* de l'ARNdb homologue à ND169. Les volumes des milieux de cultures contenant les bactéries exprimant l'ARNdb ND169 et l'ARNdb contrôle ICL mélangés sont indiqués sur l'axe horizontal. Les trois mutants 3.7, 1.24 et 5.32 et le sauvage ont été testés.

Cette expérience montre en premier lieu que la quantité d'ARNdb ingérée par les paramécies corrèle avec l'efficacité de l'ARNi. En effet, dans l'échantillon sauvage, le pourcentage de cellules tric+ est strictement corrélé à la dilution du *feeding* ND169 dans le *feeding* ICL. C'est donc un bon test pour contrôler si des mutants sont partiellement ou totalement déficients pour l'ARNi. Des trois mutants RdR2, 5.32 montre le phénotype le plus fort puisqu'il est totalement résistant au *feeding* de l'ARNdb ND169 même lorsqu'il n'est pas dilué (>90% tric+). Les deux autres allèles sont résistants au *feeding* de ND169 seulement s'il est dilué avec un autre *feeding*. Une dilution d'un quart suffit pour 1.24 alors qu'une dilution de moitié est nécessaire pour 3.7.

6) Effet de la mutation 5.32 à l'autogamie

Si la mutation 5.32 est celle qui a le phénotype le plus fort, il est alors probable que les phénotypes observés au cours des évènements sexuels soient aussi plus marqués. Les trois mutants RdR2 ont donc été amenés à l'autogamie et si au cours de la croissance végétative aucune différence entre les mutants et le sauvage n'a pu être détectée, il en est tout autrement pendant et après l'autogamie.

La même expérience que celle décrite dans la figure supplémentaire 7C de l'article a été réalisée avec le mutant 5.32. Il apparaît clairement qu'après fragmentation du mac et 4 jours de carence alimentaire, la quantité de cellules mortes (visible par l'accumulation de cadavres au fond du milieu de culture) est bien plus importante dans les mutants RdR2 que dans les sauvages et légèrement plus importante dans le mutant 5.32 que dans les mutants 3.7 et 1.24. Même si l'effet observé semble significatif, il est malheureusement difficile de quantifier le nombre de cellules mortes, l'expérience devra être répétée dans des conditions différentes pour permettre une analyse plus précise. De plus, les clones 5.32 fragmentés et en carence alimentaire depuis moins de 4 jours semblent incapables de recommencer une croissance végétative et meurent après seulement quelques divisions (données non présentées). Enfin, il est apparu que les paramécies 5.32 provenant du milieu stock, tombent autogames (ou tout au moins fragmentent le MAC) après seulement quelques divisions (maximum 4) ce qui suggère que les cellules sont déjà âgées (au moins 16 divisions) dans le stock. Ce résultat est particulièrement étonnant puisque les stocks de paramécies sont maintenus à 14°C et nourris avec un volume de 2 ml de milieu riche une fois par mois, permettant aux paramécies de réaliser environ 2 à 4 divisions (selon le nombre de cellules dans le tube) avant d'entrer en carence alimentaire à nouveau. Une longue période de carence alimentaire induit l'autogamie avant les 20 divisions théoriques, les cellules en stock sont donc en général jeunes (entre 1 et 10 divisions). Tous ces résultats suggèrent donc que le mutant 5.32 contrairement aux allèles 1.24 et 3.7 est simplement incapable de réaliser une autogamie.

La première hypothèse est que le mutant 5.32 est incapable de développer un nouveau mac après fragmentation de l'ancien. Afin d'observer les noyaux au cours de l'autogamie, les cellules 5.32 ou sauvages ont été fixées au paraformaldéhyde puis marquées au DAPI aux différents stades de l'autogamie (Figure 45). Contrairement à l'hypothèse, le débobinage de l'ancien MAC, la fragmentation et le développement des deux nouveaux mac dans le mutant 5.32 semblent tout à fait normaux. Aucune différence significative avec des cellules sauvages n'a pu être observée. Cependant, l'aspect normal des nouveaux mac ne signifie pas qu'ils sont fonctionnels.

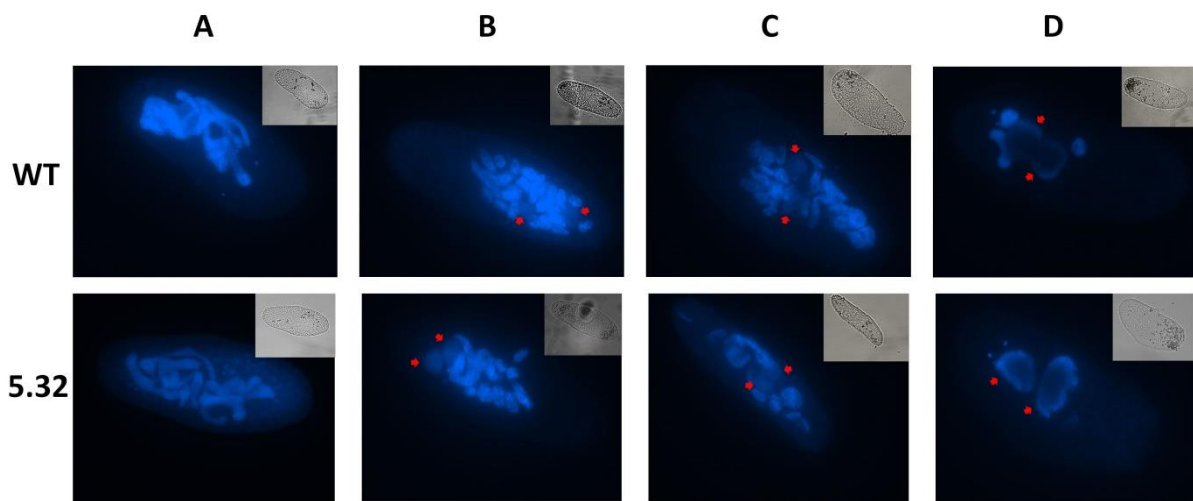


Figure 45: Evolution du mac au cours de l'autogamie du mutant 5.32. Les différents stades de l'autogamie sont présentés de gauche à droite. A) Débobinage de l'ancien MAC. B et C) Fragmentation de l'ancien mac et développement des nouveaux mac (indiqués par des flèches rouges). D) Cellule avec deux nouveaux mac avant la division caryonidale. Les fragments de l'ancien mac ont été progressivement lysés.

7) Croisement du mutant 5.32

L'analyse des allèles 1.24 et 3.7 indique que les macronoyaux n'ont pas de défaut de réarrangements au cours de leur développement, il est cependant possible que les micronoyaux germinaux du mutant 5.32 soient défectueux au moment de l'autogamie ce qui les rendraient incapables de produire de nouveaux mac fonctionnels. RDR2 pourrait ainsi être impliqué dans la méiose des mic ou la fécondation. Il est aussi possible que RDR2 ait un rôle en croissance végétative et soit impliqué par exemple dans la ségrégation des chromosomes lors de la mitose du mic. L'implication d'une RdRP dans la ségrégation des chromosomes a été montré chez *S. pombe* où l'absence de RdR1 et Cid12 perturbe la formation de l'hétérochromatine et la position des télomères au cours de la mitose et de la méiose (Hall et al. 2003^[111]). Si c'est le cas chez *P. tetraurelia*, ça pourrait expliquer que le mutant 5.32 ait une croissance végétative normale (puisque que le mic est transcriptionnellement inactif en croissance végétative), soit capable de faire une autogamie, mais que sa progéniture ait des problèmes de viabilité au moment où le nouveau mac se développe à partir du mic. Cette

hypothèse proposée dans l'article expliquerait également les phénotypes stochastiques de létalité observés sur les deux mutants RdR2 partiels (3.7 et 1.24).

Le mutant 5.32 a été croisé pour tenter de comprendre quels types de dommages pourraient affecter les mic mais aussi pour maintenir en vie le mutant 5.32. En effet, si le mutant est incapable de réaliser une autogamie depuis 3 ans, les cellules doivent être très âgées (entre 72 et 144 divisions) et proches d'entrer en sénescence. Les paramécies utilisées pour le croisement avec 5.32 portent une mutation récessive dans le gène ND7 les rendant incapables d'éjecter les trichocystes à l'état homozygote mutant. Sur 42 couples testés, 2 couples sont morts des deux côtés du croisement et 35 couples n'ont survécu que d'un côté et sont incapables d'éjecter les trichocystes ce qui indique que ces F1 proviennent du parent ND7 et n'ont pas reçu le micronoyau de la cellule 5.32. Enfin, 5 couples ont survécu des deux côtés et sont toujours de phénotype tric- d'un côté et tric+ de l'autre. La cellule F1 du côté ND7 n'a donc pas reçu le micronoyau 5.32 et, l'échange unidirectionnel des mic étant un événement peu fréquent, il est probable que la cellule 5.32 n'ai pas non plus reçu le micronoyau de ND7 (ceci n'a pour autant pas été testé). Au contraire le croisement contrôle sauvage par ND7 a montré une bonne efficacité d'échange des mic et une faible mortalité (sur 24 couples, 18 ont échangés les mic au moins dans une direction, seulement 3 n'ont pas échangés et 3 sont morts). Ce résultat montre que les mutants 5.32 ne peuvent donner une progéniture viable et sont donc stériles.

Pour tenter de résoudre le problème, les mutants 5.32 ont été injectés avec l'allèle sauvage du gène RdR2 dans l'optique de compléter le phénotype mutant et de réaliser le croisement. Un plasmide exprimant la GFP a été co-injecté afin de contrôler l'efficacité de l'injection. Il est à noter que cellules injectées perdent leur plasmide avec le macronoyau parental après fragmentation, les cellules complémentées redeviennent donc mutantes après autogamie ou conjugaison à moins qu'elles n'aient reçu un allèle sauvage. 9 clones ont survécus à l'injection et 3 d'entre eux ont présentés une fluorescence verte indiquant que le plasmide contrôle portant la GFP s'est correctement exprimé. Etonnamment, les 3 clones correctement injectés n'étaient pas complémentés pour le phénotype de résistance au *feeding* d'ARNdb. Deux de ces trois clones ont tout de même été croisés avec des cellules ND7. Sur 21 couples testés, 19 ont survécu des deux côtés et 11 des 19 couples survivants ont présentés un phénotype tric+ de chaque côté indiquant que ND7 a bien récupéré le micronoyau du mutant 5.32 et a généré des F1 hétérozygotes (Figure 46). Il est difficile de savoir si c'est véritablement l'injection qui a permis de réaliser ce croisement (puisque le phénotype de sensibilité à l'ARNi n'a pas été restauré) ou simplement le fait que les cellules aient été croisées dans des conditions différentes par rapport à l'expérience précédente (culture à 18°C pour l'injection, réactivité en dépression en non en tube,...)

Les cellules F1 de chaque côté d'un des couples ont été amenées à l'autogamie et les phénotypes des F2 ont été analysés. Sur 60 cellules autogames isolées de chaque côté du

croisement, entre 22% et 42% meurent, entre 20% et 30% régénèrent un mac à partir d'un fragment de l'ancien mac et sont donc toujours hétérozygotes. Enfin 40% et 48% de cellules deviennent de vraies F2 homozygotes. Parmi ces cellules, malgré des effectifs trop faibles pour un test statistique, on peut observer une ségrégation 2:2 du phénotype de l'éjection des trichocystes et un biais en faveur du phénotype RNAi+ et donc de l'allèle sauvage de RDR2 (entre 67% et 72% des F2 homozygotes sont RNAi+) qui suggère qu'une partie des F2 portant l'allèle 5.32 ne sont pas capables de développer un nouveau mac fonctionnel et meurent ou régénèrent l'ancien mac hétérozygote (Figure 46).

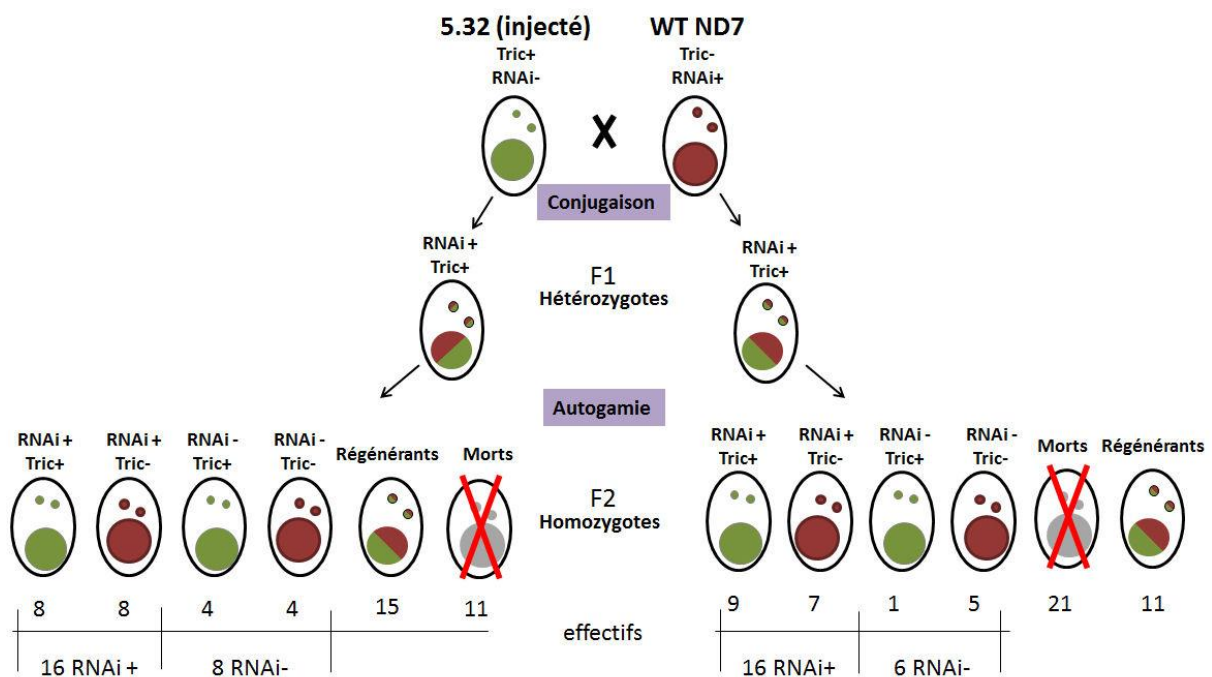


Figure 46 : Représentation schématique du croisement 5.32 par ND7. Les phénotypes pour l'ARNi (RNAi- ou RNAi+) et pour la capacité à décharger les trichocystes (tric+ ou tric-) sont indiqués

Trois hypothèses peuvent expliquer qu'une cellule hétérozygote pour un gène donné présente une déficience partielle à l'autogamie. 1) La mutation est partiellement dominante (effet dominant négatif ou quantité de protéine sauvage insuffisante) et le gène, exprimé par l'ancien mac, est impliqué dans l'autogamie ou dans le développement des nouveaux macs. 2) Le gène est exprimé par les nouveaux mac en développement (homozygotes mutants) et agit directement sur le développement du nouveau MAC. 3) Si les mic ont perdu un ou plusieurs chromosomes dans le parent 5.32, les mic des F1 pourraient être aneuploïdes (certains chromosomes diploïdes et d'autres haploïdes) et donc souvent (en fonction du nombre de chromosomes haploïdes) donner des produits méiotiques déficients pour la génération F2. Des expériences complémentaires seront nécessaires pour trancher entre ces trois hypothèses. La construction de la protéine de fusion RdR2 nous permettrait de vérifier si le gène est au moins partiellement exprimé à partir de l'ancien mac en regardant l'autogamie de clones injectés. De

plus, la comparaison entre le croisement de jeunes ou de vieilles cellules homozygotes mutantes pour RDR2 nous permettrait de savoir si le mutant RdR2 accumule ou non des défauts dans le mic au cours des divisions en croissance végétative. Pour vérifier l'hypothèse de la perte de chromosomes mic, il est possible de séquencer les scnRNAs, qui couvrent normalement tout le génome mic, en début d'autogamie. Leur absence sur un ou plusieurs scaffold nous indiquerait la perte de chromosomes. Enfin, il n'est pas exclu que RDR2 soit impliqué dans les réarrangements du génome, il faudrait donc séquencer le nouveau mac de cellules post-autogames non viables.

E. Identification des mutations induites par les UV.

Les pools de F2 réalisés après croisement avec sauvage permettent de sélectionner la mutation responsable du phénotype de résistance à l'ARNi. Cependant, ce pool de F2 ne permet pas d'éliminer les autres mutations du génome induites par les UV qui seront présentes sur 50% des lectures couvrant la position en moyenne (Figure 1 de l'article). Un programme a été développé pour estimer le nombre de mutation induite par les UV dans chacun des sept clones dont le mac a été séquencé. Une position dans le génome mutant est considéré comme mutée si elle respecte ces trois conditions : 1) La base est couverte par un minimum de 30 ou 50 lectures. 2) Entre 20% et 80% des lectures sont différentes de la base de référence 3) La position est identique à la base de référence dans tous les autres génomes mutants et sauvages.

La couverture minimale requise pour analyser la position a été choisie en fonction de la couverture moyenne du génome. Ainsi, une couverture minimale de 30x a été choisie pour les génomes avec une couverture moyenne par base inférieure à 50x (mutants 5.6, 1.20 et 1.8) et une couverture minimale de 50x pour les génomes avec une couverture moyenne par base supérieure à 50x (mutants 5.2, 3.1, 3.18 et 1.28). Il en résulte que seule une portion du génome (entre 15% et 95% selon les mutants) a été analysée par le programme. Le nombre de mutations induites par les UV a donc été extrapolé linéairement pour obtenir une estimation du nombre de mutations sur le génome entier. Enfin, 3 des mutants (3.1, 1.8 et 3.18) ont été croisés deux fois avec le sauvage. Pour retrouver le nombre de mutations induites par les UV, le nombre de positions identifiés dans le génome a donc été multiplié par deux (Figure 47).

Mutant	Gène muté	couverture moyenne par base	Couverture minimale choisie	Pourcentage du génome analysé	mutations trouvées	Extension au génome complet	Nombre de backcross	Estimation du nombre de mutation induite par les U.Vs
5.6	Dcr1	33,2x	30x	70,9%	15	21,1	1	21,1
5.2	Rdr1	87,7x	50x	95,3%	12	12,6	1	12,6
3.1	Rdr1	58,6x	50x	60,1%	9	15,0	2	30,0
1.20	Rdr1	26x	30x	60,5%	19	31,4	1	31,4
1.8	Cid1	20,1x	30x	15,9%	4	25,2	2	50,4
3.18	Pds1	103,1x	50x	68,0%	9	13,2	2	26,0
1.28	Pds2	52,2x	50x	64,8%	40	61,7	1	61,7

Figure 47: Estimation du nombre de mutations induites par les UV. Les 7 clones mutants dont le

génomique mac a été séquencé ont été analysés pour estimer le nombre de positions mutées après la mutagenèse aux UV.

Entre 12 et 63 positions mutées ont été estimées dans les différents mutants. Ce nombre étant relativement faible, et la probabilité qu'une mutation soit sans impact phénotypique étant élevée (mutations silencieuses, dans des régions intergéniques,...), il est assez peu probable que les autres mutations induites par les UV affectent phénotypiquement les mutants de l'ARN interférence. De plus, la probabilité d'avoir des mutations induites par les UV génétiquement liées à la mutation responsable du phénotype mutant pour l'ARN interférence est très faible. Il suffit donc de quelques croisements avec une cellule sauvage pour être sûr d'éliminer les éventuels effets des mutations induites par les UV.

II. Deuxième partie : Etude du rôle des gènes de l'ARNi dans la production des siRNAs

A. Primary and secondary siRNA biosynthesis involving distinct RdRP activities in Paramecium tetraurelia

Pour étudier le rôle de ces gènes dans la biosynthèse ou l'action des siRNAs, ces derniers ont été séquencés à partir de cellules sauvages ou mutantes, préalablement nourries d'un ARNdb homologue à un gène non essentiel. L'analyse bio-informatique a permis de montrer que des siRNAs dits 'primaires' sont produits à partir de l'ARNdb bactérien, tandis que des siRNAs dits 'secondaires' sont produits à partir de la totalité de l'ARNm endogène ciblé, et sont majoritairement de polarité anti-sens. Alors que la production des siRNAs primaires dépend de tous les gènes trouvés dans le crible, celle des siRNAs secondaires requiert seulement la RdRP RDR2 et, comme chez *C. elegans*, semble indépendante de Dcr1. Une fraction de ces siRNAs (environ 10%) est modifiée par l'addition d'une ou plusieurs uridines non génomiques en 3'. On a également montré que certains clusters de siRNAs endogènes (non induits expérimentalement) dépendent de Rdr1 et de Cid1, tandis que d'autres dépendent de Rdr2. Enfin, la présence de siRNAs anti-sens aux ARN ribosomiaux (rRNA) bactériens suggère que la paramécie est capable de réprimer les ARNs produits naturellement chez les bactéries qu'elle ingère.

Ces résultats sont présentés dans un article qui est soumis dans le journal *Nucleic Acid Research*. Les documents supplémentaires de l'article se trouvent dans **l'annexe III**

Genetic dissection of primary and secondary siRNA biosynthesis in the ciliate *Paramecium tetraurelia*

Quentin Carradec^{1,2}, Ulrike Götz³, Olivier Arnaiz⁴, Juliette Pouch¹, Martin Simon³, Eric Meyer^{1*} and Simone Marker^{1,3*}

¹Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS; Inserm, U1024; CNRS, UMR 8197; Paris, F-75005 France.

²Sorbonne Universités, UPMC Univ., IFD, 4 place Jussieu, 75252 Paris cedex05, France.

³Saarland University, Centre for Human and Molecular Biology, Molecular Cell Dynamics, Campus A2 4, 66123 Saarbrücken, Germany.

⁴CNRS UPR3404 Centre de Génétique Moléculaire, Gif-sur-Yvette F-91198 cedex, France; Université Paris-Sud, Département de Biologie, Orsay, F-91405, France.

*To whom correspondence should be addressed. Tel : +49 681 302 4938, E-mail: simone.marker@uni-saarland.de; Tel : +33 1 44 32 39 48, E-mail: emeyer@biologie.ens.fr

Abstract

In the ciliate *Paramecium tetraurelia*, an efficient RNAi response can be triggered by feeding cells with bacteria producing double-stranded RNA, a mechanism that is widely used to silence endogenous genes. A mutagenesis screen has shown that some of the genes involved also control the production of endogenous siRNAs, but the natural functions of these interconnected pathways are unknown. Here, we used high-throughput sequencing to characterize the small RNAs synthesized in wild-type cells or RNAi-deficient mutants upon dsRNA feeding. We show that high levels of primary siRNAs of both strands are processed from the dsRNA trigger by a Dicer-dependent mechanism which also requires the RNA-dependent RNA polymerases Rdr1 and Rdr2 and other factors. Targeting a portion of an endogenous mRNA with primary siRNAs further induces the synthesis of secondary siRNAs spreading along the entire mRNA, demonstrating the occurrence of both 3'-to-5' and 5'-to-3' transitivity in this ciliate. Secondary siRNAs depend on Rdr2 but not on other factors required for primary siRNAs. They show a strong antisense bias, have a 5'-monophosphate end, and are produced at much lower levels than primary siRNAs; our data indicate that they do not contribute much to dsRNA-induced RNAi efficiency. Clusters of endogenous siRNAs are shown to depend either on Rdr1 or on Rdr2, but not on both. We further provide evidence that the *Paramecium* RNAi machinery can process single-stranded RNAs from its bacterial food, producing Rdr2-dependent, 23-nt siRNAs that are antisense to bacterial rRNAs. The possible natural functions of these complex and interconnected small-RNA pathways are discussed in the light of these results.

Introduction

Small RNAs (sRNAs) are important regulators of transcript processing in most eukaryotes. Loaded onto the Argonaute subunit of effector complexes, they can target a variety of complementary RNAs such as messenger RNAs, non-coding or nascent transcripts (GHILDIYAL and ZAMORE 2009; KETTING 2011). RNA interference (RNAi)-related mechanisms have diversified into complex networks. RNA interference (RNAi)-related mechanisms have diversified into complex networks of interconnected pathways. They control gene expression either post-transcriptionally, through mRNA cleavage or translational inhibition, or transcriptionally, through the deposition of epigenetic modifications on chromatin (CASTEL and MARTIENSSSEN 2013; MEISTER and TUSCHL 2004). In addition to their endogenous functions, they can provide efficient control of exogenous parasitic elements, such as of transposons and viruses.

In short interfering RNA (siRNA) pathways, double-stranded RNA (dsRNA) precursors are processed into 21-28nt siRNAs by Dicer-mediated endonucleolytic cleavage (MEISTER and TUSCHL 2004). These primary siRNAs can target cleavage of complementary transcripts. In plants, nematodes and fungi, the targeting of an mRNA by primary siRNAs induces a second round of siRNA synthesis through the recruitment of RNA-dependent RNA polymerases (RdRPs), leading to amplification of the silencing response (for reviews see (BAULCOMBE 2007; BRODERSEN and VOINNET 2006; VOINNET 2008)). In *Arabidopsis thaliana*, Rdr6 synthesizes long dsRNA molecules from mRNA cleavage products, providing new substrates for the same Dicer (BRODERSEN and VOINNET 2006; VOINNET 2008). This leads to 5'-to-3' and 3'-to-5' transitivity (VAISTIJ *et al.* 2002), *i. e.* production of secondary siRNAs beyond the region of the transcript matching the initial dsRNA trigger. In *C. elegans*, transitivity is mechanistically different: secondary siRNAs are synthesized in a Dicer-independent manner by an unprimed, discontinuous mode of RdRP activity (Rrf-3; Ego-1) which produces short antisense molecules with 5'-triphosphate ends (AOKI *et al.* 2007; MANIAR and FIRE 2011; PAK and FIRE 2007; SIJEN *et al.* 2007). *C. elegans* secondary siRNAs are in excess to primaries and are the major effectors of silencing (CLAYCOMB *et al.* 2009; GU *et al.* 2009; SIJEN *et al.* 2007). Their synthesis is a highly controlled process: target mRNA cleavage products, but not exogenously introduced dsRNA, are copied by the RdRP, which protects the cell from feed-forward amplification of siRNAs and potential off-target effects (PAK *et al.* 2012).

SiRNA pathways interact in a complex manner in various organisms. In nematodes for instance, RNAi is inducible exogenously by feeding the worms with dsRNA-producing bacteria (TIMMONS *et al.* 2001; TIMMONS and FIRE 1998), which appears to rely on an antiviral defence pathway (LU *et al.* 2005; WILKINS *et al.* 2005). This shares components with endogenous RNAi pathways (LEE *et al.* 2006), is negatively regulated by the endogenous Eri

pathway (KENNEDY *et al.* 2004), and competes with endogenous RNAi targets (SARKIES *et al.* 2013).

Remarkable complexity of sRNA pathways has also been observed in ciliates, unicellular eukaryotes separating their genome into somatic functions (macronucleus (MAC)), responsible for all gene expression, and a germline lineage (micronucleus). *Paramecium tetraurelia* mounts an RNAi response to exogenous dsRNA produced by food bacteria (GALVANI and SPERLING 2002). A forward genetic screen for mutants deficient in dsRNA-induced silencing, yielding a collection of complete and partial loss-of-function alleles (MARKER *et al.* 2014), showed that this pathway is non-essential and that it involves the non-essential genes *RDR1* (RdRP), *CID1* (nucleotidyl-transferase) and *PDS1* (unknown function), the essential genes *DCR1* (Dicer), *RDR2* and *CID2* (LEPÈRE *et al.* 2009; MARKER *et al.* 2014; MARKER *et al.* 2010). In addition, three proteins of the Piwi subclade of Argonaute proteins (*PTIWI12*, *PTIWI15* and *PTIWI13*) have been implicated in dsRNA-induced RNAi (BOUHOUCHE *et al.* 2011). Small-scale sRNA sequencing in conjugating cells suggested that two distinct siRNA species are produced within the dsRNA target region, one accumulating on both strands and likely representing primary siRNAs and another, showing a strong antisense bias, suggestive of secondary siRNAs (LEPÈRE *et al.* 2009). Intriguingly, two distinct RdRPs, Rdr1 and Rdr2, were implicated in exogenously triggered primary siRNA production (MARKER *et al.* 2010). However, secondary siRNA synthesis has not yet been confirmed in this or any other ciliate. Silencing by dsRNA in *P. tetraurelia* seems to operate at the post-transcriptional level, resulting in mRNA cleavage within the dsRNA target region (JAILLON *et al.* 2008). However, the natural function of this pathway is unknown; apart from antiviral defence (no *Paramecium* viruses are known so far), it may target other non-self RNAs (SARKIES and MISKA 2013) to which bacterial feeders are typically exposed.

Another constitutively expressed RNAi pathway can be experimentally induced by high-copy, non-translatable transgenes (GALVANI and SPERLING 2001; RUIZ *et al.* 1998). Transgene-induced silencing requires some of the genes involved in dsRNA-induced RNAi, such as *RDR2*, *CID2*, *DCR1* and *PTIWI13* (BOUHOUCHE *et al.* 2011; LEPÈRE *et al.* 2009; MARKER *et al.* 2014), but also additional genes such as *RDR3* and *PTIWI14*. *RDR3* was further shown to have endogenous functions, regulating a multigene family (MARKER *et al.* 2010). Furthermore, *RDR2* and *RDR3* are required for the production of endogenous siRNA from an intergenic locus of unknown function (MARKER *et al.* 2014; MARKER *et al.* 2010).

In this study, we have deep-sequenced the small RNAs produced in wild-type and mutant *Paramecium* strains subjected to dsRNA feeding with the aims (i) to test the previous hypothesis that a transitivity mechanism produces secondary siRNAs, (ii) to shed light on the mechanisms of primary and secondary siRNA synthesis by dissecting their genetic requirements, and (iii) to provide new insight into the natural functions of these pathways. Using dsRNA triggers that contain either an intron or a substitution to distinguish it from the endogenous target mRNA, we showed that although the vast majority of siRNAs are primary

siRNAs processed from the dsRNA, a minor fraction of secondary siRNAs are indeed produced from the entire mRNA, clearly demonstrating the existence of transitivity for the first time in any ciliate.

Materials and Methods

Paramecium strains, cultivation, induction of RNAi and phenotypic analyses

Experiments were carried out with wild-type strain 51 of *P. tetraurelia* and strain 51-derived RNAi deficient mutants (MARKER *et al.* 2014), backcrossed at least once. For standard cultures, cells were grown at 27°C in wheat grass powder (Pines International Co., Lawrence, KS, USA) infusion medium bacterized with *Klebsiella pneumoniae* the day before use and supplemented with 0.8 µg/ml β-sitosterol. *NDI69*, a non-essential, single-copy gene required for the discharge of secretory granules (trichocysts) was used as RNAi reporter gene. Silencing of *NDI69* results in a quantitative phenotype that allows distinguishing full and partial RNAi deficiencies. Production of dsRNA in *E. coli* strain HT115DE3, feeding to *Paramecium* cells and monitoring of trichocyst discharge phenotypes were carried out as described previously (MARKER *et al.* 2014; MARKER *et al.* 2010; SIMON *et al.* 2006). To reveal partial loss-of-function phenotypes, *NDI69* dsRNA producing *E. coli* were typically mixed in 1:5 ratio with *ICL7a* dsRNA producing bacteria, as described (MARKER *et al.* 2014)), unless otherwise stated. In order to reduce the amount of RNAi-unrelated small RNA fragments possibly present within dsRNA-producing *E. coli* food bacteria or *Paramecium* food vacuoles at the time of sampling, cells were first allowed to eat the majority of food bacteria. Then, prior to harvest for RNA preparation, cells were washed for 30min in *Klebsiella*-conditioned, exhausted medium (free of bacteria).

Plasmid constructs

To induce silencing by dsRNA feeding, fragments of the coding region had been cloned into the plasmid L4440. Plasmid constructs of the genes *NDI69*, *ICL7a*, *RDR3*, *PTIWI08* and *PTIWI14* were previously described (BOUHOUCHE *et al.* 2011; MARKER *et al.* 2010). DsRNA was synthesized in *E. coli* HT115 DE3 from T7 promoters as described (MARKER *et al.* 2010; SIMON *et al.* 2006).

Small RNA analysis by northern blot and high-throughput sequencing

Total RNA extraction and small RNA Northern blots were carried out as described (MARKER *et al.* 2010). To detect *ND169* siRNAs two adjacent 50nt sense oligonucleotide probes matching to the dsRNA region were used, as most of the siRNAs produced from this region are antisense to the target transcript.

For small RNA library construction, the ~19-28nt fraction was purified from 50µg of total RNA by polyacrylamide gel electrophoresis (15%, 19:1 acrylamide:bisacrylamide) and gel-eluted with 0.3M sodium chloride, followed by ethanol precipitation. The eluate was used for library construction using standard Illumina protocols, with the minimum recommended number of 11 PCR cycles.

Small RNA mapping and data analyses

Trimmed sequence data sets were mapped with the Burrows-Wheeler Alignment tool (BWA) (LI and DURBIN 2009) using default settings. Reads were first aligned to a set of bacterial sequences, as well as to mitochondrial and ribosomal DNA sequences, tRNA and snoRNA, according to the recent *Paramecium tetraurelia* genome annotation (ParameciumDB, <http://paramecium.cgm.cnrs-gif.fr/>), allowing one mismatch per read. Then, cleared data sets were aligned on the reference genome of strain 51 (ARNAIZ *et al.* 2012; AURY *et al.* 2006) without mismatch.

Read counts were normalized to the total number of sequence reads obtained per sample, expressed as reads per million. For determination of siRNA levels mapping within the dsRNA trigger region, the average of 22nt and 24nt normalized small RNA counts was defined as a threshold level to be passed for unambiguous siRNA accumulation. SiRNA levels represented in Fig.3C and D are corrected by this method. Outside the dsRNA region, 22nt, 23nt and 24nt size fractions were considered as siRNAs, as they were detectably dependent on Rdr2. SiRNAs mapping to the L4440 polylinker sequences, to the *ICL7a* mismatch, to introns and to exon-exon junctions were defined as those overlapping the respective sequence/nucleotide with a minimum of 3nt. SiRNAs carrying 3'-untemplated nucleotides were identified in sequential rounds of removing the last 3' nucleotide of each non-mapped small RNA, and re-aligning of these clipped reads to contaminants, and then on the MAC genome, as described above. Non-mapping reads of an alignment round were processed until 8 3' nucleotides were clipped.

For identification of MAC regions producing endogenous siRNAs, reads were aligned as described above, allowing no mismatch. Among those regions producing predominantly 23nt siRNAs, only clusters were considered that showed minimum siRNA coverage of 60x and a

minimum size of 250bp. Representations of siRNA distributions were made with R scripts, ggbio and Rsamtools packages (LAWRENCE *et al.* 2009; YIN *et al.* 2012).

Results

A large fraction of small RNAs map to genes targeted by dsRNA feeding

A previous small-scale analysis of siRNAs associated with dsRNA-induced silencing of the *ND7* gene in conjugating cells suggested the existence of two distinct classes: one that appeared to be cleaved from both strands of the dsRNA trigger and another one, exclusively antisense to the mRNA and carrying untemplated 3' polyA tails, which may represent secondary siRNAs (LEPÈRE *et al.* 2009).

To characterize dsRNA-induced siRNAs, we first carried out high-throughput sequencing of small (~15-30 nt) RNAs from 3 cultures of the wild-type strain 51 in which one or two non-essential genes were silenced. The single-copy gene *ND169*, involved in trichocyst discharge, was targeted either alone (WT-2 and WT-3 samples) or together with *ICL7a* (WT-1 sample), by feeding cells with a mix of the bacterial strains producing each dsRNA (Fig.1a and Fig.S1). *ICL7a* is a centrin gene which has an 93%-identical ohnolog from WGD1, the most recent whole-genome duplication. Small-RNA libraries were constructed to specifically capture 5'-monophosphorylated molecules, since previous northern blot analyses indicated that detectable siRNAs from the dsRNA sequence are monophosphorylated (MARKER *et al.* 2010). We obtained between 10 and 26 million reads for each library. After removal of reads derived from bacterial, mitochondrial and structural RNAs such as rRNA and tRNA (together representing between 54 and 88% of the total), the remaining ones were mapped to the MAC genome of strain 51, allowing no mismatch. This yielded numerous small RNAs matching the genes targeted by dsRNA in the relevant cultures (Table S1), but not in a control culture fed with the standard *Klebsiella pneumoniae* (not shown). The numbers of target-gene small RNAs varied between 15 000 and 94 000 reads per million (rpm), which represents a large fraction of MAC-mapping small RNAs in each case: 36% for *ND169* and 20% for *ICL7a* in sample WT-1 (Table S1). To broaden the analysis, we used small-RNA sequencing datasets from other dsRNA-feeding experiments: the single silencing of *RDR3* or *PTIW14*, both specifically involved in transgene- but not in dsRNA-induced RNAi (BOUHOUCHE *et al.* 2011; MARKER *et al.* 2010), *PTIW108* (a WGD1 paralog of *PTIW14*), or *ICL7a* (for gene details see Fig.S1). Similarly, large numbers of small RNAs specifically mapped to the cognate target gene in each case (Table S1).

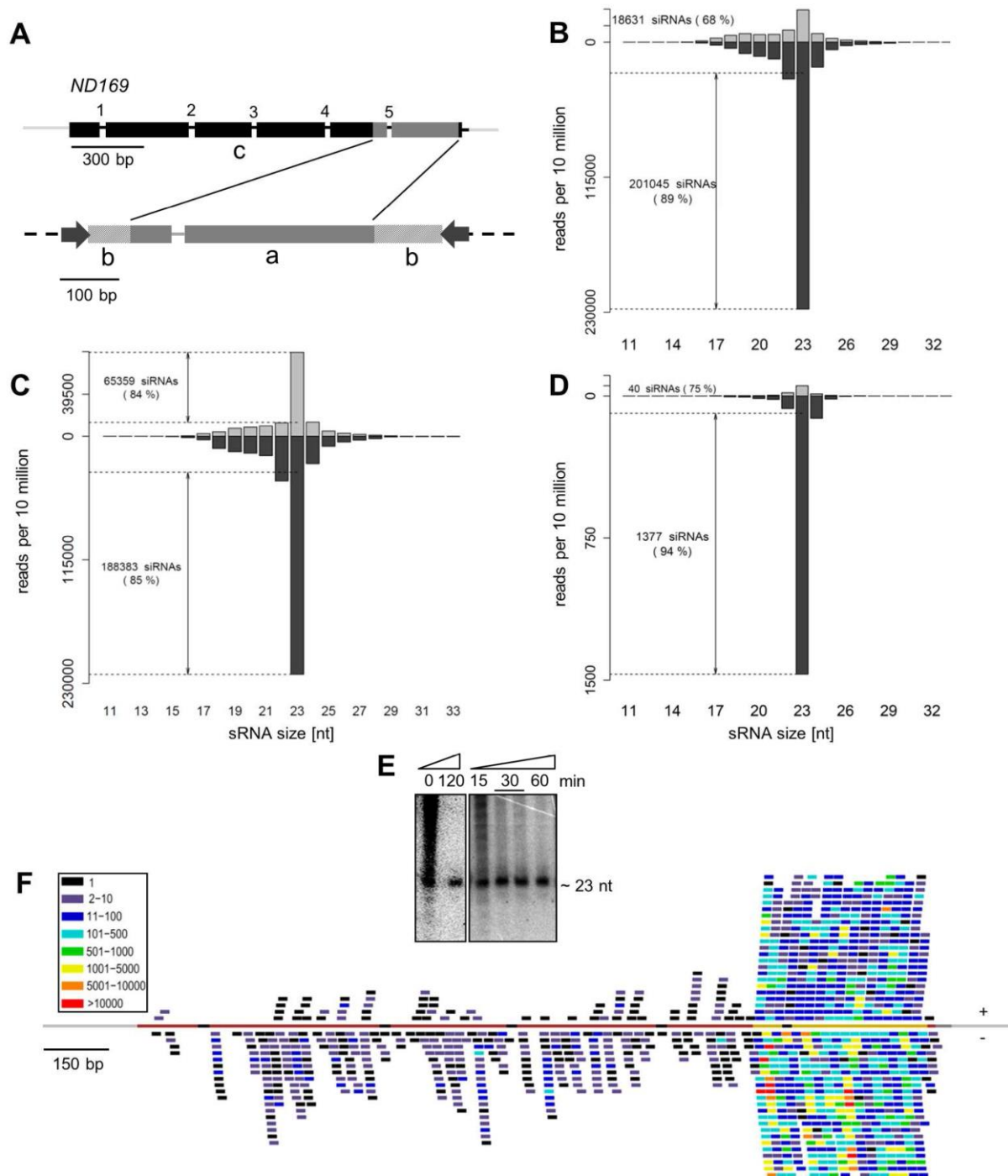


Fig.1: 23nt siRNAs are produced from both strands of the dsRNA region and spread along the target mRNA. (A) The *ND169* gene (top) was silenced using a 410bp dsRNA fragment, located in the 3' part of the coding sequence (c). Introns are numbered 1-5. The dsRNA fragment was cloned into the polylinker part of the L4440 vector (b), flanked by convergent T7 promoters (black arrows). The intron 5 sequence was present in the dsRNA (a). (B) Normalized size distribution of sense (grey) and antisense (black) small RNAs mapping to the *ND169* dsRNA sequence (region a in subfigure A) in reads per 10 million. Dashed lines represent threshold levels for accumulation of unambiguous 23nt siRNAs above the background (see text). (C) Normalized size distribution of small RNAs mapping to the top (grey) and bottom strand (black) of the L4440 5' and 3' vector sequence (b in subfigure A). (D) Normalized size distribution of sense (grey) and antisense (black) small RNAs mapping outside the *ND169* dsRNA sequence (c in subfigure A). No threshold level for unambiguous siRNAs was defined (see text). (E) Total RessaiNA from vegetative wild type cells subjected to *ND169* dsRNA

feeding was hybridized with a sense probe corresponding to the dsRNA region. Smear is detectable when cells were not allowed to complete digestion of dsRNA-containing food bacteria prior to RNA isolation. Samples of cultures washed and incubated for 0-120 minutes in *Klebsiella* standard food are shown. **(F)** Distribution of 23nt siRNAs on the *NDI69* genomic sequence (red line; black: introns; light grey: flanking intergenic regions; dark grey: 3'UTR; yellow: dsRNA target region) (not normalized). Individual siRNAs are represented as trait, the colour indicating its abundance. siRNAs represented above and below the line are sense and antisense to the *NDI69* mRNA, respectively. All subfigures show sample WT-1 (other samples see Fig.S6). 3'U-siRNAs are not included in the figure.

The vast majority of siRNAs map to the dsRNA trigger region of target genes

The 411-bp segment of the *NDI69* gene used for production of the dsRNA trigger is located near the 3' end of the coding sequence (Fig.1A). In addition to this segment, the dsRNA trigger contains short sequences derived from the plasmid vector, between the convergent T7 promoters and flanking the restriction sites used for cloning (71 and 114 bp at the left and right ends, respectively). By definition, siRNAs matching these vector sequences, which are not present in the *P. tetraurelia* genome, can only be primary siRNAs, *i. e.* siRNAs cleaved by Dicer from the dsRNA trigger.

The mapping of small RNAs perfectly matching *NDI69* revealed that 99% of them are located in the dsRNA trigger region (details see Table S2A), and their size distribution shows a strong peak at 23 nt (54% of *NDI69* small RNAs on average) (Fig.1B). In addition, a large number of small RNAs was found to map to the vector parts of the dsRNA (54% of dsRNA-mapping reads in WT-1, 57% in the 23-nt peak) (Fig.1C). Small RNAs of other sizes between 16 and 29 nt, as well as a fraction of 23-nt molecules, may not be siRNAs but likely represent random degradation products from the dsRNA being digested in food vacuoles. They were readily detected on northern blots when cells were not allowed to complete digestion of ingested bacteria before RNA extraction (Fig.1E). To count the 23-nt siRNAs in a stringent manner, the number of RNAi-unrelated 23-nt RNAs was estimated to be the mean of the numbers of 22- and 24-nt molecules, and only those in excess of this value were counted as unambiguous siRNAs.

Both sense and antisense siRNAs were found in the *NDI69* dsRNA region, though an antisense bias was noted (88% of 23-nt siRNAs on average). Strikingly, some antisense siRNAs were highly overrepresented, with more than 560 rpm, whereas most others were represented by less than 30 rpm (Fig.1F). These copy-number variations are highly reproducible for a given sequence since the same *NDI69* siRNAs were overrepresented in the three wild-type samples (Fig.S2), and they fully account for the global antisense bias in the dsRNA region (Fig.1F). Among other genes tested, dramatic overrepresentation of some siRNAs was also found to result in a global strand bias in the *PTIWI08* and *PTIWI14* dsRNA regions, whereas overrepresented siRNAs occur equally on both strands for *RDR3* and *ICL7a* (Fig.S3 and Table S2A). Similarly, siRNAs mapped in equivalent numbers to both strands of

the right vector part of the *NDI69* dsRNA trigger, whereas the left vector part showed an excess of reads on the bottom strand (85%) (Fig.S4). Thus, read-count heterogeneity appears to be a local, sequence-dependent effect that does not necessarily result in any strand bias. No particular feature (nucleotide position biases, GC- or purine content) was found to be associated with high-copy reads, and it remains unclear whether the observed heterogeneity is of technical or biological origin, or both. We conclude that siRNAs accumulate at high levels on both strands of the entire dsRNA.

Secondary siRNAs are produced from the entire target transcript

About ~1% of *NDI69* small RNAs mapped outside the dsRNA region, covering the entire mRNA including the 33-nt 3'UTR (Fig.1F), but none was found in the 500 bp immediately upstream and downstream of the *NDI69* transcript, even in the genes located there. Thus, dsRNA feeding induces the production of secondary siRNAs that spread beyond the trigger region but seem to be produced exclusively from the target transcript. Their lower abundance and spreading in both directions was also observed for the other target genes, in which the dsRNA trigger regions were located at different positions relative to the coding sequence (*RDR3*, central dsRNA fragment; *PTIWI14* and *PTIWI08*, near the 5' end; and *ICL7a*, covering the entire cds, but not the 3'UTR) (Fig.S3 and Table S2A). 86% of *NDI69* small RNAs outside the dsRNA region are 23 nt in length, indicating that they represent *bona fide* siRNAs, and the 22-nt and 24-nt molecules (12%) are also likely to be secondary siRNAs since they depend on Rdr2 (see below; Fig.S6B). Very few molecules were found at other sizes between 16 and 29 nt, confirming that the high background observed in the trigger region is due to dsRNA degradation products.

NDI69 secondary siRNAs exhibit a strong antisense bias (97% on average in WT samples) (Fig.1D and Table S2A). Unlike the weaker biases seen in the dsRNA region, which can favor either strand, the same strong antisense bias was observed for all genes tested outside the dsRNA region (Table S2A). In all cases, some secondary siRNAs were highly overrepresented, as seen in the dsRNA regions (Fig.S2). The antisense bias was observed both 5' and 3' of the dsRNA region and appears to be a characteristic feature of secondary siRNAs, which is reminiscent of the situation in *C. elegans* (PAK and FIRE 2007; SIJEN *et al.* 2007). Yet the small but significant fraction found on the sense strand, which was not observed when cells were grown on *Klebsiella* or upon feeding of other dsRNAs, suggests that at least some secondary siRNAs may derive from a dsRNA precursor. Both sense and antisense secondaries show a random frequency of each nucleotide at the 5' end (not shown).

Outside the dsRNA region, many *NDI69* siRNAs contained exon-exon junctions (213 rpm on average in WT samples), indicating that they were produced from the spliced mRNA. Secondary siRNAs containing intron sequences were much rarer (7 rpm on average) (Table 1). From these analyses we conclude that dsRNA feeding induces the accumulation of

predominantly antisense secondary siRNAs over the entire length of the target transcript, at much lower levels than in the dsRNA trigger region, and mostly produced from the spliced mRNA.

siRNAs from the dsRNA region are predominantly primary siRNAs

SiRNAs mapping in the dsRNA region may be either primaries or secondaries. To determine the proportions of each, we took advantage of the fact that the *NDI69* dsRNA contains one 23-nt intron (intron 5) (Fig.1A). In contrast to introns outside the dsRNA region, intron 5 was highly covered by siRNAs (1,311 and 4,989 rpm on average for sense and antisense strands, respectively), whereas very few siRNAs contained the exon-exon junction (1.3 and 1.7 rpm on average) (Table 1 and Fig.2A). Intron 5 is very efficiently spliced, as determined by the comparison of RNAseq data from wild-type and NMD knockdown conditions (Saudemont et al., unpublished), which would suggest that the vast majority of siRNAs in the dsRNA region are primaries. However, the intron 5 region is poorly covered compared to other sequences of the dsRNA trigger. To confirm this result independently of any possible effect of dsRNA feeding on splicing, we counted siRNAs covering a point substitution present in the *ICL7a* dsRNA fragment, and those containing the endogenous base (Fig.2B and C). 98.6% (11,445 rpm sense and antisense on average in two replicates) of siRNAs contained the substitution versus 1.4% (164 rpm) that matched the endogenous transcript. Taken together, these results indicate that secondary siRNAs are indeed produced from the dsRNA region of the mRNA, where they accumulate at the same relatively low level as over the rest of the transcript; the vast majority of siRNAs mapping in the dsRNA region are primary products of dsRNA cleavage by Dicer.

Sample	outside dsRNA region				dsRNA region			
	intron 1 - 4		junction 1 - 4		intron 5		junction 5	
	sense	antisense	sense	antisense	sense	antisense	sense	antisense
WT-1	1	5	0	116	936	3.858	1	1
WT-2	0	5	1	135	1.087	8.665	1	2
WT-3	0	9	2	385	1.910	2.443	2	2
WT- <i>Kleb</i>	0	0	0	0	0	0	0	0

Table 1: Number of *NDI69* siRNAs (in reads per million) overlapping exon-exon junctions and mapping introns in wild type samples (WT) samples fed with *E. coli* (*NDI69* dsRNA) or *Klebsiella* (no dsRNA).

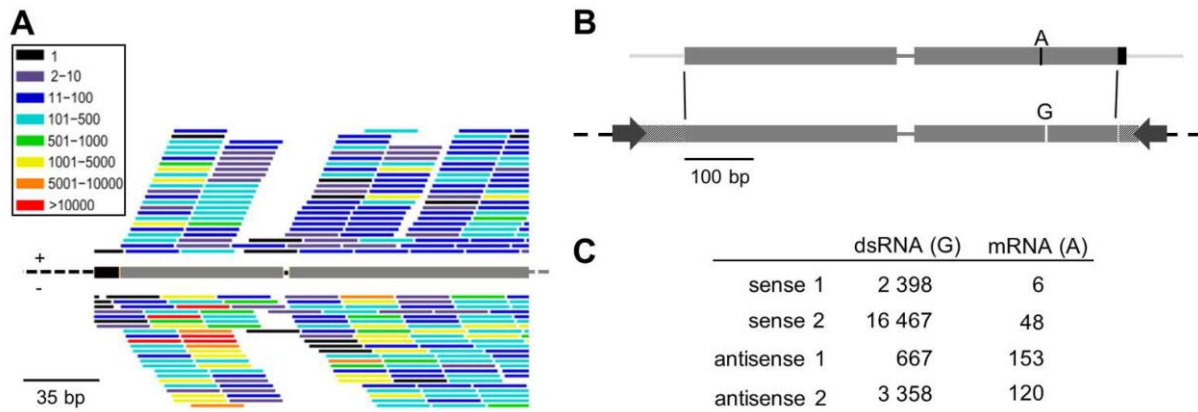


Fig.2: Predominantly primary siRNAs are produced from the dsRNA target region.

(A) Very few siRNAs overlap the exon-exon junction 5 (black gap) located in the *ND169* dsRNA region (grey bar). The graphic representation is identical to Fig.1F. The black bar represents the 5' vector. (B) An A to G mismatch is located in the *ICL7a* dsRNA (lower grey bar; dashed bars correspond to the 5' and 3' vector); the dsRNA corresponds to the bulk of the *ICL7a* cds (upper bar, black parts are not represented in the dsRNA). (C) SiRNAs mapping to the mismatch represent primary siRNAs, whereas those overlapping the respective genomic sequence represent secondaries, produced from the endogenous transcript. Numbers in rpm are given for replicate 1 (WT-1) and 2 (*ICL7a* single dsRNA feeding). 3'U-siRNAs are not included in the figure.

Primary siRNA accumulation is strongly affected in mutants

We then sequenced small RNA pools in mutants of the dsRNA-induced RNAi pathway fed with *ND169* dsRNA (Fig.S5). Only hypomorphic missense alleles were available for the essential genes *DCR1* (*dcr1-5.5* and *dcr1-5.27*) and *RDR2* (*rdr2-1.24* and *rdr2-3.7*); these mutations only partially impair silencing efficiency, requiring dilution of the dsRNA trigger to detect the phenotype (MARKER *et al.* 2014). For *RDR1*, we analysed two strong alleles resulting in complete silencing deficiency (*rdr1-5.28*, nonsense; *rdr1-3.16*, catalytic missense) and a hypomorphic allele (*rdr1-1.22*). We further studied complete-deficiency missense alleles of the nucleotidyl-transferase *CID1* and the RNAi factor *PDS1* (*cid1-1.8*, *cid1-5.14*, *pds1-3.18* and *pds1-5.18*) as well as a nonsense *PDS1* allele (*pds1-3.17*) and a hypomorphic *CID1* allele (*cid1-3.20*) (Table S3). Northern blots previously showed that the phenotypic strength of these mutations correlates with the reduction of siRNA levels in the dsRNA region (MARKER *et al.* 2014) (Fig.3A).

Accordingly, the 23-nt peak of *ND169* siRNAs mapping to the dsRNA region collapsed more or less strongly in the different mutants (Fig.3B and Fig.S6). Of note, normalized siRNA levels varied about 3-fold among wild-type samples (Fig.3C), although silencing was fully efficient in all 3 cases. This may be due to differences in the amount of dsRNA produced by bacterial cultures after IPTG induction and/or to differences in the washing and starvation of cells before RNA extraction (see Mat. & Met.). Compared to the average of wild-type samples, 23-nt siRNA counts were reduced between 17- (*pds1-5.18*) and 250-fold (*rdr1-5.28*)

in mutants with a full phenotype, and between 2- (*rdr2-3.7*) and 11-fold (*rdr1-5.5*) in hypomorphic mutants (Fig.3C). Consistent with the conclusion that most dsRNA-region siRNAs are primaries, siRNAs mapping to the vector parts of the dsRNA were reduced to a similar extent in each mutant (Fig.3D). The 2-fold reduction in *rdr2-3.7*, the most hypomorphic of the two *RDR2* alleles, seems not significant given the variations observed among wild-type samples, but this is consistent with the weak effect of the mutation on silencing efficiency, which requires dsRNA dilution to see the phenotype (MARKER *et al.* 2014). Overall, these results confirm that all factors are involved in the accumulation of primary siRNAs, including the two RdRPs. We further note that the 23-nt peak is completely abolished in the nonsense *RDR1* allele (*rdr1-5.28*) (Fig.3B), indicating that primary siRNAs cannot be produced in the absence of the Rdr1 protein.

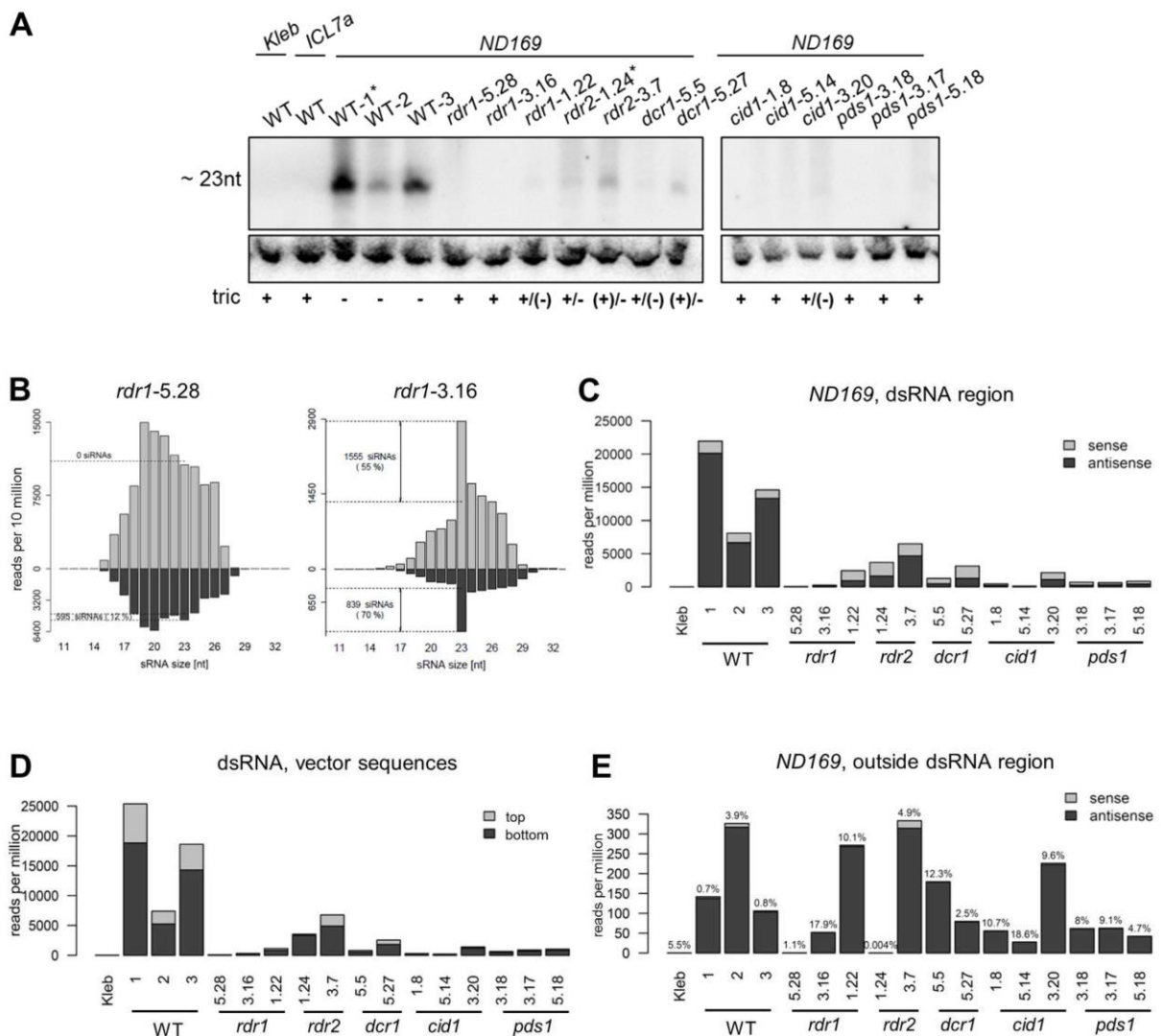


Fig.3: SiRNA accumulation is strongly affected in RNAi mutants. (A) Northern blot analysis of siRNAs produced in vegetative wild type cells and RNAi-deficient mutants subjected to *ND169* (and *ICL7a* in WT-1 and *rdr2-1.24* (*) dsRNA feeding. A probe corresponding to a 100bp fragment of the dsRNA and oriented sense to the mRNA was used. ~2.5 fold more siRNAs mapped to the dsRNA in WT-1 compared to WT-2, although *ND169* food bacteria were diluted with *ICL7a* bacteria

(*ND169:ICL7a 2.5:1*). Of note, the two strongly hypomorphic alleles *rdr2-3.7* and *dcr1-5.27* produced nearly wild type levels of siRNAs. Trichocyst discharge phenotypes are given below. **(B)** Normalized size distribution of small RNAs mapping the *ND169* dsRNA region in wild type (WT-1) and *rdr1-5.28*. Note that the absolute number of 23nt small RNAs is higher in the *rdr1* null mutant, but no unambiguous siRNAs accumulating above the background threshold level were found. **(C)** Normalized counts of sense (grey) and antisense (black) 23nt siRNAs mapping the dsRNA region in wild type (WT) and mutant samples (in reads per million). Counts were corrected by the threshold definition for unambiguous siRNAs described in the text. Corrected siRNA numbers and their percentage relative to the total number of 23nt small RNA reads are indicated. **(D)** Normalized counts of 23nt siRNAs mapping the 5' and 3' vector part of the dsRNA sequence. The same correction as described for (C) was used. **(E)** Normalized counts of 23nt siRNAs mapping outside the *ND169* dsRNA region in WT and mutant samples (in reads per million). Read counts were not corrected, as no background level due to experimentally induced dsRNA degradation was expected. The percentage of 23nt secondary siRNAs (i.e. mapping outside the dsRNA region) relative to the corrected total number of 23nt siRNAs is indicated on top of the bars. Note that the proportion of secondary siRNAs varied 4-fold between wild type samples (0.4% in WT-1 and WT-3; 1.6% in WT-2).

Rdr2 is required for accumulation of secondary siRNAs

In contrast to primary siRNAs, secondary siRNAs were hardly affected in most mutants. In *dcr1* mutants (*dcr1-5.5* and *dcr1-5.27*, both missense mutations in the helicase domain), the number of *ND169* 23-nt secondary siRNAs was not significantly reduced and the antisense bias was not affected. Likewise, in *cid1*, *pds1* and two of the *rdr1* alleles (*rdr1-3.16* and *rdr1-1.22*), secondary siRNA levels were not or only moderately (3-7-fold) reduced (Fig.3E and Table S4A). Thus these genes may not be directly involved in the synthesis of secondary siRNAs; where observed, the modest reduction may simply be due to a stronger depletion of primary siRNAs, which are needed to trigger production of secondaries. Only in the case of the nonsense mutant *rdr1-5.28* were secondary siRNAs completely lost, which may be linked to the complete absence of 23-nt primaries. Under this hypothesis the substantially reduced levels of primary siRNAs in other mutants, though precluding efficient silencing of *ND169*, would still be sufficient to induce near-wild-type levels of secondaries.

The strongest allele of *RDR2* (*rdr2-1.24*) behaved differently: despite a considerable level of primaries, secondary siRNAs were almost completely absent (250 fold compared to the average level of WT samples) (Fig.3E, Fig.S6B and Fig.S7A). The few reads counted in *rdr2-1.24* and *rdr1-5.28* are within the range of sample contamination, as determined from the control sample WT-K, not fed with *ND169* dsRNA (10 reads in WT-K, 9 in *rdr1-5.28* and 4 in *rdr2-1.24*). Thus, *Rdr2* appears to be responsible for the synthesis of secondary siRNAs.

The occurrence of 3' transitivity suggests that primary siRNAs do not act as primers for the synthesis of a long dsRNA precursor, which is reminiscent of the Dicer-independent, 5' triphosphate secondaries in *C. elegans* (AOKI *et al.* 2007; MANIAR and FIRE 2011; PAK and FIRE 2007; PAK *et al.* 2012; SIJEN *et al.* 2007). Although previous analyses only detected 5' monophosphate siRNAs in the *ND169* dsRNA region (MARKER *et al.* 2010), we asked

whether a significant amount of 5' triphosphate molecules, which would not be included in our libraries, can be detected in other regions of the mRNA. Northern blots probed with *ND169* sequences upstream of the dsRNA region revealed at best trace amounts of siRNAs (Fig.S8), ruling out an important fraction of 5' triphosphate molecules and suggesting that the sequenced secondaries represent the actual pool.

A fraction of primary and secondary siRNAs is 3'-uridylylated

To check for possible 3' addition of untemplated nucleotides, reads that could not be mapped in the above analysis were trimmed at the 3' end by one nucleotide and mapped again, allowing no mismatch. The procedure was repeated up to 8 times. This revealed the presence of one or several non-genomic 3' uridines (U) in a small fraction of small RNAs mapping within and outside the *ND169* dsRNA, and of small RNAs mapping to dsRNA vector sequences (8%, 4% and 5% on average in WT cells, respectively). The tail size varied from one to eight Us, but more than 50% were modified with a single U. Their size distribution shows a peak at 23 genomic nucleotides (Fig.S9), indicating that they represent modified primary and secondary siRNAs. Other non-templated nucleotides (mono- or poly-G,-C, or -A) were observed at much lower frequencies compared to Us (Fig.4A). The very low proportion of A deviates from the previous observation that a substantial fraction of dsRNA-induced and other siRNAs in conjugating cells (50%) carried 1-7 nt 3' polyA tails (LEPÈRE *et al.* 2009). Curiously, primary and secondary 3'U-siRNAs were reproducibly more abundant on the sense strand of *ND169* (77% and 72% of the total 23nt-3'U primary and secondary siRNAs on average in WT samples, respectively) (Fig.4 and Table S4B), indicating that they are not subjected to the strong antisense bias observed for non-modified siRNAs. However, 3'U-siRNAs of other target genes occurred in similar strand distribution as non-uridylylated siRNAs (Table S2B and C), and showed lower abundances ($\leq 7\%$ of 23nt-3'U siRNAs vs. 9.7% on average for *ND169*) (Table S2C), indicating a potential sequence-specific effect. In mutants, the levels of primary 3' uridylylated siRNAs (*ND169* dsRNA region and vector sequences) were reduced to a similar extent as non-modified primary siRNAs (Fig.S9). Likewise, 3'U secondary siRNAs show the same genetic requirements as non-modified ones, i.e. are not dependent on Rdr1 catalytic activity, Cid1, Dcr1 and Pds1, but on Rdr2. However, Cid1 does not seem to be involved directly in the uridylation of these siRNAs, since the proportion of 3'U siRNAs in the residual fraction of primary and secondary siRNAs was not reduced in *cid1* mutants (Fig.S9).

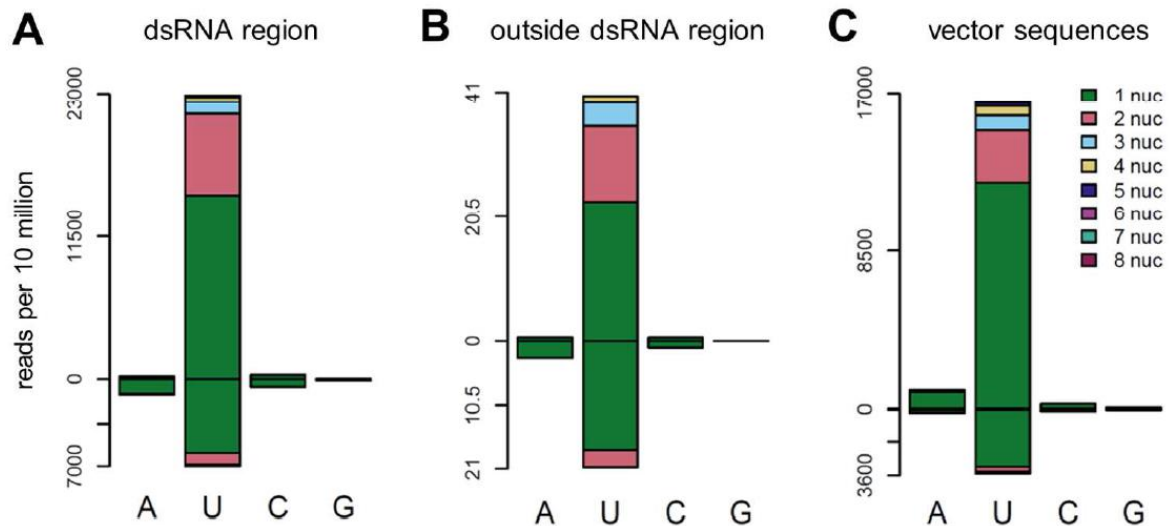


Fig.4: A subclass of dsRNA-induced siRNAs carries 3'-untemplated uridines. Number of sense siRNAs (top) or antisense siRNAs (bottom) in the WT-1 sample modified with one or several 3' untemplated nucleotides (A, U, C or G) mapping the *ND169* dsRNA region (A), outside the dsRNA (B) or the 5' and 3' L4440 vector sequences (C). Colours indicate the number of 3' untemplated nucleotides. Only modified 23nt siRNAs are shown, i.e. >23nt.

Endogenous siRNA clusters depend either on Rdr1 and Cid1 or on Rdr2

Rdr2 was previously shown to be involved in the production of endogenous siRNAs of unknown function, clustering in an intergenic region (cluster22) (MARKER *et al.* 2014). To gather insight into the endogenous roles of proteins involved in the exogenously induced RNAi pathway we looked for spontaneous production of 23-nt siRNAs from the MAC genome in wild-type and mutant cells, comparing only samples subjected to dsRNA feeding. We considered only clusters of small RNAs showing a 23-nt peak, regions longer than 250bp, and a minimum coverage of 60x on average in WT samples to allow for comparison with mutants. This revealed a non-exhaustive list of 10 clusters (Table 2 and Fig.S10) which showed strong reduction of small RNA count in at least one mutant. This, together with their predominant size of 23 nt, indicates that they represent *bona fide* siRNAs.

Clusters fell mainly in two categories. In the first one the number of siRNAs was significantly reduced in the strongest *rdr2* allele (*rdr2*-1.24), but not in other mutants (the known cluster 22, as well as clusters 17 and 51 (Fig.5A, B and C)). Two of these (cluster 17 and 51) overlap exactly annotated mRNAs, with a strong antisense bias (>95%) similar to Rdr2-dependent *ND169* secondaries. Thus, Rdr2 seems to be able to produce siRNAs independently of Rdr1-dependent primary siRNAs. Clusters of the second group were only located in intergenic regions, and for five of them transcripts were detected by directional RNA-Seq (Table 2). siRNAs mapped equally to both strands or showed a minor bias to one strand. SiRNAs from these clusters were less abundant in *rdr1* and *cid1* mutants (*rdr1*-5.28, *rdr1*-3.16, *cid1*-1.8,

cid1-5.14) (clusters 79, 47, 121, 110, 160, 112 and 143) (Fig.5 D-J), indicating a concerted action of the two proteins.

Interestingly, some Rdr1/Cid1-dependent siRNA clusters are clearly overrepresented in the *rdr2-1.24* mutant (clusters 79, 47, 112). This may indicate that Rdr2 plays a role antagonistic to that of Rdr1 in the accumulation of these siRNAs.

Similarly to dsRNA-induced siRNAs, a minor fraction of endogenous cluster siRNAs was found to carry untemplated 3'Us (7.7-28% in intergenic clusters and 1.3-2.6% in gene mapping clusters) (Fig.S10K). We conclude that the genes involved in the dsRNA-induced RNAi mechanism possess endogenous functions in siRNA accumulation at specific cluster regions. Endo-siRNAs exhibit similar properties to experimentally-induced ones, as they consist of 3' unmodified and a minor fraction of 3'U siRNAs, and have overlapping genetic requirements. However, no cluster appeared to be affected by *pds1* mutations.

Cluster	length [bp]	gene annotation (ParameciumDB)	transcript count		mean siRNA coverage/bp		% siRNAs on strand -
			+	-	strand +	strand -	
17	982	GSPATT00007066001	516	8	13	443	97
51	3021	GSPATT00017018001	474	0	0	20	98
22	331	intergenic	466	2	1545	3540	70
79	440	intergenic	28	0	70	106	60
47	611	intergenic	0*	0*	7	38	84
121	1007	intergenic	24	0	403	2567	86
110	275	intergenic	16	0	515	1801	78
160	412	intergenic	22	0	153	166	52
112	303	intergenic	4*	4*	56	105	65
143	545	intergenic	128	0	16	18	53

Table 2: Genomic cluster loci producing endogenous 23nt siRNAs. Overlapping poly(A)-containing transcripts were identified using directional RNA-Seq data sets (normalized values). + and – strands of siRNAs were defined according to these transcripts, but only represent top and bottom strands for the 2 clusters with no clearly defined single-stranded transcripts (*). SiRNA coverages are mean values of WT1-3.

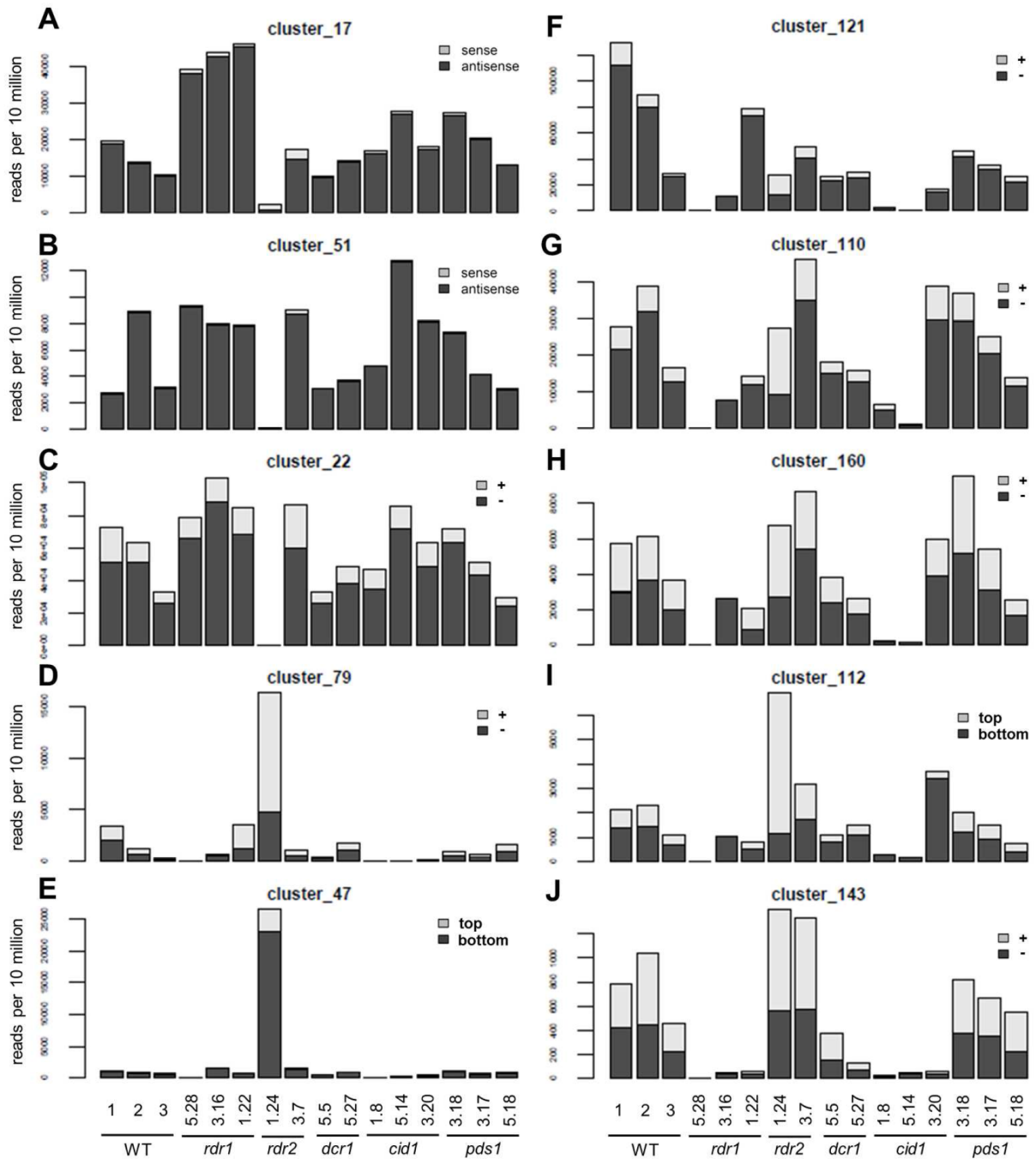


Fig 5: Endogenous siRNA clusters dependent of either Rdr2 or Rdr1 and Cid1. (A and B) Abundance of endogenously produced 23nt siRNAs mapping to coding genes in wild type cells and RNAi-deficient mutants. (C-J) Endogenous 23nt siRNAs mapping to ncRNAs (see Table 2). SiRNAs oriented sense/+ to mRNAs/nc-transcripts are represented in grey, those oriented antisense/- are shown in black. SiRNA size distributions for each cluster are shown in Fig.S10.

As a preliminary test to determine whether the uptake of RNA from food vacuoles and its processing by RNAi is restricted to dsRNA, the reads were mapped to *E. coli* 16S and 23S rRNA sequences without pre-mapping to other sequences, allowing no mismatch. In wild-type samples, variable quantities of small RNAs corresponding to the sense rRNA transcripts were identified (Fig.6). Variations are likely due to differences in cell starvation and washes prior to RNA extraction (see Mat. & Met.), leading to variable densities of food bacteria. The size distribution of sense-rRNA small RNAs did not show any specific pattern, consistent with most of these molecules representing rRNA fragments being degraded. Intriguingly however, a number of rRNA matches were antisense to the rRNAs, and their size distribution showed a clear peak at 23 nt. These antisense small RNAs were much less abundant than sense fragments, but they likely represent *bona fide* siRNAs synthesized by the *Paramecium* RNAi machinery: indeed, the same analysis performed on the mutant data sets revealed that the antisense rRNA count is strongly reduced in the *rdr2*-1.24 mutant, and the remaining molecules do not show a peak at 23 nt (Fig.S12). It is more difficult to conclude about a possible effect of other mutations: although antisense rRNA counts are also reduced in *rdr1*-5.28 and *dcr1*-5.27, all mutants except *rdr2*-1.24 still show a clear peak at 23 nt, specifically in the antisense distribution, calling for more carefully controlled experiments to interpret these quantitative variations.

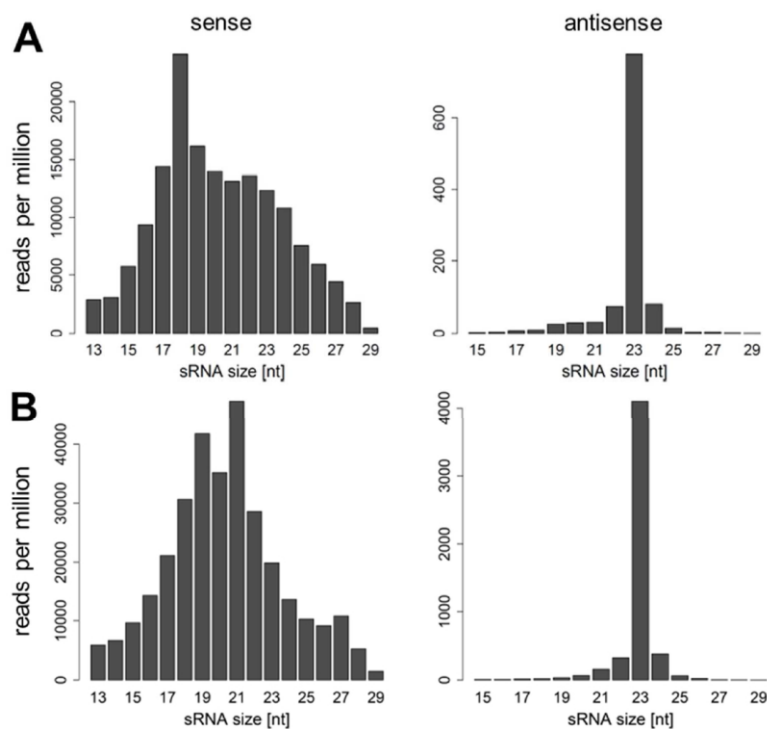


Fig. 6: 16S and 23S *E. coli* rRNA-derived siRNAs accumulate in wild type *P. tetraurelia* cells. Size distribution and normalized abundance of sense (left) and antisense (right) oriented small RNAs are shown (sample WT-1). (A) 16S-rRNA mapping, (B) 23S-rRNA mapping small RNAs.

Discussion

Mechanisms of primary siRNA synthesis

The sequence data sets from various mutants, including both complete loss-of-function and hypomorphic alleles, allowed us to confirm previous conclusions from northern blot analyses that siRNAs matching the dsRNA trigger, here shown to be overwhelmingly primaries, require the *DCR1*, *RDR1*, *RDR2*, *CID1*, and *PDS1* genes for their synthesis or accumulation at wild-type levels. While it seems reasonable to assume that Dcr1 cleaves the dsRNA to yield 23-nt primary siRNAs, the reason why two different RdRPs should also be required is intriguing. Primary siRNAs were completely absent in the nonsense mutant *rdr1-5.28*, indicating that the Rdr1 protein is strictly required, and that Rdr2 alone is not sufficient. The residual amount of primary siRNAs observed in other *RDR1* alleles therefore indicates that the mutant proteins are still partially functional in this respect. In the *rdr1-3.16* mutant, a highly conserved Asp residue thought to be required for RdRP catalytic activity (CURABA and CHEN 2008; MAKEYEV and BAMFORD 2002; SUGIYAMA *et al.* 2005) is changed to Tyr, raising the possibility that Rdr1 plays a structural role rather than a catalytic one. It could for instance associate with Dcr1, as shown for the homologous proteins in *Tetrahymena* (LEE and COLLINS 2007; TALSKY and COLLINS 2010), in a complex that would be required for Dcr1 activity or recruitment to the dsRNA substrate. However, we cannot exclude that *rdr1-3.16* retains some catalytic activity; RdRP activity might be required to amplify the dsRNA trigger, if it is imported from the food vacuole in very small amounts. This would be in contrast to *C. elegans*, where ingested dsRNA is not amplified (PAK *et al.* 2012). Furthermore, the silenced phenotype of dsRNA-fed cells rapidly reverts when they are returned to normal *Klebsiella* food, which argues against long-lasting dsRNA replication. Alternatively, the bacterial dsRNA could be imported from the food vacuole in single-stranded form; RdRP activity would then be required to produce a genuine dsRNA trigger, as demonstrated or inferred in other systems (DALMAY *et al.* 2000; GENT *et al.* 2010; LEE *et al.* 2010; MOURRAIN *et al.* 2000; VASALE *et al.* 2010).

The requirement for *RDR2* may be less stringent, since only the most compromised allele *rdr2-1.24* showed reduced primary siRNA levels. However, only hypomorphic alleles were available for this essential gene; further work will be required to test primary siRNA production in the absence of any Rdr2 protein, and to determine the respective roles of the two RdRPs in this process.

Primary siRNAs were severely depleted in *cid1* full-deficiency mutants. Nucleotidyl-transferases have similarly been implicated in RNAi pathways in *S. pombe* (MOTAMEDI *et al.* 2004; VOLPE *et al.* 2002), *C. elegans* (CHEN *et al.* 2005; JOSE *et al.* 2011), and *T. thermophila* (COUVILLION *et al.* 2009; LEE and COLLINS 2007; LEE *et al.* 2009), where they are required

for accumulation of specific siRNAs and in several cases have been shown to physically interact with RdRPs in RNA-dependent RNA polymerase complexes (RdRCs) (LEE *et al.* 2009; MOTAMEDI *et al.* 2004; TALSKY and COLLINS 2010; VAN WOLFSWINKEL *et al.* 2009). *In vitro* tests of *T. thermophila* RdRCs with ssRNA templates indicated that RdRP activity is primed by looping of the 3' end, and that addition of polyU tails by the Rdn nucleotidyl-transferases enables full-length conversion to dsRNA (TALSKY and COLLINS 2010). Such a mechanism might explain the role of Cid1 in primary siRNA synthesis in *Paramecium*, if RdRP catalytic activity is required. However it may be, the Cid1 nucleotidyl-transferase activity is apparently not responsible for the observed uridylation of siRNAs (see below), and the parallel effects of *cid1* and *rdr1* mutations on all siRNAs examined, including endogenous siRNAs, strongly suggest that the two proteins are obligate partners in an RdRC complex (Fig.7).

Mutations in the *Paramecium*-specific *PDS1* gene also strongly decreased the levels of dsRNA-induced primary siRNAs, but unlike *cid1* and *rdr1* mutations they did not affect any of the clusters of endogenous siRNAs examined. The Pds1 protein thus appears to be specifically involved in exogenously triggered RNAi, suggesting it may participate in the import of bacterial dsRNA from the food vacuole.

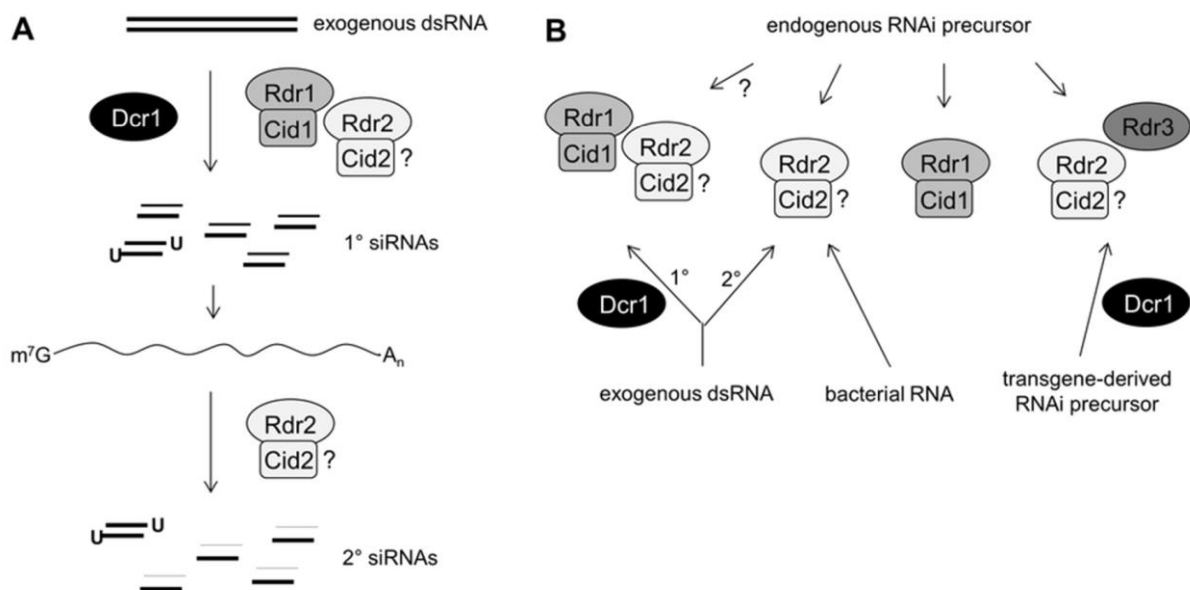


Fig. 7: RdRP functions in *P. tetraurelia* exogenous and endogenous RNAi pathways. (A) Two distinct RdRPs co-operate and act sequentially in the biosynthesis of siRNAs from exogenous dsRNA. Whereas accumulation of both, primary and secondary siRNAs requires Rdr1 and Rdr2, the production of secondaries is only dependent on Rdr2, provided primary siRNAs accumulate above a threshold level. Note that complex formation of RdRPs with Cid nucleotidyl transferases (RdRCs) (see text) is hypothetical. The role of the essential Cid2 was not tested in this study. (B) SiRNAs produced from different types of endogenous RNAi precursors require distinct, but overlapping RdRP activities, creating limiting conditions in the pathway network. Rdr2 and the putatively non-catalytic RdRP Rdr3 act together with Dcr1 in transgene-induced silencing (Marker *et al.* 2010, Marker *et al.* 2014). Some endo-siRNAs seem to be produced independently of Rdr2 by Rdr1-Cid1, or vice versa. Note that Rdr1-Rdr2 co-operation may also occur in endo-siRNA biogenesis. Only confirmed roles of Dcr1

(LEPÈRE *et al.* 2009; MARKER *et al.* 2014) and Rdr3 (MARKER *et al.* 2010) are shown.

Mechanisms of secondary siRNA synthesis

By definition, secondary siRNAs are synthesized from an mRNA as a consequence of its targeting by primary siRNAs. Any mutation that completely abolishes primary siRNA production should therefore also result in the lack of secondaries, as observed for the nonsense mutant *rdr1-5.28*. However, most other mutants, including the other two *rdr1* alleles, still produce a residual amount of primary siRNAs. This allowed us to show that these mutations have little or no impact on secondary siRNA levels: in cases where a small reduction was observed, it may be explained by a stronger depletion of primaries. The only exception is *rdr2-1.24*: despite a significant amount of residual primaries, secondary siRNAs were completely absent in this mutant. Our results thus implicate only Rdr2 in their production, likely through its catalytic activity (Fig.7A). Spreading of siRNA production beyond the dsRNA trigger involves RdRP activity in other systems (BAULCOMBE 2007; VOINNET 2008). In both plants and nematodes, primer-independent RdRP initiation permits 5'-to-3' transitivity (SIJEN *et al.* 2007; VAISTIJ *et al.* 2002). Similarly, the occurrence of 5'-to-3' transitivity in *P. tetraurelia* implies that the synthesis of antisense RNAs by the Rdr2 protein is not primed by primary siRNAs. Rdr2 activity may also require association with a nucleotidyl-transferase in an RdRC complex; previous recursive RNAi tests of 4 close paralogs of *CID1* identified *CID2* as a possible candidate: like *RDR2*, this gene is involved in both dsRNA-induced and transgene-induced RNAi, and may be an essential gene since no viable mutant was recovered from the screen (MARKER *et al.* 2014).

Unlike primary siRNAs, secondary siRNAs show a strong antisense bias for all target genes tested and were not significantly depleted in the two hypomorphic *dcr1* alleles tested. This suggests that a large fraction could be produced by a Dicer-independent mechanism similar to that described in *C. elegans* (AOKI *et al.* 2007; PAK and FIRE 2007; SIJEN *et al.* 2007), *i. e.* multiple unprimed Rdr2 initiation events resulting in the direct synthesis of 23-nt antisense molecules. Although the small but significant fraction of sense secondary siRNAs could conceivably be due to the same mechanism acting on natural antisense transcripts, which can be detected at low levels for much of the macronuclear genome (LEPÈRE *et al.* 2008), it is unclear whether the latter would be sufficiently abundant. Furthermore, the resulting 5'-triphosphate siRNAs would have to be rapidly converted to 5'-monophosphate by an RNA phosphatase such as PIR-1 (DUCHAINE *et al.* 2006).

A different possibility is suggested by the absolute antisense bias reported for 5'-monophosphate endogenous siRNAs in *T. thermophila* (COUVILLION *et al.* 2009; LEE and COLLINS 2006). The only RdRP in this ciliate, which is closely related to *P. tetraurelia*'s Rdr1 and Rdr2, has a processive activity on ssRNA templates *in vitro*, synthesizing long dsRNAs

that can be cleaved by Dicer (LEE and COLLINS 2007; TALSKY and COLLINS 2010). To explain the antisense bias, it was proposed that physical interaction of Dicer with the RdRC dictates the polarity of Dicer processing, resulting in preferential loading of the antisense strand onto Piwi proteins (LEE and COLLINS 2007). A similar mechanism, with a less stringent strand selection, could account for the synthesis of *P. tetraurelia* secondary siRNAs by Rdr2. Although secondary siRNA levels were not reduced in the two *dcr1* alleles tested, both have missense mutations in the helicase domain but wild-type RNaseIII domains, and the possibility remains that the helicase domain is dispensable for the synthesis of secondary, but not primary, siRNAs. Similarly, the helicase domain of the *C. elegans* Dicer was shown to be required only for certain classes of siRNAs (PAVELEC *et al.* 2009; WELKER *et al.* 2010). Alternatively, another protein may be responsible for dsRNA dicing. The other two Dicer genes in *Paramecium* appear to have non-catalytic RNaseIII domains, but at least one potentially functional Dicer-like gene (lacking a helicase domain, *DCL4*) has no known function (LEPÈRE *et al.* 2009).

The role of primary and secondary siRNA uridylation

A minor fraction of *NDI69* primary and secondary siRNAs was found to carry 3' non-templated uridines, a modification catalyzed by nucleotidyl-transferases (KIM *et al.* 2010). This fraction is not reduced in residual primaries nor in the secondaries from *cid1* mutants (though absolute numbers of 3'U secondaries are low and therefore less informative). This suggests that one or several of the other ≥ 22 nucleotidyl-transferases encoded in the *P. tetraurelia* genome (MARKER *et al.* 2014) are responsible for their uridylation.

The possible function of 3' uridylation of dsRNA-induced siRNAs remains unknown. It may be a transient modification that promotes siRNA degradation, as observed in plants in the absence of 3' methylation (LI *et al.* 2005; REN *et al.* 2012; ZHAO *et al.* 2012) and in the *C. elegans* germline (VAN WOLFSWINKEL *et al.* 2009). In contrast to non-uridylated siRNAs, we observed an absence of any antisense bias for *NDI69* 3'U primary and secondary siRNAs, whereas 3'U siRNAs of other target genes showed the same biases observed for their non-uridylated counterparts. This, and the fact that the level of uridylation also varies among different RNAi target genes may indicate that uridylation depends on multiple factors. In support of this idea, (poly)adenylation is the predominant 3' modification of dsRNA-induced and endogenous siRNAs during conjugation (LEPÈRE *et al.* 2009), suggesting that nucleotidyl transferase activity underlies entirely different constraints when meiosis-specific small RNAs (scnRNAs) are active or cells undergo prolonged starvation. However, the fact that 3' uridylation occurs on vector-specific siRNAs suggests that at least uridylation of primary siRNAs does not reflect the activity of RNAi-unrelated exosome-mediated degradation of

aberrant transcripts, as observed in *S. pombe* (BUHLER *et al.* 2007; BUHLER *et al.* 2008; HALIC and MOAZED 2010).

The functions of primary and secondary siRNAs in dsRNA-induced silencing

In *C. elegans*, secondary siRNAs induced by dsRNA feeding accumulate at much higher levels than primary siRNAs (PAK and FIRE 2007; PAK *et al.* 2012; SIJEN *et al.* 2007). In contrast, we found *Paramecium* secondary siRNAs to be much less abundant than primary siRNAs. Although we sequenced only 5'-monophosphate molecules, our northern blot analyses ((MARKER *et al.* 2010) and this study) indicate that triphosphorylated secondary siRNAs, if present at all, are at best a very minor fraction of the total siRNA pool. Unlike primaries, secondary siRNAs are produced from the target mRNA, so that mRNA abundance may be a limiting factor. Continuous dsRNA import and production of primary siRNAs, together with mRNA slicing (JAILLON *et al.* 2008) and eventually degradation, could thus explain the observed ratio. Furthermore, secondary siRNAs might affect transcription of the target gene in the macronucleus (see below), resulting in a negative feedback. A non-exclusive possibility is that a limiting amount of the Rdr2 protein, which is involved in the synthesis of both primaries and secondaries, is titrated away from its role in secondary siRNA production by continuous dsRNA feeding.

Interestingly, another difference with *C. elegans* is that in the latter, secondary siRNAs were reported to be the actual silencing effectors (CLAYCOMB *et al.* 2009; GU *et al.* 2009; SIJEN *et al.* 2007), whereas our results suggest that primary siRNAs are the major actors in *Paramecium*. Indeed, a considerable silencing deficiency is seen in mutants with significantly reduced amounts of primary siRNAs, but wild-type or only slightly reduced levels of secondaries (*rdr1-3.16*, *dcr1-5.5*). Conversely, complete absence of secondaries and slightly reduced amounts of primaries only lead to a very partial silencing deficiency, as seen in the *rdr2-1.24* mutant. Thus, secondary siRNAs do not appear to contribute much to the observed silencing phenotype, suggesting they may have distinct functions. It should be noted, however, that all our experiments were carried out with the widely used strain 51 of *P. tetraurelia*, which has a long history of maintenance in stock collections, and we cannot exclude that this strain has fixed a mutation affecting the levels and/or functions of dsRNA-induced secondary siRNAs.

In *C. elegans*, primary and secondary siRNAs are loaded onto different Argonaute proteins (YIGIT *et al.* 2006). Among the three *P. tetraurelia* Piwi proteins implicated in dsRNA-induced silencing by recursive RNAi experiments (BOUHOUCHE *et al.* 2011), Ptiwi13 was shown to have the strongest effect: its knockdown completely suppressed silencing of a reporter gene. The protein, which contains the typical slicer catalytic triad DDH, is a good candidate to load primary siRNAs and cleave cytoplasmic mRNAs within the dsRNA region

(JAILLON *et al.* 2008). A weaker effect was seen with the closely related ohnologs Ptiwi12 and Ptiwi15, as reporter silencing was only partially suppressed, and only after joint knockdown of the two ohnologs. Secondary siRNAs could be loaded onto these proteins, which contain EDH triads and may not have slicer activity but otherwise resemble Ptiwi09, a protein implicated in the scnRNA pathway and shown to be nuclear (BOUHOUCHE *et al.* 2011). These observations suggest that a possible function of secondary siRNAs may be to relay information about cytoplasmic RNAi to the nucleus, as observed for some secondary siRNAs in *C. elegans*. The latter associate with the non-slicing Ago protein Nrde-3 (GUANG *et al.* 2008), inducing heritable nuclear silencing by pre-mRNA-targeting (BURTON *et al.* 2011; GU *et al.* 2012; GUANG *et al.* 2010), but contribute only partially to silencing in the generation exposed to the dsRNA trigger. In support of this hypothesis, Rdr2 is also involved in transgene-induced silencing (MARKER *et al.* 2014), a pathway initiated in the nucleus and possibly associated with nuclear functions of Rdr2-dependent siRNAs (MARKER *et al.* 2010).

Endogenous siRNA clusters: RdRPs may compete as well as co-operate

Mutants deficient in exogenously triggered RNAi were found to have altered levels of endogenous siRNAs clustering at specific macronuclear loci, suggesting that the corresponding genes normally participate in transcriptome regulation. Perhaps because the mutagenesis screen selected for fast-growing lines, no obvious phenotype was observed during vegetative growth, and only *rdr2* mutants showed defects in sexual reproduction (MARKER *et al.* 2014). Among the 10 siRNA clusters analysed, 3 were almost completely absent in the strongest *rdr2* allele. These mapped to known transcripts (2 protein-coding mRNAs, 1 ncRNA) with a strong antisense bias and appear to be mechanistically similar to dsRNA-induced secondary siRNAs, although they do not depend on upstream Rdr1 action. One was known to depend on the putatively non-catalytic Rdr3, indicating that Rdr2 and Rdr3 can co-operate for endogenous clusters, as they do in transgene-induced silencing (MARKER *et al.* 2014; MARKER *et al.* 2010) (schematized in Fig.7B).

In the other 7 clusters (including 5 that appear to map to ncRNAs with variable antisense biases), siRNA levels were significantly reduced in full-deficiency *rdr1* and *cid1* mutants, although in contrast to dsRNA-induced primary siRNAs they were not affected in *pds1* mutants. Since these Rdr1/Cid1-dependent siRNAs cover the entire length of the ncRNAs, it is impossible to tell whether they also give rise to a minor fraction of Rdr2-dependent siRNAs. Interestingly, in at least two cases siRNA levels were in fact markedly increased in the strongest *rdr2* allele, suggesting that Rdr2 may normally down-regulate transcripts that are used as precursors for Rdr1/Cid1-dependent siRNAs. Thus, Rdr1 and Rdr2 may compete for such transcripts. Alternatively, Rdr2 may compete with Rdr1 for pathway components, such as unknown RdRC co-factors. If the Rdr1-Rdr2 co-operation on exogenous dsRNA is

physically interdependent, loss of Rdr2 during dsRNA feeding may result in increased overall availability of Rdr1. However, it remains unclear why increased siRNA levels are not observed at other Rdr1-Cid1-dependent cluster loci in *rdr2* mutants. For all types of clusters, siRNA levels did not appear to be significantly reduced in the two hypomorphic *dcr1* alleles, suggesting that the helicase domain is dispensable. However, as noted for dsRNA-induced secondary siRNAs, the Dcr1 RNaseIII domains may still be involved.

The natural functions of exogenously triggered RNAi

We have provided evidence that *P. tetraurelia* synthesizes siRNAs that are antisense to the ribosomal RNAs of its bacterial food, in an Rdr2-dependent manner. Thus, the coupling of RNA import from food vacuoles to RNAi processing is not restricted to dsRNA. In previous tests, feeding cells with bacteria producing ssRNA homologous to a *Paramecium* gene did not result in phenotypically detectable silencing (GALVANI and SPERLING 2002), but this may be due to the higher stability of dsRNA in the HT115 strain of *E. coli*, and the need for high levels of primary siRNAs to completely suppress a *Paramecium* mRNA.

This finding raises the possibility that single-stranded bacterial RNAs compete with the dsRNA trigger for RNAi factors (Fig.7B), which could explain the important variations observed in primary siRNA levels between different wild-type cultures fed the same dsRNA-producing bacterial strain. In addition to intrinsic variations in the amounts of dsRNA in IPTG-induced bacterial cultures, relative to the rRNA, supposedly monoxenic *Paramecium* cultures are frequently contaminated by other bacterial species, which could affect the efficiency of silencing. More fundamentally, the finding considerably broadens the possible natural functions of this mechanism. dsRNA-induced silencing could obviously provide protection against dsRNA viruses, but no virus infecting *Paramecium* has been described so far; the mechanism may now be envisioned to protect against the horizontal transfer of single-stranded retroelements. Furthermore, it may provide control of bacterial RNAs impacting the cell's endogenous gene expression, as observed recently in *C. elegans* (LIU *et al.* 2012). Further studies have to show how this unicellular eukaryote adapts RNAi responses to competing exogenous and endogenous RNAi triggers.

Acknowledgements

We thank Marcel Schulz for critical discussion of data and all lab members for discussions and continuous support. The sequencing of small RNAs benefited from the facilities and expertise of the high throughput sequencing platform of IMAGIF (Centre de Recherche de Gif—www.imagif.cnrs.fr) and the technical assistance of Jasmin Kirch, Department of Epigenetics, University of Saarbrücken.

Funding

Agence Nationale de la Recherche [‘Investissements d’Avenir’ program ANR-10-LABX-54 MEMO LIFE/ANR-11-IDEX-0001–02 Paris Sciences et Lettres* Research University, ANR-08-BLAN-0233 ‘ParaDice’, ANR-12-BSV6–0017 ‘INFERNO’ to E.M.]; Fondation pour la Recherche Médicale [‘Equipe FRM’ to E.M.] and Fondation Pierre Gilles de Gennes [post-doctoral fellowship to S.M.]. This study was carried out in the context of the CNRS-supported European Research Group ‘Paramecium Genome Dynamics and Evolution’ and the European COST Action BM1102. Funding for open access charge: Agence Nationale de la Recherche.

References

- AOKI, K., H. MORIGUCHI, T. YOSHIOKA, K. OKAWA and H. TABARA, 2007 In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J* **26**: 5007-5019.
- ARNAIZ, O., N. MATHY, C. BAUDRY, S. MALINSKY, J. M. AURY *et al.*, 2012 The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet* **8**: e1002984.
- AURY, J. M., O. JAILLON, L. DURET, B. NOEL, C. JUBIN *et al.*, 2006 Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**: 171-178.
- BAULCOMBE, D. C., 2007 Molecular biology. Amplified silencing. *Science* **315**: 199-200.
- BOUHOUCHE, K., J. F. GOUT, A. KAPUSTA, M. BETERMIER and E. MEYER, 2011 Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* **39**: 4249-4264.
- BRODERSEN, P., and O. VOINNET, 2006 The diversity of RNA silencing pathways in plants. *Trends Genet* **22**: 268-280.
- BUHLER, M., W. HAAS, S. P. GYGI and D. MOAZED, 2007 RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**: 707-721.
- BUHLER, M., N. SPIES, D. P. BARTEL and D. MOAZED, 2008 TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. *Nat Struct Mol Biol* **15**: 1015-1023.
- BURTON, N. O., K. B. BURKHART and S. KENNEDY, 2011 Nuclear RNAi maintains heritable gene silencing in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **108**: 19683-19688.
- CASTEL, S. E., and R. A. MARTIENSSEN, 2013 RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* **14**: 100-112.
- CHEN, C. C., M. J. SIMARD, H. TABARA, D. R. BROWNELL, J. A. MCCOLLOUGH *et al.*, 2005 A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in *C. elegans*. *Curr Biol* **15**: 378-383.
- CLAYCOMB, J. M., P. J. BATISTA, K. M. PANG, W. GU, J. J. VASALE *et al.*, 2009 The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**: 123-134.
- COUVILLION, M. T., S. R. LEE, B. HOGSTAD, C. D. MALONE, L. A. TONKIN *et al.*, 2009 Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes Dev* **23**: 2016-2032.

- CURABA, J., and X. CHEN, 2008 Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem* **283**: 3059-3066.
- DALMAY, T., A. HAMILTON, S. RUDD, S. ANGELL and D. C. BAULCOMBE, 2000 An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**: 543-553.
- DUCHAINED, T. F., J. A. WOHLSCHEGEL, S. KENNEDY, Y. BEI, D. CONTE, JR. *et al.*, 2006 Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124**: 343-354.
- GALVANI, A., and L. SPERLING, 2001 Transgene-mediated post-transcriptional gene silencing is inhibited by 3' non-coding sequences in Paramecium. *Nucleic Acids Res* **29**: 4387-4394.
- GALVANI, A., and L. SPERLING, 2002 RNA interference by feeding in Paramecium. *Trends Genet* **18**: 11-12.
- GENT, J. I., A. T. LAMM, D. M. PAVELEC, J. M. MANIAR, P. PARAMESWARAN *et al.*, 2010 Distinct phases of siRNA synthesis in an endogenous RNAi pathway in *C. elegans* soma. *Mol Cell* **37**: 679-689.
- GHILDYAL, M., and P. D. ZAMORE, 2009 Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**: 94-108.
- GU, S. G., J. PAK, S. GUANG, J. M. MANIAR, S. KENNEDY *et al.*, 2012 Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44**: 157-164.
- GU, W., M. SHIRAYAMA, D. CONTE, JR., J. VASALE, P. J. BATISTA *et al.*, 2009 Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell* **36**: 231-244.
- GUANG, S., A. F. BOCHNER, K. B. BURKHART, N. BURTON, D. M. PAVELEC *et al.*, 2010 Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465**: 1097-1101.
- GUANG, S., A. F. BOCHNER, D. M. PAVELEC, K. B. BURKHART, S. HARDING *et al.*, 2008 An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321**: 537-541.
- HALIC, M., and D. MOAZED, 2010 Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell* **140**: 504-516.
- JAILLON, O., K. BOUHOUCHE, J. F. GOUT, J. M. AURY, B. NOEL *et al.*, 2008 Translational control of intron splicing in eukaryotes. *Nature* **451**: 359-362.
- JOSE, A. M., G. A. GARCIA and C. P. HUNTER, 2011 Two classes of silencing RNAs move between *Caenorhabditis elegans* tissues. *Nat Struct Mol Biol* **18**: 1184-1188.
- KENNEDY, S., D. WANG and G. RUVKUN, 2004 A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**: 645-649.
- KETTING, R. F., 2011 The many faces of RNAi. *Dev Cell* **20**: 148-161.

- KIM, Y. K., I. HEO and V. N. KIM, 2010 Modifications of small RNAs and their associated proteins. *Cell* **143**: 703-709.
- LAWRENCE, M., R. GENTLEMAN and V. CAREY, 2009 rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**: 1841-1842.
- LEE, H. C., A. P. AALTO, Q. YANG, S. S. CHANG, G. HUANG *et al.*, 2010 The DNA/RNA-dependent RNA polymerase QDE-1 generates aberrant RNA and dsRNA for RNAi in a process requiring replication protein A and a DNA helicase. *PLoS Biol* **8**.
- LEE, R. C., C. M. HAMMELL and V. AMBROS, 2006 Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. *RNA* **12**: 589-597.
- LEE, S. R., and K. COLLINS, 2006 Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev* **20**: 28-33.
- LEE, S. R., and K. COLLINS, 2007 Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol* **14**: 604-610.
- LEE, S. R., K. B. TALSKEY and K. COLLINS, 2009 A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *RNA* **15**: 1363-1374.
- LEPÈRE, G., M. BETERMIER, E. MEYER and S. DUHARCOURT, 2008 Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev* **22**: 1501-1512.
- LEPÈRE, G., M. NOWACKI, V. SERRANO, J. F. GOUT, G. GUGLIELMI *et al.*, 2009 Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res* **37**: 903-915.
- LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.
- LI, J., Z. YANG, B. YU, J. LIU and X. CHEN, 2005 Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr Biol* **15**: 1501-1507.
- LIU, H., X. WANG, H. D. WANG, J. WU, J. REN *et al.*, 2012 *Escherichia coli* noncoding RNAs can affect gene expression and physiology of *Caenorhabditis elegans*. *Nat Commun* **3**.
- LU, R., M. MADURO, F. LI, H. W. LI, G. BROITMAN-MADURO *et al.*, 2005 Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*. *Nature* **436**: 1040-1043.
- MAKEYEV, E. V., and D. H. BAMFORD, 2002 Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol Cell* **10**: 1417-1427.
- MANIAR, J. M., and A. Z. FIRE, 2011 EGO-1, a *C. elegans* RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. *Curr Biol* **21**: 449-459.

- MARKER, S., Q. CARRADEC, V. TANTY, O. ARNAIZ and E. MEYER, 2014 A forward genetic screen reveals essential and non-essential RNAi factors in *Paramecium tetraurelia*. *Nucleic Acids Res* **23**.
- MARKER, S., A. LE MOUËL, E. MEYER and M. SIMON, 2010 Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res.* **38**: 4092-4107. Epub 2010 Mar 4093.
- MEISTER, G., and T. TUSCHL, 2004 Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**: 343-349.
- MOTAMEDI, M. R., A. VERDEL, S. U. COLMENARES, S. A. GERBER, S. P. GYGI *et al.*, 2004 Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**: 789-802.
- MOURRAIN, P., C. BECLIN, T. ELMAYAN, F. FEUERBACH, C. GODON *et al.*, 2000 Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**: 533-542.
- PAK, J., and A. FIRE, 2007 Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**: 241-244.
- PAK, J., J. M. MANIAR, C. C. MELLO and A. FIRE, 2012 Protection from feed-forward amplification in an amplified RNAi mechanism. *Cell* **151**: 885-899.
- PAVELEC, D. M., J. LACHOWIEC, T. F. DUCHAINE, H. E. SMITH and S. KENNEDY, 2009 Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. *Genetics* **183**: 1283-1295.
- REN, G., X. CHEN and B. YU, 2012 Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. *Curr Biol* **22**: 695-700.
- RUIZ, F., L. VAYSSIE, C. KLOTZ, L. SPERLING and L. MADEDDU, 1998 Homology-dependent gene silencing in *Paramecium*. *Mol Biol Cell* **9**: 931-943.
- SARKIES, P., A. ASHE, J. LE PEN, M. A. MCKIE and E. A. MISKA, 2013 Competition between virus-derived and endogenous small RNAs regulates gene expression in *Caenorhabditis elegans*. *Genome Res* **23**: 1258-1270.
- SARKIES, P., and E. A. MISKA, 2013 RNAi pathways in the recognition of foreign RNA: antiviral responses and host-parasite interactions in nematodes. *Biochem Soc Trans* **41**: 876-880.
- SIJEN, T., F. A. STEINER, K. L. THIJSSSEN and R. H. PLASTERK, 2007 Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**: 244-247.
- SIMON, M. C., S. MARKER and H. J. SCHMIDT, 2006 Posttranscriptional control is a strong factor enabling exclusive expression of surface antigens in *Paramecium tetraurelia*. *Gene Expr* **13**: 167-178.
- SUGIYAMA, T., H. CAM, A. VERDEL, D. MOAZED and S. I. GREWAL, 2005 RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling

- heterochromatin assembly to siRNA production. *Proc Natl Acad Sci U S A* **102**: 152-157.
- TALSKY, K. B., and K. COLLINS, 2010 Initiation by a eukaryotic RNA-dependent RNA polymerase requires looping of the template end and is influenced by the template-tailing activity of an associated uridylyltransferase. *J Biol Chem* **285**: 27614-27623.
- TIMMONS, L., D. L. COURT and A. FIRE, 2001 Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**: 103-112.
- TIMMONS, L., and A. FIRE, 1998 Specific interference by ingested dsRNA. *Nature* **395**: 854-854.
- VAISTIJ, F. E., L. JONES and D. C. BAULCOMBE, 2002 Spreading of RNA Targeting and DNA Methylation in RNA Silencing Requires Transcription of the Target Gene and a Putative RNA-Dependent RNA Polymerase. *The Plant Cell* **14**: 857-867.
- VAN WOLFSWINKEL, J. C., J. M. CLAYCOMB, P. J. BATISTA, C. C. MELLO, E. BEREZIKOV *et al.*, 2009 CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell* **139**: 135-148.
- VASALE, J. J., W. GU, C. THIVIERGE, P. J. BATISTA, J. M. CLAYCOMB *et al.*, 2010 Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci U S A* **107**: 3582-3587.
- VOINNET, O., 2008 Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci* **13**: 317-328.
- VOLPE, T. A., C. KIDNER, I. M. HALL, G. TENG, S. I. GREWAL *et al.*, 2002 Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833-1837.
- WELKER, N. C., D. M. PAVELEC, D. A. NIX, T. F. DUCHAINE, S. KENNEDY *et al.*, 2010 Dicer's helicase domain is required for accumulation of some, but not all, *C. elegans* endogenous siRNAs. *RNA* **16**: 893-903.
- WILKINS, C., R. DISHONGH, S. C. MOORE, M. A. WHITT, M. CHOW *et al.*, 2005 RNA interference is an antiviral defence mechanism in *Caenorhabditis elegans*. *Nature* **436**: 1044-1047.
- YIGIT, E., P. J. BATISTA, Y. BEI, K. M. PANG, C. C. CHEN *et al.*, 2006 Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**: 747-757.
- YIN, T., D. COOK and M. LAWRENCE, 2012 ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol* **13**: 2012-2013.
- ZHAO, Y., Y. YU, J. ZHAI, V. RAMACHANDRAN, T. T. DINH *et al.*, 2012 The Arabidopsis nucleotidyl transferase HESO1 uridylylates unmethylated small RNAs to trigger their degradation. *Curr Biol* **22**: 689-694.

B. Deux classes de siRNAs uridylés

Parmi les sRNAs uridylés induits par le feeding de ND169 et alignés sur l'ARNdb inducteur, deux différentes catégories peuvent être isolées. La première catégorie est composée de sRNAs de 23 nt génomiques additionnés d'une ou plusieurs uridines (catégorie appelée "23 + U"). La seconde catégorie de siRNAs est composée de sRNAs dont la partie génomique est plus courte que 23 nt (entre 15 et 22 nt) et qui sont complétés à 23 nt avec des uridines non génomiques (catégorie appelée "complété à 23") (Figure 47C).

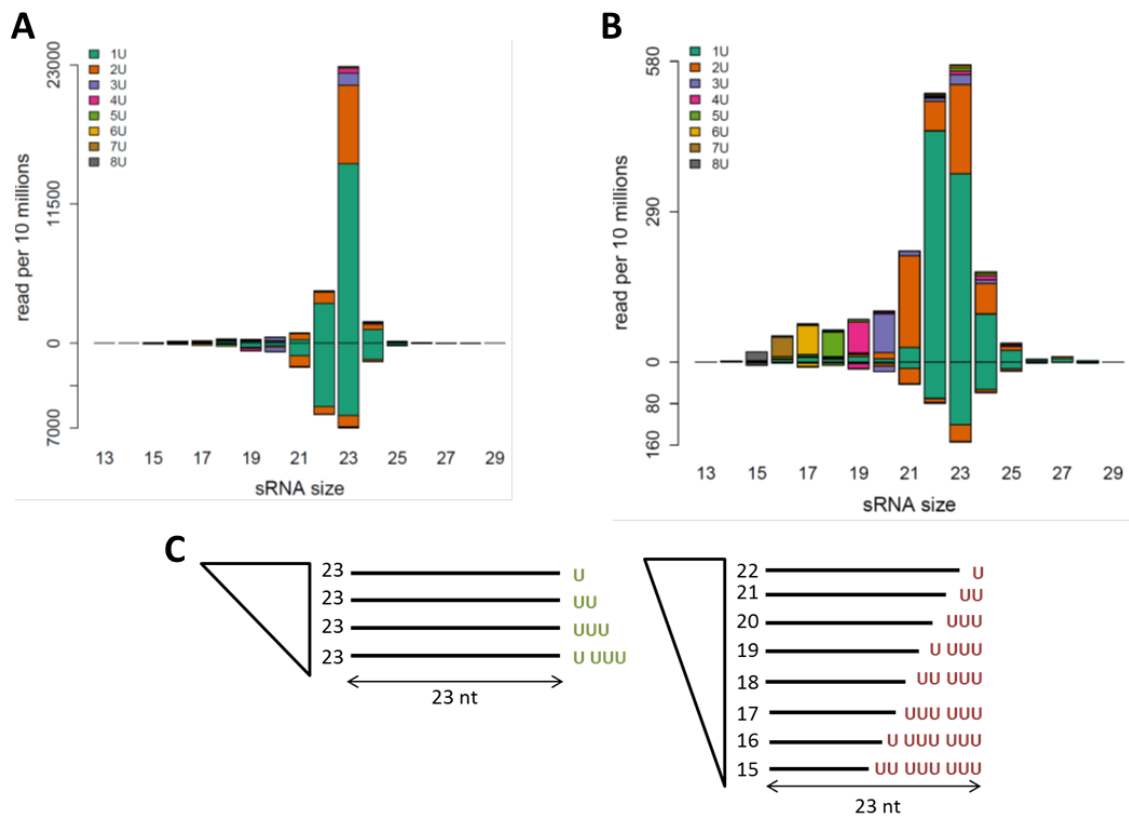


Figure 47 : Deux classes de siRNAs uridylés. Distribution de taille des siRNAs modifiés avec une ou plusieurs uridines normalisée en reads par 10 millions. Les couleurs indiquent le nombre d'uridine non génomique en 3'. A) échantillon sauvage, B) mutant *dcr1-5.5*. C) Représentation schématique des deux classes d'uridylation. A gauche, 23 nt + U ; à droite, complété à 23 nt.

Dans des conditions sauvages, les sRNAs "23 + U" sont fortement majoritaires et les sRNAs complétés à 23, bien que présents, sont peu visible sur la distribution de taille (Figure 47A). Au contraire, dans des conditions mutantes pour dicer (allèles *dcr1-5.5* et *dcr1-5.27*), la catégorie "23 + U" est fortement réduite et la catégorie "complété à 23" est bien plus visible (entre 15 et 22 nt) (Figure 47B et Figure S9 de l'article). Les proportions de chacune de ces deux catégories dans les autres mutants semblent similaires au sauvage (données non montrées). Il semble donc que les sRNAs complétés à 23 soient indépendant de Dicer. En ce qui concerne les sRNAs secondaires uridylés alignés sur ND169 en dehors de la région *feeding*, les sRNAs "23+U" sont bien présents mais les quantités sont trop faibles pour que la catégorie "complétée à 23" soit détectable. Cependant, dans les mutants Dicer qui n'affectent pas la production des siRNAs secondaires (uridylés ou non), la catégorie complétée à 23 est aussi absente (Figure S9). On peut donc suggérer que cette seconde classe de siRNAs uridylés est spécifique des siRNAs primaires. Une hypothèse pour expliquer l'origine de ces sRNAs plus courts complétés à 23 nt par des uridines non génomiques est que les produits de dégradation de l'ARNdb, qui forment le bruit de fond pour les siRNAs non uridylés (Figure 3B de l'article), peuvent aussi être uridylés. L'uridylation pourrait ajouter un nombre aléatoire d'uridine quel que soit la taille du produit de dégradation mais le fait que les sRNAs de 23 nt puissent être chargés dans les protéines Piwi et donc être stabilisés pourrait expliquer que l'on retrouve principalement des sRNAs complétés à 23 nt.

Résumé

L'ensemble de ce travail réalisé sur l'espèce *Paramecium tetraurelia* a permis d'étudier les mécanismes d'ARN interférence. En particulier, la voie d'ARNi induite par *feeding* d'ARNdb a pu être en partie décryptée grâce à la découverte de 4 nouveaux gènes et par l'analyse des petits ARN synthétisés après induction par l'ARNdb ingéré. Le point essentiel de ce travail fut la combinaison de deux techniques de biologie qui nous ont apporté un grand nombre d'informations sur le fonctionnement de cette voie cellulaire.

La première technique mise en place a été un crible de génétique direct basé sur une mutagenèse par exposition des paramécies aux UV. Le crible imaginé repose sur le détournement de la fonction biologique de l'ARNi qui consiste probablement à se défendre contre les parasites moléculaires présents dans l'environnement. En effet, les paramécies ont été soumises au *feeding* d'un ARNdb complémentaire à un gène endogène essentiel de la paramécie. Les paramécies sauvages capables de faire de l'ARNi sont donc piégées par leur propre mécanisme de défense et meurent rapidement après avoir réprimé l'expression de leur gène essentiel. Au contraire, les paramécies mutantes incapables de faire de l'ARNi survivent au *feeding* d'ARNdb et ont donc été sélectionnées. Ce crible a permis d'isoler 79 mutants totalement ou partiellement déficients pour l'ARNi. L'obtention de ces mutants mendéliens représente un avantage certain pour l'analyse des voies d'ARNi en comparaison aux méthodes précédemment établies dans le laboratoire qui consistaient à éteindre par ARN interférence un gène potentiellement impliqué dans l'ARNi. En effet, cette méthode avait comme inconvénient de nécessiter un *a priori* sur le gène mais aussi de recourir à de l'ARNi récursif qui ne permet de conclure qu'en cas de résultat positif et apporte très souvent des réponses partielles. Cette nouvelle méthode d'identification sans *a priori* des mutants mendéliens pour l'ARNi a aussi ses propres limites : elle ne permet de trouver des mutations nulles que dans des gènes non essentiels et non redondants. Il est cependant théoriquement possible d'isoler des allèles hypomorphes dans des gènes essentiels. Les gènes mutés dans les différents clones ont d'abord été identifiés par des techniques de biologie moléculaires, extraction d'ADN, PCR et enfin séquençage des gènes déjà connus. Les clones déficients pour l'ARNi mais sauvages pour tous ces gènes ont été croisés avec une paramécie sauvage et des pools de F2 ont été réalisés pour permettre l'identification de la mutation responsable du phénotype de déficience pour l'ARNi parmi les mutations induites aux UV. La seconde limite de cette technique est le temps et le coût nécessaire pour identifier tous les gènes impliqués dans un processus. En effet, plus le nombre de gènes identifiés dans la voie augmente plus il y a de gènes à séquencer individuellement par clone déficient avant de séquencer le génome complet. De plus, le croisement et le séquençage du génome MAC des mutants est une procédure relativement longue (au minimum 2 mois et demi) et coûteuse.

Les résultats de cette analyse ont permis l'obtention de 49 allèles dans 3 gènes déjà connus : DCR1, RDR1 et RDR2 mais aussi dans 3 nouveaux gènes : CID1, PDS1 et PDS2.

Etant donné que 41, 16 et 10 clones sont respectivement mutés dans RDR1 CID1 et PDS1 avec un grand nombre d'allèles différents, il semble alors que le crible soit saturé pour les gènes non essentiels. Au contraire, pour les gènes DCR1 et RDR2, peu d'allèles ont été trouvés et ces allèles sont hypomorphes suggérant que ces gènes sont essentiels. CID2 a été découvert dans un second temps par l'analyse par ARNi récursif des 22 gènes de la famille CID du génome de *P.tetraurelia*. Aucun allèle n'ayant été trouvé dans le crible pour ce gène, il est probable qu'il soit également essentiel. De manière intéressante, les gènes essentiels DCR1, CID2 et RDR2 sont également impliqués dans la voie d'ARNi induite par transgène, tandis que les gènes non essentiels PDS1, CID1 et RDR1 sont spécifiquement impliqués dans le traitement de l'ARNdb ingéré. Ces résultats montrent que les deux voies d'ARNi partagent certains facteurs mais aussi que la voie par transgène est essentielle alors que la voie par *feeding* d'ARNdb ne l'est pas. Si cette voie exogène d'ARNi semble inutile à la paramécie en conditions de laboratoire et à court terme, il est très probable qu'il en soit autrement dans des conditions environnementales naturelles bien plus complexes puisque qu'une pression de sélection s'exerce sur les gènes impliqués dans cette voie depuis des millions d'années (conservation de la voie au sein du groupe *P.aurelia*). La question de la fonction naturelle de cette voie d'ARN interférence reste donc entière après cette analyse.

Une seconde technique a été mise en place dans le laboratoire pour étudier plus précisément le rôle de ces gènes dans la biosynthèse ou l'action des siRNAs. Les mutants obtenus dans le crible sont un matériel idéal pour étudier l'impact de l'absence d'un gène sur l'accumulation des siRNAs de la voie d'ARNi induite par ARNdb. Les siRNAs de cellules sauvages ou mutantes, préalablement nourries d'un ARNdb homologue au gène non essentiel ND169 ont donc été séquencés en grande quantité. 14 banques mutantes et 3 banques sauvages réalisées dans le laboratoire ainsi que 5 banques réalisées dans le laboratoire de Martin Simon en Allemagne ont été analysées. L'analyse bio-informatique a consisté, après avoir éliminé les séquences contaminants (séquences bactériennes, ribosomales, mitochondriales), à aligner sans mismatch les siRNAs sur l'ARNm ciblé, sur l'ARNdb inducteur ou bien sur le génome complet puis à les compter en suivant différents critères : la taille, l'orientation, la position et les modifications non génomiques en 3'.

La distribution de taille des sRNAs a révélé un pic très net à 23 nt, taille attendue pour les siRNAs. Afin d'isoler les siRNAs produits par le mécanisme d'ARNi des sRNAs résultants de la dégradation de l'ARNdb, le nombre sRNAs de 23 nt a été retranché de la moyenne des sRNAs de 22 et 24 nt. Des siRNAs ainsi définis ont été identifiés sur l'ARNdb inducteur mais aussi sur la totalité de l'ARNm endogène ciblé, en plus faible quantité (1% du total), et très majoritairement de polarité anti-sens. Cette observation démontre la capacité de *P.tetraurelia* à synthétiser des siRNAs secondaires à partir de l'ARNm ciblé et donc la capacité de propagation de l'ARNi dans les deux orientations le long de l'ARNm. La production des siRNAs primaires (produits à partir de l'ARNdb) dépendent de tous les gènes trouvés dans le

crible y compris les deux RdRPs. Le fait que des RdRPs soient nécessaires à la production de siRNAs primaires peut suggérer qu'ils proviennent d'un ARN simple brin dont le brin complémentaire a été synthétisé par la RdRP et non d'un ARNdb importé de la vacuole digestive. Cependant il est aussi possible que les RdRPs agissent en complexe avec Dicer et soient nécessaires à son activité (voir la première partie de la discussion). Au contraire, la synthèse des siRNAs secondaires requiert seulement RDR2 et, comme chez *C. elegans*, semble indépendante de Dcr1. Cependant, environ 4% des siRNAs secondaires sont dans l'orientation sens, ce qui suggère qu'un certain nombre de siRNAs secondaires sont tout de même produits à partir d'un long ARNdb précurseur clivé par Dicer. Ces deux résultats contradictoires suggèrent que soit une autre protéine Dicer non identifiée dans le crible est impliquée dans la synthèse de ces siRNAs soit le domaine hélicase de Dcr1 n'est pas requis pour la synthèse des siRNAs secondaires sens (les deux mutants Dcr1 séquencés ayant une mutation faux-sens dans le domaine hélicase). Les siRNAs secondaire sont en quantité équivalente dans le mutant Dcr1 (incapable de faire de l'ARNi) et dans le sauvage ce qui suggère qu'ils n'ont pas la capacité d'éteindre l'expression de l'ARNm (voir la deuxième partie de la discussion). Une fraction de ces siRNAs (environ 10%) est modifiée par l'addition d'une ou plusieurs uridines en 3'. Ces siRNAs modifiés semblent dépendants des mêmes gènes que les siRNAs non modifiés mais ne présentent pas la même distribution sur les deux brins. En effet, les siRNAs uridylés primaires ou secondaires ont un biais sens contrairement aux siRNAs secondaires qui sont très majoritairement anti-sens. Deux hypothèses peuvent expliquer cette observation. 1) L'uridylation est une trace du processus de synthèse des siRNAs (formation d'une boucle poly U selon le modèle *Tetrahymena* nécessaire à la RdRP pour synthétiser le brin complémentaire) qui ne reste présente que sur une faible quantité d'entre eux. 2) L'uridylation est simplement une marque indiquant le début du processus de dégradation du siRNA (voir la troisième partie de la discussion). Enfin, les siRNAs synthétisés sur des loci endogènes ont été recherchés dans le génome macronucléaire. Les 10 clusters de siRNAs les plus couverts ont été analysés, ils présentent tous un pic à 23nt et une fraction d'entre eux est uridylé. 3 de ces locus sont dépendants de Rdr2 et les autres sont dépendants de Rdr1 et Cid1 indiquant que les deux RdRPs ont bien des fonctions différentes et que Cid1 semble interagir (directement ou non) avec Rdr1. L'ensemble de ces résultats sont résumés dans la (Figure 48).

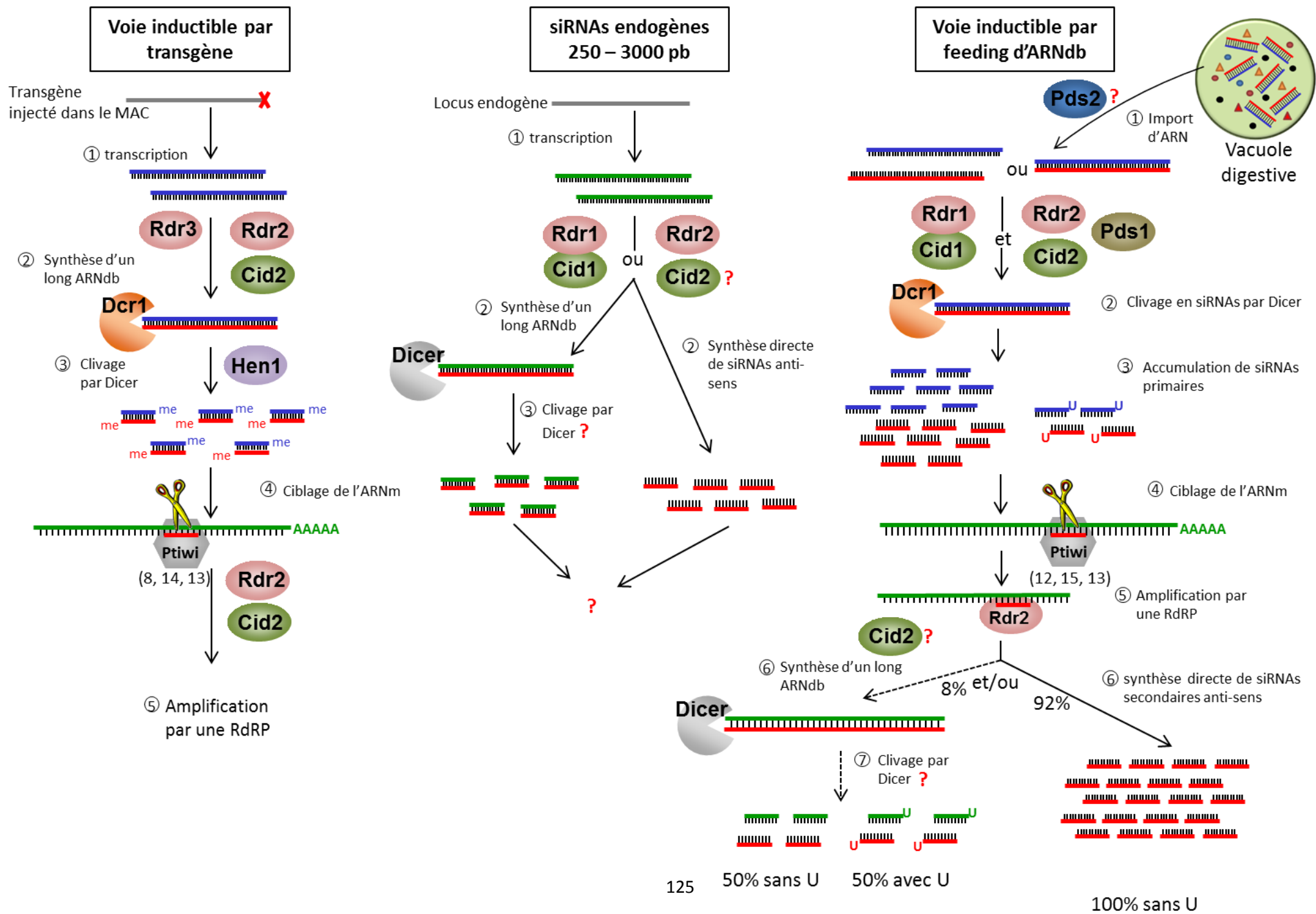


Figure 48 : Schéma des voies de siRNAs végétatives chez *P. tetraurelia*. La voie induite par injection de transgènes dans le macronoyau de la paramécie nécessite deux RdRPs (RDR2 et RDR3) ainsi que la nucléotidyl transférase Cid2 pour générer de long ARNdb. Les ARNdb sont clivés par Dicer puis méthylés par Hen1 avant d'être chargés sur les protéines Piwi (Marker, Simon *et al*, données non publiées). La synthèse des siRNAs secondaires a pu être montré pour cette voie et nécessiterait les protéines Rdr2 et Cid2 (Marker, Simon *et al*, données non publiées). Deux classes de siRNAs endogènes requièrent soit Rdr1 et Cid1 soit Rdr2 et possiblement Cid2. Il est possible qu'un long ARNdb soit généré à partir des transcrits par les RdRPs puis soit clivé par Dicer pour former des siRNAs en quantité équivalente sur les deux brins ou bien que des siRNAs anti-sens soient générés directement à partir des transcrits à l'aide d'une RdRP. Enfin, la voie inductible par *feeding* d'ARNdb (à droite) est la mieux caractérisée à l'heure actuelle. La protéine Pds2 est la protéine candidate pour être impliquée dans l'import des ARN des vacuoles digestives vers le cytoplasme. Les ARN importés pourraient l'être sous forme simple ou double brin mais dans les deux cas les protéines Rdr1 et Rdr2, Cid1 et Cid2 et Pds1 sont nécessaire avec Dcr1 pour accumuler des siRNAs primaires (voir la première partie de la discussion). Les protéines Rdr1, Cid1 et Pds1 ne sont pas essentielles à la survie de la paramécie indiquant que cette voie d'ARNi est dispensable. Des siRNAs secondaires, en faible quantité en comparaison aux siRNAs primaires (voir la deuxième partie de la discussion), nécessitant Rdr2, semblent être synthétisés par deux mécanismes distincts. Le mécanisme principal pourrait être la synthèse directe de siRNAs anti-sens, indépendamment d'un clivage par Dicer (comme dans le modèle *C. elegans*). Le second mécanisme pourrait être la synthèse d'un long double brin qui serait clivé par une protéine Dicer pour produire des siRNAs qui s'accumulent sur les deux brins. L'addition d'une ou plusieurs uridines non génomique en 3' concerne environ 10% des siRNAs avec un léger biais sens. Cette modification pourrait être nécessaire à la synthèse d'un long ARNdb par une RdRP ou être simplement une marque indiquant le début du processus de dégradation des siRNAs (voir la troisième partie de la discussion). Les siRNAs de ces trois voies présentent tous un pic très fort à 23nt qu'il s'agisse des siRNAs primaires ou secondaires. Les points d'interrogations indiquent les étapes qui demanderaient plus d'analyses pour être confirmées.

Conclusion Discussion

Les résultats obtenus au cours de ma thèse m'ont amené à me poser un certain nombre de questions concernant les mécanismes de cette voie d'ARN interférence. Dans cette discussion, 3 points principaux seront abordés: la forme sous laquelle l'ARN est importé dans le cytoplasme puis traité par la machinerie d'ARNi, la fonction de l'uridylation des siRNAs et enfin le rôle des siRNAs secondaires. Dans la dernière partie, des hypothèses sur la fonction endogène de la voie d'ARN interférence induite par ARNdb seront proposées.

A. Sous quelle forme l'ARN est-il importé dans le cytoplasme ?

La ou les protéines impliquées dans l'import des ARN des vacuoles digestives sont pour le moment inconnues chez la paramécie. Même si la protéine Pds2 semble être une candidate potentielle, son implication n'a pas encore pu être démontrée. La localisation intracellulaire de Pds2 devrait nous permettre de répondre à cette question. La question suivante serait de connaître l'état sous lequel l'ARN est importé de la vacuole vers le cytoplasme. Que ce soit chez le nématode ou chez la paramécie, seul le *feeding* d'ARNdb permet d'activer les voies d'ARNi (Fire et al. 1998^[51], Galvani & Sperling 2002^[17]). Il est donc possible que les ARN simple brin soit incapables d'entrer dans le cytoplasme de l'organisme soit parce qu'ils ne sont pas suffisamment stable dans les vacuoles digestives soit parce que la protéine capable d'importer les ARN est spécifique des formes double-brin. Chez le nématode, le *feeding* d'un ARN simple brin est capable d'être pris en charge par la protéine Cas9 pour introduire des mutations spécifiques de l'ADN (Liu et al. 2014^[112]). Cet exemple montre qu'un ARN simple brin ingéré peut être activement importé dans le cytoplasme et être suffisamment stable pour être utilisé par l'organisme. De plus, j'ai pu montrer que les paramécies sont capables de produire des siRNAs anti-sens à partir des ARN ribosomiaux des bactéries *E. coli* qu'elles ingèrent, ce qui montre qu'un ARN simple brin est capable d'entrer dans le cytoplasme de la paramécie. La stabilité des ARN en général dépend de leur capacité intrinsèque à former des structures secondaires, stabilisées par des appariements intramoléculaires. Les ARNr étant des molécules particulièrement structurées et en très grande quantité, elles sont peut-être plus stables dans les vacuoles digestives. La présence ou l'absence de siRNAs anti-sens sur les ARNm bactériens permettrait de répondre à cette question.

Si les ARNs peuvent entrer dans le cytoplasme ce n'est pas pour autant qu'ils sont pris en charge par la machinerie d'ARN interférence. En effet, les conditions nécessaires pour qu'une RdRP ait la capacité de synthétiser le brin complémentaire d'un ARN sont assez mal connus. La première idée était que les RdRPs ont besoin d'une amorce simple brin qui en s'appariant avec sa cible déclencherait la synthèse du brin complémentaire. C'est le cas par exemple lorsqu'un ARNm est ciblé par un sRNA, les RdRPs peuvent alors utiliser ce duplex pour synthétiser des siRNAs secondaires. Cependant on a pu montrer que les RdRP (de paramécie notamment) sont capables de synthétiser des siRNAs en 3' de la région induite par

feeding d'ARNdb et donc d'amplifier sans utiliser de siRNA comme amorce. Cette activité des RdRP a aussi été montrée chez le nématode et les plantes (Sijen et al. 2007^[113]). Cette transivité en 3' n'est pourtant pas une démonstration de l'activité amorce-indépendante des RdRPs. Chez *T.thermophila* il a été suggéré par des expériences *in vitro* que l'extrémité 3' d'un ARN peut-être uridylé pour former une boucle et servir d'amorce à la RdRP. L'idée plus générale serait que ce sont les ARN simple brin aberrants qui sont pris en charge par les RdRPs. Cette notion assez vague peut inclure des transcrits non codants, des ARNs tronqués générés par des cassures double brin, des transcrits naissants, des ARNm mal épissés,... Les mécanismes permettant aux RdRPs de se fixer sur ces ARNs aberrants sont pour le moment très peu connus, il est donc imaginable qu'un ARN simple brin en provenance des vacuoles digestives soit reconnu par une RdRP s'il est par exemple déjà chargé dans la protéine qui aurait servi à l'importer des vacuoles digestives vers le cytoplasme.

Enfin, il est aussi possible que la quantité de siRNAs produite par l'activité éventuelle des RdRPs sur un ARN simple brin soit insuffisante pour induire l'ARNi en comparaison à un clivage par Dicer. C'est d'ailleurs l'impression laissée par la faible quantité de siRNAs secondaires produite chez *P. tetraurelia* (voir le paragraphe suivant). Mais si les RdRPs ne sont pas efficaces pour produire des siRNAs secondaires, comment expliquer que les deux RdRPs soient requises pour synthétiser les siRNAs primaires chez la paramécie. 2 hypothèses pourraient expliquer la nécessité des RdRPs pour induire la production de siRNAs primaires à partir d'un ARNdb importé dans le cytoplasme. La première hypothèse est que les RdRPs pourraient former un complexe avec Dicer et être indispensables à son activité. L'interaction entre une protéine Dicer et une RdRP a déjà été montrée *in vitro* chez le cilié *T. thermophila* (Lee & Collins 2007^[114]). Les auteurs dans cet article proposent que les RdRPs puissent être indispensables pour synthétiser un brin complémentaire tri-phosphorylé après chaque coupure par Dicer du long ARNdb (ceci suggère que les deux brins de l'ARN doivent se dissocier à chaque coupure). En effet, l'efficacité de clivage par Dicer semble beaucoup plus grande lorsque l'ARNdb est tri-phosphorylé. La seconde hypothèse est que le nombre de molécules d'ARNdb capables d'entrer dans le cytoplasme est très faible, les RdRPs seraient alors essentielles pour amplifier le dsRNA et ainsi générer un nombre suffisant de siRNAs primaires pour réprimer l'ARNm. Chez le nématode, une étude a montré cependant que l'ARNdb n'est pas amplifié (Pak et al. 2012^[53]). Pour autant, ces hypothèses n'expliquent pas pourquoi deux RdRPs sont nécessaires.

B. Le rôle de l'uridylation des siRNAs

Les modifications des siRNAs semblent assez complexes puisque plusieurs études, dans différentes conditions ont révélés des modifications 3' variables. Le séquençage des siRNAs induits par *feeding* au cours de la conjugaison a montré que 50% des siRNAs sont modifiés par une ou plusieurs adénines non génomiques en 3' (Lepere et al. 2009^[27]), alors que ceux induits en croissance végétative sont uridylés pour 10% d'entre eux. De plus, des bibliothèques réalisées dans le laboratoire de M. Simon ont révélé un taux non négligeable de mono-C non génomique en plus de l'uridylation. Il semble alors que les modifications des siRNAs soit dépendantes de l'état des cellules au moment de l'extraction. Il est aussi possible que des biais soient introduits selon les différentes conditions d'extraction des petits ARN, de construction des bibliothèques ou de séquençage des siRNAs (Raabe et al. 2014^[115]). Cependant, au moins deux hypothèses peuvent expliquer l'observation d'un certain taux d'uridylation dans les siRNAs induits par l'ARNdb.

La première hypothèse, proposée dans le second article, est que l'uridylation est un signal pour la dégradation des siRNAs. En effet, seule une fraction des siRNAs primaires et secondaires est modifiée. De plus, il apparaît qu'un certain nombre de siRNAs uridylés ont leur séquence génomique plus courte que 23 nucléotides ce qui suggère qu'une partie des siRNAs uridylés ont été partiellement dégradés. L'uridylation des siRNAs entraînant leur dégradation a auparavant été montrée chez les plantes (Zhao et al. 2012^[116]) et chez *C. elegans* (van Wolfswinkel et al. 2009^[117]), cependant cette hypothèse n'explique pas pourquoi le biais anti-sens des siRNAs non modifiés disparaît pour les siRNAs uridylés. Il semble étonnant que le mécanisme de dégradation affecte plus en proportion les siRNAs sens (environ 50% d'entre eux) que les siRNAs anti-sens (environ 1% d'entre eux). De plus, les deux uridyl-transférases qui sont impliquées dans la voie sont essentielles à son fonctionnement et nécessaire pour la synthèse des siRNAs uridylés et non uridylés. Il faut alors suggérer que l'uridylation observée, si elle entraîne la dégradation des siRNAs, est ajoutée par une autre uridyl-transférase qui ne serait pas utile à l'accumulation des siRNAs et qui n'aurait donc pas été trouvée dans le crible.

La seconde hypothèse est que l'uridylation réalisée par Cid1 ou Cid2 pourrait être une trace du mécanisme de synthèse des siRNAs et être essentielle à la production de tous les siRNAs (uridylés ou non). Cette possibilité a déjà été évoquée dans le modèle *Tetrahymena thermophila* où, *in vitro*, l'addition d'uridines permettrait la formation d'une boucle en 3' et servirait d'amorce à la RdRP (voir introduction p31) (Talsky & Collins 2010^[91]). En considérant cette information, le modèle suivant peut être proposé pour la paramécie (Figure 49).

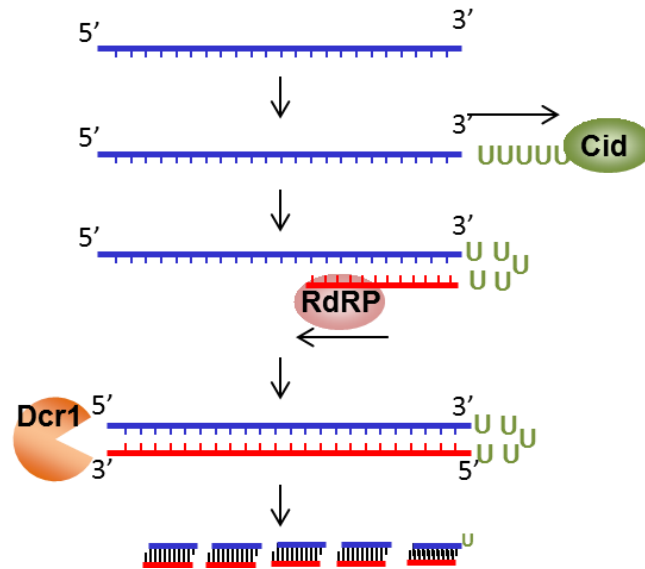


Figure 49: Hypothèse sur le rôle de l'uridylation. L'ARN sens inducteur est en bleu, l'ARN anti-sens synthétisé par la RdRP en rouge.

L'ARN simple brin pourrait être uridylé par une nucléotidyl-transférase Cid et ainsi permettre la synthèse du brin complémentaire par la RdRP. Si Dicer clive ensuite ces longs ARNdb, on peut supposer que seul le siRNA en 3', à proximité de la boucle poly U, serait uridylé. Ce modèle à l'avantage d'expliquer pourquoi seulement une fraction des siRNAs est uridylé et pourquoi les Cid sont essentiels à la synthèse de tous les siRNAs. Pour les siRNAs primaires, la quantité équivalente de siRNAs uridylés sur chacun des 2 brins est concordante avec ce modèle puisque que l'ARNdb n'est pas orienté mais selon ce modèle, pour expliquer la présence des quelques siRNAs secondaire anti-sens uridylés, il faudrait imaginer qu'un long ARN anti-sens à l'ARNm (et en dehors de la région de l'ARNdb inducteur) puisse servir de matrice à la RdRP.

C. Pas d'amplification de l'ARNi induite par ARNdb chez *P. tetraurelia*

La répression des gènes par ARNi chez la paramécie nécessite un apport continu de l'ARN inducteur pour être maintenu. L'injection d'ARNdb dans le cytoplasme de la paramécie est suffisamment efficace pour éteindre l'expression d'un gène seulement pendant 48h (Galvani & Sperling 2002^[17]). En effet, *P. tetraurelia* réalisant 4 divisions par jour, la quantité d'ARNdb initialement injecté est divisée par 256 en 48h. Cette observation suggère que l'ARNdb injecté n'est pas amplifié ou en tout cas pas suffisamment pour maintenir la répression mais aussi que l'amplification de l'ARNi par la production de siRNAs secondaires n'est pas suffisante au-delà de 48h. Le même effet est observé par *feeding* d'ARNdb: tan que les paramécies sont nourries avec les bactéries transformées la répression du gène ciblé est maintenue, mais le phénotype réverte rapidement lorsque le milieu de culture est remplacé par

un milieu contenant des bactéries non transformées. J'ai pu cependant observer qu'il existe une production de siRNAs secondaires dépendante de la RdRP RDR2 chez la paramécie. Contrairement à *C. elegans* où la quantité de siRNAs secondaires est plus grande que la quantité de siRNAs primaires, chez *P. tetraurelia*, les siRNAs secondaires ne représentent que 1% de tous les siRNAs produits sur le gène ciblé. De plus ces siRNAs secondaires semblent non fonctionnels puisque le mutant *dcr1-5.5* en produit une quantité équivalente au sauvage mais n'est plus capable de faire d'ARN interférence.

Deux hypothèses sont possibles pour expliquer ces observations. La première est que le *feeding* d'ARNdb en continu apporte dans le cytoplasme une grande quantité d'ARNdb et donc une grande quantité de siRNAs primaires qui vont efficacement réprimer les ARNm. Si c'est le cas, la faible quantité de siRNAs secondaires pourrait être simplement expliquée par l'absence de l'ARNm nécessaire à leur production. De plus, étant donné que Rdr2 est impliqué à la fois dans la synthèse des siRNAs primaires et des siRNAs secondaires, il est possible que la quantité de protéines soit limitante pour la synthèse des secondaires lorsque des siRNAs primaires sont produits massivement. Dans cette hypothèse la voie de synthèse des siRNAs secondaires serait donc inutile lors du *feeding* massif d'ARNdb mais pourrait être essentielle dans d'autres conditions par exemple si la quantité d'ARNdb initiale est très faible ou que l'ARN inducteur est sous forme simple brin.

La seconde hypothèse (non exclusive avec la première) est que la voie de production des siRNAs secondaires ne soit plus fonctionnelle. Il existe plusieurs exemples chez les eucaryotes où les voies d'ARNi ont partiellement ou totalement disparues. C'est le cas par exemple de *Saccharomyces cerevisiae*, *Trypanosoma cruzi* ou encore *Plasmodium falciparum* (Shabalina & Koonin 2008^[6]). L'absence des voies d'ARNi chez ces eucaryotes unicellulaires peut s'expliquer par l'isolation de l'espèce dans un milieu environnemental très peu stringent, c'est le cas des parasites tel que *Trypanosoma* ou *Plasmodium* qui sont dépendants d'un organisme hôte qui leur confère une protection contre le milieu extérieur. Pour *Saccharomyces cerevisiae*, les levures déficientes pour l'ARNi auraient été sélectionnées naturellement par un virus à ARNdb qui confère à la levure infecté la capacité de tuer les levures non infectés (système "killer virus") (Drinneberg et al. 2011^[118]). Cet exemple montre un cas où la voie d'ARNi antivirale apporte un net désavantage sélectif. Dans le cas des paramécies, la souche 51 de *Paramecium tetraurelia* est conservée en laboratoire depuis 1939 et l'absence de pression de sélection sur la voie d'ARN interférence induite par *feeding* (voie non essentielle dans les conditions de laboratoire) aurait pu conduire à l'introduction de mutations dans RDR2 rendant inefficace la production des siRNAs secondaires ou dans une protéine Piwi affectant leur stabilité. Chez *Tetrahymena thermophila*, le *feeding* d'ARNdb ne fonctionne pas et le gène *Pds1* (spécifique de la voie induite par *feeding* chez la paramécie) est absent suggérant que cette voie d'ARNi aurait pu être perdue au cours de l'évolution.

En conclusion, le mécanisme de production des siRNAs secondaires existe chez *Paramecium tetraurelia* mais, au moins pour la voie induite par *feeding*, ne permet pas

d'amplifier la quantité de siRNAs et ne semble pas participer à la dégradation de l'ARNm ciblé.

D. La fonction endogène des siRNAs induits par *feeding* d'ARNdb.

Si la fonction endogène de cette voie reste mystérieuse, la compréhension des mécanismes de production des siRNAs nous permet d'imaginer plusieurs raisons qui auraient contribué à la conservation de cette voie d'ARN interférence au cours de l'évolution.

L'idée générale est que les voies d'ARNi sont utiles pour la défense contre les parasites moléculaires tels que les transposons et les virus. Jusqu'à présent aucun virus n'infectant la paramécie n'a été trouvé contrairement à plusieurs protozoaires pour lesquels des virus ont été décrits (Hyman & Abedon 2012^[119]). Si effectivement aucun virus n'est capable d'infecter la paramécie, on peut se demander si ce sont ses voies d'ARNi qui lui offrent une protection efficace contre les infections virales. Cependant, la plupart des virus de plantes ou d'animaux ont développé des mécanismes pour contourner ou inactiver les voies d'ARNi de l'hôte, il est donc peu probable que ce soit les voies d'ARNi de la paramécie qui empêchent à elles seules les infections virales. En ce qui concerne les transposons, il est théoriquement possible que l'ARNi empêche l'intégration de transposons qui proviendraient de la nourriture en réprimant les transcrits, cependant le code génétique de la paramécie étant différent de celui des bactéries (un seul codon stop : TGA), il est peu probable qu'un transposon bactérien puisse facilement intégrer le génome de la paramécie.

En revanche, plusieurs endosymbionts ont été décrits pour la paramécie. Par exemple l'algue verte unicellulaire *Chlorella* (Fujishima & Kodama 2012^[120]) ou encore la bactérie *Caedibacter caryophilus* affiliée aux *Alphaproteobacteria* (Beier et al. 2002^[121]). Cette dernière donne à son hôte la capacité de tuer ces congénères qui ne sont pas infectés par la bactérie. Par définition, un endosymbiote doit interagir avec son hôte pour un certain nombre de processus cellulaires. Il est alors possible que l'ARNi ait un rôle important à jouer dans ces relations endosymbiotiques. Certains ARN de la bactérie infectieuse (ARN ribosomiaux ou même ARNm) pourraient être réprimés par les voies d'ARNi de la paramécie pour empêcher l'infection ou simplement pour limiter sa prolifération dans l'hôte. Si c'est le cas, cela suggère que des ARN seraient excrétés de la bactérie dans le cytoplasme de la paramécie pour être traités par l'ARNi puis que les siRNAs produits pourraient entrer dans la bactérie pour éteindre l'expression des ARNm. L'étude de la prolifération des endosymbionts dans des mutants déficients pour l'ARNi pourrait être particulièrement intéressante.

Références bibliographiques

Primary Sources

Secondary Sources

Uncategorized References

1. Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell*, **2**, 279-289.
2. van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N. and Stuitje, A.R. (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *The Plant cell*, **2**, 291-299.
3. Guo, S. and Kemphues, K.J. (1995) par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*, **81**, 611-620.
4. Baulcombe, D.C. (1996) RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant molecular biology*, **32**, 79-88.
5. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806-811.
6. Shabalina, S.A. and Koonin, E.V. (2008) Origins and evolution of eukaryotic RNA interference. *Trends in ecology & evolution*, **23**, 578-587.
7. Gantier, M.P. and Williams, B.R. (2007) The response of mammalian cells to double-stranded RNA. *Cytokine & growth factor reviews*, **18**, 363-371.
8. Sarkies, P. and Miska, E.A. (2013) RNAi pathways in the recognition of foreign RNA: antiviral responses and host-parasite interactions in nematodes. *Biochemical Society transactions*, **41**, 876-880.
9. Hohn, T. and Vazquez, F. (2011) RNA silencing pathways of plants: silencing and its suppression by plant DNA viruses. *Biochimica et biophysica acta*, **1809**, 588-600.
10. Maillard, P.V., Ciaudo, C., Marchais, A., Li, Y., Jay, F., Ding, S.W. and Voinnet, O. (2013) Antiviral RNA interference in mammalian cells. *Science*, **342**, 235-238.
11. Tabara, H., Grishok, A. and Mello, C.C. (1998) RNAi in *C. elegans*: soaking in the genome sequence. *Science*, **282**, 430-431.
12. Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A. and Mello, C.C. (1999) The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell*, **99**, 123-132.
13. Sundaram, P., Echaliier, B., Han, W., Hull, D. and Timmons, L. (2006) ATP-binding cassette transporters are required for efficient RNA interference in *Caenorhabditis elegans*. *Mol Biol Cell*, **17**, 3678-3688.
14. Sundaram, P., Han, W., Cohen, N., Echaliier, B., Albin, J. and Timmons, L. (2008) *Caenorhabditis elegans* ABCRNAi transporters interact genetically with rde-2 and mut-7. *Genetics*, **178**, 801-814.
15. Winston, W.M., Sutherlin, M., Wright, A.J., Feinberg, E.H. and Hunter, C.P. (2007) *Caenorhabditis elegans* SID-2 is required for environmental RNA interference. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10565-10570.
16. Nuez, I. and Felix, M.A. (2012) Evolution of susceptibility to ingested double-stranded RNAs in *Caenorhabditis* nematodes. *PLoS one*, **7**, e29811.
17. Galvani, A. and Sperling, L. (2002) RNA interference by feeding in *Paramecium*. *Trends in genetics : TIG*, **18**, 11-12.

18. Zhou, R., Mohr, S., Hannon, G.J. and Perrimon, N. (2014) Inducing RNAi in Drosophila cells by soaking with dsRNA. *Cold Spring Harbor protocols*, **2014**.
19. Dunoyer, P., Brosnan, C.A., Schott, G., Wang, Y., Jay, F., Alioua, A., Himber, C. and Voinnet, O. (2010) An endogenous, systemic RNAi pathway in plants. *The EMBO journal*, **29**, 1699-1712.
20. Axtell, M.J. (2013) Classification and comparison of small RNAs from plants. *Annual review of plant biology*, **64**, 137-159.
21. Johanson, T.M., Lew, A.M. and Chong, M.M. (2013) MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. *Open biology*, **3**, 130144.
22. Macrae, I.J., Li, F., Zhou, K., Cande, W.Z. and Doudna, J.A. (2006) Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harbor symposia on quantitative biology*, **71**, 73-80.
23. Kim, K., Lee, Y.S., Harris, D., Nakahara, K. and Carthew, R.W. (2006) The RNAi pathway initiated by Dicer-2 in Drosophila. *Cold Spring Harbor symposia on quantitative biology*, **71**, 39-44.
24. Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S. and Bass, B.L. (2010) Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. *Rna*, **16**, 893-903.
25. Cook, A. and Conti, E. (2006) Dicer measures up. *Nature structural & molecular biology*, **13**, 190-192.
26. Ji, X. (2008) The mechanism of RNase III action: how dicer dices. *Current topics in microbiology and immunology*, **320**, 99-116.
27. Lepere, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duhaucourt, S. and Meyer, E. (2009) Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. *Nucleic acids research*, **37**, 903-915.
28. Dlakic, M. (2006) DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics*, **22**, 2711-2714.
29. Liu, Q., Feng, Y. and Zhu, Z. (2009) Dicer-like (DCL) proteins in plants. *Functional & integrative genomics*, **9**, 277-286.
30. Xie, Z. and Qi, X. (2008) Diverse small RNA-directed silencing pathways in plants. *Biochimica et biophysica acta*, **1779**, 720-724.
31. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J. and Carthew, R.W. (2004) Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, **117**, 69-81.
32. Kuhn, C.D. and Joshua-Tor, L. (2013) Eukaryotic Argonautes come into focus. *Trends in biochemical sciences*, **38**, 263-271.
33. Meister, G. (2013) Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*, **14**, 447-459.
34. Steiner, F.A., Hoogstrate, S.W., Okihara, K.L., Thijssen, K.L., Ketting, R.F., Plasterk, R.H. and Sijen, T. (2007) Structural features of small RNA precursors determine Argonaute loading in Caenorhabditis elegans. *Nature structural & molecular biology*, **14**, 927-933.
35. Vaucheret, H. (2008) Plant ARGONAUTES. *Trends in plant science*, **13**, 350-358.
36. Lee, H.C., Chang, S.S., Choudhary, S., Aalto, A.P., Maiti, M., Bamford, D.H. and Liu, Y. (2009) qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature*, **459**, 274-277.
37. Grishok, A. (2013) Biology and Mechanisms of Short RNAs in Caenorhabditis elegans. *Advances in genetics*, **83**, 1-69.
38. Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C. *et al.* (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, **133**, 116-127.
39. Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R. and Tuschl, T. (2005) Identification of novel argonaute-associated proteins. *Current biology : CB*, **15**, 2149-2155.

40. Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T. and Ishikawa, M. (2010) In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Molecular cell*, **39**, 282-291.
41. Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T. and Tomari, Y. (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Molecular cell*, **39**, 292-299.
42. Johnston, M., Geoffroy, M.C., Sobala, A., Hay, R. and Hutvagner, G. (2010) HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Mol Biol Cell*, **21**, 1462-1469.
43. Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. and Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740-744.
44. 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.
45. Nishida, K.M., Miyoshi, K., Ogino, A., Miyoshi, T., Siomi, H. and Siomi, M.C. (2013) Roles of R2D2, a cytoplasmic D2 body component, in the endogenous siRNA pathway in Drosophila. *Molecular cell*, **49**, 680-691.
46. Tomari, Y., Matranga, C., Haley, B., Martinez, N. and Zamore, P.D. (2004) A protein sensor for siRNA asymmetry. *Science*, **306**, 1377-1380.
47. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199-208.
48. Nykanen, A., Haley, B. and Zamore, P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, **107**, 309-321.
49. Kwak, P.B. and Tomari, Y. (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature structural & molecular biology*, **19**, 145-151.
50. Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V. and Liu, Q. (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science*, **325**, 750-753.
51. Maiti, M., Lee, H.C. and Liu, Y. (2007) QIP, a putative exonuclease, interacts with the Neurospora Argonaute protein and facilitates conversion of duplex siRNA into single strands. *Genes & development*, **21**, 590-600.
52. Steiner, F.A., Okihara, K.L., Hoogstrate, S.W., Sijen, T. and Ketting, R.F. (2009) RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in *Caenorhabditis elegans*. *Nature structural & molecular biology*, **16**, 207-211.
53. Pak, J., Maniar, J.M., Mello, C.C. and Fire, A. (2012) Protection from feed-forward amplification in an amplified RNAi mechanism. *Cell*, **151**, 885-899.
54. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.
55. Grewal, S.I. (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Current opinion in genetics & development*, **20**, 134-141.
56. Taliaferro, J.M., Aspden, J.L., Bradley, T., Marwha, D., Blanchette, M. and Rio, D.C. (2013) Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression. *Genes & development*, **27**, 378-389.
57. Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G. and Qi, Y. (2012) A role for small RNAs in DNA double-strand break repair. *Cell*, **149**, 101-112.
58. Cogoni, C. and Macino, G. (1999) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature*, **399**, 166-169.
59. Sardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N. and Maine, E.M. (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Current biology : CB*, **10**, 169-178.

60. Dalmay, T., Hamilton, A., Rudd, S., Angell, S. and Baulcombe, D.C. (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, **101**, 543-553.
61. Zong, J., Yao, X., Yin, J., Zhang, D. and Ma, H. (2009) Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene*, **447**, 29-39.
62. Willmann, M.R., Endres, M.W., Cook, R.T. and Gregory, B.D. (2011) The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. *The Arabidopsis book / American Society of Plant Biologists*, **9**, e0146.
63. Wang, X., Laurie, J.D., Liu, T., Wentz, J. and Liu, X.S. (2011) Computational dissection of Arabidopsis smRNAome leads to discovery of novel microRNAs and short interfering RNAs associated with transcription start sites. *Genomics*, **97**, 235-243.
64. Fei, Q., Xia, R. and Meyers, B.C. (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant cell*, **25**, 2400-2415.
65. Axtell, M.J., Jan, C., Rajagopalan, R. and Bartel, D.P. (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell*, **127**, 565-577.
66. Chen, H.M., Li, Y.H. and Wu, S.H. (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3318-3323.
67. Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K. and Tabara, H. (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *The EMBO journal*, **26**, 5007-5019.
68. Pak, J. and Fire, A. (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science*, **315**, 241-244.
69. Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Conte, D., Jr. and Duchaine, T.F. (2012) Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nature structural & molecular biology*, **19**, 90-97.
70. Makeyev, E.V. and Bamford, D.H. (2002) Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Molecular cell*, **10**, 1417-1427.
71. Calo, S., Nicolas, F.E., Vila, A., Torres-Martinez, S. and Ruiz-Vazquez, R.M. (2012) Two distinct RNA-dependent RNA polymerases are required for initiation and amplification of RNA silencing in the basal fungus *Mucor circinelloides*. *Molecular microbiology*, **83**, 379-394.
72. Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E. and Zamore, P.D. (2011) Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Molecular cell*, **44**, 572-584.
73. Seila, A.C. and Sharp, P.A. (2008) Small RNAs tell big stories in Whistler. *Nature cell biology*, **10**, 630-633.
74. Ji, L. and Chen, X. (2012) Regulation of small RNA stability: methylation and beyond. *Cell research*, **22**, 624-636.
75. Kim, Y.K., Heo, I. and Kim, V.N. (2010) Modifications of small RNAs and their associated proteins. *Cell*, **143**, 703-709.
76. Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Current biology : CB*, **12**, 1484-1495.
77. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, **307**, 932-935.
78. Yang, Z., Ebright, Y.W., Yu, B. and 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 research*, **34**, 667-675.

79. Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P. and Zamore, P.D. (2007) The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current biology : CB*, **17**, 1265-1272.
80. Billi, A.C., Alessi, A.F., Khivansara, V., Han, T., Freeberg, M., Mitani, S. and Kim, J.K. (2012) The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. *PLoS genetics*, **8**, e1002617.
81. Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J., Roovers, E.F., Ladurner, P., Berezikov, E. and Ketting, R.F. (2010) Hen1 is required for oocyte development and piRNA stability in zebrafish. *The EMBO journal*, **29**, 3688-3700.
82. Kamminga, L.M., van Wolfswinkel, J.C., Luteijn, M.J., Kaaij, L.J., Bagijn, M.P., Sapetschnig, A., Miska, E.A., Berezikov, E. and Ketting, R.F. (2012) Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of *Caenorhabditis elegans*. *PLoS genetics*, **8**, e1002702.
83. Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J. and Tewari, M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome research*, **21**, 1450-1461.
84. Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S., Baba, T. and Suzuki, T. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes & development*, **23**, 433-438.
85. Jones, M.R., Quinton, L.J., Blahna, M.T., Neilson, J.R., Fu, S., Ivanov, A.R., Wolf, D.A. and Mizgerd, J.P. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nature cell biology*, **11**, 1157-1163.
86. Ibrahim, F., Rymarquis, L.A., Kim, E.J., Becker, J., Balassa, E., Green, P.J. and Cerutti, H. (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 3906-3911.
87. Wedeles, C.J., Wu, M.Z. and Claycomb, J.M. (2013) A multitasking Argonaute: exploring the many facets of *C. elegans* CSR-1. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, **21**, 573-586.
88. Jose, A.M., Garcia, G.A. and Hunter, C.P. (2011) Two classes of silencing RNAs move between *Caenorhabditis elegans* tissues. *Nature structural & molecular biology*, **18**, 1184-1188.
89. Lee, S.R. and Collins, K. (2006) Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes & development*, **20**, 28-33.
90. Lee, S.R., Talsky, K.B. and Collins, K. (2009) A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *Rna*, **15**, 1363-1374.
91. Talsky, K.B. and Collins, K. (2010) Initiation by a eukaryotic RNA-dependent RNA polymerase requires looping of the template end and is influenced by the template-tailing activity of an associated uridylyltransferase. *The Journal of biological chemistry*, **285**, 27614-27623.
92. Adl, S.M., Simpson, A.G., Lane, C.E., Lukes, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V. *et al.* (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol*, **59**, 429-493.
93. Coyne, R.S., Lhuillier-Akakpo, M. and Duharcourt, S. (2012) RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biology of the cell / under the auspices of the European Cell Biology Organization*, **104**, 309-325.
94. Singh, D.P., Saudemont, B., Guglielmi, G., Arnaiz, O., Gout, J.F., Prajer, M., Potekhin, A., Przybos, E., Aubusson-Fleury, A., Bhullar, S. *et al.* (2014) Genome-defence small RNAs exapted for epigenetic mating-type inheritance. *Nature*, **509**, 447-452.
95. Le Mouel, A., Butler, A., Caron, F. and Meyer, E. (2003) Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in *paramecia*. *Eukaryotic cell*, **2**, 1076-1090.

96. Duharcourt, S., Keller, A.M. and Meyer, E. (1998) Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Molecular and cellular biology*, **18**, 7075-7085.
97. Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.M., Wilkes, C.D., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouel, A. *et al.* (2012) The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS genetics*, **8**, e1002984.
98. Baudry, C., Malinsky, S., Restituuto, M., Kapusta, A., Rosa, S., Meyer, E. and Betermier, M. (2009) PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes & development*, **23**, 2478-2483.
99. Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aich, N. *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**, 171-178.
100. Coleman, A.W. (2005) *Paramecium aurelia* revisited. *J Eukaryot Microbiol*, **52**, 68-77.
101. Sandoval, P.Y., Swart, E.C., Arambasic, M. and Nowacki, M. (2014) Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Developmental cell*, **28**, 174-188.
102. Mochizuki, K. and Gorovsky, M.A. (2005) A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes & development*, **19**, 77-89.
103. Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M. and Meyer, E. (2011) Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic acids research*, **39**, 4249-4264.
104. Marker, S., Le Mouel, A., Meyer, E. and Simon, M. (2010) Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic acids research*, **38**, 4092-4107.
105. Ruiz, F., Vayssie, L., Klotz, C., Sperling, L. and Madeddu, L. (1998) Homology-dependent gene silencing in *Paramecium*. *Mol Biol Cell*, **9**, 931-943.
106. Galvani, A. and Sperling, L. (2001) Transgene-mediated post-transcriptional gene silencing is inhibited by 3' non-coding sequences in *Paramecium*. *Nucleic acids research*, **29**, 4387-4394.
107. Galvani, A. and Sperling, L. (2000) Regulation of secretory protein gene expression in *paramecium* role of the cortical exocytotic sites. *European journal of biochemistry / FEBS*, **267**, 3226-3234.
108. Garnier, O., Serrano, V., Duharcourt, S. and Meyer, E. (2004) RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Molecular and cellular biology*, **24**, 7370-7379.
109. Simon, M.C., Marker, S. and Schmidt, H.J. (2006) Posttranscriptional control is a strong factor enabling exclusive expression of surface antigens in *Paramecium tetraurelia*. *Gene expression*, **13**, 167-178.
110. Lepere, G., Betermier, M., Meyer, E. and Duharcourt, S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes & development*, **22**, 1501-1512.
111. Hall, I.M., Noma, K. and Grewal, S.I. (2003) RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 193-198.
112. Liu, P., Long, L., Xiong, K., Yu, B., Chang, N., Xiong, J.W., Zhu, Z. and Liu, D. (2014) Heritable/conditional genome editing in *C. elegans* using a CRISPR-Cas9 feeding system. *Cell research*, **24**, 886-889.
113. Sijen, T., Steiner, F.A., Thijssen, K.L. and Plasterk, R.H. (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, **315**, 244-247.

114. Lee, S.R. and Collins, K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nature structural & molecular biology*, **14**, 604-610.
115. Raabe, C.A., Tang, T.H., Brosius, J. and Rozhdestvensky, T.S. (2014) Biases in small RNA deep sequencing data. *Nucleic acids research*, **42**, 1414-1426.
116. Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B. and Chen, X. (2012) The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Current biology : CB*, **22**, 689-694.
117. van Wolfswinkel, J.C., Claycomb, J.M., Batista, P.J., Mello, C.C., Berezikov, E. and Ketting, R.F. (2009) CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell*, **139**, 135-148.
118. Drinnenberg, I.A., Fink, G.R. and Bartel, D.P. (2011) Compatibility with killer explains the rise of RNAi-deficient fungi. *Science*, **333**, 1592.
119. Hyman, P. and Abedon, S.T. (2012) Smaller fleas: viruses of microorganisms. *Scientifica*, **2012**, 734023.
120. Fujishima, M. and Kodama, Y. (2012) Endosymbionts in paramecium. *European journal of protistology*, **48**, 124-137.
121. Beier, C.L., Horn, M., Michel, R., Schweikert, M., Gortz, H.D. and Wagner, M. (2002) The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the Rickettsiales (Alphaproteobacteria) and to *Francisella tularensis* (Gammaproteobacteria). *Applied and environmental microbiology*, **68**, 6043-6050.

Annexes

I. Annexe I : Documents supplémentaires de l'article 1.

A forward genetic screen reveals essential and non-essential RNAi factors in *Paramecium tetraurelia*

Simone Marker, Quentin Carradec, Véronique Tanty, Olivier Arnaiz and Eric Meyer

Supplementary Figures S1-S8

Supplementary Tables S1-S7

Supplementary references

– Figure S1 –

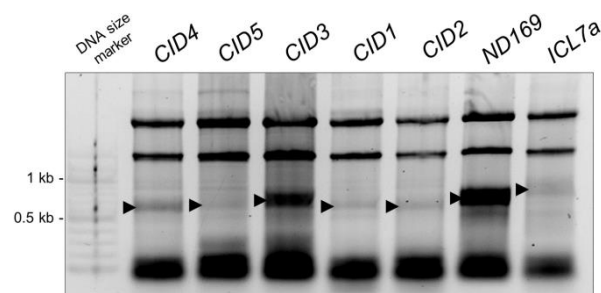


Figure S1: DsRNA isolated from *E. coli* HT115 DE3 food bacteria to induce silencing of Cid1-like genes, *ND169* (reporter) and *ICL7a* (control). Bacterial total RNA was extracted by hot phenol (pH 4.0) and purified by ethanol-0.4M sodium chloride precipitation. High molecular weight bands represent 16S and 26S ribosomal RNA. Note that the amount of dsRNA (black arrowheads) produced from different gene fragments is highly variable, but reproducible for each fragment.

- Figure S2 -

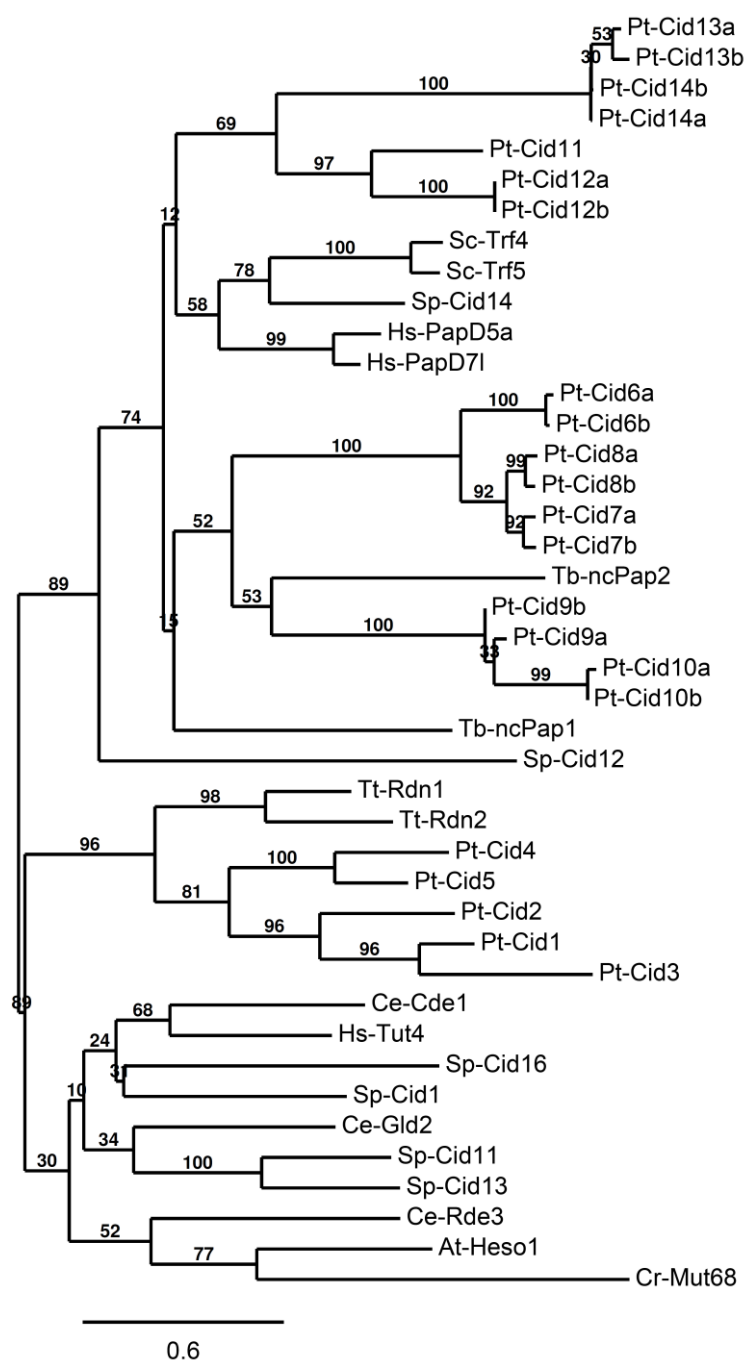


Figure S2: Non-canonical poly(A) RNA polymerases (Cid1-like nucleotidyl transferases) identified in the *P. tetraurelia* MAC genome. Alignments of protein sequences corresponding to the conserved domain were done using the MUSCLE software. Phylogenetic analysis was carried out using PhyML (bootstrapping procedure, 100 bootstraps) with default parameters and trees were created by TreeDyn (<http://www.phylogeny.fr>) (78). Pt, *Paramecium tetraurelia*; Tt, *Tetrahymena thermophila*; Sp, *Schizosaccharomyces pombe*; Sc, *Saccharomyces cerevisiae*; Tb, *Trypanosoma brucei*; Ce, *Caenorhabditis elegans*; At, *Arabidopsis thaliana*; Hs, *Homo sapiens* (accession numbers see Table S7B).

– Figure S3 –

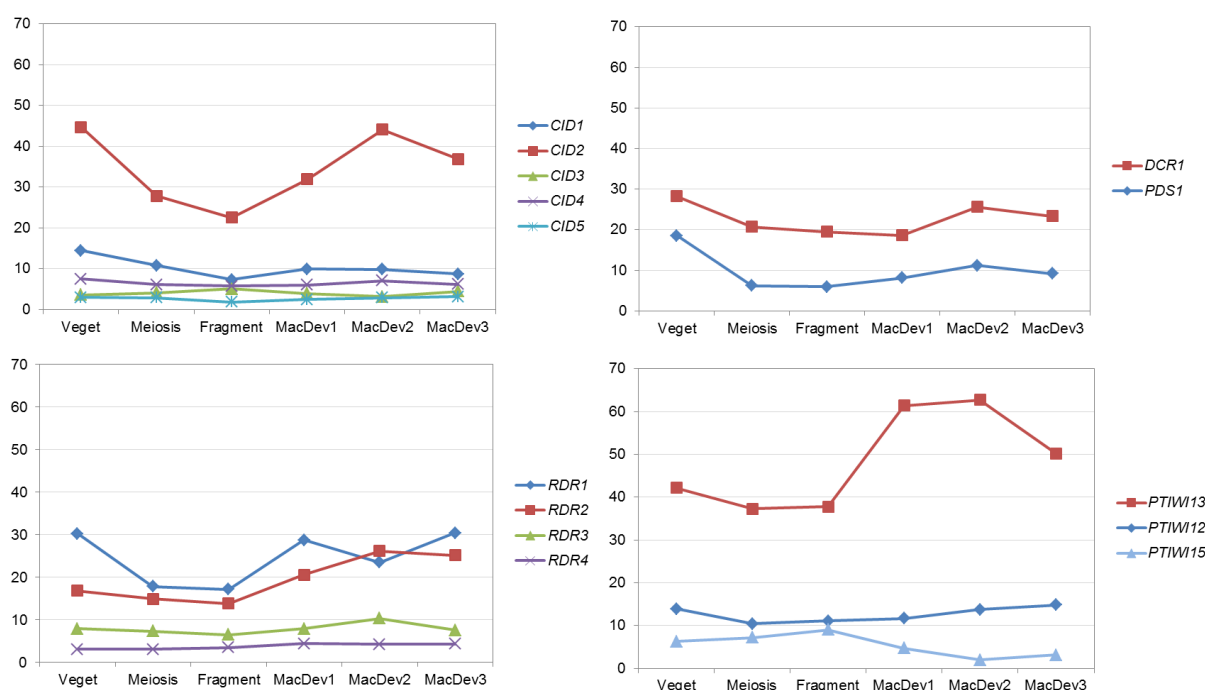


Figure S3: Expression levels of genes involved in dsRNA-induced RNAi during vegetative growth (veget) and autogamy (meiosis of MICs, fragmentation of the old MAC, developmental stages of the zygotic new MAC) according to microarray experiments (4-6 replicates each (25,79)). Genes specifically involved in dsRNA-RNA-induced silencing are represented in blue and light blue, genes also involved in transgene-induced silencing are represented in red. All changes in expression are not significant. Expression data are publicly available at the Gene Expression Omnibus database (80), under accession numbers GSE17996, GSE17997, GSE17998 and GSE18002.

- Figure S4 -

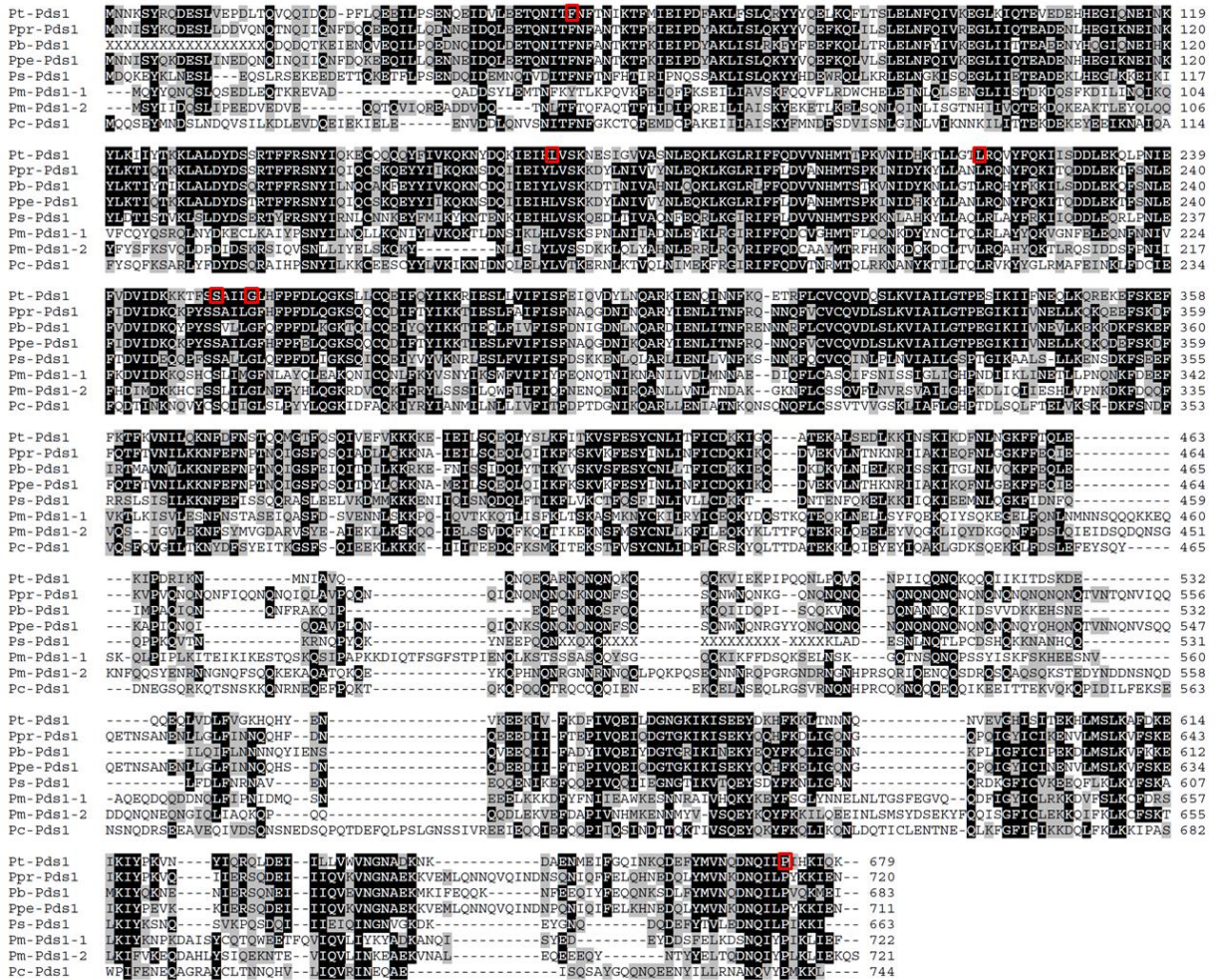


Figure S4: Pds1 homologs identified in the MAC genome of *Paramecium* species. Alignments of complete protein sequences were done using the MUSCLE software. Amino acids for which missense alleles were obtained by the screen for RNAi-deficient mutants are highlighted in red. Ppr, *P. primaurelia*; Pb, *P. biaurelia*; Pt, *P. tetraurelia*; Ppe, *P. pentataurelia*; Ps, *P. sexaurelia*; Pm, *P. multimicronucleatum*, Pc, *P. caudatum*.

- Figure S5 -

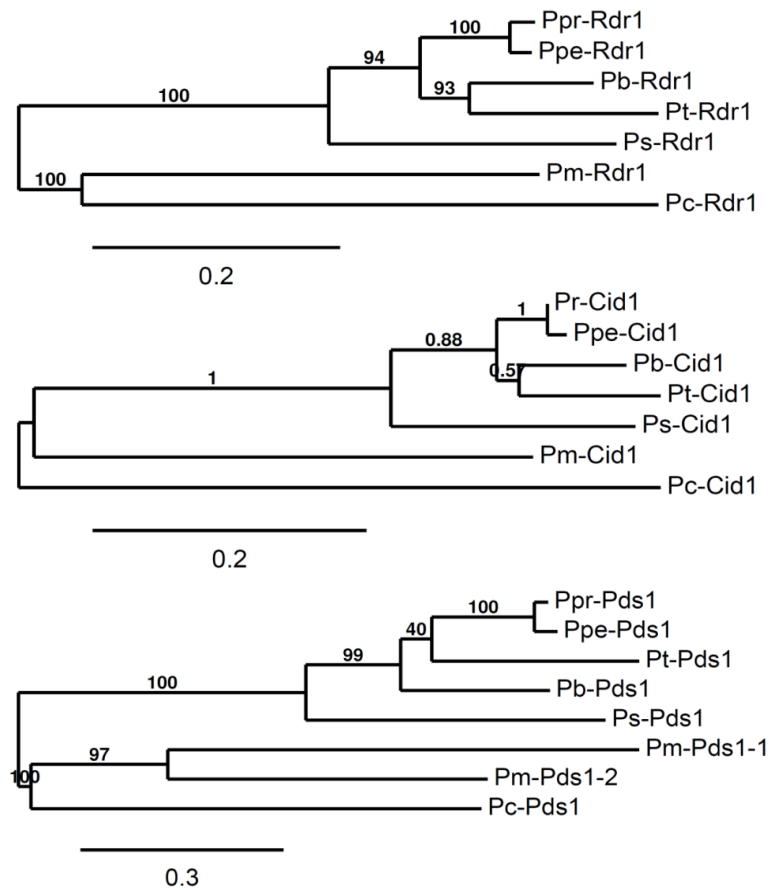


Figure S5: Conservation of RNAi factors in *Paramecium* species. Generation of alignments and phylogenetic trees see Fig. S2. Complete protein sequences were used to create the alignments. Ppr – *P. primaurelia*, Pb – *P. biaurelia*, Pt – *P. tetraurelia*, Ppe – *P. pentaurelia*, Ps – *P. sexaurelia*, Pm – *P. multimicronucleatum*, Pc – *P. caudatum*.

– Figure S6 –

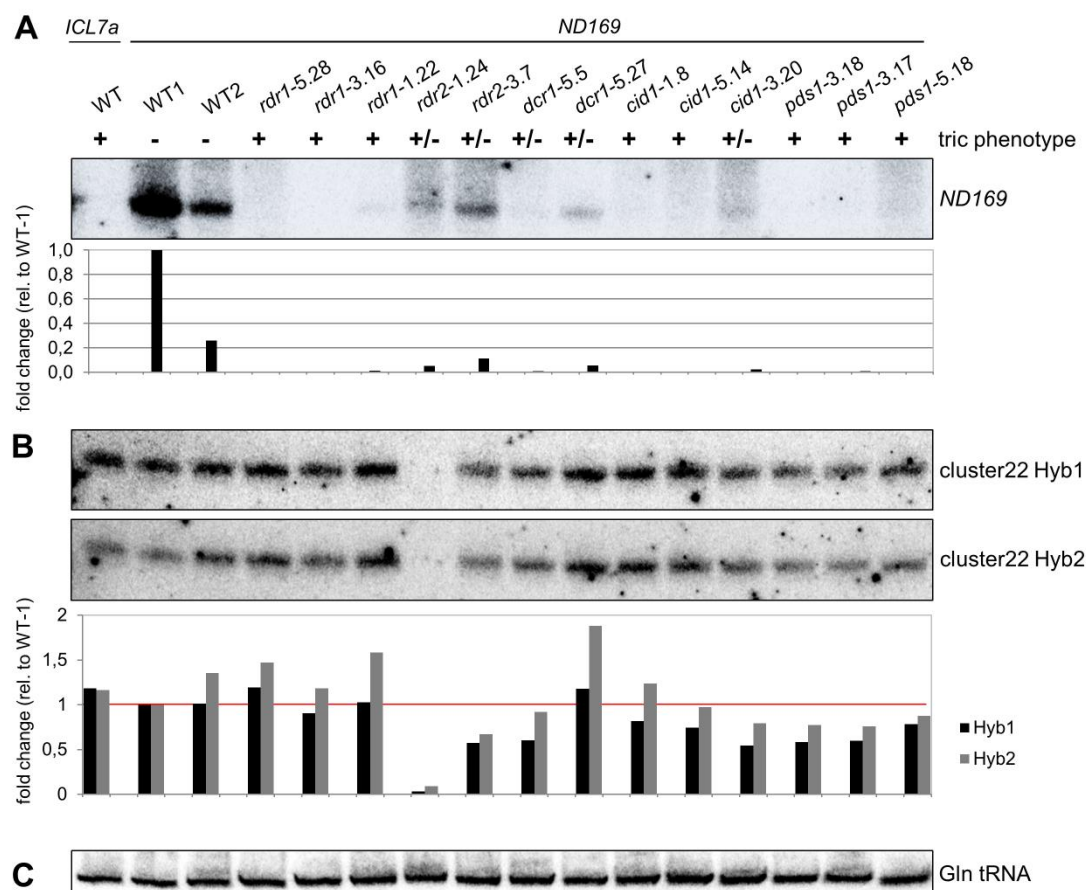


Figure S6: Quantification of small RNA northern blots shown in Fig. 4A and 4B (main article). (A) Detection of dsRNA *feeding*-associated siRNAs (*ND169*) in wild type cells and RNAi mutants. siRNA quantities are relative to the level in wild type-1. (B) The hybridisation with the cluster 22 probe (top strand) shown in Fig. 4 was repeated to compensate for quantification artefacts due to non-specific background signals (spots). siRNA quantities are relative to the level in wild type-1 (red line). (C) Glutamine tRNA levels were used as normalisation reference.

– Figure S7 –

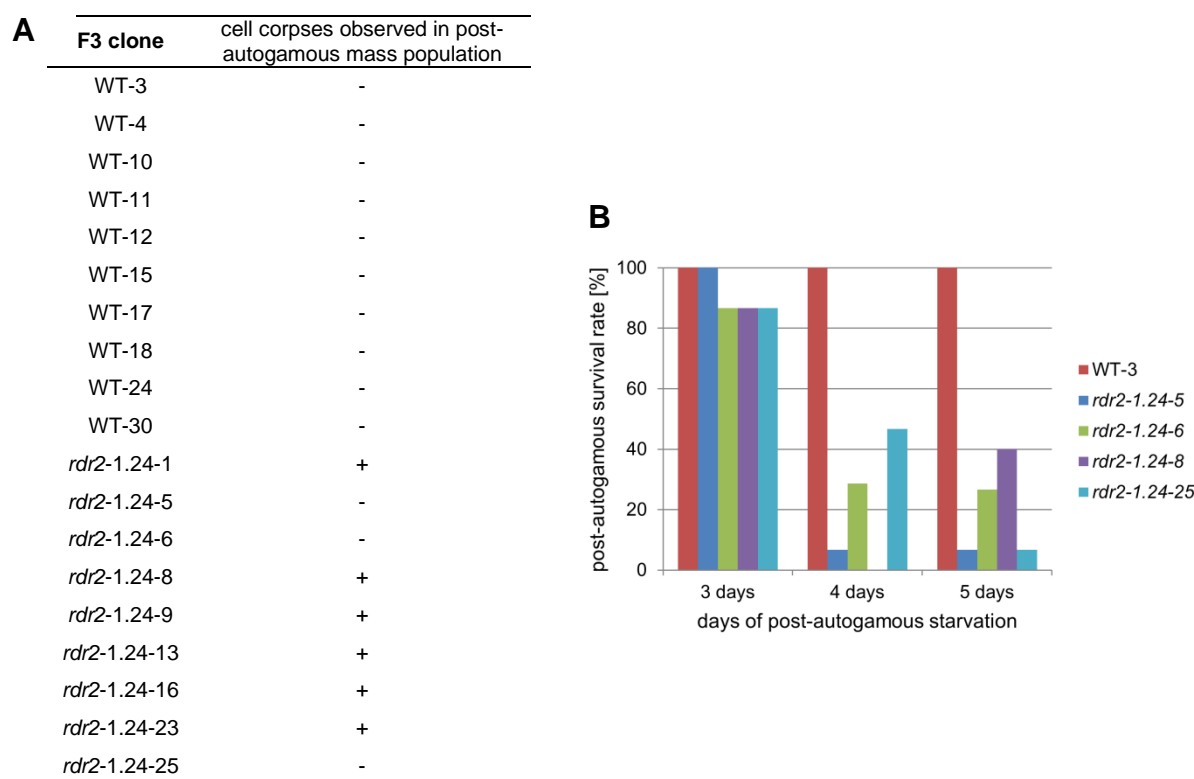
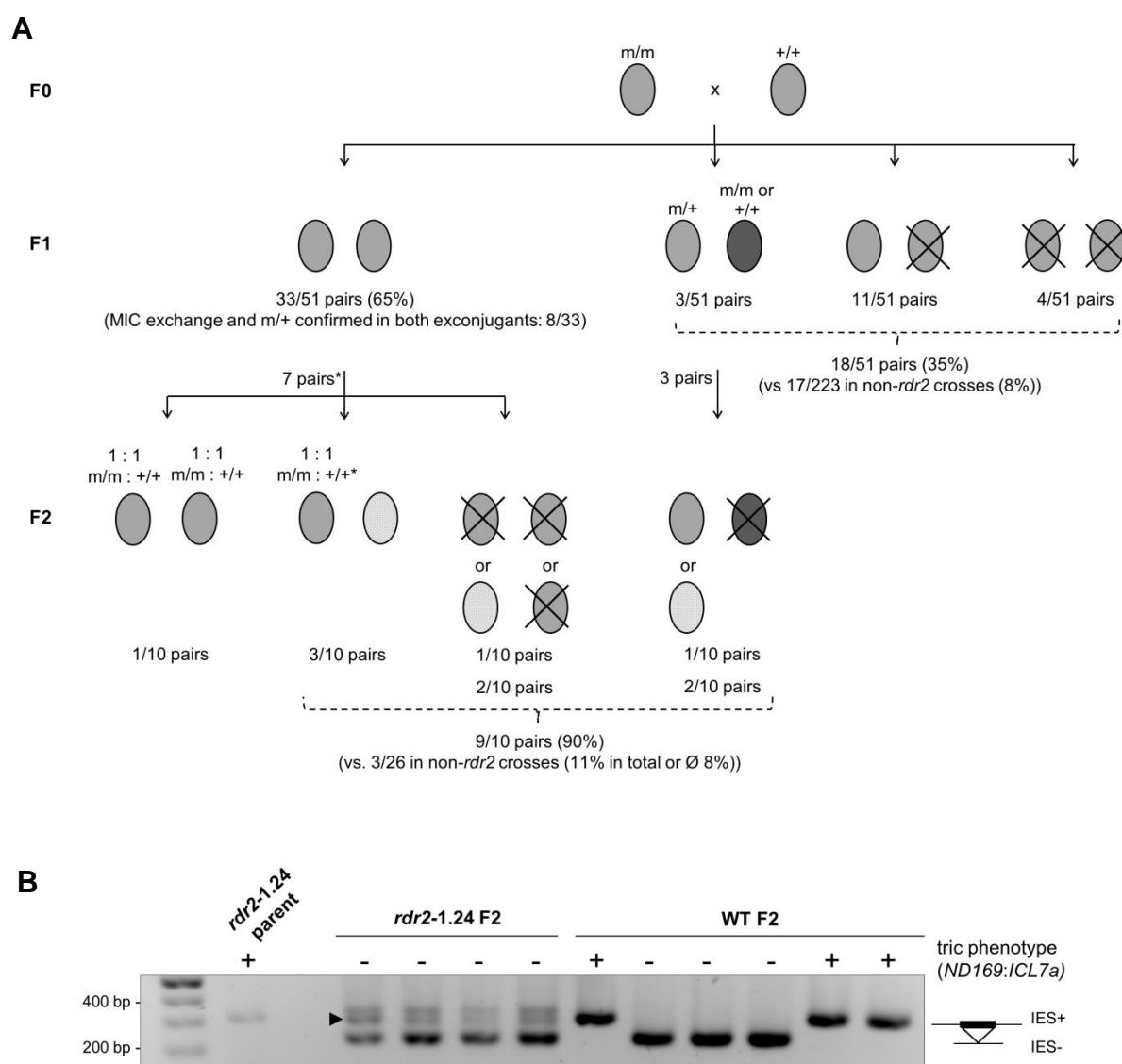


Figure S7: The *rdr2* mutant shows reduced post-autogamous survival rates after prolonged starvation. Post-autogamous survival rates of selected F3 clones descending from a cross of *rdr2*-1.24 (survival of F2 clones occurred only on the wild type side of the cross (29/30)). After ~25 vegetative divisions individual F3 clones were grown to starvation (~1000 cells) to induce autogamy. **(A)** After 5 days of starvation, occurrence of dead cells on the bottom of the depression slides was monitored. Lethality co-segregates with the *rdr2*-1.24 mutation, but is not completely penetrant, probably due to its hypomorphic nature. **(B)** Surviving post-autogamous cells of selected clones from experiment (A) were grown for ~25 vegetative divisions until starvation to induce another autogamy. 15 individual post-autogamous cells were isolated in fresh *K. pneumoniae* medium after three, four and five days of starvation and monitored for survival/normal growth. Note that clones *rdr2*-1.24-5, -6 and -25 did not show post-autogamous lethality in the previous generation. The partial post-autogamous lethality after prolonged starvation was also observed for non-crossed *rdr2*-1.24 mutants (data not shown). Furthermore, young post-autogamous cells (i.e. not able to fall autogamous again) isolated after short starvation, and thus being able to grow vegetatively, did not show lethality when re-isolated after a second starvation of 5 days. Thus, we conclude that starvation does not reduce viability of vegetative *rdr2* mutants, but that lethality after prolonged starvation is specific for cells that had undergone autogamy.

– Figure S8 –

**Figure S8: Summary of phenotypic effects observed in *rdr2* mutants during conjugation.**

(Absolute numbers of analysed pairs and phenotypes, and segregation of mutations see Table S4A). Conjugation between two cells of different mating types implicates meiosis of the MICs, followed by reciprocal exchange of one of two gametic nuclei between conjugants. Karyogamy with the stationary haploid MIC results in heterozygous MICs and MACs in each partner in F1. However, exchange of gametic nuclei quite frequently fails in the wild type; this results in each conjugant undergoing autogamy, which does not change a homozygous genotype.

(A) Phenotypes of progeny of conjugating *rdr2* × WT pairs are shown. Fully viable cells are represented in grey, whereas lethality is indicated by crossed grey cells. Exconjugants were considered as viable if at least one karyonide (cell arisen from first cell division after development of two new MACs) grew normally. All F1 pairs are listed (rate of MIC exchange for each cross see Table S4A). F1 cells showed lethality or anomalies (see below) in 35% of the pairs, occurring on one or both exconjugants. In *rdr2* × WT pairs producing viable F1s, the rate of successful exchange appeared to be lower compared to most other crosses (22%

(8/36; counting also pairs with only one heterozygous exconjugant; tested molecularly by sequencing or PCR). In 3 pairs, only one F1 MAC was heterozygous while the other was homozygous (dark grey), suggesting unidirectional failure of exchange and development of one F1 MAC from a haploid nucleus. It should be noted that MAC homozygosity in one or both F1s can also be explained by successful fertilization followed by abortion of heterozygous MAC development, which can result in regeneration of the parental MAC. In one of these 3 cases, homozygosity was observed on the wild type cytoplasmic side, in one case on the mutant cytoplasmic side, and in the third case the cytoplasmic side could not be determined (details see Table S4A).

F2 clones arisen from autogamy of selected viable F1 pairs (*) that had exchanged the MICs showed significant lethality or sickness with reduced growth rate and abnormal cell shape (represented in dotted light grey) (9/10; $p=0.01$). Each F2 depicted as one cell represents the phenotype of 30 individual post-autogamous clones. The lethal or sick phenotype occurred on at least 14% and up to 100% of the clones descending from one exconjugant, whereas the mutation segregated normally when F2 clones were viable (tested in at least one pair where indicated). Results from both *rdr2*-alleles (1.24 and 3.7) are merged in the figure.

(B) Some F2 clones seemed to have regenerated the old MAC during autogamy, resulting in heterozygosity. In this particular pair (*rdr2*-1.24 x WT), 29/30 F2 clones descending from the wild type cytoplasmic side survived, showing 1:1 segregation of the mutation, i.e. retention of the *RDR2* IES and RNAi deficiency. The mutation in the IES boundary was verified by sequencing in some of these clones. Only 6 clones arisen from the mutant cytoplasmic side were viable, but still heterozygous like the F1 generation, suggesting MAC regeneration. Phenotypes of ten selected clones are shown in the figure. The reduced band intensity of the IES⁺ form in heterozygous clones is due to a PCR amplification bias favouring the IES⁻ form when both forms are present (not shown). The upper band above the IES⁺ form represents a IES⁺/IES⁻-heteroduplex due to saturation of the PCR.

– Supplementary tables –

Table S1: DsRNA fragments used for RNAi “by feeding” and genetic complementation constructs. For dsRNA *feeding* DNA fragments were cloned into the plasmid vector L4440. For genetic complementation experiments by microinjection gene sequences containing the endogenous promoter and 3’UTR were cloned into the plasmid vector pUC18.

Gene	accession number (ParameciumDB)	Orf size ATG-TGA [bp]	dsRNA fragment (position within orf)	Reference	complementation construct (rel. to ATG)
<i>NSF</i>	GSPATG00010950001	2334	19-2329	(28)	
<i>ND169</i>	GSPATG00008337001	1879	1450 - 1860	(22)	
<i>ICL7a</i>	GSPATG00021610001	582	1 - 580	(24)	
<i>PDS1</i>	PTETG600032001	2084	922 - 1454		-270 - +2561
<i>CID1</i>	PTETG9100013001	1051	34 - 553		-1309 - +1121
<i>CID2</i>	PTETG13400003001	1020	390 - 883		
<i>CID3</i>	GSPATG00025353001	1076	330 - 875		
<i>CID4</i>	PTETG600025001	1074	420 - 909		
<i>CID5</i>	GSPATG00002857001	1030	172 - 689		
<i>RDR1</i>	PTETG8500012001	4319	2670 - 3409	(22)	
<i>RDR2</i>	GSPATG00036857001	4107	1289 - 1559	(22)	
<i>RDR3</i>	GSPATG00006401001	3374	1789 - 2462	(22)	
<i>PTIWI13</i>	PTETG4800007001	2483	105 - 787	(25)	
<i>DCR1</i>	GSPATG00021751001	5394	3019 - 3987	(15)	

Table S2: Whole genome re-sequencing of pools of F2 mutant clones to identify mutations in RNAi factor genes. Pools of wild type F2 clones were sequenced for the first two mutants subjected to this method (*). For subsequent mutants the *rna1* mutation could be identified by comparing the data sets to the previous ones.

	RNAi phenotype	number of F2 clones for pool	sequencing read length	sequencing method	number of mapped reads	genome coverage	mean genome coverage	mutation coverage
<i>cid1-1.8</i>	<i>rna1-</i>	17	75 bp	single end	15x10 ⁶	96.0%	16-X	8-X
<i>CID1-1.8*</i>	<i>RNAi+</i>	20	75 bp	single end	4x10 ⁶	86.3%	4-X	2-X
<i>rdr1-3.1</i>	<i>rna1-</i>	23	75 bp	single end	24x10 ⁶	98.5%	26-X	16-X
<i>RDR1-3.1*</i>	<i>RNAi+</i>	22	75 bp	single end	35x10 ⁶	96.2%	37-X	28-X
<i>rdr1-5.2</i>	<i>rna1-</i>	24	100 bp	paired end	63x10 ⁶	95.0%	88-X	102-X
<i>pds1-3.18</i>	<i>rna1-</i>	23	50 bp	paired end	174x10 ⁶	95.9%	103-X	38-X

Table S3: Experimental setup of the screen for RNAi deficient mutants and observed mutation frequency for genes involved in the dsRNA-induced RNAi pathway. UV treatment was done on three replicates (*); each culture showed similar division rates and mutation frequencies. 20-25% post-autogamous lethality was observed in irradiated cultures, however, the final number of post-autogamous cells after 5 days of starvation was identical to the non-irradiated control, due to higher frequency of karyonidal divisions (†). Survivors of the screen were counted 36 hours after transfer of the cells from standard *Klebsiella* to *E. coli* HT115 *NSF*-dsRNA medium. For the 71 analysed mutants, the experimentally determined allele redundancy is 35% (25 redundant alleles, see Table S6).

experimental steps		UV+ *	UV- (control)
UV irradiation	initial number of cells subjected to irradiation	200,000	200,000
↓			
starvation, autogamy	number of cells after autogamy (~2 cell divisions)	900,000†	900,000
↓			
starvation, re-feeding	number of cells after re-feeding (~2 cell divisions), subjected to <i>NSF</i> -dsRNA screening	2,500,000	2,500,000
↓			
<i>NSF</i> dsRNA feeding	surviving cells in the mass culture after <i>NSF</i> -dsRNA screening (~7 cell divisions of fully RNAi deficient mutants)	108,000	3,300
↓			
determination of survivors / ml and isolation of 150 surviving cells	<i>NSF</i> survival rate relative to number of cells subjected to screening (considering 7 divisions, i.e. relative to 844 cells (vs 26 in control))	0.03%	0.001%
↓			
experimental verification: 79 viable RNAi mutants (53%)	experimentally verified frequency of viable RNAi mutants among isolated survivors of the <i>NSF</i> screen	53%	-
	surviving cells in the mass culture after <i>NSF</i> -dsRNA screening corrected by experimentally verified mutant frequency	57,240	-
	<i>NSF</i> survival rate relative to number of cells subjected to screening corrected by experimentally verified mutant frequency	0.016%	-
	estimated number of originally induced <i>mai</i> mutations (based on ~11 cell divisions since irradiation and 53%-corrected)	28	-
	hypothetical mutation frequency relative to originally irradiated cells	1/7142 (0.014%)	
	theoretical chance to isolate daughter cells	1/28 (3.6%)	

Table S4: Survival rate and phenotypes of post-conjugational progeny of *rnaI* mutants and segregation of *rnaI* phenotypes.

(A) Back crosses were carried out with wild type strain 51 or, in some cases, with the *nd7* mutant, carrying the *nd7-1* mutation (33) as a genetic marker for successful reciprocal MIC exchange in conjugating pairs. Numbers in potency indicate the number of back crosses performed with this particular allele. Generally, the rate of MIC exchange can vary strongly between different crossing experiments in *P. tetraurelia*.

For F1, pairs with and without MIC exchange are listed. At least six pairs were isolated for each cross experiment. A proportion of isolated F1 pairs of *rdr2* crosses showed lethality on one or both cytoplasmic sides of the cross. Survival is defined as at least one karyonide of the exconjugants (ex) giving rise to viable progeny. F1 RNAi phenotypes (not listed) were tested by dsRNA-induced *feeding* as described for F2 (see below). RNAi wild type phenotypes observed for both exconjugants indicated successful MIC exchange and complementation, due to recessive behavior of the mutated allele. Since the *rdr2-3.7* phenotype shows low penetrance, MIC exchange in *rdr2-3.7* x WT pairs was verified by sequencing of the macronuclear *RDR2* gene. In three of these pairs only one exconjugant MAC was heterozygous for the 3.7 mutation, whereas the other was homozygous wild type or mutant, as indicated by an asterisk (*) (see also F2 generation). Cytoplasmic parental origin was determined by mating type in F1 or F2, as this trait is inherited cytoplasmically during conjugation and autogamy.

At least one of the pairs that had exchanged the MICs was brought to F2 by autogamy. F2 clones showed lethality or slow growth phenotypes in a significant fraction of *rdr2* pairs (9/10; binomial distribution, $p = 0.01$). This was not observed in control pairs that had not exchanged MICs. Healthy F2 progeny is defined as $\geq 86\%$ survival rate and normal growth of post-autogamous clones (black) on both sides of the cross when isolated into fresh *K. pneumoniae* medium (deduced from the mean survival rate of $90.0\% \pm 2.89$ in wild type crosses by normal test using a confidence interval of 90%). At least 30 individual F2 clones were isolated for each pair. Pairs with reduced F2 survival rate and/or occurrence of sick cells below the 86% threshold are labelled in red. Segregation of *rnaI* mutations in F2 were tested by silencing of *NSF*, *ND7* or *ND169:ICL7a* dsRNA *feeding*, or by PCR to detect the *rdr2-1.24* (IES retention, see Fig. S7B) and the *rdr1-5.28* mutation. As the *RDR2*-IES is a non-maternally controlled IES, the 1.24 mutation segregates in a Mendelian manner. In most pairs, the Chi-square test confirmed the expected Mendelian segregation of mutant and wild type RNAi phenotypes (1:1 ratio in single mutants and 3:1 in double mutants, respectively). Deviation from the expected 1:1 ratio was only observed in favor of phenotypically wild type cells, which may be mis-scored mutants due to lethality independent of *NSF feeding*. Clones indicated with an asterisk (*) were homozygous on this cytoplasmic side in F1. Two asterisks (**) indicate heterozygosity in F2 after autogamy (regenerants, see Fig. S7B). In pairs for which both exconjugants were mating type 8 due to occasional mating type switch during conjugation, the cytoplasmic origin could not be determined (#). For some pairs, mating types and thus cytoplasmic origin was not determined (+). In the *rdr2-3.7* pair that showed viable F2 progeny (‘), the RNAi phenotype could not be determined unambiguously due to the strong hypomorphic nature of the allele. n.t., not tested.

(B) Double mutants were produced by crossing *rdr1* x *cid91* and *rdr1* x *rdr2*. The 5.28 mutation in *rdr1-rdr2* double mutants was only tested (by sequencing) in 17 clones and only on the *rdr1* mutant side (†). In *rdr1-cid91* double mutants, a phenotypic test was not performed. The 1.8 mutation was tested by MnlI digest and the 3.1 mutation by TaqI digest, as these sites are mutated in the respective mutant; only 15 and 8 clones were analysed in the two pairs tested, respectively. Note that no significant lethality was observed in the F2 generation of the *rdr1-rdr2* double mutant.

A

cross	F1 survival			F1 MIC exchange	F1 total MIC exchange	pairs with healthy F2 progeny	F2 segregation							total RNAi+ : <i>mai</i> -	χ^2 p-value
	both ex.	one ex.	no ex.				phenotype test	RNAi+ descendant			<i>mai</i> - descendant				
								survival	RNAi+	<i>mai</i> -	survival	RNAi+	<i>mai</i> -		
Wild type					12/20 (60%)	3/3									
WT x WT	2/3	1/3	-	-		1/1	-	26/30	-	-	27/30	-	-	-	-
WT x <i>nd7</i>	11/12	1/12	-	4/11		1/1	-	27/30	-	-	26/30	-	-	-	-
	9/9	-	-	8/9		1/1	-	29/30	-	-	27/30	-	-	-	-
<i>cid91</i>					30/88 (34%)	9/9									
<i>cid91</i> -1.8 x WT	6/6	-	-	3/6		1/1	<i>NSF</i>	29/30	18	11	28/30	17	11	35 : 22	>0.05
<i>cid91</i> -1.8 ¹ x WT	10/10	-	-	3/10		1/1	<i>ND7</i>	29/30	10	19	28/30	14	14	24 : 33	>0.05
<i>cid91</i> -1.8 ² x WT	32/36	2/36	2/36	12/32		1/1	<i>NSF</i>	29/30	12	17	n.t.	n.t.	n.t.	12 : 17	>0.05
						1/1	<i>NSF</i>	30/30	10	20	n.t.	n.t.	n.t.	10 : 20	>0.05
<i>cid91</i> -1.8 ³ x <i>nd7</i>	8/8	-	-	2/8		1/1	<i>NSF</i>	29/30	12	17	29/30	18	11	30 : 28	>0.05
<i>cid91</i> -1.23 x WT	6/6	-	-	3/6		1/1	<i>NSF</i>	29/30	19	10	30/30	17	13	36 : 26	>0.05
<i>cid91</i> -5.14 x WT	6/6	-	-	2/6		1/1	<i>NSF</i>	30/30	17	13	27/27	7	20	24 : 33	>0.05
<i>cid91</i> -3.20 x WT	5/6	1/6	-	3/5		1/1	<i>NSF</i>	27/30	16	11	28/30	16	12	32 : 23	>0.05
<i>cid91</i> -3.4 x WT	5/6	-	1/6	2/5		1/1	<i>NSF</i>	29/30	26	3	23/30	18	5	44 : 8	6e-7
<i>rdr1</i>					26/57 (46%)	6/6									
<i>rdr1</i> -3.1 x WT	6/6	-	-	2/6		1/1	<i>NSF</i>	30/30	14	16	28/30	12	16	26 : 32	>0.05
<i>rdr1</i> -3.1 ¹ x WT	7/7	-	-	5/7		1/1	<i>NSF</i>	30/30	19	11	30/30	16	14	35 : 25	>0.05
<i>rdr1</i> -3.1 ² x <i>nd7</i>	7/8	1/8	-	1/7		1/1	<i>NSF</i>	29/30	16	13	29/30	21	8	37 : 21	0.04
<i>rdr1</i> -1.22 x WT	6/6	-	-	3/6		1/1	<i>NSF</i>	29/30	12	17	29/30	19	10	31 : 27	>0.05
<i>rdr1</i> -5.2 x WT	22/24	2/24	-	11/22		1/1	<i>NSF</i>	48/60	33	15	59/60	34	25	67 : 40	0.009
<i>rdr1</i> -5.28 x WT	9/9	-	-	4/9		1/1	<i>NSF</i>	29/30	13	16	29/30	16	13	29 : 29	1
<i>pds1</i>					24/35 (69%)	5/6									
<i>pds1</i> -3.18 x WT	6/6	-	-	6/6		1/1	<i>NSF</i>	29/30	12	17	30/30	15	15	27 : 32	>0.05
<i>pds1</i> -3.18 ¹ x WT	6/7	1/7	-	4/6		1/1	<i>NSF</i>	29/30	18	11	30/30	15	15	33 : 26	>0.05
<i>pds1</i> -3.13 x WT	3/4	1/4	-	2/3		1/1	<i>NSF</i>	30/30	14	16	29/30	19	10	33 : 26	>0.05
<i>pds1</i> -3.17 x WT	8/8	-	-	4/8		0/1	<i>NSF</i>	8/30	5	3	17/30	11	6	16 : 9	>0.05
<i>pds1</i> -5.4 x <i>nd7</i>	8/9	1/9	-	6/8		1/1	<i>NSF</i>	30/30#	19	11	28/30#	19	9	38 : 20	0.02
<i>pds1</i> -5.17 x WT	4/6	1/6	1/6	2/4		1/1	<i>NSF</i>	28/30	14	14	30/30	15	15	29 : 29	1
<i>dcr1</i>					2/14 (14%)	2/2									
<i>dcr1</i> -5.5 x WT	14/15	1/15	-	2/14		1/1	<i>ND169:ICL7a</i>	30/30	18	11	29/30	14	16	32 : 27	>0.05
						1/1		30/30	n.t.	n.t.	30/30	n.t.	n.t.		
<i>rdr2</i>					8/36 (22%)	1/10									
<i>rdr2</i> -3.7 x WT exp.1	9/9	-	-	2/9		0/1	-	4/30	-	-	0/30	-	-	-	-
						0/1	-	0/30	-	-	5/30	-	-	-	-
<i>rdr2</i> -3.7 x WT exp.2	8/9	-	1/9	1*/8		0/1	-	0/30* #	-	-	2/30#	-	-	-	-
						1/1 (control)	-	30/30	-	-	30/30	-	-	-	-

<i>rdr2-3.7</i> x WT exp.3	12/18	4/18	2/18	2*/12; 4/12	0/1	-	0/30*	-	-	26/30	n.t.	n.t.	n.t.	n.t.
					0/1		0/30	-	-	10/30*	-	-	-	-
					0/1	<i>ND169:ICL7a</i>	28/30	(9)'	(12)'	22/30	(9)'	(8)'	(18 : 20)'	(>0.05)'
<i>rdr2-1.24</i> x WT	7/15	7/15	1/15	2/7	1/1	<i>RDR2-IES</i>	29/30+	n.t.	n.t.	30/30+	n.t.	n.t.	n.t.	n.t.
					0/1	<i>ND169:ICL7a</i>	29/30	18	11	6/30**	6	0	-	>0.05
					0/1		26/30	n.t.	n.t.	9/30	-	-	18 : 11	>0.05
					0/1		0/30	-	-	1/30	-	-	-	-
					1/1 (control)		29/30	-	-	27/30	-	-	-	-
					1/1 (control)		30/30	-	-	29/30	-	-	-	-

B

cross	F1 survival			F1 MIC exchange	F1 total MIC exchange	pairs with healthy F2 progeny	F2 segregation									
	both ex.	one ex.	no ex.				phenotype test	survival	RNAi+	<i>mai</i> -	survival	RNAi+	<i>mai</i> -	total RNAi+ : <i>mai</i> -	χ^2 p-value	
<i>cid91</i> x <i>rdr1</i>					6/7 (86%)											
<i>cid91-1.8</i> ³ x <i>rdr1-3.1</i> ²	7/9	1/9	1/9	6/7		1/1		<i>cid91-1.8</i> ³ descendant			<i>rdr1-3.1</i> ² descendant					
								27/30+			28/30+					
								MnII digest	2/7	5/7		5/8	3/8	-	-	
								TaqI digest	4/7	3/7		2/8	6/8	-	-	
								double mut.	-	2/7		-	1/8	-	-	
						0/1			11/30+		22/30+					
								MnII digest	3/4	1/4		2/4	2/4	-	-	
								TaqI digest	3/4	1/4		1/4	3/4	-	-	
								double mut.	-	1/4		-	1/4	-	-	
						1/1			30/30+	n.t.	n.t.	30/30+	n.t.	n.t.	-	-
<i>rdr1</i> x <i>rdr2</i>					1/10 (10%)											
<i>rdr1-5.28</i> ¹ x <i>rdr2-1.24</i>	10/15	2/15	3/15	1/10		1/1	<i>ND169:ICL7a</i>	28/30	9	21	28/30	7	22	16 : 43	>0.05	
							<i>rdr2-IES</i>		13	15		12	15	-	-	
							<i>rdr1-5.28</i> †		7/17	10/17		n.t.	n.t.	-	-	
							double mut.		-	7/17		n.t.	n.t.	-	-	

Table S5: Extrapolated number of UV-induced SNPs per genome. (A) Sequencing reads of each mutant were aligned on the *P. tetraurelia* (strain 51) reference MAC genome using the BWA software (38) allowing 2 mismatches per read. A pipeline was developed to estimate the number of single nucleotide polymorphisms after UV mutagenesis. A SNP in the genome was defined as UV-induced mutation (and not responsible for the RNAi deficient phenotype) if it matched three criteria: i) The SNP position is covered by at least 30 (1.8 mutant) or 50 (other mutants) reads. ii) More than 20% and less than 80% of reads covering the position carry the mutation. iii) The position is not defined as UV-induced mutation in another mutant. Because the minimum coverage of 30-X or 50-X restricted the analysis only to a fraction of the genome, the total number of mutations was deduced by linear extrapolation to the whole genome. Wild type and mutant pool data sets were pooled where available (*). Note that due to technical reasons only a small number of sequence reads of high quality was obtained for mutant 1.8, leading to insufficient coverage and therefore inaccurate estimation of the total number of mutations (†). However, the SNP responsible for the *rmai* mutation could readily be identified (see Table S2). (B) Frequency of allelic UV-induced mutations (excluding the silent substitutions 3.15-2 (A to T) and 1.27-2, a substitution downstream of the 1.27-1 frameshift mutation). Substitutions refer to the gene coding sequence. Non-SNP mutations are either multiple adjacent substitutions or insertions/deletions, or both (*) (see Table S6).

A

	number of F2 clones for pool and their phenotype	genome coverage	mutations identified when covered by 20-80% of the reads	total number of mutations in pool after extension to complete genome	number of back crosses	number of mutations before back cross
<i>cid1-1.8; CID1-1.8*</i>	17 <i>rmai</i> -; 20 <i>RNAi</i> +*	15.9 % [†]	8	51	2	102 [†]
<i>rdr1-3.1; RDR1-3.1*</i>	23 <i>rmai</i> -; 22 <i>RNAi</i> +*	60.1 %	8	19	2	38
<i>rdr1-5.2</i>	24 <i>rmai</i> -	95.3 %	25	26	1	26
<i>pds1-3.18</i>	23 <i>rmai</i> -	68.0 %	13	13	2	26

B

	Transitions	Transversions	Non-SNP mutations*
C → T	7	C → A	0
G → A	17	G → T	3
T → C	4	A → C	2
A → G	0	T → G	0
		G → C	0
		C → G	1
		A → T	4
		T → A	4

Table S6: Alleles obtained from the screen for mutants deficient in dsRNA-induced RNAi. Nucleotide bases in capital letters are coding, introns or IESs are represented in lower case. Insertions or deletions of amino acids are abbreviated as in and del, respectively. RNAi phenotypes are recorded as fully (-) or partially (+/-) RNAi deficient, according to similar observations in dsRNA-induced silencing against the genes *NSF*, *ND169* and *ND7*.

Allele	Mutation	Intron/IES	Mutant with same mutation	AA substitution	RNAi phenotype
<i>rdr1-1.1</i>	GGT → GAT		1.10, 1.21	G772D	-
<i>rdr1-1.3</i>	CGA → TGA			R975stop	-
<i>rdr1-1.4</i>	GAT → AAT			D1021N	-
<i>rdr1-1.9</i>	TAG AGA AAA ACA TAA TGG GAT TAA TTG TAC G → TAC G		1.16	del445-454 QRKTQWDQL	-
<i>rdr1-1.11</i>	GGA → GAA		1.18, 5.15	G687E	-
<i>rdr1-1.13</i>	GGA → AGA			G687R	-
<i>rdr1-1.14</i>	TGG → TGA			W657stop	-
<i>rdr1-1.15</i>	GGC → AGC			G1089S	-
<i>rdr1-1.17</i>	CTT → CCT			L55P	-
<i>rdr1-1.22</i>	Agta...tagA → Agtt...tagA	Int5 (3n+1)		in(1277)(stop1314)	-(+/-)
<i>rdr1-1.25</i>	AGT → CGT			S1360R	-
<i>rdr1-1.26</i>	GAT → GAA			D1123E	-
<i>rdr1-1.29</i>	TTG GAA → TTC AAA			del(20) (stop 43)	-
<i>rdr1-3.1</i>	CGA → TGA		3.3, 3.9	R1361stop	-
<i>rdr1-3.2</i>	AAT → AAA		3.6	N822K	-
<i>rdr1-3.5</i>	TAT → AAT		3.8	Y1025N	-
<i>rdr1-3.10</i>	GAC → AAC			D1019N	-
<i>rdr1-3.11</i>	AGA → AGC			R703S	-
<i>rdr1-3.12</i>	CTC → CCC			L1234P	-
<i>rdr1-3.15-1</i>	Agta...tagA → Tata...tagA	Int5 (3n+1)		in(1277)(stop1314)	-
<i>rdr1-3.16</i>	GAT → TAT			D1021Y	-
<i>rdr1-5.2</i>	ATG → ATA		5.16, 1.2	M1I	-
<i>rdr1-5.7</i>	Ggta...tagA → Ggta...tatA	Int4 (3n+2)	5.25	in(941)(stop953)	-
<i>rdr1-5.8</i>	Ggta...tagA → Ggta...taaA	Int4 (3n+2)		in(941)(stop953)	-
<i>rdr1-5.11</i>	GGT → TGT		5.22, 5.24, 5.33	G772C	-
<i>rdr1-5.28</i>	TCA TTG → A AT TGA			del(374)/stop	-
<i>rdr2-1.24</i>	Tta...tgT A → Taa...tgT A	IES (3n+2)		1276 (stop1330)	+/-
<i>rdr2-3.7</i>	GGA → GAA			E859K	+/-

<i>dcr1-5.5</i>	GAA → AAA		E87K	+/-
<i>dcr1-5.27</i>	CAT → GAT		H622D	+/-
<i>cid1-1.6</i>	GAT → AAT	1.12	D70N	-
<i>cid1-1.8</i>	TCC → TTC	1.23, 5.3, 5.12	S174F	-
<i>cid1-3.4</i>	GAT → AAT		D68N	-
<i>cid1-3.19</i>	TGA → TAA		23aa extension	-
<i>cid1-3.20</i>	Tgta...aagG → Tgta...aaaG	Int2 (3n+1)	in(232)(stop253)	- (+/-)
<i>cid1-5.14</i>	AAT → TAT	5.13, 5.19, 5.29	N131Y	-
<i>cid1-5.21</i>	GGA → GAA	5.30, 5.31	G155E	-
<i>pds1-1.5</i>	CCT → TCT		P673S	-
<i>pds1-1.27-1</i>	GTA AAG AAA AAA AAG → GTA ATA AAA AAA AAA G		in(391)(stop 426)	-
<i>pds1-3.13</i>	TCT → TTT		S252F	-
<i>pds1-3.17</i>	CGA → TGA		R468stop	-
<i>pds1-3.18</i>	TTA → TTT		L218F	-
<i>pds1-5.4</i>	TTC → TCC		F52S	+/-
<i>pds1-5.9</i>	AAT ATG → A TA TGA		del(463)/stop	-
<i>pds1-5.17</i>	TTG → TCG		L170S	-
<i>pds1-5.18</i>	GGA → GAA		G256E	-
<i>pds1-5.23</i>	AAA → A AC		del(348) (stop 406)	-

Table S7: Cid1-like non-canonical poly(A)/(U) RNA polymerases. Conserved regions were used to create the alignment and phylogenetic tree in Fig. S2. **(A)** 22 Cid1-like genes were identified in the *P. tetraurelia* MAC genome. **(B)** Selected Cid1-like proteins of other organisms.

A

<i>Gene</i>	accession (ParameciumDB)	scaffold (ParameciumDB)	Pfam A domains (low cut off)	conserved region used in alignments	ohnolog	WGD
<i>CID1</i>	PTETG9100013001	91	NTP_transf_2	55-304	-	-
<i>CID2</i>	PTETG13400003001	134	NTP_transf_2	61-323	-	-
<i>CID3</i>	GSPATG00025353001	88	NTP_transf_2	57-301	-	-
<i>CID4</i>	PTETG600025001	6	NTP_transf_2	50-285	-	-
<i>CID5</i>	GSPATG00002857001	6	NTP_transf_2	48-282	-	-
<i>CID6a</i>	GSPATG00022434001	74	TACC; YscO; TolA; Vicilin_N; dsDNA_bind; NTP_transf_2	522-738	GSPATG00036288001	1
<i>CID6b</i>	GSPATG00036288001	159	CobT; SSDP; Vicilin_N; Drf_GBD; NTP_transf_2	518-734	GSPATG00022434001	1
<i>CID7a</i>	PTETG1500013001	15	DUF612; Borrelia_orfA; TFIIF_alpha; NTP_transf_2	638-869	PTETG2500017001	1
					PTETG3000015001	2
					PTETG3100017001	2
<i>CID7b</i>	PTETG2500017001	25	TFIIF_alpha; DUF1754; NTP_transf_2; PAP_assoc	641-872	PTETG1500013001	1
					PTETG3000015001	2
					PTETG3100017001	2
<i>CID8a</i>	PTETG3000015001	30	NTP_transf_2; PAP_assoc	314-545	PTETG3100017001	1
					PTETG1500013001	2
					PTETG2500017001	2
<i>CID8b</i>	PTETG3100017001	31	TFIIF_alpha; DUF1754; TPR_MLP1_2; NTP_transf_2; PAP_assoc	640-871	PTETG3000015001	1
					PTETG1500013001	2
					PTETG2500017001	2
<i>CID9a</i>	PTETG1000003001	10	NTP_transf_2	108-304	PTETG5900002001	1
<i>CID9b</i>	PTETG5900002001	59	NTP_transf_2; PAP_assoc	108-304	PTETG1000003001	1
<i>CID10a</i>	GSPATG00016065001	47	NTP_transf_2; PAP_assoc	108-305	GSPATG00024416001	1
<i>CID10b</i>	GSPATG00024416001	83	NTP_transf_2; PAP_assoc	108-305	GSPATG00016065001	1
<i>CID11</i>	GSPATG00018140001	55	NTP_transf_2	107-307	-	-
<i>CID12a</i>	GSPATG00018141001	55	NTP_transf_2; PAP_assoc	300-506	GSPATG00029019001	1

CID12b	GSPATG00029019001	108	NTP_transf_2; PAP_assoc	309-515	GSPATG00018141001	1
CID13a	PTETG14000019001	140	Ank; Ank; NTP_transf_2	366-572	PTETG13000008001	1
					GSPATG00036244001	2
					GSPATG00025039001	2
CID13b	PTETG13000008001	130	Ank; Ank; Ank; NTP_transf_2	365-571	PTETG14000019001	1
					GSPATG00036244001	2
					GSPATG00025039001	2
CID14a	GSPATG00036244001	158	Ank; Ank; Ank; NTP_transf_2; tRNA-synt_2c	362-567	GSPATG00025039001	1
					PTETG13000008001	2
					PTETG14000019001	2
CID14b	GSPATG00025039001	86	Ank; Ank; Ank; NTP_transf_2; Ribosomal_S8	363-568	GSPATG00036244001	1
					PTETG13000008001	2
					PTETG14000019001	2

B

Protein	organism	accession (NCBI/GenBank)	conserved region used in alignments
Ce-Rde3	<i>C. elegans</i>	NP_491834	92-359
Ce-Gld2	<i>C. elegans</i>	NP_001021845	287-570
Ce-Cde1	<i>C. elegans</i>	NP_498099	1054-1291
Sp-Cid1	<i>S. pombe</i>	NP_594901	88-331
Sp-Cid11	<i>S. pombe</i>	NP_595210	93-314
Sp-Cid12	<i>S. pombe</i>	NP_588273	64-258
Sp-Cid13	<i>S. pombe</i>	NP_593157	97-326
Sp-Cid14	<i>S. pombe</i>	NP_001018181	285-487
Sp-Cid16	<i>S. pombe</i>	NP_593571	927-1156
Tb-ncPap1	<i>T. brucei</i>	ACN51269.1	368-587
Tb-ncPap2	<i>T. brucei</i>	ACN51270.1	158-391
Hs-PapD5	<i>H. sapiens</i>	AGE92663.1	243-442
Hs-PapD7 I	<i>H. sapiens</i>	AGE92663.1	264-463
Hs-Tut4	<i>H. sapiens</i>	NP_001009881	996-1232
Sc-Trf4	<i>S. cerevisiae</i>	CAA99134.1	223-426
Sc-Trf5	<i>S. cerevisiae</i>	NP_014100.2	220-423

At-Heso1	<i>A. thaliana</i>	NP_181504	53-301
Cr-Mut68	<i>C. reinhardtii</i>	XP_001695423	56-337
Tt-Rdn1	<i>T. thermophila</i>	ABS32300	239-502
Tt-Rdn2	<i>T. thermophila</i>	ABS32301	254-511

– Supplementary references –

78. Dereeper,A., Guignon,V., Blanc,G., Audic,S., Buffet,S., Chevenet,F., Dufayard,J.-F., Guindon,S., Lefort,V., Lescot,M. *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.*, **36**, W465-W469.
79. Arnaiz,O., Gout,J.F., Betermier,M., Bouhouche,K., Cohen,J., Duret,L., Kapusta,A., Meyer,E. and Sperling,L. (2010) Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics*, **11**, 547.
80. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207-210.

II. Annexe 2: Autres documents de la partie I des résultats

S4-51 1 TTATGAAAAAATTCAAATGAA-----TATAAATTTACTTCTACTATATCAAAACGCTGAGTTCAGTTTTGAATAAA---
S2-Ca1 1 -----AGTAATAAAAACATAAATTTTACTTCTACTTAAATATATAATGAGATAAATAGTATTAGATAAAT
S6-AZ84 1 ATATCTTTATAAAAAATAAATGTTGAA-----ATTACTAAAATATATGATAAAATTAAGAAATATTAAAGTTTG
S1-Ir42 1 ATC-----TTAGTAATGAAAGCTATAAGTTTACTTCTACTTGATTTATAAATGAATTAATAGTATTAAATAAAT
Pcau 1 CATAGTTGTTGTTTCAGTAATGAAATGTACAACATCTCTCATTAAATGTTTAAACAAATTTGATAAATTAAGAAATC
Pmul 1 -----GTTGTTTAGTAATGAAATCATACAATAATAATGTTAATAATTTTAAATGAATTAATAGTATTAGTAAAT
S4-51oh 1 -----GTTATAAGTTTACGATTTCTAAAATATACAAAAGAAATTAATAGCATTAAAGAAAAT
S2-Ca1oh 1 -----GTTATAAGTTTACGATTTCTAAAATATACAAAAGAAATTAATAGCATTAAAGAAAAT
S6-AZ84oh 1 -----TATAAGTTTACTTTACTAAAATATACAAAAGAAATTAATAGTATTAAAGAAAAT
S1-Ir42oh 1 -----TATAAATACTTTACTAAAATATACAAAAGAAATTAATAGTATTAAAGAAAAT

S4-51 73 -----AAATAATATTCGTTTGATTTTGAGTAGCAATCGAAATCAAAATTTGTTATGTTGATTGTAATAA
S2-Ca1 76 AGGGTTTTACAAATAATTTTTAAAGATGAGGTCATAGAAAATA-----ATTTGTTTTAAATTAATGAGAAT
S6-AZ84 70 GGATAGC-----AATATTTAGATGATAAATAAGAAATTTAGATATGATT-----TTAAATCAATAGAGAACT
S1-Ir42 72 GGGTATTTGGTAATATGAGAAATAATCCTTAGAGATGAATAATAATTGAT-----TTTAAATGATGAGAAAT
Pcau 81 AGCTTACCTTTAATATGAGAAATCATCATTAGAGGACATCAAGAAATAAAACATGGAAATGAAGCGCTATTGAAAAT
Pmul 74 AGCATATTAGTAGTATAAAAATAATTCATTTGAAATGAAATGAATTAAGATGAAATGCAATTAATGATTGAAAAT
S4-51oh 56 GGGATTTTGTAAATATGAAAAATAATCCTTTGAGGATGATGCAATTTGAATATGATTTGATTTAATGATAGAGAAAT
S2-Ca1oh 56 GGGATTTTGTAAATATGAAAAATAATCCTTTGAGGATGATTAATAATTTGAATATGATTTGAGTTAAATTAAGAGAAAT
S6-AZ84oh 54 GGGATTTTACAAATATGAAAAATAATCCTTTGAGGATGATTAACAATTTGAATATGACTATTTTAATTAAGAGAACT
S1-Ir42oh 54 GGGATTTTGTAAATATGAAAAATAATCCTTTGAGGATGATTAATAATTAATGAATTTAGTTAATGATAGAAAAT

S4-51 137 AGGGGAGTTT---GATAAAGGATATAATATGATTGTTATTATTTATTAATAT--ATT-GATCATAATCAAAATAAGG
S2-Ca1 135 ATGAGGATATG---GATGAAGAGAATAATATGATTGTTATTATTATTAATATATAT--GATCAAAATAAAGG
S6-AZ84 136 CGGAAGGGATT---TATAAAGAGAATAATATGATTGTAATTTATTTATTAATAT--ATT-GATGATAATCAAAACAAAGG
S1-Ir42 141 ATGGTATG---AAAGAGGAATAATATGATTGTTATTATTTATTAATATATA--TTT-GATCATAATCAAAATAAGG
Pcau 161 ATGGAGGATA---GATGATGTAATAATAATGATAATTTGCTCATTG-----TCAGATGAGA
Pmul 154 ATGGAGGATT---GAATAATGATTTTATGATTT-----TATT--TTGATAAATCAAAATGAGG
S4-51oh 136 ACGGTGGTATT---GATGAAGAGAATAATTTGATGTTATTATTATTATTAATATA--TTT-GATCATAATCAAAATAAGG
S2-Ca1oh 136 ATGGTGGTATT---GATGAAGAGAATAATGATGATTGTTATTATTATTATTGATATA--TTT-GATCATAATCAAAATAAGG
S6-AZ84oh 134 ATGGTGGTGTAAATGAAGAAGAGAATAAATATGATAGTTATTATTTT--TTAATATA--TTT-GATCATAATCAAAATGAGG
S1-Ir42oh 134 ATGGTGGTATT---GATGAAGAGAATAAATATGATTGTTATTATTATTATTAATATATATTTGATCATAATCAAAATAAGA

S4-51 211 TCTTGAAACCAAGTCTCTTTCTAAAACCTTATCACCCTTCTCGCTTGATAATGTCTATCATTCTGTATTAGATTTTCATC
S2-Ca1 211 TCTTGAAATCCAAGTCTCTTTCTAAAAGTATACACTTCTCGCTTGATAATGTACTATCATTCTATATAGATTTTCATC
S6-AZ84 211 TCTTGAAATCCAAGTCTTTGCTAAAGTTGATCACCCTTCTCGCAATGTGCTATCATTCTATATTAGATTTTCATC
S1-Ir42 211 TCTTGAAATCCAAGTCTCTAATAAGTTTACTTCTTCACTTGATAATGTACTATCATTCTGTATTAGATTTTCATC
Pcau 218 TCTAGAAAGCCAAGTCTCTACGAGGCAATCCTTCTACTAGAGAGAGACTGTCAATTTCTGTACTCAATTTCTCCTC
Pmul 211 TCTTGAAATCCAAGTCTCTTTCTTAAGAAATCACCCTTCTCGCTTGATAATACGGCTATCATTCTATATATATTTTCATC
S4-51oh 211 TCTTGAAATCCAAGTCTCTTTCTAAAGTTTAACTTCTTCTCGCTTGATAATGTCTATCATTCTGTATTAGATTTTCATC
S2-Ca1oh 211 TCTTGAAATCCAAGTCTTTTCTAAAGTTTATCACCCTTCTCGCTTGATAATATACATCATTCTATATTAATTTTCATC
S6-AZ84oh 211 TCTTGAAATCCAAGTCTCTATCTGAAGCTTAACTTCTTCACTTGATAATGTCTGTCAATTTCTATATTAGATTTTCATC
S1-Ir42oh 211 TCTTGAAATCCAATCTCTTTCTAAAGTTTAACTTCTTCACTTGATAATGTACTATCATTCTGTATTAGATTTTCATC

S4-51 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTGATTTTCTTCTCTATTTGTTGAGGAAATGTTGTC
S2-Ca1 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTATATTTCTTCTCTATTTAGGAAATGTTGTC
S6-AZ84 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTATTTTCTTCTCTATTTAAGGAAATGTTGTC
S1-Ir42 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTATATTTTCTTCTCTATTTGTTAAGGAAATGTTGTC
Pcau 298 GAAATCTAATATTA--GTATTA-----TACTTCTTCTGCTGTAATTTCTTCTACGATTGTTTAGGAAATGTTGTC
Pmul 291 GAAATCTATATATATTATATGG-----TACTTCTTCTTCTATATTTTCTTCTCTATTTAAGGAAATGTTGTC
S4-51oh 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTGACTTTCTTCTCTATTTGTTGAGGAAATGTTGTC
S2-Ca1oh 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTATATTTTCTTCTCTATTTGTTAAGGAAATGTTGTC
S6-AZ84oh 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTATATTTTCTTCTCTATTTAAGGAAATGTTGTC
S1-Ir42oh 291 AAAATCTATAGTAAA--TTCAAACCTTTTATTTCCTTCTTCTCTATATTTTCTTCTCTATTTGTTAAGGAAATGTTGTC

S4-51 369 CCAATCTGGATAACCTAATAAT--TCAAGATATTATGTTACTTGGCCAACATCAATATGGGTACTGTGAAAAATAGTGACA
S2-Ca1 369 CCAATCTGGATAACCTAATAACA--ATTCAACAATATATAATTACTTGGCAACATCAATATAGGGACTGTGAAAAATAGTGACA
S6-AZ84 369 CCAATCTAGGATAACCTTTAAAATTTGAGAGTTGA--ATTACTTGTAAAAGAGATATGGGTACTGTGAAAAATAGTGATA
S1-Ir42 369 CCAATCTGGATAACCTAATAAA--ATTCATAAAGATATATTACTTGGCAATATTAATATGAGGAAATGTGAAAAATAGTGACA
Pcau 369 CTAATGCTGGATAACCTAAAATAAATTAAGGTAT----ACTTGGCAAGAGATGATAGGAACAGTGAATATAGTGATA
Pmul 363 CTACTTCTGGATAACCTTATATAGTA----TATAAATTAATATATCAAAAATATATTTGGTGTACAAAAGATAGTAAACA
S4-51oh 370 CCAATCTGGATAACCTAGACA--ATCTAAGAAATCATGTACTTGGCAATATCAATGGGACTGTGAATATAGTGACA
S2-Ca1oh 371 CCAATCTGGATAACCTTTAAAATTTCAATTAATATATTACTTGGCAACATTAATATGGGACTGTGAAAAATAGTGATA
S6-AZ84oh 369 CCAATCTGGATAACCTTTAAAATCTCAAT--AATTAATTAATTTACTTGGCAAAAACAGTATGGGACTGTGAAGATAGTGACA
S1-Ir42oh 369 CCAATCTGGATAACCTATAC--AATTCATAAATTAATTTACTTGGCAACATAGTATGGGTACTGTGAAAAATAGTGACA

S4-51 447 TCGATGCCTAATCTAACATACCAAGGAATAGGCTCAT--TG--ATAATATTTAATAAATTAATGTAATAAATAATTTGTTT M Q N N N C L
S2-Ca1 448 TCAATGCCTAATCTAACATACCAAGGAATAGGCTCATTTG--ATAATATTTAATAAATTAATGTAATAAATAAATTTGTTT
S6-AZ84 448 TCAATGCCAGTCTAACAAACCAAGGTATAGGCTCAT--TTAATAATATTTAATAAATAATGTAATAAATAAATTTGTTT
S1-Ir42 448 TCAATGCCAATCTAACATACCAAGGAATAGGCTCATTTG--ATAATATTTAATAAATAATGTAATAAATAAATTTGTTT
Pcau 444 TCGATGCCTAATCTCACATACCAATGGAATGGGCTCATTAATTAATTAATAATAAATTAAGCT-----AATTTGCTT
Pmul 439 TCTATCTCTAATCTAACAAACCAAGGAAATGGTCTCATTTTAAATTTATTAATATAT--ATG-----ATTAATGTTT
S4-51oh 449 TCAATGCCTAATCTGACAACCATTGGGATGGTCTCAT--TTAATAATATTTAATAAATTAATGTAATAAATAAATTTGTTT
S2-Ca1oh 451 TCAATGCCAATCTAACATACCAATGGTATAGGCTCAT--TTAATAATATTTAATAAATAAATGTAATAAATAAATTTGTTT
S6-AZ84oh 448 TCAATCTCTAATCTAACAAACCAAGGATAGGCTCATTTAATAAATAAATAAATAAATAAATAAATAAATTTGTTT
S1-Ir42oh 448 TCAATGCCAATCTAACATACCAATGGGATGGTCTCAT--TTAATAATATTTAATAAATAAATGTAATAAATAAATTTGTTT

F C E K Q R I E G T I H Y P T L L Q T

S4-51 525 GTTTTGCGAAAAACAAGAATAGAAAGGACCATCCATTATCCCACTCTATTACAGACAGTACCAATCTAATTT-----AT
S2-CA1 527 GTTTTGCGAAAAATAAAGATTAGAAGGAGCAATCCATTATCCCTACTCTATTATAAACAGTATCAATCCAATCT-----AT
S6-AZ84 527 GTTTTGCGAAAAATAAAGATTAGAAGGAGGCTCCATCCATTATCCATCTCTTTTCAACTCAGTATAAATAAATTT-----AT
S1-Ir42 527 GTTTTGCGAAAAATAAAGATTAGAAGGAGGCTCCATTATCCCACTCTATTATAAACAGTATCAATCCAATCT-----AT
Pcau 516 CTTTTGTGAGAAGACAAAGACTAGAGGGAGCATACAATATCCATAGTTATTGCGAACTGTATCTATTATGTG-----AG
Pmul 511 ATTTTGTGAGAAACAGAAAGCTGATGGTAATATCAATTATCCATAATTACTAAGAACAAGTAAAGATATATTAAATAATTT--
S4-51oh 528 GTTTTGCGAAAAATAAAGATTAGAAGGAGGCTCCATCCATTATCCCACTCTATTATAAACAGTATCAATTTAATATATTTAAT
S2-CA1oh 530 GTTTTGTGAGAAATAAAGATTAGAAGGAGGCTCCATCCATTATCCCACTCTATTATAAACAGTATCAATTTAATATATTTAAT
S6-AZ84oh 528 GTTTTGCGAAAAATAAAGATTAGAAGGAGGCTCCATCCATTATCCCACTCTATTATAAACAGTATCAATTTAATATATTTAAT
S1-Ir42oh 527 GTTTTGTGAGAAATAAAGATTAGAAGGAGGCTCCATTATCCATCTTTATTATAAACAGTATTTACTTAATCT-----AT

A L Y L K Y N S Q D E M S I D D L I K H A H K

S4-51 600 A----AAGGCACCTTTACTGAAATATAATCTCAAGATGAAATGTCAACTCGATGATCTGATTAACATGCCCATAAAGA
S2-CA1 602 C----AAGGCACCTTTATTTAAAGTACAATCTTAAGATGAAATGACAATAGATGATCTCATTAAACATTCCCATAAAA
S6-AZ84 602 G----AAGGCACCTTTATTTAAAGTACAATCTTAAGATGAAATGACAATAGATGATCTGATTAACATTCCTATAAAGA
S1-Ir42 602 C--AAAGGCACCTTTATTTAAAGTACAATCTTAAGATGAAATGACAATAGATGATCTTATTAACATTCCTATAAAGA
Pcau 591 T----CAGGCCCTCTACCTCAAATATAACCGTGCAGAGAATAACCTTTGGAGGACATCTCAAACATGCCCATAAAA
Pmul 589 -----AGGCATTTATTTAAAGTACAATCTCAAGATGAAATGACATTTAGATGATTTAATTAACATTCCTATAAAGA
S4-51oh 603 C----AAGGCACCTTTACTTAAAGTACAATCTCAAGATGAAATGACATTTAGATGATCTCATCAACATTCCCATAAGA
S2-CA1oh 610 T----AAGGCACCTTTATTTGAAATACACTCTCAAGATGAAATGACAATAGATGATCTTATCAACATTCCTATAAAGA
S6-AZ84oh 603 AAAAAAGGCCCTATATTTGAAATACAATCTCGAGATGAGATGACGATAGATGATCTCATTAACATTCCTATAAAGA
S1-Ir42oh 602 A----AAGGCATTTATTTAAAGTACAATCTCAAGATGAAATGAAATATAGATGATCTCATTAAACATTCCCATAAGA

K W V I D F Y E K S Q E I Q D Q E F L K K Y Y D R E H

S4-51 674 AATGGGTTATAGATTTCTATGAGAAATCATAAGAAATATAGGATTAAGAGTTTCTTAAGAAATCTATGACAGAGAACAT
S2-CA1 676 AATGGGTTATAGACTTCTATGAGAAATCATAAGAAATAAAGACTAAGAGTTTCTCAAGTAATATTTAGTACTAT---CAT
S6-AZ84 676 AATGGGTTATAGATTTTATGAGAAACTTAAGAAATTAAGACTCAGAAATTTCTATAGAAATTTATGGTTGC-----CAT
S1-Ir42 678 AATGGGTGATAGATTTATGAGAAATCATAAGGACATATAGATTAAGATTTCTTAAGAAATATTTAGGATAT---CAT
Pcau 665 AGTGGGTGATGACTTTTTCGAGAAAGCACTTTGAAGTGAATGAT---GAATTTCTTAAGAAACGATCTCTAAT---CCT
Pmul 661 AATGGGTTATAGACTTTTATGAAATAAATCAATAAGAACTTAGAGAGAAGAAATTTATGAAATAAAGTGTATAAAT---TAT
S4-51oh 677 AATGGGTCATAGATTTTATGAGAAATCACAAGAAATCAAGACCAAGAGTTTCTCAAGAAATACCATGATTTG---CAT
S2-CA1oh 684 AATGGGTTATAGATTTCTATGAGAAATCTTAAGAAATCAAGACCAAGAGTTTCTCAAGAAATACCATGATTTG---CAT
S6-AZ84oh 681 AATGGGTTATAGATTTTATGAGAAATCTTAGGAAATTAAGATGAAGAGTTTCTTAAGAAATATTTAGCTTT---CAT
S1-Ir42oh 676 AATGGGTTATAGATTTCTATGAGAAATCATAAGAAATATAAGACTAAGAAATTTCTCAAGAAATCTATGATTTG---CAT

G I V P K L I D L T E Y Y K Y S L D T P N L F M L P L

S4-51 754 GGGATCGTTCCATAACTAATTTGATTTAACTGAATCACTACAATATAGTTTAGATACACCTAACTATTTATGCTCCCAT
S2-CA1 753 GAAATCAACTCTAAATTAATTTGATTTAACTGAATCACTACAATATAGTTTAGATACACCTAACTATTTATGCTCCCAT
S6-AZ84 753 GAGATCACTCTAAATTTGATTTAACTGAATCACTACAATATAGTTTAGATACACCTAACTATTTATGCTCCCAT
S1-Ir42 755 TAGATCACATTTAAATTAATTTGATTTAACTGAATCACTACAATACAGATTAGATACACCTAACTATTTATGCTCCCAT
Pcau 739 GAAATCTACTCTAAATTTAGTATGATTTGACAGAGTCTAAGATATAGACTTGACACTCTTATATGATGATGCTTACCCTT
Pmul 738 GAGATAGCATAAAACCTATCTAATTTGACTGAATTTATAAATATAGATTAGACACACCCACATTTATATATGTTACCATT
S4-51oh 754 GAGATCATCTCCAAATTAATTTGATTTGACGGATTTTACAATATAGATTTGGACACACCTAACTGTTTATGCTCCCAT
S2-CA1oh 761 GAGATCACTCTAAATTAATTTGATTTAACTGAATCACTACAATATAGATTAGACACACCTAACTATTTATGCTCCCAT
S6-AZ84oh 758 GAAATCAGTCCATAACTAATTTGATTTAACTGAATTTACAATATAGTTTAGAATCTCCTAACTGTTTATGCTCCCAT
S1-Ir42oh 753 GAGATCATCTCTAAATTAATTTGATTTAACTGAATTTACAATATAGATTTGGACACACCTAACTATTTATGCTCCCAT

F K L I K N Y N K L R K E Y H Y T K I K

S4-51 834 GTTTAAAGCTAATAAAAACTACATAAATTAAGAAAGTATCATCTTCAATATTTAGAGAGTATCATTACACAAAATCAA
S2-CA1 833 ATTTAAAGCTAATGAAAAATTACAATAAATACGAAAGTATCATATTTCAATATTTAGAGAGTATCATTACACATAAATCAA
S6-AZ84 833 ATTTAAAGTATGAAAACTACATAAATTAAGAAAGTATCATCTTCAATATTTAGAGAGTATTTATTACATTTAAATAAA
S1-Ir42 835 ATTTAAAGTTAATGAAAAATTACAATAAATACGAAAGTATCATCTTCAATATTTAGAGAGTATCATTACACATAAATCAA
Pcau 819 GAGTAAATTTGTTTCCAAATTACAATAAATACGAAAGTATCATGGCGACTATTTAGAGATTTACACTATCTCCATCAA
Pmul 818 ATGTAATTTAATAACAATCACATAAATTAAGAAAGTATCATTATATATTTAGAGAAATATCATTATACAAAATAAA
S4-51oh 834 ATTTAAATTTGATCAAAAATTACAATAAATACGAAAGTATCATTTTCAATATTTAGAGAAATACCATTAAATCAA
S2-CA1oh 841 ATTTAAATTTGATAAAAATTACAATAAATACGAAAGTATCATTTTCAATATTTAGAGATTTATCATTACACGCTCAATCAA
S6-AZ84oh 838 ATTTAAATTAATCAGAACTTATAATAAATACGAAAGTATCATCTTCAATATTTAGAGAGTATCATTACACATAAATTA
S1-Ir42oh 833 ATTTAAATTTGATAAAAATTACAATAAATACGAAAGTATCATTTTCAATATTTAGAGATTTATCATTACACTTAAATCAA

Y K L K I L V P Q N G I K V D K K Y T Q L L Q L T

S4-51 914 ATACAAATTAATAATTTGGTGCCTCAAAATGGCATTAAAGGTT---GATAAAAAATATACTCAATTTGCAATTGACGA
S2-CA1 913 GTATAAATTAATAATTTGATACCTCATAATAGCAATAAAGTT---GATATACAAATTTACACAAATTTACAAATTAACAA
S6-AZ84 913 GTATAAATTAATAATTTGGTACCTCAAAATACCAATAATGTC---GATGTTAAATTTTCCATACTATTAATCAACTAA
S1-Ir42 915 ATATAAATTAATAATTTGGTACCTTAAATTAATAAATAAAGTATGATATAAATAACACAAATTTACAAATTAACAA
Pcau 899 ATATAGACTAGGTATCTTGACTCCTTCCGCTCCACCTAAGAAATGAGATACCACCTACTCATAAATGCTTCAAATAACCT
Pmul 898 ATACAAACTGGATTATTAACATAAACATAAAAACT-----AGATGTACCATTACATAAATTTATAAATTAATG
S4-51oh 914 ATATAAATTAATAATTTGTTACCTTAAATTCGCAATAATGTT---GATATCAAAATTTACGCAATTTGCAATTAACAA
S2-CA1oh 921 ATATAAATTAATAATTTGTTACCTTAAATTCGCAATAATGTT---AATGTTCAAATTTACATAAATTTAATTAACAA
S6-AZ84oh 918 GTATAAATTAATAATTTGGTACCTCAAAATCAACTCATGTT---GATATACAAATTTAATTAATTTGTAATTAACAA
S1-Ir42oh 913 ATATAAATTAATAATTTGTTACCTTAAATTCGTAATAATATTT---GATGTTCAAATTTACATAAATTTGTAATTAACAA

M S N Q E D E Q Q S L I S L L N K F Q P Q Q I A I P

S4-51 991 TGAGCAACCAAGAAGATGAAT---AGTAATCTTTGATTTCACTCTTAAATAAATTTCTAACCATAAATAAATAGCTATCCCT
S2-CA1 990 TGAGTAAACAAGATGAGGAAT---AGTAGTCTTTGATTTCTCTCTTAAATAAATTTCTAACCATAAATAAATAGCTATCCCT
S6-AZ84 990 TGAGTAAACAAGAAAGATGAAT---ACAGTCTTTGATTTCTCTCTTAAATAAATTTCTAACCATAAATAAATAGCTATCCCT
S1-Ir42 995 TGTCTAAAATAAGAAGAAGAAAT---AGTAATCTTTGATTTCTCTCTTAAATAAATTTCTAACCATAAATAAATAGCTATCCCT
Pcau 979 CGAATGAAAGTTAATAGGATTTCTCTTACTTTAGTTTGAAGAACTTAGTGCCTTACCATAAAGAACAAAGTCTATAGG
Pmul 967 ATGATGAAATCA-----TAACTCAATTTAGTCTTATACATAAATAAATTTAGCAAAAATAAATAGTAAAATTTT
S4-51oh 991 TGAGTAAACAAGAAAGAAGAACT---ACCATCTTTGATTTCTCTCTCAATAAATTTGTAACCAACAAACCTAAAATGCCT
S2-CA1oh 998 TGAGTAAATAAGAAGAAGAACT---AATAGTCTTTGATTTCTCTCTTAAATAAATTTTAACTAACCAACAAACCTAAAATGCCT
S6-AZ84oh 995 CTAGTAAAGAAGAAGAAGAACT---AATAGTCTTTGATTTCTCTCTTAAATAAATTTTAACTAACCAACAAACCTAAAATGCCT
S1-Ir42oh 990 TGAGTAAATAAGAAGAAGAACT---TAACTCAATTTAGTCTTATACATAAATAAATTTCTAACCATAAATAAATAGCTATCCCT

K Q L K S Y T Q Q S C I Q T T K L H T K P L I K K S
S4-51 1068 AAATAACTAAAATCATATACATAAATCAAGTATGACTACTAAAGTTGCAT---ACGAAACCCCTGATTAATAAGAG
S2-CA1 1067 AAATAAAATAAATCATATACATAAATCAAAATGTTAGACTACTAAATATAT---CCGAATCCCTTAATCAAAATAGAG
S6-AZ84 1067 AAATAAAATGAATCTTACTCAATAAATCAAGTTTATCGACTACTAAATTAATT---TCGAAAGCCCTACAATCAAATTTAG
S1-Ir42 1072 AAATAAAATAAATCATATACATAAATCAAAATTAATCTACTACTAAATTAATCA---AAATAACCATTAATCAAATAAA
Pcau 1059 -----TAAACAAATTTAACTGCATCTAAATGTATACTCAACCTAAATTTGATCAAATAAAG
Pmul 1034 -----AAATTAATACACAATAATCTAGTTTATCAACTACTAAATTAAT---TAAAACCATTAATTCGATAAAG
S4-51oh 1068 AAATAAAATAAATCATATACATAAATCGAGTTTATCGCAACTAAATTAATTT---CAAAGGCCCTTGATTAATAAGAG
S2-CA1oh 1075 AAATAAAATAAATCATATACATAAATCAAAATTTGTCGACTCTAAATTAATTT---TAGAAACCATTAATTAATAAGAG
S6-AZ84oh 1075 AAACAATTTAAATCATATACATGTAATCAAGTTTATCGCAACTAAATTAAT---CAGAAATCTTTAATTAACAAGAA
S1-Ir42oh 1067 AAATAAAATTAATCATATACATAAATCAAGTTTATCAACATCTAAATTAATTT---TCGAAACCCCTTGATTAATAAAG

E Q A S Y K L K S I V N G D I V Q Q Q P Q L L S Q Q
S4-51 1145 TGAATAAGCAAGCTACAAACTTAAAGTCAATTGTAATGGGGATATAGTGTAAACAACACCTTAATTTATGTCCTCAATAAT
S2-CA1 1144 TGAATAAGCAAGCAAAAACTTAAATCTCTTTTAAATGGGGATATATTTAGTAACCAATAATTAATCACT-----
S6-AZ84 1144 TGAGCAACCAAGTAAAAAACTAAAATCCATTTTAAATGGGGACATTCGTTAACCTTAA---CAATTAATTAACATAATAAT
S1-Ir42 1149 TGAATAAGCAAGCAAAAACTTAAATCTCTTTTAAATGGAGAAATCTATAACAATAA---TAATTTATGCTCAATAAT
Pcau 1115 CGAGTAAGTCAGTCGCAAAATTAAGGCCATTTCAAATAATGATGTAAACCACTAGT-----GT
Pmul 1102 TGAATAAGTAAGTCGTAATTAAGGCCATTTTAAATGGTGAATTCGTTCTCCACCTTAAATAATCCATCCAAATCTC
S4-51oh 1145 TGAATAATTAAGTAGAAAACTTAAATCCATTTTAAATGGGGATATTCGTTAACCAATACCACAATAATAGCC-----AC
S2-CA1oh 1152 TGAATAATTAAGTAGAAAACTTAAATCCATTTTAAATGGGGATATTCGTTAAATAATACCATAATCATTGCC-----AC
S6-AZ84oh 1152 TGAATAAGTAAGTAGAAAACTTAAATCCATTTTAAATGGGGATATTCGTTAACCAATAACTATACTA---ATTTCC-----AC
S1-Ir42oh 1144 TGAATAATTAAGTAGAAAACTTAAATCTATTTTAAATGGAGATATTCGTTAAATAATACCATAATCATTGCC-----AT

Q Q Y Q L Q R R T Q S K S H S N I K F H T L Q S V N
S4-51 1225 AATAAATTAATTTATAAAGGGAACATAAATCAAAATCTCATTGCAATATTAATAATTCATACACTTCAAGGTGATAA---T
S2-CA1 1214 --CAATAATAAAT---AAAGAGAGTATAAATCAAAATCTCATTGCAATATTAATAATTCATACCTTCAAGGTGATAA---T
S6-AZ84 1221 AATAAATAAATAAAT---AAAAAGACTTAAATCCAAATCTCATTCTACTATTAAATTTTACTACTTTAATAGTGTATA---T
S1-Ir42 1226 AATAAATAAATAAAT---TAAAGGAGGATAAATCAAAATCTCATTGCAATATTAATAATTCATACCTTCAAGGTGATAA---T
Pcau 1174 AATAAATAAATAAAT---TAAAGGAGGATAAATCAAAATCTCATTGCAATATTAATAATTCATACCTTCAAGGTGATAA---T
Pmul 1182 AATCAACACTTAACTTTTACATAGTACATAT-----
S4-51oh 1219 AACCAACCAATAAAT---GAAGAGGACATAAATCAAAATCTCATTGCAATATTAATAATTCATACACTTCAAGGTGATAA---T
S2-CA1oh 1226 AACCAATAAATAAAT---AAAGAAACATAAATCAAAATCTCATTGCAAGCATTAAATTTTACTACTTCCATAGTGTATAA---T
S6-AZ84oh 1223 AATAAATAAATAAAT---TAAAGGAGGATAAATCAAAATCTCATTGCAATATTAATAATTCATACCTTCAAGGTGATAA---A
S1-Ir42oh 1218 AACCAATAAATAAAT---AAAGAGGACATAAATCAAAATCTCATTCTACTATTAAATTTTACTACTTCCACAGTATAA---T

N N S K K Q S I A Y N K S N L Q P I K T L G Q S K
S4-51 1302 AACAAAT-----TCAAAAAAAGCAGAGTATGCTTACAATAAATCAAAATCTGTAACCCATTTAAAATCTGGGACAAAGCAA
S2-CA1 1286 AATAAAT-----TCAAAAAAATGAGCATTTCAATGGAATAAATTAATCTATAACCAATAAATACTTTAGGATAAAGCAA
S6-AZ84 1295 AACACA-----ACCAAAAAATGAGTATCCAATCTAGTAAATCAAAATCTTTAGCCAAATCAAACCTTTGGGATAAAGCAA
S1-Ir42 1303 AATAAAT-----TCAAAAAAATGAGTATGCTTACAATAAATCAAAATCTGTAACCCATTTAAAATATAAAGGATAAAGCAA
Pcau 1251 GTTTATGGTCCATACCAAGAAGTTGAGCATCCAATAACCACTTATGAAACAACCTCT-----
Pmul 1214 -----AAAAAATGAGCATTTAAATAAAGTATGTAACCACTGAAAAA-----
S4-51oh 1293 AACAAAT-----TCAAAAAAAGCAGAGTATGTAACAAGTAAATCAAAACCTGTAATCAATCAAACCTTTGGGATAAAGCAA
S2-CA1oh 1300 AACAAAT-----ACAAAAAAGCAGAGTATGTAACAAGTAAATCAAAATCTGTAACCAATCAAACCTTTGGGATAAAGTAA
S6-AZ84oh 1297 ATCAGT-----TCAAAAAAATGAGTATGTAAGAAGCAGTAAATCAATCATGCAACCAATAAATACTTTAGGATAAAGTAA
S1-Ir42oh 1292 AATAAAT-----TCAAAAAAATGAGTATGTAACAAGTAAATCAAAATTTGTAATCAATTTAAAATCTTTGGGATAAAGCAA

I G S Q A D I I S N I Q I Q N T N S I G K T R S R R
S4-51 1376 AATAGGATCTCAAGCGGATATAATCAGCAATATTTAAATATAGAACCAATCTTATAGGTAACCAAGATCAAGGAA
S2-CA1 1360 AATAGGATCTTAAGCAGAAATAAATTAACAACATTTATCTATAGAAAGCAATCTTAATGTAACCAAGGAAACAAATAGAA
S6-AZ84 1369 AATCGGATCATAGGCTGATATCATTAGCAGCATATATTTATAGAAAAGTAATCCCATCGTTAAAAACAGATAGCAAAGAA
S1-Ir42 1377 ATTTGGATCTTATAGTTGACATAAATTAACAATAATTTAATATAGAAAAGTAATCTATTGTTAAAAACAAGATAACAAGAA
Pcau 1308 -----TAGGCAGATATCATTAGTATGATTTTACTTCAAAAAAAGCAATGTTAGAAAGTAG-----AA-----AA
Pmul 1255 -----ATCATCAGCAGACATAAATAGTATGATTTATTTACAAAAAAGTGGATCCAAAGCTCGATAAATATAA-----AA
S4-51oh 1367 AATCGGATCTCAGGCAGATATAATTAGTAGCATATCTTACAGAAAAGCAACCCATTGTTAAAAATAGATAACAAGAA
S2-CA1oh 1374 AATAGGATCATAAGCAGACATAAATTAGTAGTATTTACTTTATAGAAAACAATCTTATGTTAAAAATAGATAACAAGAA
S6-AZ84oh 1371 AATTTGGATCATAAGCTGATATTTATAGCAGCATATTTTACAGAAAATCAATCCAAATTTGTTAAAAAGCAGATAAATAAGGA
S1-Ir42oh 1366 AATTTGGATCATAAGCAGATATAATTAGTAGTATTTATTTATAGAAAACAATCTTATTGTTAAAAATAGATAACAAGAA

K L I Q L F Y K * M I I L F A Q F L L S V Q A I Y I L H
S4-51 1456 AATTAATTTAATTTATTTTATAAATGATTTATATTTTGC-ATAATTTTACTTTTCAGTTTAAAGCAATTTACATCTTACAC
S2-CA1 1440 AATTAATTTAATTTATTTTATAAATGATTTATATTTATGTC-ATTATTTTACTTTTCAGTTTAAAGCCATTTACATCTTACAT
S6-AZ84 1449 AATTAATTTAATTTATTTTATAAATGATTTATATTTTGC-ATTAATTTTACTTTTCAGTTTAAAGCAATTTACATCTTACAT
S1-Ir42 1457 AATTAATTTAATTTATTTTATAAATGATTTATATTTTGC-ATTATTTTACTTTTCAGTTTAAAGCCATTTATTTTACAT
Pcau 1366 AATTAATGTAATTTTATAAATGATTTATATTTTGC---TTGCTTCTCACATCAGTATTTCCATTTACATCTTCCAT
Pmul 1323 AAATTAACACAATTTTTTATAAATGATTTATATTTTGC-ATTATTTACTGCAAGTCAGAAAGT---ATCTATATTTCTTCAT
S4-51oh 1447 AATTAATTTAATTTATTTTATAAATGATTTACATTTTAC-ATTATTTTACTCGCAGTTTAAAGCCATTTACATCTTACAT
S2-CA1oh 1454 AATTAATTTAATTTATTTTATAAATGATTTATATTTTAC-ATTATTTACTTCCAGTTTAAATCCATTTACATCTTACAT
S6-AZ84oh 1451 AATTAATTTAATTTATTTTATAAATGATTTATATTTTGC-ATTATTTTACTTTTCAGTTTAAAGCCATTTTAAATTTACATCTTACAT
S1-Ir42oh 1446 AATTAATTTAATTTATTTTATAAATGATTTATATTTTAAATTTCAATTTAATTTTACTTTCAATTTAGGTTTTCGAAATTCGAAATTT

T K G I S I S Q L D Y F V I E L N Q I D Y K V K S Q
S4-51 1535 ACCA-AAGGAATTTCAATAGTCAATTAGATTTTGTAAATGAGTTGAATCAAATGATTA-TAAAGTGAAGTCACAA
S2-CA1 1519 ACTA-AAGGAATTTCAATCAGTCAATTAGATTTTGTATTGAAATGAAATCAAAGTGATTA-TAAAGTGAAGTCACAA
S6-AZ84 1528 TCCA-AAGGAATTTCAATCAATAAATAGAGTATTTTGTATTGAACTGAATCAATTTAATTA-TAGGGAAGTCCGAA
S1-Ir42 1536 ACCA-AAGGAATTTCAATTAATCAATTAGATTTTGTAAATGAAATCAAATCAAAGTGATTA-TAAAGTGAAGTCACAA
Pcau 1442 ACTA-AGGGTATCTCAATGAAAAAAGCAGTACTTTGTAGTGAATCAATTAGTCAAATTA-CAACTTAAATTCATAA
Pmul 1399 ACCA-AAGGAATTTCAATTAATAGAGATGATTTACTTTCTAATTAATTAATTAACAGATTA-TAATGCAAAATCATAA
S4-51oh 1526 GCTAAAAAGAATATCAGTCAATCAATTTGATTGCTCATCATCTTAAATGAATTAGGATTTGCGTTA-
S2-CA1oh 1533 GAATAAAAGAATGTCAAACAAAT-----ATTTTATAGCTAATTTAATTAGAGTAAATTTTAAATTTGATATTCACAA
S6-AZ84oh 1531 AAATTTGTTTCAATAAAGGTTTGAACAAAAGAGGTTTCAAT-----
S1-Ir42oh 1512 ----AAAGGAATTTCAATCAATCAATTCATTATTTTATCATCTAATAGAATTAAGTAAACTCAATG-----

I A H N G S F T I S K L Y Q D S A L F D I F I L D Q

S4-51 1613 ATGGCACATAATGGATCATTACAAATATCCAAKCTTATAGGATTCTGCTTTATTTGATATATTCATCCTA--GATTAA
S2-CA1 1597 ATGGCACATAATGGATCATTACAAATATCCAAAGCTTATAATGATTCTGCTTTATTTGATATATTCATCCTA--GATTAA
S6-AZ84 1606 ATGGCACATAATGGATCATTACAAATATCCAAACTTTATAATGACTCTGCTTTATTTGATATATTCATCCTT--GATTAA
S1-Ir42 1614 ATGGCACATAATGGATCATTACAAATATCCAAACTCTATAATGATTCTGCTTTATTTGATATATTCATCCTA--GATTAA
Pcau 1520 ATAGTACACAAATGGATCAATCGCAATCACTAAATGTATGATGACTCAGCTCTCTTTGATGTTTTTTACTT--GATTAA
Pmul 1477 ATCTCACATAATGCTTCATTCAACATACAAAGCTTATAATGATTCAAGCATATTTGATATATTTGATTA--GATGAA
S4-51oh 1594 -----ATAATGATTCTCCTTTTTCCTCGTATATATTTTGTATTA-----
S2-CA1oh 1604 ATTTCAATAATTTGATTATCCACAAATCAACTATAATAACTAAATATGATATCAATTAATTTACTAAATCAATTGTAC
S6-AZ84oh 1576 -----
S1-Ir42oh 1576 -----ATTCAAATATTCATTTTATAATGA-----TTATTGATATATTTATCTTATATCTAC

A in 1.28 mutant

P G Y T C Y L Q A Y Y N S S W E T E C F F L F N Q S Q

S4-51 1691 CCTGGTTATACCTGTTATTTATAAGCCTATTACAAATAGTCTTGGGAAACAGAATGTTTTTCTTGTAAATAGAGTTA
S2-CA1 1675 CCAAGCTATACCTGTTATTTATAAGCATATTACAAATAGTCTTGGGAAACAGAATGTTTTTCTTGTAAATAGAGTTA
S6-AZ84 1684 ATCGGATATACATGCTACCAATAGGCTTTTATTAATAGTCTTGGGAAACCAATGCTTTTTCTTATTTAACTAGTCA
S1-Ir42 1692 ACTGGATATAATGCTATTTATAAGCCTATTATTAATAGTCTTGGGAAACAGAATGTTTTTCTTATTTAACTAGTCA
Pcau 1598 GTGGGATATCAATGTTATTTGTAAGCTTATTACAAATAGCTCATGGGAGTAGGAGTCTTTTTCTTATTAATAGTTA
Pmul 1555 AGTGGACTATAATGTTATTTATAAGCATTTTACAAATAGTCTTGGGAAACAGAATGTTTTTCTTATTTAACTAGTCA
S4-51oh 1637 ACTGCATATACCTAGTTATTTATAATCTTATTAAATAGTAAACATAATGTTTTCTATTTTCGAGTAGAGTCAACCAAT
S2-CA1oh 1684 TAACTCTGT-----
S6-AZ84oh 1576 -----
S1-Ir42oh 1632 TAATTAGAGCAAAAATATAAAAAATTAATACATAATATTTTAGATGTACTTAATATTTAATATGAAAAATACATAA

T S L N N F E Y E Q K Y I V A D A Q R Y L V I D N T

S4-51 1771 AACCAAGTCTCAATAATTTGAAATAGGCAAAAGTATATAGTAGCAGATGCACACAGATATTTGGTTATAGATAACACTC
S2-CA1 1755 AACCAAGCTTAATAATTTGAAATATCAAAAACAATACTGTTAGCAGATGCACAAAGATATTTAGTTATTTGATAACACTC
S6-AZ84 1764 AACCAAGCTTAATAATTTGAAATATCAAAAACAATACTGTTAGCAGATGCACAAAGATATTTAGTTATTTGATAACACTC
S1-Ir42 1772 AGCAAGTCTCAATAATTTGAAATATCAAAAAAGCTATCCTGTAGCAGATGCACAAATATATTTAGTTATTTGATAACACTC
Pcau 1678 AACCAAGCTTAATAATTTGAAATATCAAAAAAGCTATCCTGTAGCAGATGCACAAAGATATTTAGTTATTTGATAACACTC
Pmul 1635 AACCTAGTTTGAATAATTTTCAATACCAAAAATATATAGTGTAGAAATGCATTAAGATATATTAATTTGATAACACTC
S4-51oh 1717 GGATATCATAAACAGTATCATTGAT-----AGATGCACAAATGTATTCTATTTATTGTAGAGCTC
S2-CA1oh 1693 -----
S6-AZ84oh 1576 -----
S1-Ir42oh 1712 CTAATTTAATGTGATTTTAACTACTATATTAATAATCTAGAACTATCTTTGA-----

P F P Q N G A Y F A D R L Y L Q L D Y I O L I O P I P

S4-51 1851 CATTCCCTCAAAATGGAGCTT-ATTTGCTGATAGATTGTATTGCAATGGATTATATATAATTTGATCTAACCAATACC
S2-CA1 1835 CTTTCCCTCAAAATGGAGCTT-ATTTGCTGATAGATTATATTTATAATTTAGATTATATTTAATTTGATCTAACCAATACC
S6-AZ84 1844 CATTTCCCTAGAAATGGTGCAT-TTTTGGCTGATAGATTATATTTATAATTTAGATTATATTAAGTGTGAACCAATACC
S1-Ir42 1852 CTTTCCCTCAAAATGGAGCTT-ATTTGCTGATAGATTGTATTTACAAATTTAGATTATATTTAATTTAATCTAACCAATACC
Pcau 1758 CTTTCCCTCAAAATGGAGCTT-ATTTGCTGATAGATTGTATTTATAAGTGTGATTTCCATATAGGTTGGCAACCAATCCC
Pmul 1715 AATTTCCATAAAGATGGTGCAT-ATTTGCTGATAGATTATATTTATAAATTTGATTATACATAAGCAATTTGAACCTATACC
S4-51oh 1777 TATTACAATAAATCAACTTTTAAATTTCACTGATAGATTATATTAATAATGGAAATCACAAATATCTCTTTATTACTACCTC
S2-CA1oh 1693 -----
S6-AZ84oh 1576 -----
S1-Ir42oh 1764 -----

T Y Y Q V T I K I T I V F V G V L

S4-51 1930 TACCTATTATCAAGTTACATAAAAAAGTATATTTAGGCCA---TAAAAA---TAGTAAACAATTTGATTTGAGGAGTGTTA
S2-CA1 1914 AACCTATTATCAAGTTACATAAAAAAGTATATTTAGGCCA---TAAAAA---TAGTAAACAATTTGATTTGAGGAGTGTTA
S6-AZ84 1923 TACATATTATTAAGTTACTATAAAAAAGTATAAATAATTA---TAAAAA---TAGTAAACAATTTGATTTGAGGAGTGTTA
S1-Ir42 1931 AACCTATTATTAAGTTACATAAAAAAGTATATTTAGGCCA---TAAAAA---TAGTAAACAATTTGATTTGAGGAGTGTTA
Pcau 1837 TACCTATTATTAAGTTACTATAAAAAAGTATAAATAATTA---TAAAAA---TAGTAAACAATTTGATTTGAGGAGTGTTA
Pmul 1794 AAAATATTATTAAGTTAAACAATTTAAAGTAAATATAATGATTTAAATTT---TAGTTACAATCGTATTTGTATCAACATTA
S4-51oh 1857 AACTTAAGTTATGATATAATTTAGTACTAGGAAGT-----TAAATTTAAAGTAAATAACTAATCAAATTAATGCGTAAGAGTCTTAGAATTCATGATTAATTTAAATATA
S2-CA1oh 1693 -----TAAAGTTAAATAATACTAATCAAATTAATGCGTAAGAGTCTTAGAATTCATGATTAATTTAAATATA
S6-AZ84oh 1576 -----
S1-Ir42oh 1764 -----ATAAAGTTACAATAAAAACTAATAAAAAATTAAGTCATAGTTAGTTGTAGAATATAATTTATGACTTATA

I C F S I W C L I L S L K Y R K E T K H K E S I E R

S4-51 2007 ATTTGTTTTAGTATTTGGTGTGATATTTGAGTTTAAAGTATAGAAAAGAAACAAAACACA-AGGAATCCATAGAAAAGGA
S2-CA1 1988 ATATGTTTTAGTATTTGGTGTGATATTTGAGTTTAAAGTATAGAAAAGAAACAAAACACA-AGGAATCCATAGAAAAGGA
S6-AZ84 1997 ATATGTTTTAGTATTTGGTGTGATATTTGAGTTTAAAGTATAGAAAAGAAACAAAAGCACA-AGGAATCAATAGAAAAGGA
S1-Ir42 2005 ATTTGTTTTAGTATTTGGTGTGATATTTGAGTTTAAAGTATAGAAAAGAAACAAAACACA-AGGAATCCATAGAAAAGGA
Pcau 1911 TTTATGTTCTTTATATGTTGCTAGCTATTTGACTGTGAAATACAAAAGAAATTTAGACCTTA-AGGAACCAATTTGAAAAGGA
Pmul 1871 ATATGTTTTATAATATGGGGTATTTATTTGATATTTAAATATCGTAAATAAAAGAAACATA-AAAAATGATTTGAAAAGG
S4-51oh 1896 GTTTGTGTTAGAAATTTAGAGTATGATGGTAGTTTAAATGTAAAAGGAAATTAACAAAGCATAAGTAACCATAGAAAATAA
S2-CA1oh 1764 TAACTCAATGT-----
S6-AZ84oh 1576 -----GAATTCATAAAAATAA
S1-Ir42oh 1836 CTTTAAATAATAGAA-----

N K L N N T L S P L K Q Y K *

S4-51 2086 ATAAATTAATAATACCTTATCACCTTT-----AAAGTAATATAAATGATTTTATCGAATTCAAAATATATTTGC
S2-CA1 2067 ATAAATTAATAATACCTTATCACCTTT-----AAAGTAATATAAATGATTTTATCGAATTCAAAATATATTTGC
S6-AZ84 2076 ATAAATTAATAATACCTTATCACCTTT-----AAAGTAATATAAATGATTTTATCGAATTCAAAATATATTTGC
S1-Ir42 2084 ATAAATTAATAATACCTTATCACCTTT-----AAAGTAATATAAATGATTTTATCGAATTCAAAATATATTTGT
Pcau 1990 CCAAACTTAACACTATACATCTCCTCCAATCAAAACAAAATAATATAAATGATTTTATCGAATTCAAAATATATTTGC
Pmul 1950 ATAAAGATAAAGTACCTGCAACATGTCACCACTTAAAGGTTATAGATTTGATATTTATATATTTCTACTCAACACAA
S4-51oh 1976 ATTAATTAATAATGTTAAATATCACGCTT-----AAAGAAATATCGATAATTCATC-----
S2-CA1oh 1774 -----
S6-AZ84oh 1592 AAAAATTAATAATAAAGTATAGGATTTATGTTATTTATTTATTTATAATTTTAAAGAAATTTGAAACAAATGATATAAC
S1-Ir42oh 1851 -----TTG-----AAAGGAATATTAATGATTTATC-----

Figure 1 : Aligement de la région génomique de Pds2. Les régions génomiques de Pds2 et de son ohnologue WGD1 (oh) dans les espèces *P. tetraurelia* souche 51 (S4-51), *P. primaurelia*, souche Ir4-2 (S1-Ir42), *P. sexaurelia*, strain souche AZ8-4 (S1-Ir42) et *P. biaurelia*, souche V1-4 (S2-CA1) ont été alignés. En noir sont indiqués les nucléotides conservés. La séquence protéique en bleu correspond au gène Pds2, la séquence en vert correspond au gène en amont. Les codons initiateurs et terminateurs sont surlignés en vert.

III. Annexe 3: documents supplémentaires de l'article 2

Fig.S1

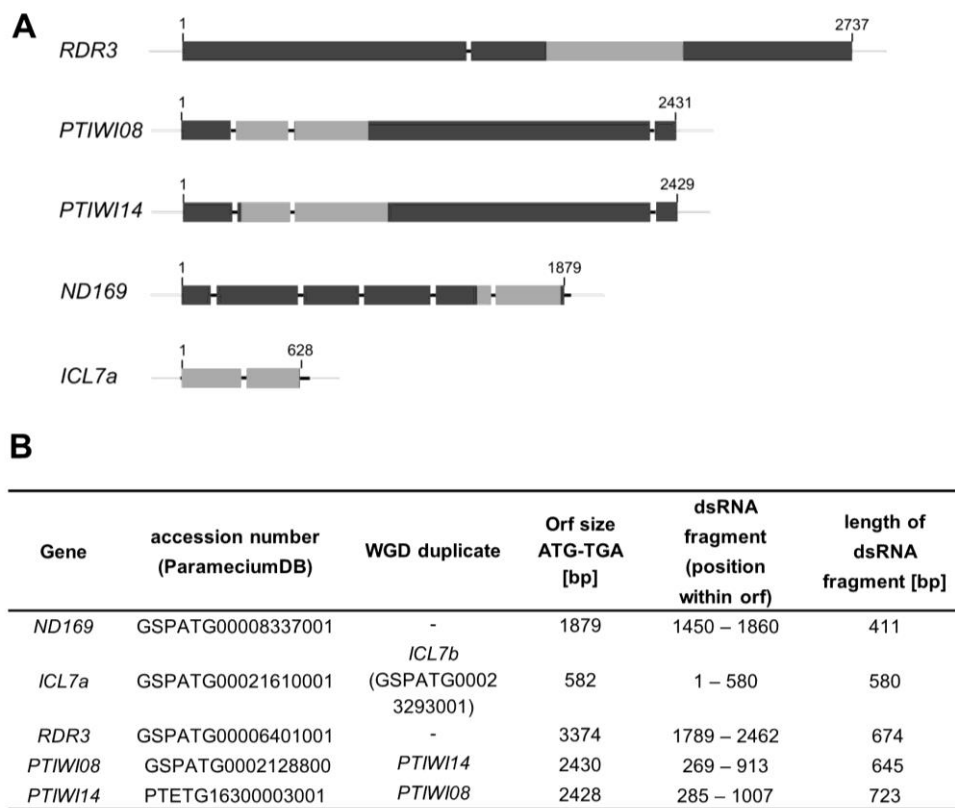


Fig. S1: (A) Relative position of dsRNA fragments (light grey) used for analysis of siRNA production from various target genes. Introns are represented as gaps. DsRNA fragments overlapping introns contained the intron sequence. **(B)** Target genes and properties of the dsRNA fragments.

Fig.S2

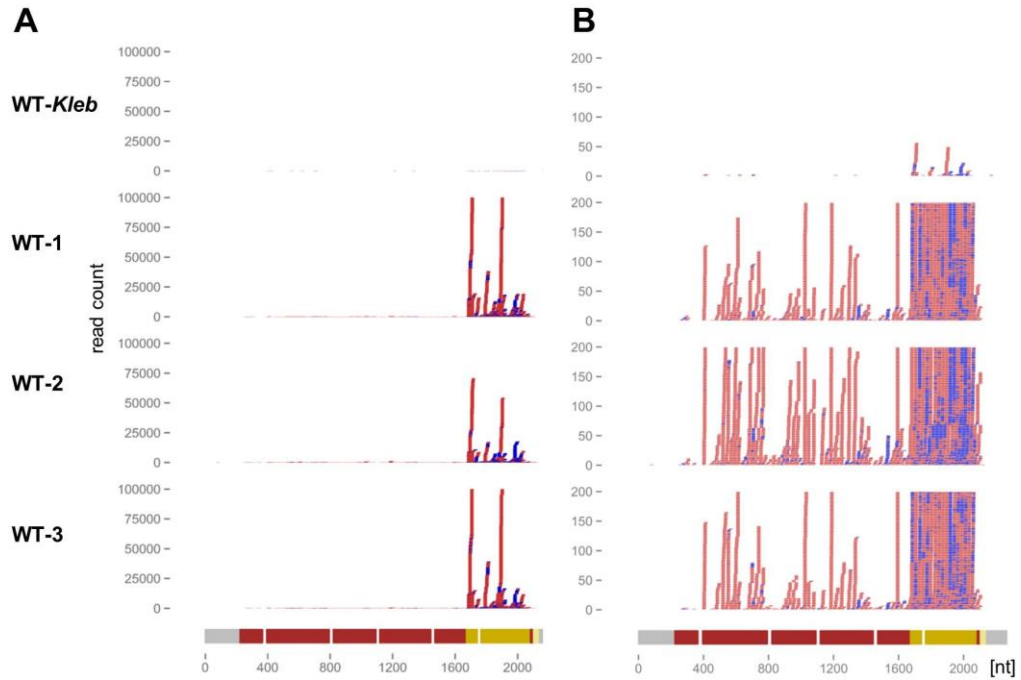


Fig. S2: 23nt siRNA distribution on the *ND169* target gene is reproducible in different biological replicates. **(A)** Some siRNAs mapping the dsRNA region (yellow bar) are highly overrepresented. Sense siRNAs are shown in blue, antisense siRNAs in red. Only 23nt siRNAs are shown. **(B)** Reduced scale reveals that secondary siRNAs spreading beyond the dsRNA target region also accumulate inhomogenously along the mRNA (dark red bar). Low levels of strongly overrepresented siRNAs are detectable in *Klebsiella* control samples, representing cross-contamination between samples. 3'U-siRNAs are not included in the figure.

Fig.S3

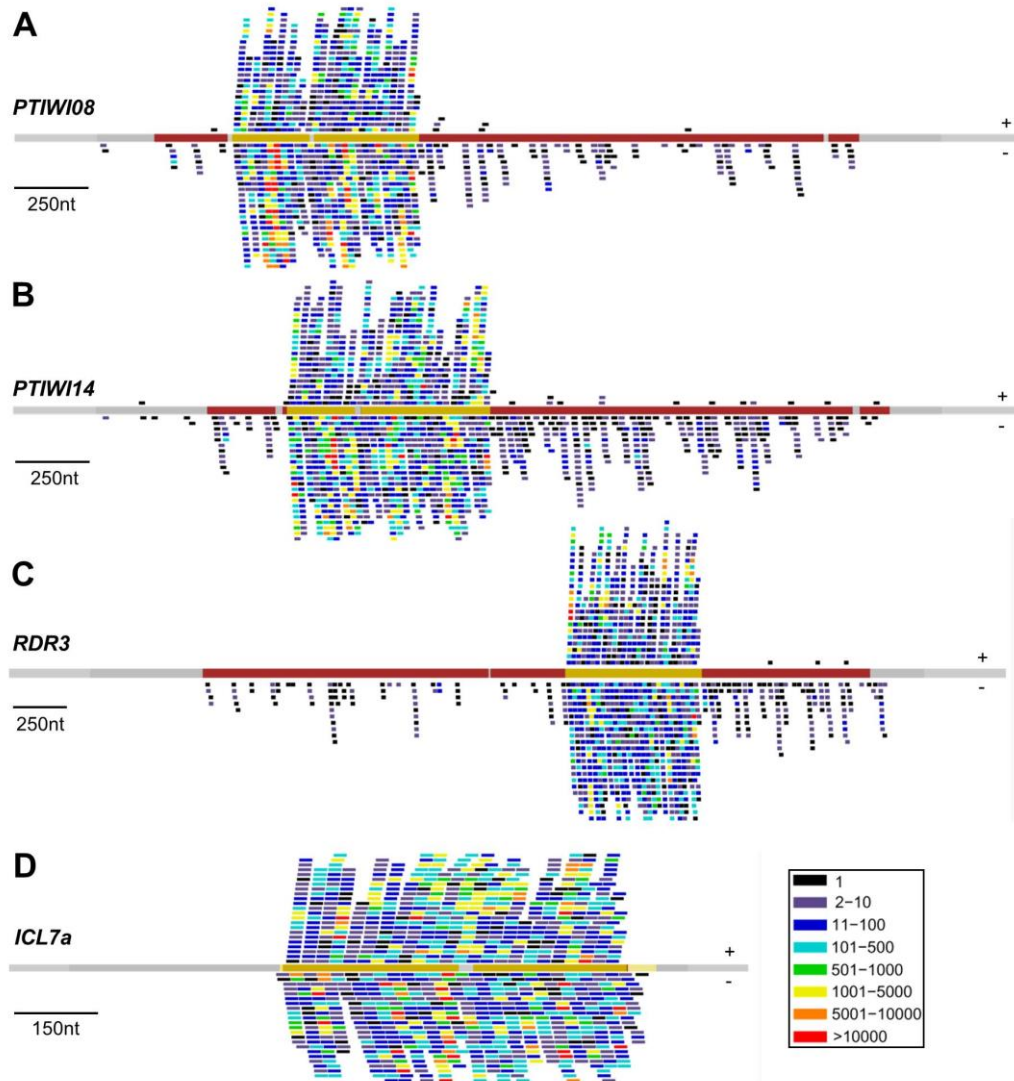


Fig. S3: Secondary siRNAs spreading beyond the dsRNA target region with a strong antisense bias is a general phenomenon of dsRNA-induced silencing. **(A, B)** Some *PTIW108*-specific and *PTIW114*-specific antisense siRNAs are highly overrepresented (orange and red traits). In contrast, *RDR3*-specific **(C)** and *ICL7a*-specific overrepresented siRNAs **(D)** accumulate from both strands. Distributions of 23nt siRNAs on the genomic sequence of target genes are shown (red line; black: introns; light grey: flanking intergenic regions; dark grey: 3'UTR; yellow: dsRNA target region). Individual siRNAs are represented as trait, the color indicating its abundance. siRNAs represented above and below the line are sense and antisense to the target mRNA, respectively. 3'U-siRNAs are not included in the figure. Note that only *PTIW108*-specific siRNAs are shown, i.e. excluding those also mapping to its WGD1-paralog *PTIW114*.

Fig.S4

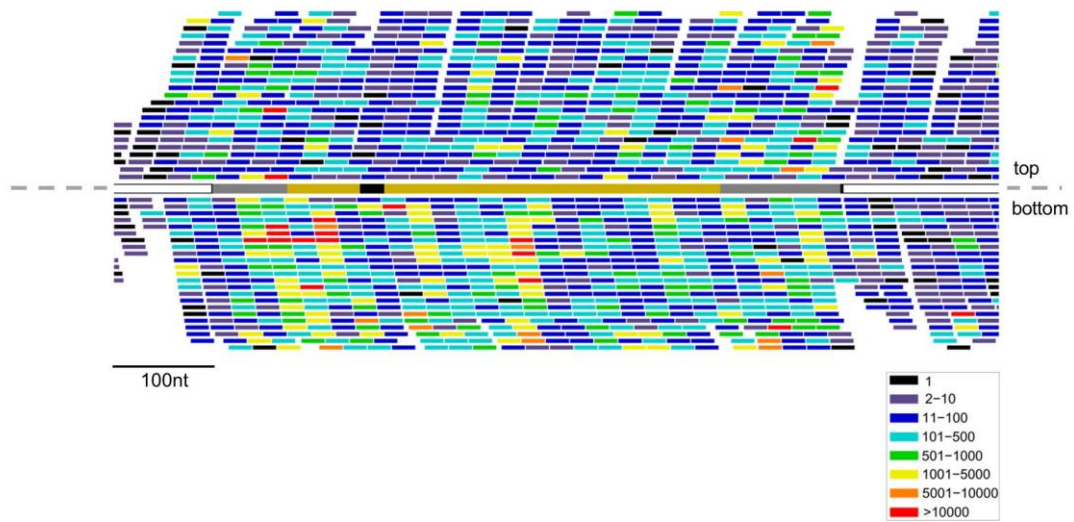


Fig. S4: L4440 vector siRNAs accumulate from both strands of the dsRNA. The distribution of 23nt siRNAs on the *NDI69* dsRNA sequence is shown (yellow line: dsRNA; black: intron; grey: 5' and 3' vector regions; white: other parts of the L4440 plasmid vector of which no dsRNA was produced). Individual siRNAs are represented as trait, the color indicating its abundance. siRNAs represented above and below the line are sense and antisense to the target mRNA, respectively. 3'U-siRNAs are not included in the figure.

Fig.S5

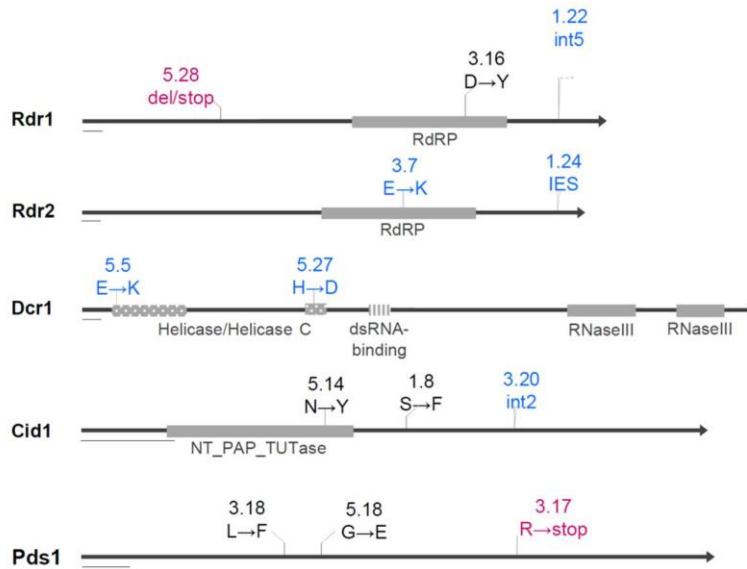


Fig. S5: Mutant alleles of the genes *RDR1*, *RDR2*, *DCRI*, *CID1* and *PDS1* used in this study. Non-ambiguous null alleles are represented in purple, other phenotypic complete loss-of-function alleles in black and hypomorphic alleles in blue. Alleles *rdr-1.22* and *cid1-3.20* retain an intron due to mutations in an intron boundary. Allele *rdr2-1.24* retains a micronucleus-specific non-coding fragment (internal eliminated sequence, IES) in the MAC-version of *RDR2*, leading to a frameshift and premature termination. Each scale bar represents 50 aa.

Fig.S6

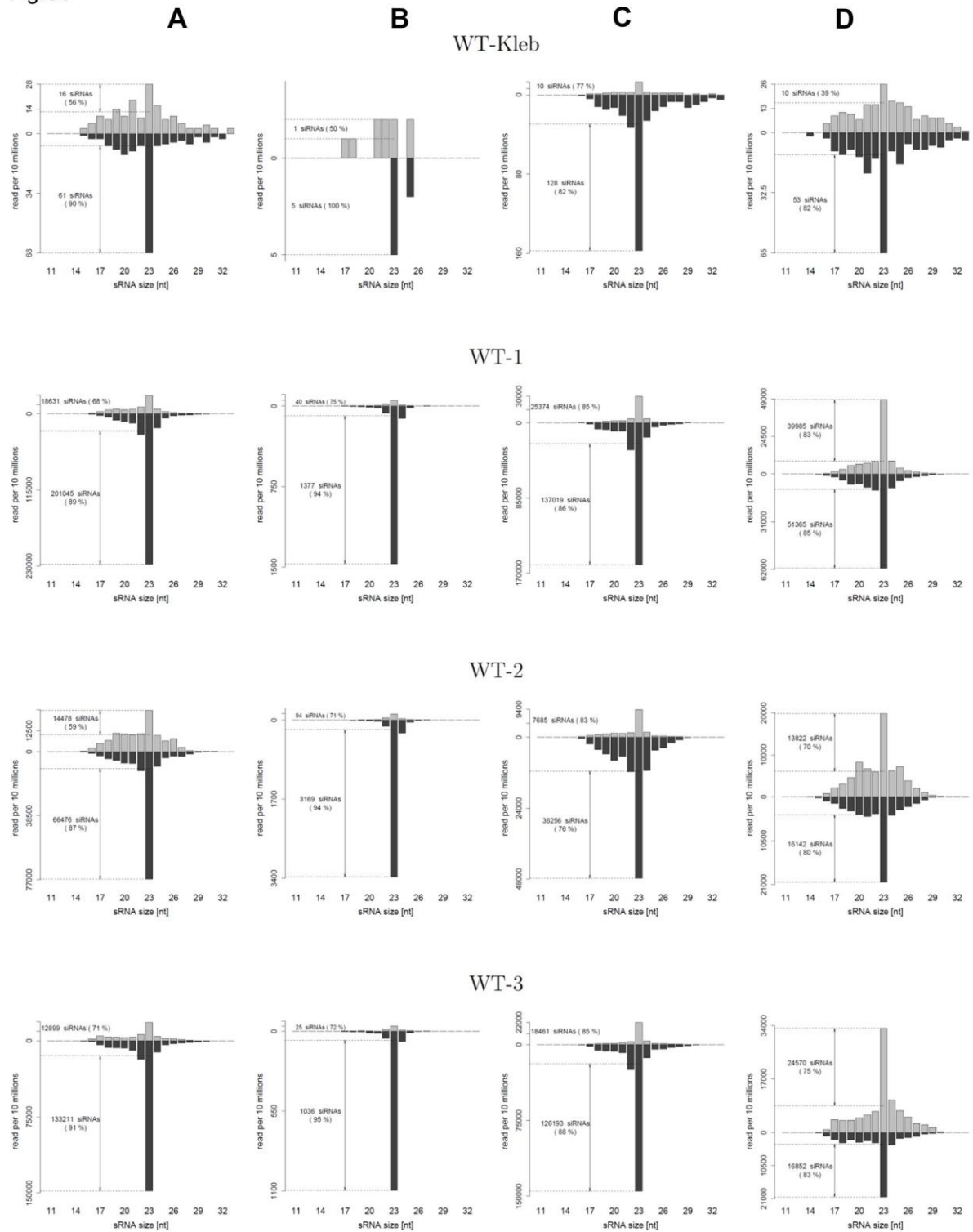


Fig.S6

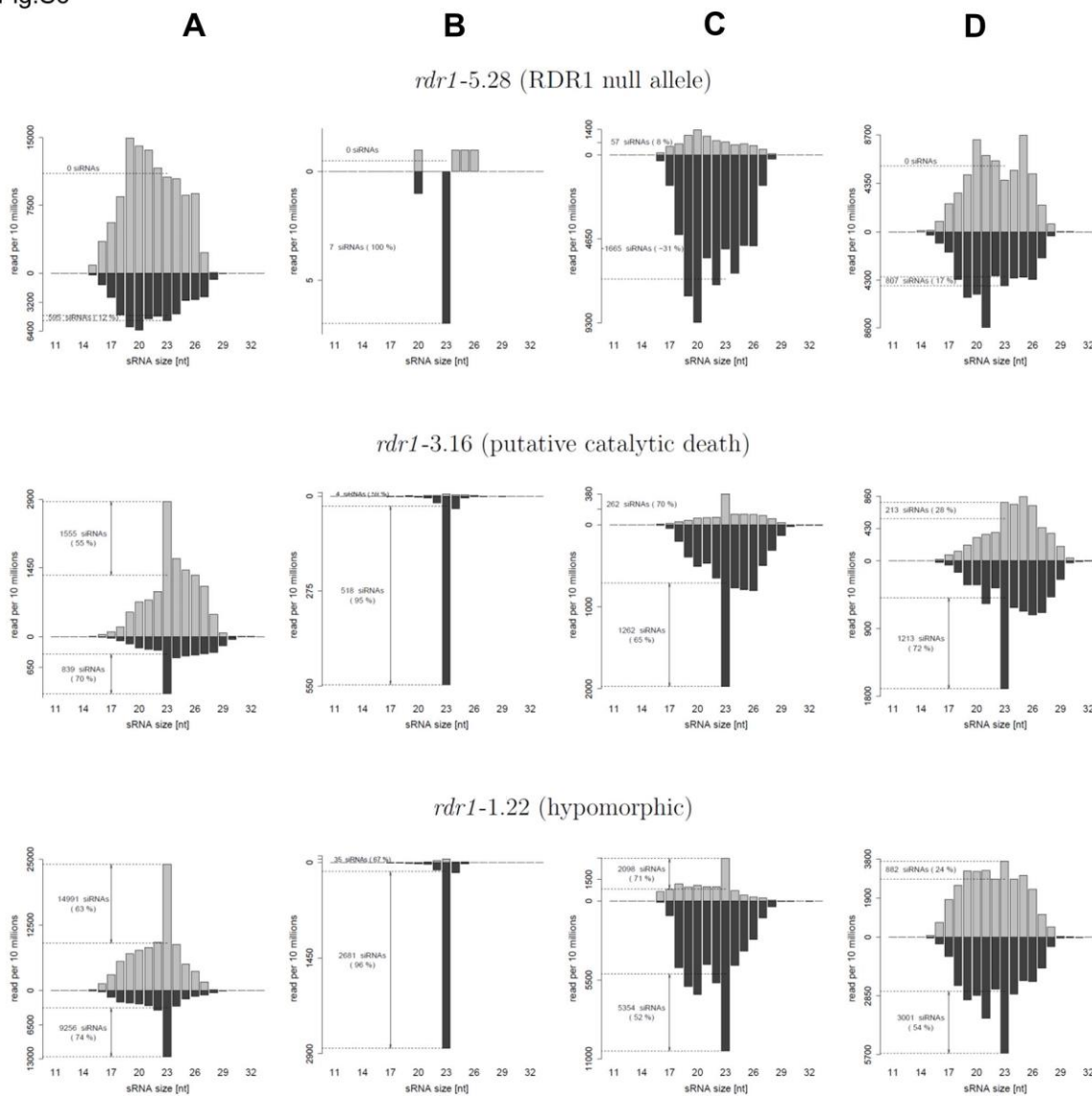


Fig.S6

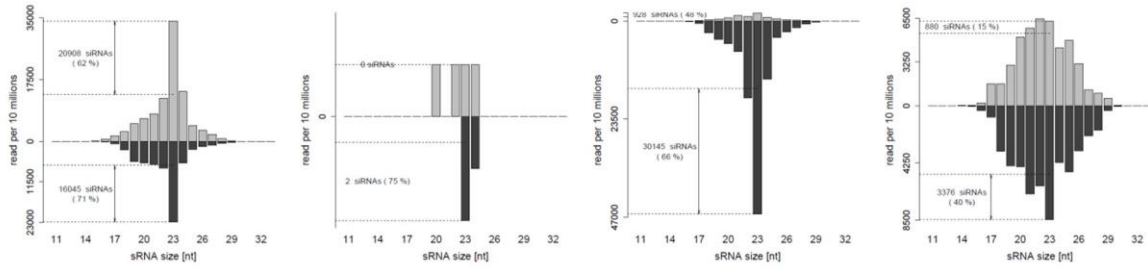
A

B

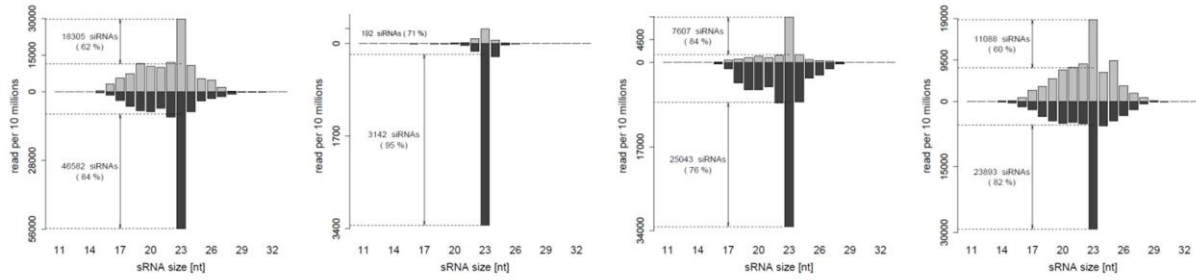
C

D

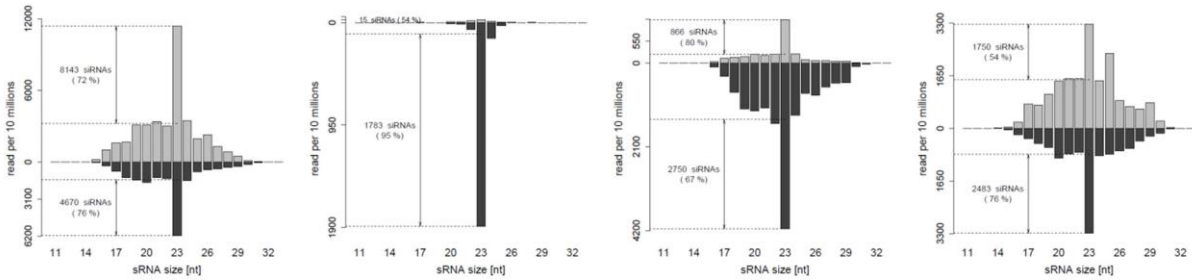
rdr2-1.24 (hypomorphic)



rdr2-3.7 (strongly hypomorphic)



dcr1-5.5 (hypomorphic)



dcr1-5.27 (strongly hypomorphic)

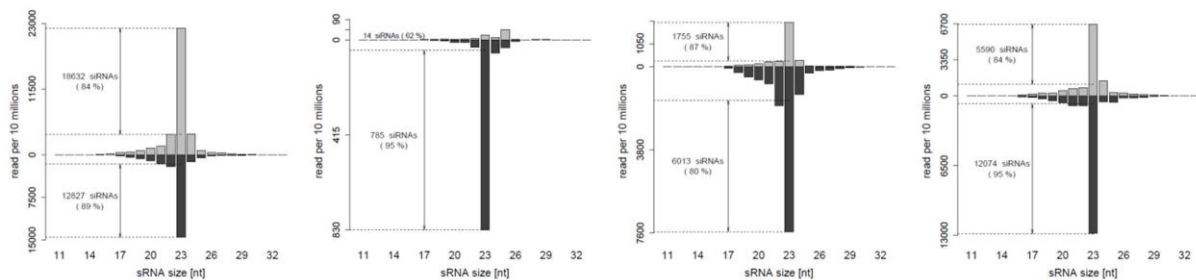


Fig.S6

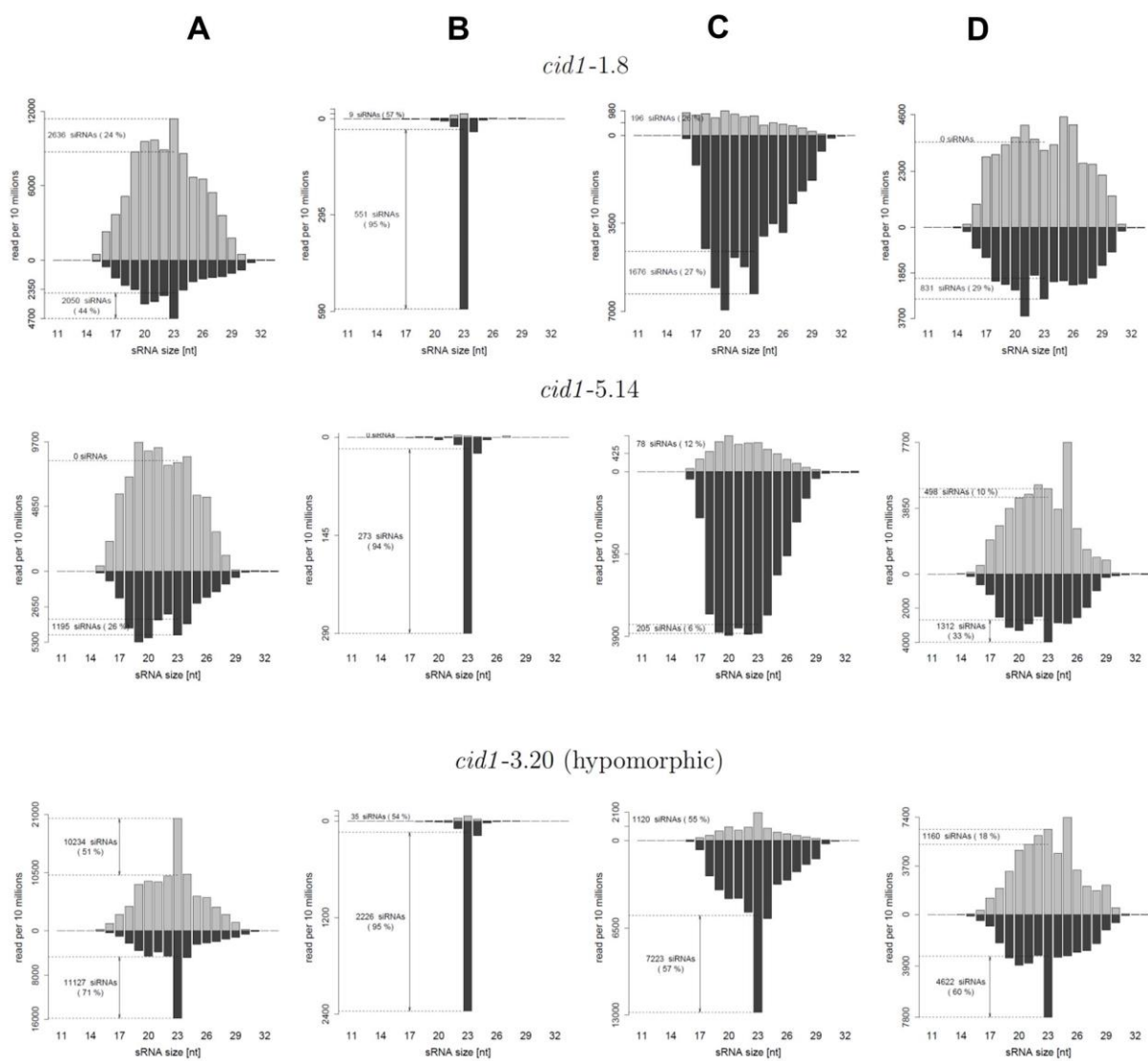


Fig.S6

A

B

C

D

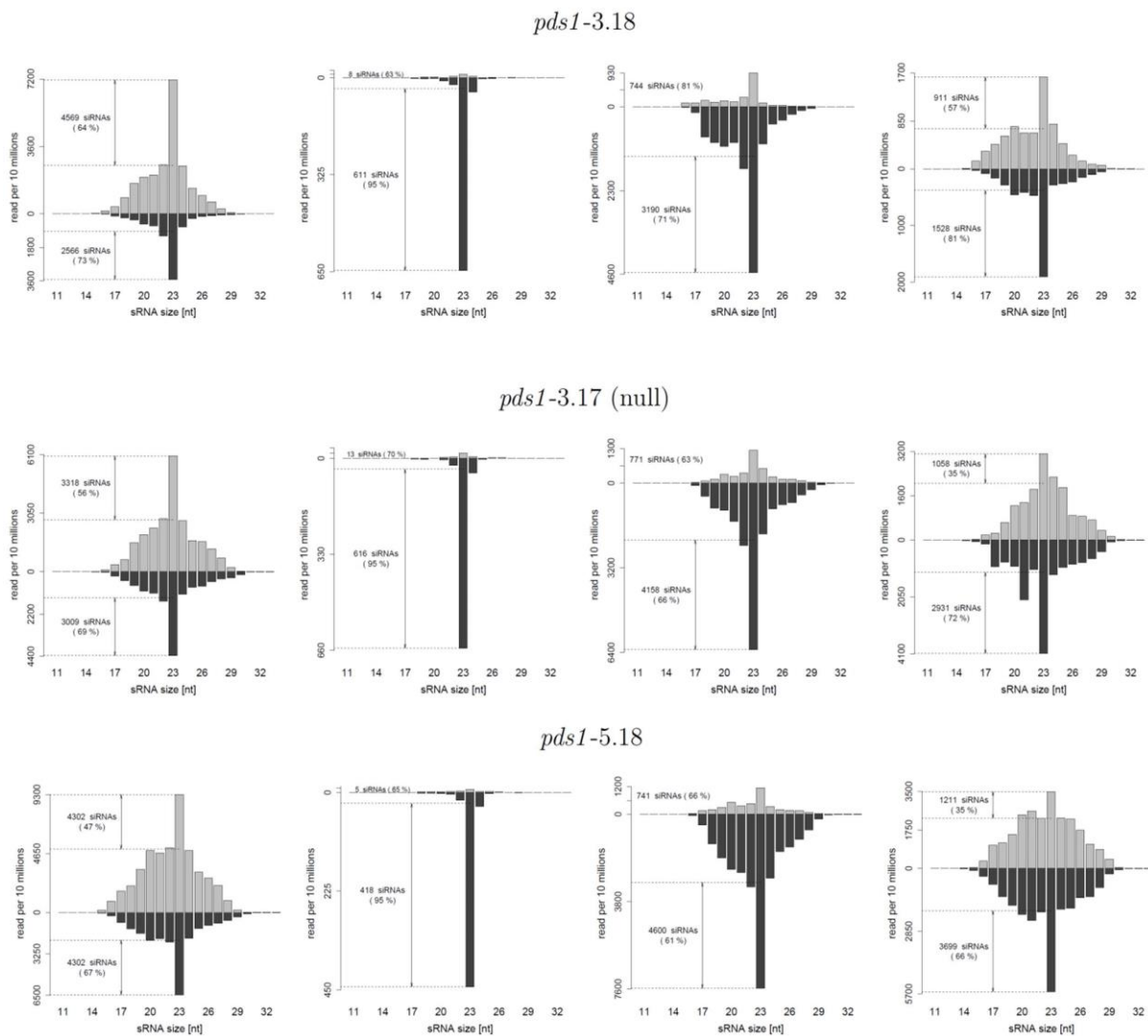


Fig. S6: Normalized size distribution of *ND169* dsRNA-induced siRNAs in wild type (WT) and RNAi-deficient mutant samples. SiRNAs mapping the *ND169* dsRNA target region (A), outside the dsRNA region (B), the 5' vector (C) and the 3' vector sequence (D) are shown. The mean level of 22 and 24nt short RNAs was used as a threshold level to calculate the abundance of unambiguous 23nt siRNAs. This strategy was applied for siRNAs mapping the dsRNA target region, including the vector. The number of unambiguous siRNAs is indicated, as well as their percentage of the total 23nt read count. Sense siRNAs are shown in grey, antisense in black.

Fig.S7

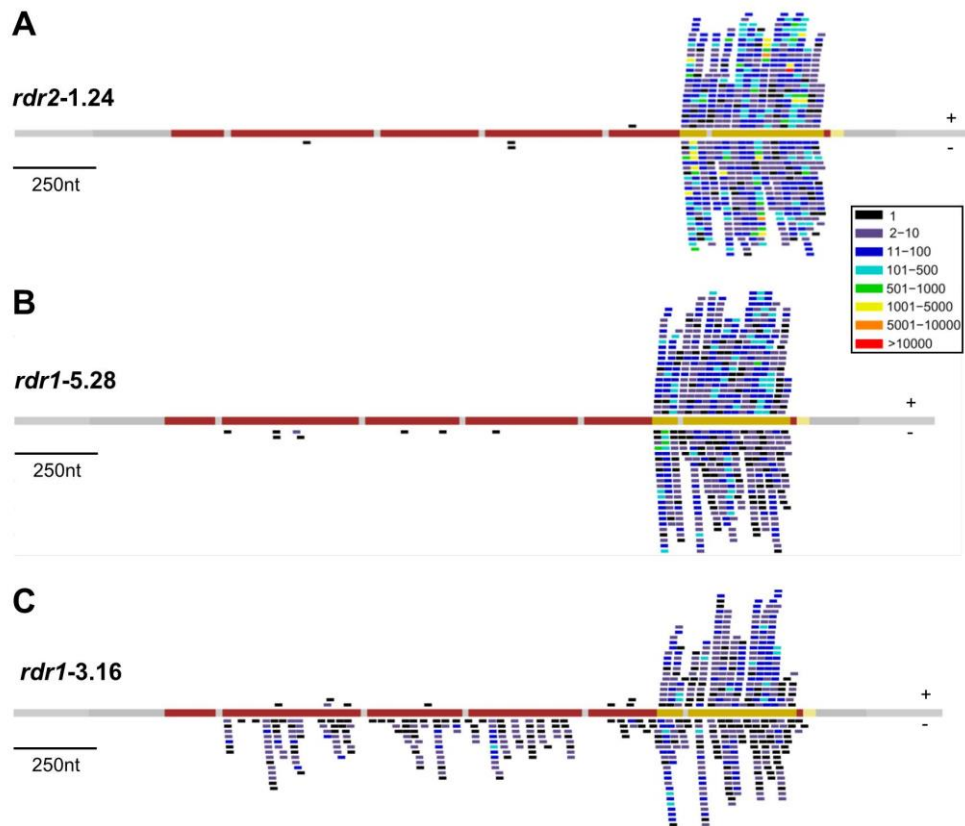


Fig. S7: Secondary siRNAs do no more accumulate in the *rdr2-1.24* mutant (**A**), and neither in the *rdr1-5.28* null mutant producing no unambiguous primary siRNAs (see text) (**B**). In contrast, secondary siRNAs are produced when primaries accumulated above the background level, as it is the case in *rdr1-3.16* (**C**). Distributions of 23nt siRNAs on the genomic sequence of the *ND169* gene are shown (red line: exons; grey: flanking intergenic regions and introns; light yellow: 3'UTR yellow: dsRNA target region). Individual siRNAs are represented as trait, the color indicating its abundance. siRNAs represented above and below the line are sense and antisense to the target mRNA, respectively. 3'U-siRNAs are not included in the figure.

Fig.S8

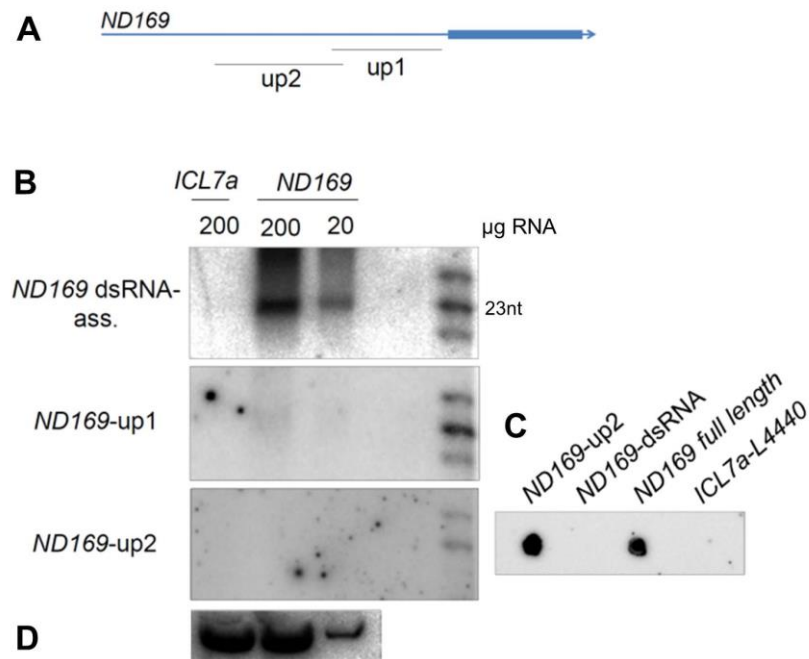


Fig. S8: Secondary siRNAs spreading beyond the dsRNA target region are hardly detectable. **(A)** Probes corresponding to the regions up1 and up2 used for northern blot analyses of siRNAs produced outside the *ND169* dsRNA region (represented as thick bar). **(B)** RNA of wild type *Paramecia* fed with *ICL7a* or *ND169* (WT-2) dsRNA were hybridized with a probe corresponding to the *ND169* dsRNA region (upper panel) and to positions upstream of the dsRNA region (middle and lower panel). A weak signal appears with one of the probes when 200 µg of total RNA were enriched for small RNAs using polyethylene glycol precipitation. No signal was obtained after prolonged exposure with the second probe (*ND169*-up2). **(C)** Functionality and specificity of this probe was confirmed by dot blot hybridization, producing signals with PCR products of the up2 region and the *ND169* full length sequence, but not with the dsRNA region and the unrelated *ICL7a* sequence. **(D)** Hybridization to glutamine tRNA served as loading control.

Fig.S9

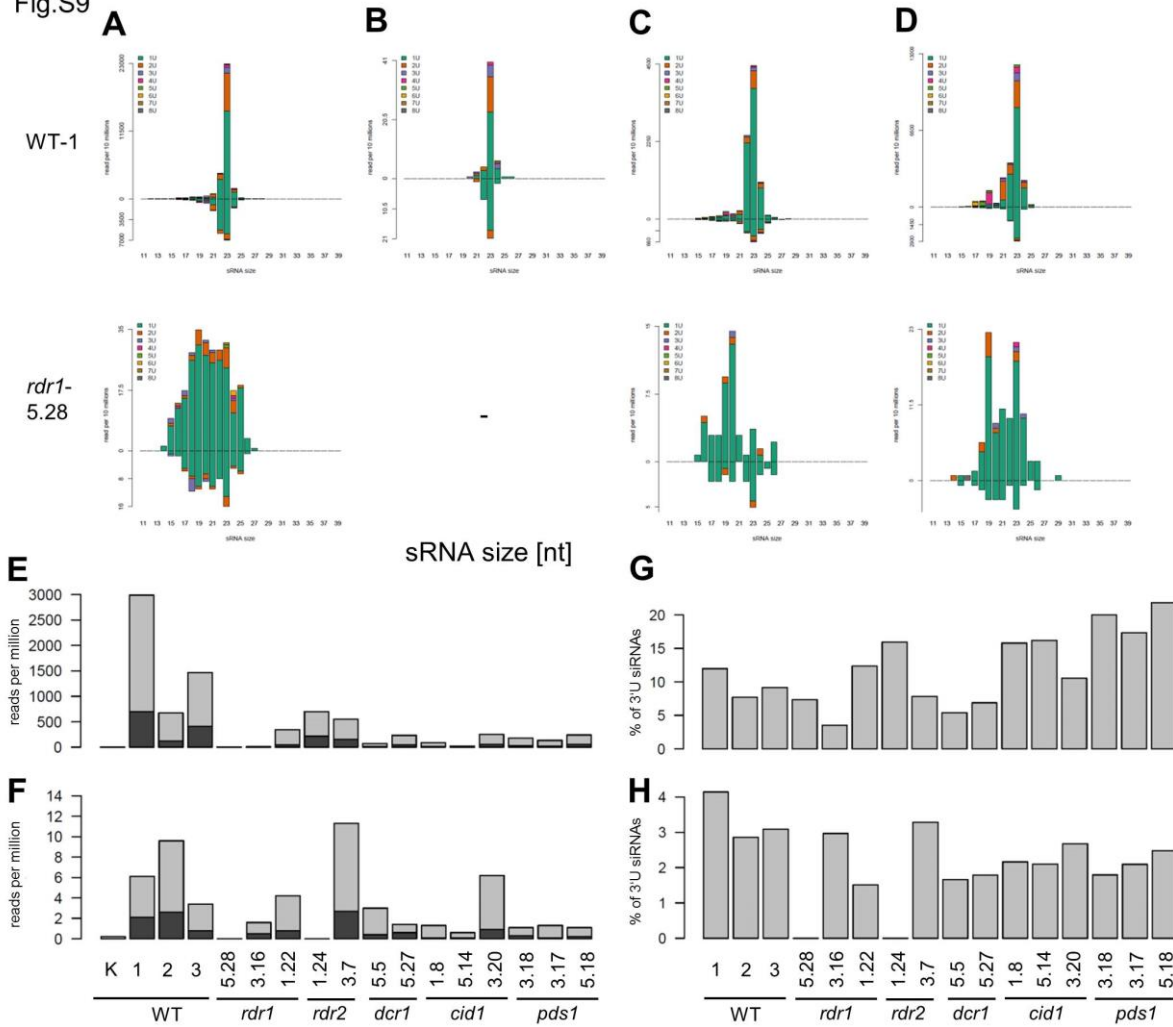


Fig. S9: Size and strand distribution of 3' uridylated siRNAs in wild type (WT-1) (top lane) and the *rdr1-5.28* mutant (bottom lane). **(A)** *ND169* dsRNA region, **(B)** outside dsRNA region, **(C)** 5' vector and **(D)** 3' vector. The number of non-genomic Us is indicated by colour. Sense reads are represented on top and antisense siRNAs on the bottom of the x-axis. **(E)** Normalized read count (in reads per million) of 3'U-siRNAs (all size fractions) mapping the *ND169* dsRNA region. Sense siRNAs are shown in grey, antisense in black. As no or only little background noise of RNAi-unrelated sRNAs was detectable within the 3'U short RNA fraction (judged by the low level of residual sRNAs in *rdr1-5.28*), background subtraction was not applied. **(F)** Normalized read count of 3'U-siRNAs (all size categories) mapping outside the *ND169* dsRNA region. The proportion of 3'U-siRNAs compared to total number of *ND169* primary siRNAs (considering background subtraction for non-3'U-siRNAs) is shown **(G)**, as well as the proportion of secondaries **(H)**. Note that the absolute level of remaining secondaries is very low in some mutants. K - *Klebsiella* control.

Fig.S10

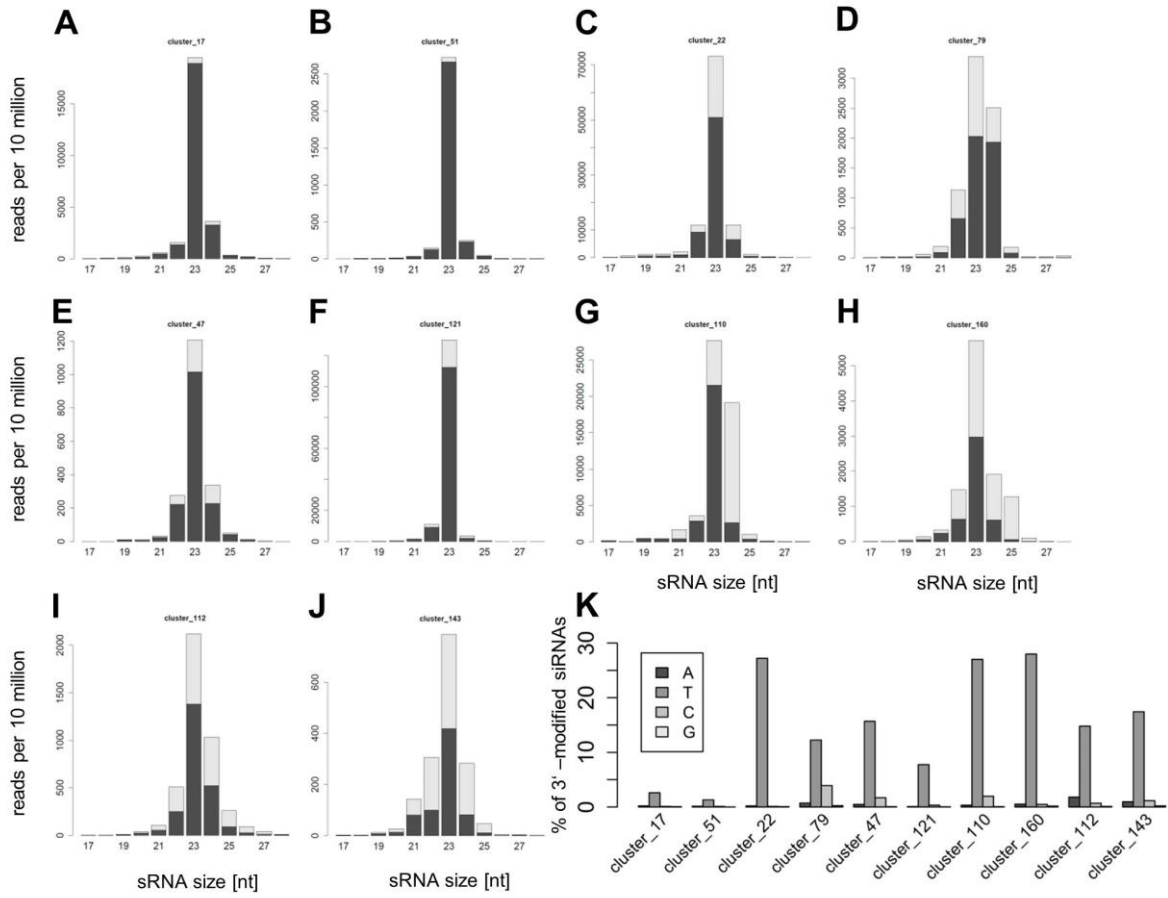


Fig. S10: (A-J) Size and strand distribution of endo-siRNA clusters shown in main Fig.5. Only wild type conditions (WT-1) are shown. SiRNAs mapping in sense orientation to poly(A)-transcripts produced from these loci (see table 2) are represented in grey, antisense in black. (K) Percentage of 3'-modified 23nt endo-siRNAs produced from selected endo-siRNA cluster regions with untemplated U, A, C or G.

Fig.S11

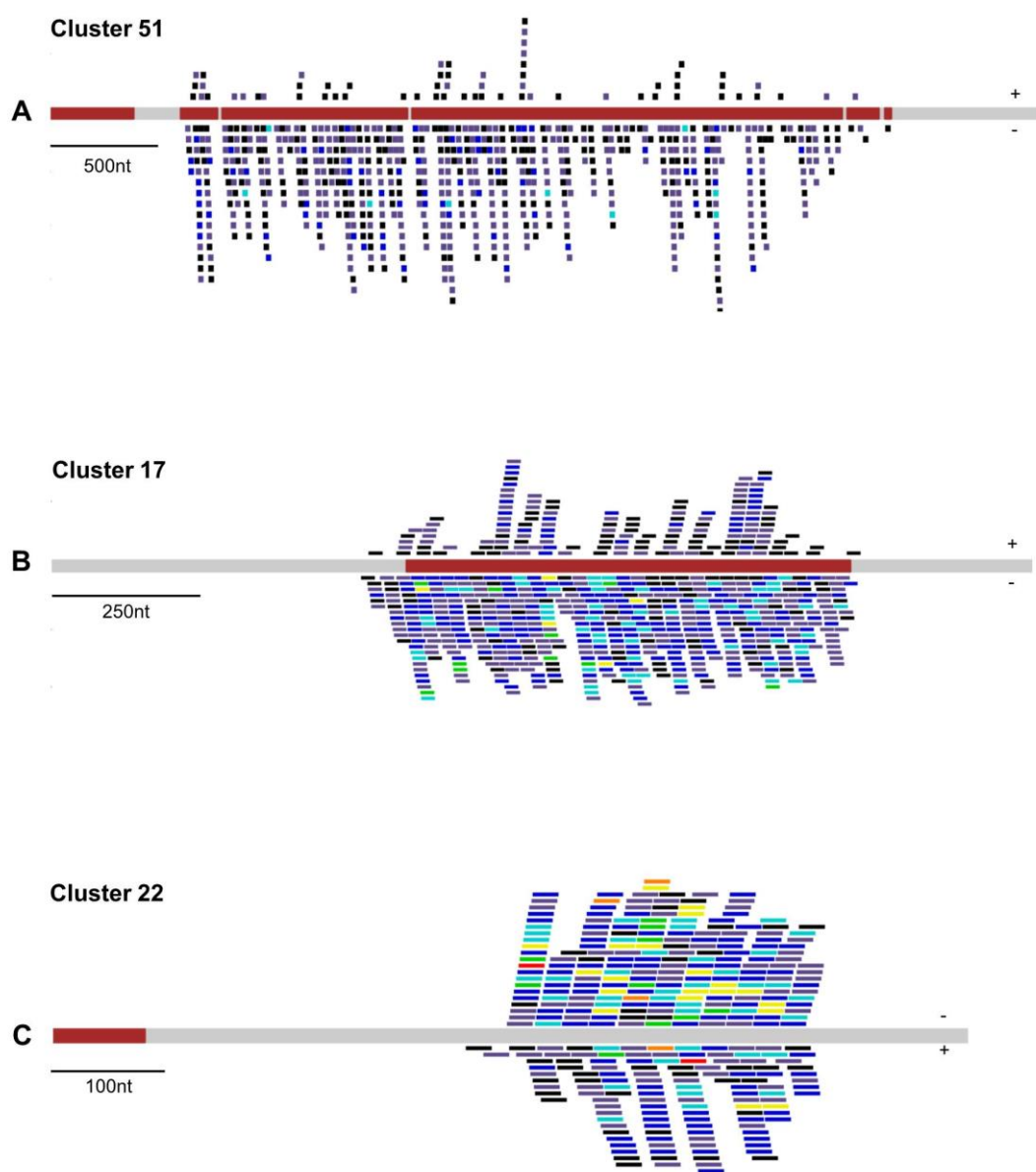


Fig.S11

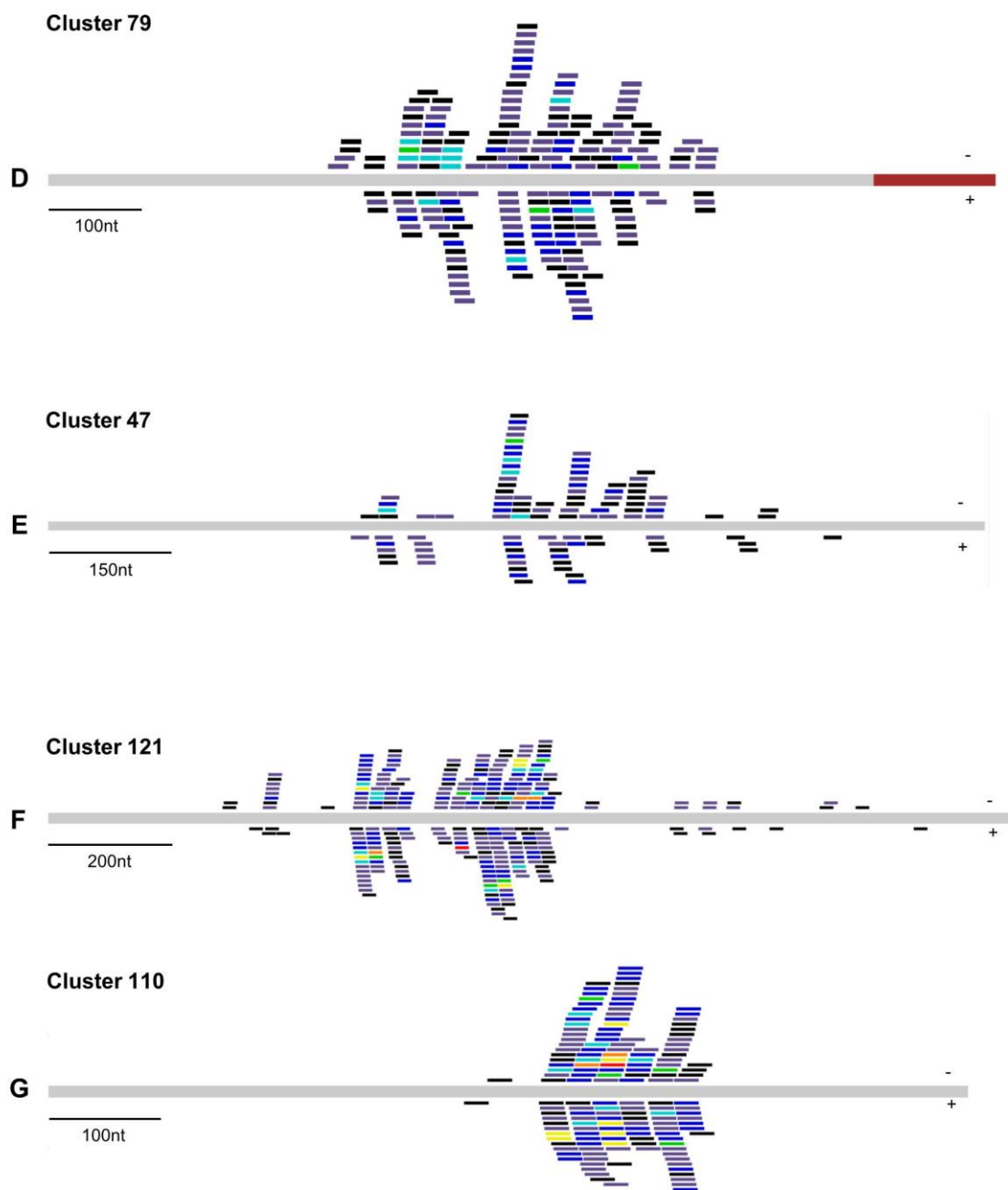


Fig.S11

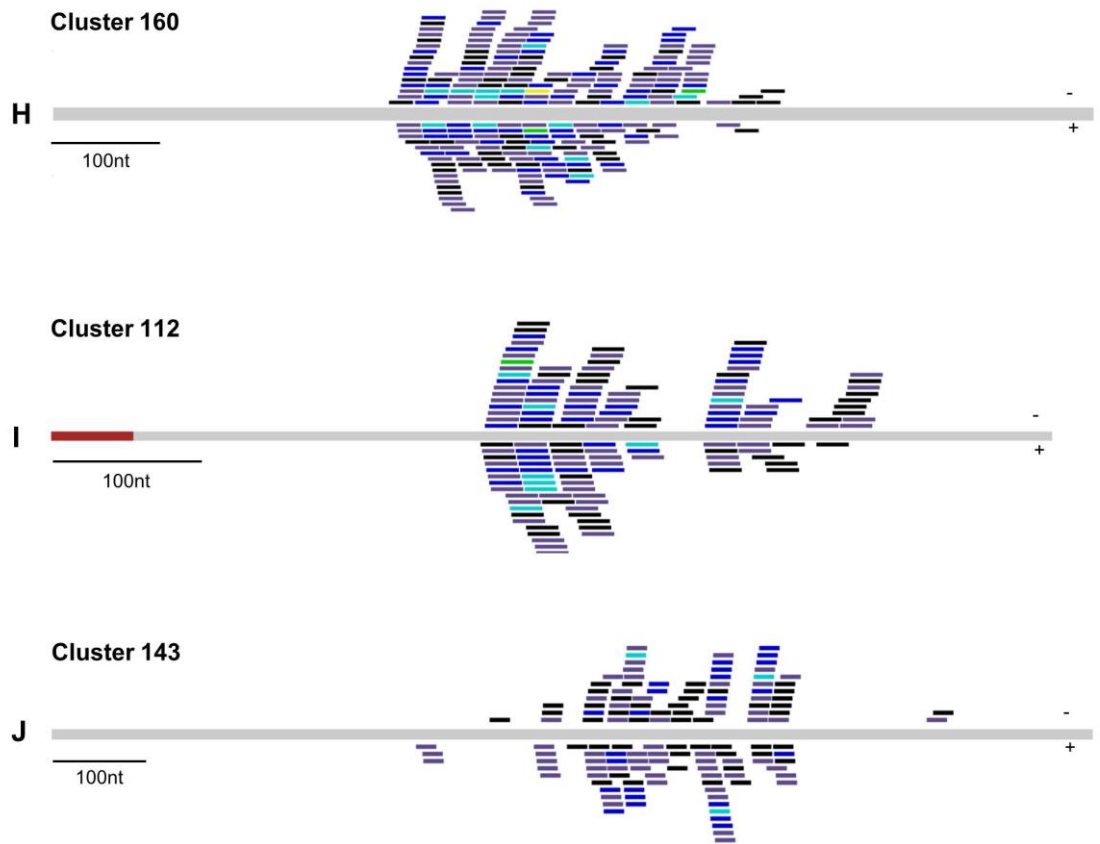


Fig. S11: Distribution of 23nt endogenous siRNAs accumulating from coding genes (A, B) and intergenic regions (C-J). Open reading frames of coding genes are represented in red, intergenic regions in grey. Strand orientations relative to the mRNA are given in A and B. Strand orientation relative to individual siRNAs are represented as trait, the color indicating its abundance. 3'U-siRNAs are not included in the figure.

Fig.S12

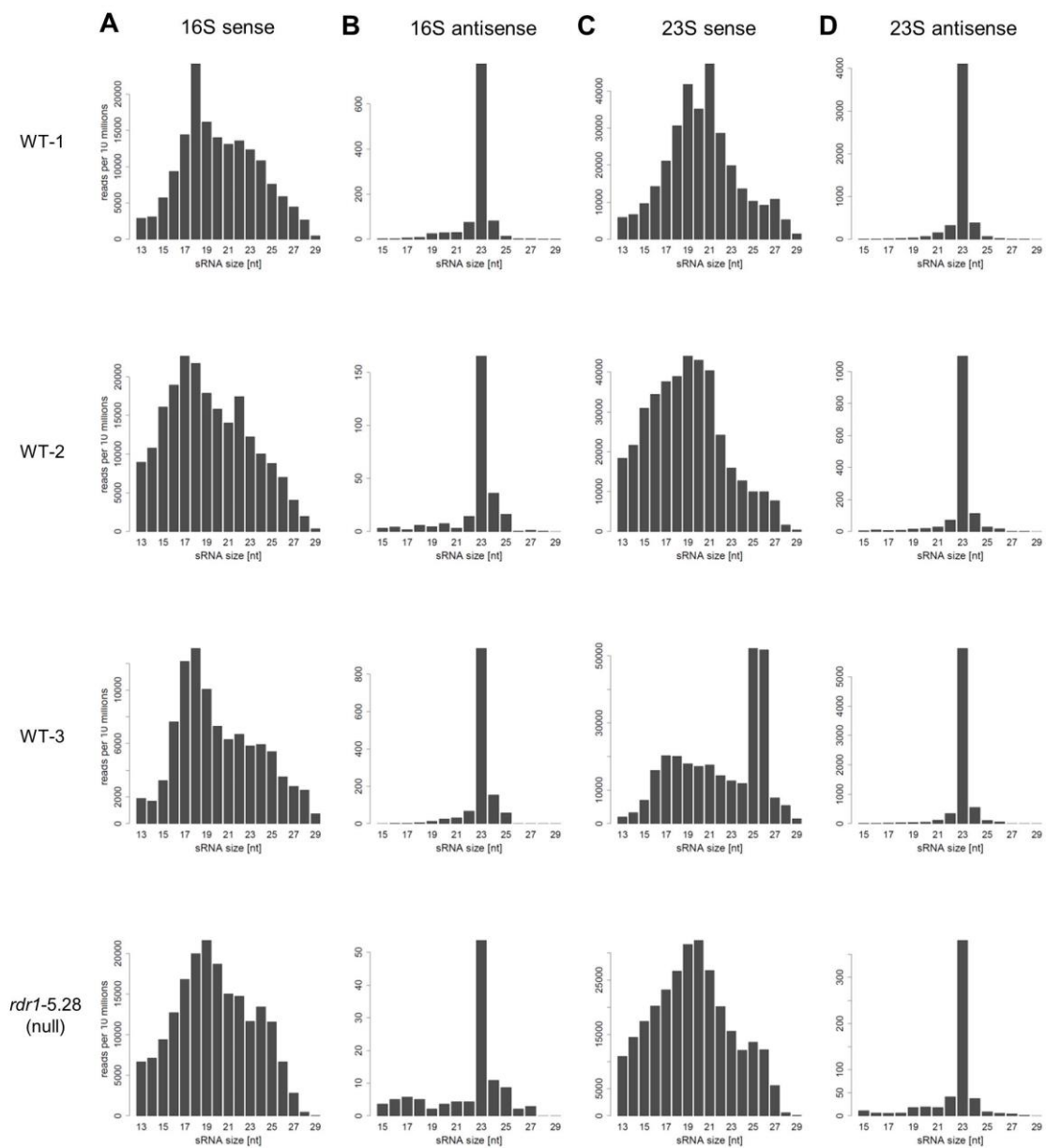


Fig.S12

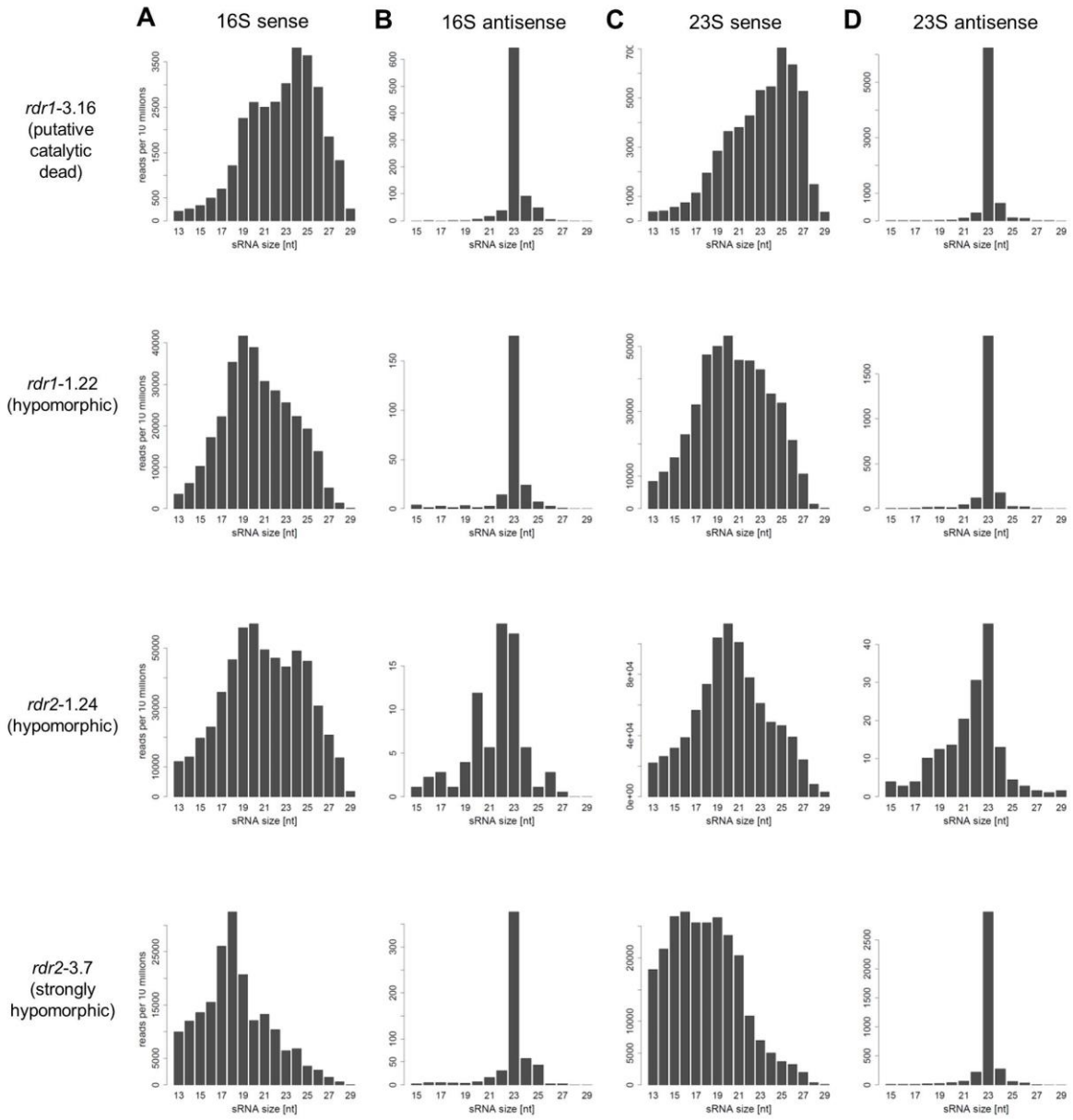


Fig.S12

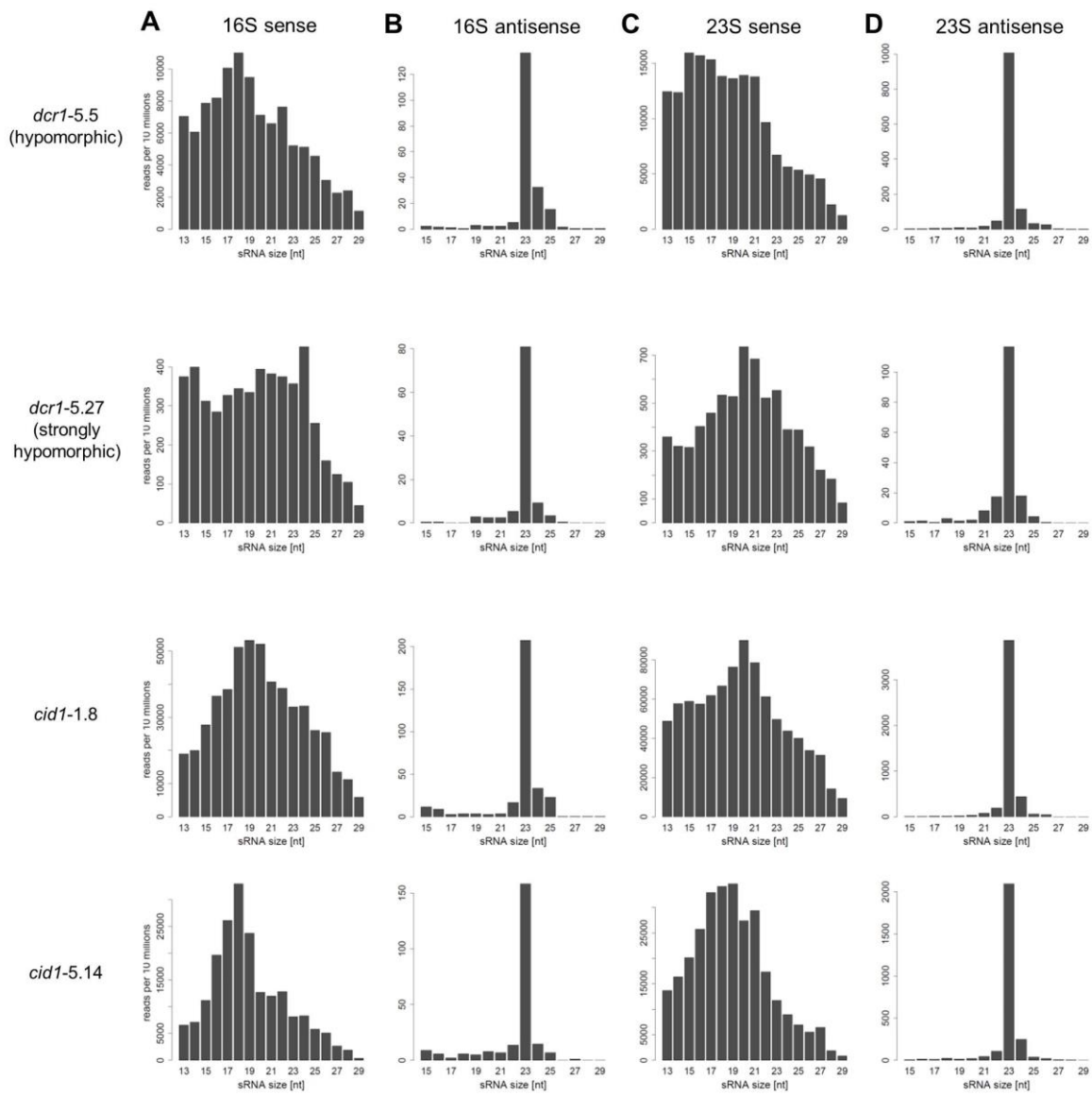


Fig.S12

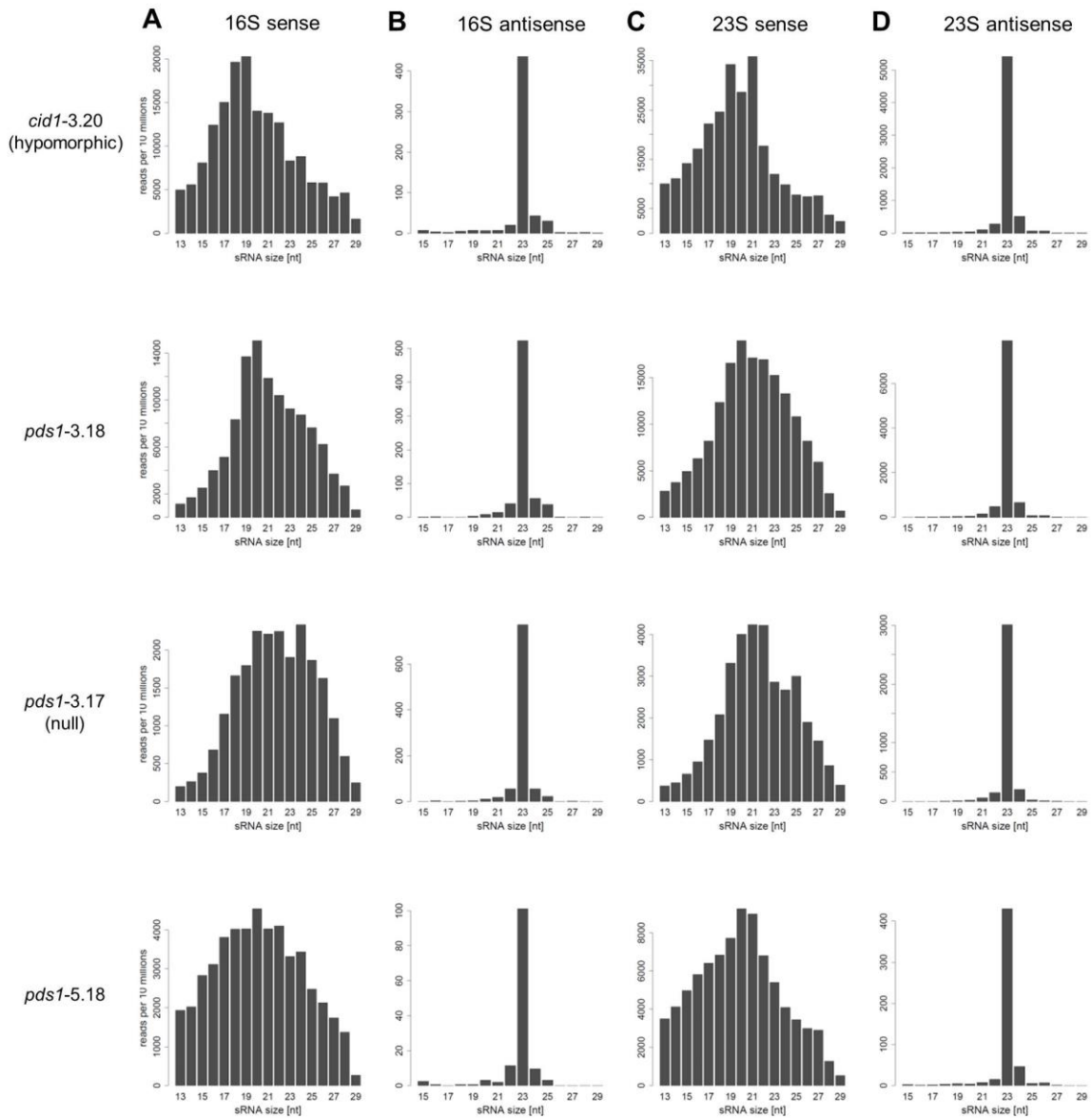


Fig. S12: Size and strand distribution of sRNAs in wild type (WT) and mutant samples aligned on bacterial ribosomal RNA. (A) 16S sense, (B) 16S antisense, (C) 23S sense, and (D) 23S antisense.

Table S1 : Outcome of small RNA sequencing performed from dsRNA feeding experiments targeting different genes (wild type cells). The total number of reads mapping without mismatch to the MAC genome, predicted exon-exon junctions and IESs (micronuclear-specific internal eliminated sequences) is shown. DsRNA feeding against *PTIWI08* and *PTIWI14*, as well as *ND169* and *ICL7a* in WT-1, respectively, were performed as double-knock down. Percentages of reads mapping the target genes are indicated. WT - wild type.

Sample	MAC + junctions + IESs		target gene
<i>RDR3</i>	2,826,366	▣	<i>RDR3</i> 20.8%
<i>PTIWI08/14</i>	5,252,716	▣	<i>PTIWI08</i> 27.9% <i>PTIWI14</i> 22.3%
WT-1	2,409,713	▣	<i>ND169</i> 35.6% <i>ICL7a</i> 19.6%
WT-2	1,542,713	▣	<i>ND169</i> 34.7%
WT-3	1,454,723	▣	<i>ND169</i> 51.7%

Table S2 : Categories of siRNAs mapping the RNAi target genes in different dsRNA feeding experiments. The dsRNA trigger consists of the 5' vector, the 3' vector and a gene specific sequence corresponding to a fragment of the target gene open reading frame (orf). siRNAs mapping the target gene orf outside the trigger dsRNA region represent secondary siRNAs. All categories are sorted by orientation (sense or antisense to the mRNA). (A) Number of 23nt siRNAs (in reads per 10 millions) corrected by the level of RNAi-unrelated small RNAs (defined by the average level of 22nt and 24nt siRNAs) (see text). (B) Number of 23nt siRNAs modified with 3' untemplated uridines (Us) (in reads per 10 millions). (C) Percentage of 23nt siRNAs modified with 3' untemplated Us. WT - wild type.

Sample	Target gene	trigger dsRNA region			outside dsRNA region			% secondary siRNAs	5' vector			3' vector		
		sense	antisense	% antisense	sense	antisense	% antisense		sense	antisense	% antisense	sense	antisense	% antisense
A														
siRNAs without 3' modification														
ICL7a	<i>ICL7a</i>	361926	572233	61%	0	900	100%	0.10%	54939	47357	46%	406	174138	99%
PTTW108/14	<i>PTTW108</i>	97383	647561	86%	6	714	99%	0.10%	32111	29061	47%	6015	10644	63%
PTTW108/14	<i>PTTW114</i>	147592	525098	78%	12	2042	99%	0.30%	34344	29359	46%	10065	11156	52%
RDR3	<i>RDR3</i>	153543	131343	46%	1	1015	99%	0.36%	25390	31215	55%	8481	6197	42%
WT-1	<i>ND1/69</i>	18630	201045	91%	40	1376	97%	0.64%	25373	137018	84%	39985	51364	56%
WT-1	<i>ICL7a</i>	68178	83321	54%	0	913	100%	0.60%	13120	35007	72%	853	47835	98%
B														
siRNAs with 3' modification														
ICL7a	<i>ICL7a</i>	21036	25822	55%	1	13	92%	0.03%	4440	3098	41%	111	6846	98%
PTTW108/14	<i>PTTW108</i>	3906	10611	73%	0	2	100%	0.01%	1425	654	31%	436	214	33%
PTTW108/14	<i>PTTW114</i>	2849	6038	68%	3	10	79%	0.15%	1475	662	31%	521	260	33%
RDR3	<i>RDR3</i>	1544	3077	67%	0	22	100%	0.46%	1055	767	42%	1177	429	27%
WT-1	<i>ND1/69</i>	22882	6995	23%	40	21	34%	0.20%	4441	658	13%	12051	2849	19%
WT-1	<i>ICL7a</i>	5329	6239	54%	2	6	79%	0.07%	4849	4642	49%	120	1963	94%
C														
Percentage of 3' untemplated siRNAs														
ICL7a	<i>ICL7a</i>	5%	4%	5%	100%	1%	2%	7%	6%	7%	7%	21%	4%	4%
PTTW108/14	<i>PTTW108</i>	4%	2%	2%	0%	0%	0%	4%	4%	2%	3%	7%	2%	4%
PTTW108/14	<i>PTTW114</i>	2%	1%	1%	19%	1%	1%	4%	4%	2%	3%	5%	2%	4%
RDR3	<i>RDR3</i>	1%	2%	2%	0%	2%	2%	4%	4%	2%	3%	12%	6%	10%
WT-1	<i>ND1/69</i>	55%	3%	12%	50%	1%	4%	15%	15%	0%	3%	23%	5%	14%
WT-1	<i>ICL7a</i>	7%	7%	7%	100%	1%	1%	27%	27%	12%	16%	12%	4%	4%

Table S3 : Total number of sequence reads obtained for each sample : *ND169* dsRNA feeding to wild type (WT) and mutants ; *RDR3*, *PTIWI08/14*, *ICL* dsRNA feeding to wild type. The numbers were used for normalization (reads per million). K - Klebsiella control (no dsRNA).

Sample	Total number of sequence reads	Ratio (reads/million)
WT-K	18,328,383	0.0546
WT-1	17,818,987	0.0561
WT-2	21,561,939	0.0464
WT-3	26,642,521	0.0375
<i>rdr1</i> -5.28	13,798,853	0.0725
<i>rdr1</i> -3.16	16,982,727	0.0589
<i>rdr1</i> -1.22	15,329,801	0.0652
<i>rdr2</i> -1.24	17,644,030	0.0567
<i>rdr2</i> -3.7	11,300,181	0.0885
<i>dcr1</i> -5.5	16,817,238	0.0595
<i>dcr1</i> -5.27	20,612,052	0.0485
<i>cid1</i> -1.8	17,825,354	0.0561
<i>cid1</i> -5.14	10,299,868	0.0971
<i>cid1</i> -3.20	17,852,587	0.0560
<i>pds1</i> -3.18	19,527,084	0.0512
<i>pds1</i> -3.17	9,686,281	0.1032
<i>pds1</i> -5.18	16,801,218	0.0595
<i>RDR3</i>	8,367,426	0.1195
<i>PTIWI08/14</i>	10,597,689	0.0944
<i>ICL7a</i>	8,679,846	0.1152

Table S4 : Normalized numbers of siRNAs (in reads per 10 million) in wild type (WT) or mutant samples aligning on the *NDI69* dsRNA (5' vector, 3' vector and trigger dsRNA region) or outside the *NDI69* dsRNA region. Categories are sorted by orientation (sense or antisense to the mRNA). (A) Number of 23nt-siRNAs corrected by the level of RNAi-unrelated small RNAs (defined by the average level of 22nt and 24nt sRNAs) (see text). (B) Number of 23nt-siRNAs modified with 3' untemplated uridines (Us) (in reads per 10 millions). No correction was applied, as the level of 3'U-small RNAs of other sizes was found to be low in RNAi mutants. (C) Percentage of 23nt-siRNAs modified with 3' untemplated Us (calculation based on the numbers given in A and B). K - Klebsiella control (no dsRNA).

Sample	sample no./ allele	trigger dsRNA region			outside dsRNA region			% secondary siRNAs	5' vector			3' vector			
		sense	antisense	% antisense	sense	antisense	% antisense		sense	antisense	% antisense	sense	antisense	% antisense	
A siRNAs without 3' modification	WT-K	15	61	80%	0	4	100%	5.4%	9	127	93%	10	53	84%	
	WT-1	18630	201045	91%	40	1376	97%	0.6%	25373	137018	84%	39985	51364	56%	
	WT-2	14477	66476	82%	93	3170	97%	3.9%	7685	36256	82%	13821	16141	53%	
	WT-3	12899	133210	91%	25	1036	97%	0.7%	18461	126193	87%	24569	16852	40%	
	<i>rdr1</i>	5.28	0	594	100%	0	7	100%	1.1%	57	0	0%	0	807	100%
		3.16	1555	838	35%	3	517	99%	17.8%	261	1262	82%	213	1212	85%
		1.22	14990	9255	38%	35	2680	98%	10.1%	2099	5354	71%	882	3000	77%
	<i>rdr2</i>	1.24	20908	16044	43%	0	1	100%	x	928	30145	97%	880	3376	79%
		3.7	18305	46581	71%	191	3142	94%	4.9%	7606	25042	76%	11087	23893	68%
	<i>der1</i>	5.5	8143	4670	36%	14	1782	99%	12.3%	865	2750	76%	1750	2483	58%
		5.27	18631	12827	40%	14	784	98%	2.5%	1754	6013	77%	5589	12074	68%
	<i>cid1</i>	1.8	2636	2049	43%	8	551	98%	10.7%	195	1676	89%	0	831	100%
		5.14	0	1195	100%	0	272	100%	18.5%	78	205	72%	497	1312	72%
		3.20	10233	11127	52%	35	2224	98%	9.6%	1120	7222	86%	1159	4622	79%
	<i>pds1</i>	3.18	4569	2565	35%	6	610	99%	8.0%	744	3190	81%	911	1527	62%
3.17		3318	3008	47%	12	615	98%	9.0%	770	4157	84%	1057	2931	73%	
5.18		4301	4302	50%	4	417	99%	4.7%	741	4600	86%	1211	3699	75%	

Sample	sample no./ allele	trigger dsRNA region			outside dsRNA region			% secondary siRNAs	5' vector			3' vector			
		sense	antisense	% antisense	sense	antisense	% antisense		sense	antisense	% antisense	sense	antisense	% antisense	
B siRNAs with 3' modification	WT-K	9	3	23%	2	0	0%	12.0%	3	0	0%	11	1	9%	
	WT-1	22882	6995	23%	40	21	34%	0.2%	4441	658	13%	12051	2849	19%	
	WT-2	5554	1198	18%	70	26	28%	1.4%	1262	178	12%	1107	508	31%	
	WT-3	10615	4072	28%	26	8	22%	0.2%	2715	306	10%	7885	1837	19%	
	<i>rdr1</i>	5.28	31	16	34%	0	0	x	x	4	5	58%	21	4	17%
		3.16	71	16	19%	11	5	33%	15.4%	14	7	33%	44	23	35%
		1.22	2980	438	13%	34	8	19%	1.2%	160	47	23%	421	110	21%
	<i>rdr2</i>	1.24	4846	2156	31%	0	0	x	x	248	71	22%	17361	362	2%
		3.7	3966	1542	28%	86	27	24%	2.0%	1271	306	19%	1488	493	25%
	<i>der1</i>	5.5	573	155	21%	26	4	14%	4.0%	64	29	31%	121	43	26%
		5.27	1888	434	19%	8	6	43%	0.6%	64	32	33%	658	125	16%
	<i>cid1</i>	1.8	775	103	12%	12	1	5%	1.4%	96	15	13%	173	41	19%
		5.14	191	40	17%	5	1	17%	2.5%	25	14	35%	40	11	21%
		3.20	1941	574	23%	53	9	14%	2.4%	232	86	27%	561	151	21%
	<i>pds1</i>	3.18	1485	296	17%	8	3	27%	0.6%	22	10	32%	547	78	13%
3.17		1115	211	16%	12	1	8%	0.1%	49	23	32%	586	98	14%	
5.18		1866	534	22%	9	2	17%	0.4%	247	108	30%	1013	145	13%	

Sample	sample no./ allele	trigger dsRNA region			outside dsRNA region			5' vector			3' vector		
		sense	antisense	total	sense	antisense	total	sense	antisense	total	sense	antisense	total
C Percentage of 3' uridylated siRNAs	WT-K	38%	4%	14%	100%	0%	27%	23%	0%	2%	53%	2%	16%
	WT-1	55%	3%	12%	50%	1%	4%	15%	0%	3%	23%	5%	14%
	WT-2	28%	2%	8%	43%	1%	3%	14%	0%	3%	7%	3%	5%
	WT-3	45%	3%	9%	51%	1%	3%	13%	0%	2%	24%	10%	19%
<i>rdr1</i>	5.28	100%	3%	7%	x	0%	0%	6%	100%	13%	100%	1%	3%
	3.16	4%	2%	4%	78%	1%	3%	5%	1%	1%	17%	2%	4%
	1.22	17%	5%	12%	49%	0%	2%	7%	1%	3%	32%	4%	12%
<i>rdr2</i>	1.24	19%	12%	16%	x	0%	0%	21%	0%	1%	95%	10%	81%
	3.7	18%	3%	8%	31%	1%	3%	14%	1%	5%	12%	2%	5%
<i>der1</i>	5.5	7%	3%	5%	65%	0%	2%	7%	1%	3%	6%	2%	4%
	5.27	9%	3%	7%	38%	1%	2%	4%	1%	1%	11%	1%	4%
<i>cid1</i>	1.8	23%	5%	16%	60%	0%	2%	33%	1%	6%	100%	5%	20%
	5.14	100%	3%	16%	100%	0%	2%	25%	6%	12%	7%	1%	3%
	3.20	16%	5%	11%	60%	0%	3%	17%	1%	4%	33%	3%	11%
<i>pds1</i>	3.18	25%	10%	20%	57%	1%	2%	3%	0%	1%	38%	5%	20%
	3.17	25%	7%	17%	50%	0%	2%	6%	1%	1%	36%	3%	15%
	5.18	30%	11%	22%	68%	0%	2%	25%	2%	6%	46%	4%	19%

