



HAL
open science

Réponse des agents non codants du génome – éléments transposables et petits ARN – à un événement d’allopolyploïdie : le génome du colza (*Brassica napus*) comme modèle d’étude

Paulina Martinez Palacios

► **To cite this version:**

Paulina Martinez Palacios. Réponse des agents non codants du génome – éléments transposables et petits ARN – à un événement d’allopolyploïdie : le génome du colza (*Brassica napus*) comme modèle d’étude. Sciences agricoles. Université Paris Sud - Paris XI, 2014. Français. ⟨NNT : 2014PA112055⟩. ⟨tel-01124300⟩

HAL Id: tel-01124300

<https://theses.hal.science/tel-01124300v1>

Submitted on 6 Mar 2015

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

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



HAL Authorization

UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : SCIENCES DU VÉGÉTAL

UMR de Génétique Végétale, Le Moulon

DISCIPLINE : BIOLOGIE

THÈSE DE DOCTORAT

Soutenance réalisée le vendredi 28 Mars 2014

par

Paulina MARTINEZ PALACIOS

Réponse des agents non codants du génome
- éléments transposables et petits ARN - à un
événement d'allopolyploidie : le génome du colza
(*Brassica napus*) comme modèle d'étude

Direction de thèse : Karine ALIX (MC, AgroParisTech)

Co-direction de thèse : Johann JOETS (IR, INRA)

Composition du jury :

Directrice de thèse :

Rapporteurs :

Examineurs :

Président du jury :

Membre invité :

Karine ALIX
Philippe LASHERMES
Christian PARISOD
Marie MIROUZE
Armel SALMON
Pierre CAPY
Johann JOETS

MC (AgroParisTech, UMR GV Le Moulon)
DR (IRD Montpellier)
Group leader P.D. (Université Neuchâtel)
CR (IRD Montpellier)
MC (Université Rennes I)
Pr (Université Paris-Sud)
IR (INRA, UMR GV Le Moulon)

Remerciements

Je voudrais remercier tout d'abord Philippe Lashermes et Christian Parisod pour avoir accepté la tâche de rapporter cette thèse. Je remercie également Marie Mirouze, Pierre Capy et Armel Salmon pour avoir accepté de faire partie des membres du jury.

Je tiens à remercier particulièrement ma directrice de thèse, Karine Alix. Merci énormément Karine, pour tout ce que tu as fait pour m'aider tout au long de la thèse, dans l'aspect professionnel mais aussi personnel. Merci d'avoir été là dans les bons moments mais aussi dans les mauvais. J'ai eu la chance de te croiser dans ma vie, d'avoir toujours ton soutien et ton amitié, et de plus j'étais aussi accueillie chez toi par ta belle famille. Merci Eric, Audrey, Laura.

Je remercie ensuite mon co-encadrant Johann Joets pour m'avoir aidé au cours de cette thèse dans l'aspect bioinformatique, merci pour ton temps et ta patience.

Merci encore Karine et Johann pour votre aide, votre disponibilité, vos encouragements et votre bonne humeur, je me suis vraiment bien sentie au cours de ces trois années de thèse !!

Je voudrais aussi remercier Philippe Brabant pour m'aider à résoudre tous mes problèmes statistiques malgré son agenda toujours débordé !

Je remercie également Dominique de Vienne pour m'avoir accueillie à l'UMR de Génétique Végétal du Moulon.

Je remercie les membres de mon comité de thèse : Eric Jenczewski, Martin Crespi, Jérôme Gouzy, Clémentine Vitte et Christine Lelandais pour avoir pris le temps de venir discuter avec moi sur mon travail et pour m'avoir donné des conseils précieux pour le bon déroulement de ma thèse !

Je remercie ensuite toutes les personnes qui m'ont aidée tout au long de la réalisation de la thèse et particulièrement celles qui m'ont aidée au laboratoire : Agnès, merci beaucoup pour ta disponibilité et pour m'aider avec toutes les extractions d'ARN ! Marpie, merci énormément pour ton aide avec la qPCR, pour m'aider à mettre en place le dispositif « Gollum » ! Marion Tapie, merci pour t'avoir autant investi dans mon projet au cours de ton stage de M1 ! Merci aussi à tous les collègues qui ont rendu mon passage au labo plus agréable, certain d'entre eux m'ont aussi donné un coup de main pour la réalisation des manips ! Merci Carine, Martine, Valérie, Hélène, Marlène, Thierry, Fabrice, Xavier, Matthieu, Delphine....

Je remercie aussi à Anne-Marie Chèvre et son équipe de l'INRA de Rennes, pour la production des colzas synthétiques, matériel sur lequel repose ma thèse entière !

Cette thèse a été financée par une bourse du Ministère de la Recherche et l'Enseignement ; je les en remercie.

Merci également au DGAP (Département de Génétique et Amélioration de Plantes), pour organiser les Journées Jeunes Chercheurs, ainsi qu'aux doctorants avec qui j'ai eu l'occasion de discuter au cours de ma thèse.

Merci aux membres de l'équipe GEAR pour avoir rendu ma thèse très agréable avec toutes nos petites réunions, merci aussi pour votre aide et votre soutien : Maud, Catherine, Dome, Philippe, Clémentine, Jean-Tristan, Margaux-Alison, Sara, Frank, Martine, Hélène, Agnès, Karine, Philippe.

Merci aussi aux membres de l'équipe ABI, pour leur aide au quotidien dans le « débogage » de mes scripts ! Pierre, Frank « Bubi », Yannick, Fabio, Gauthier.

Merci aux personnes du secrétariat pour m'aider à résoudre toute sorte de problèmes administratifs, d'imprimante, de scanner, la poste...: Rozenn, Valérie, Cyril, Khalid.

Merci à toutes les personnes qui m'ont aussi bien entouré au Moulon, et qu'ont rendu mon passage au labo très drôle,... Pedrito, Bub, Yannick, Fabio, Nico, Stef, Sara, Charlotte, Beatriz, Christophe, Pierre Rivière, Margaux-Alison, Héloïse, Mariangela, Aude, Jean-Tristan, Carine, Marlene, Helène, Jérôme ... Merci pour votre bonne humeur, pour les moments de détente et de rigolade !!! Je vous remercie du fond du cœur pour votre amitié, vos sourires !

Un merci spécial à mes « colocs » de bureau : Véro, Stéf, Aurélie, Sara...

A mes copines de gym : Stef, Sara, Charlotte, et un peu Héloïse ;)

Merci à tous les gens qui font du Moulon, un endroit aussi agréable où j'ai pu profiter pendant un peu plus de trois ans ! Merci spécialement pour la bonne humeur, et la bonne ambiance !

Je remercie enfin à tous mes amis en dehors du labo...pour leurs encouragements et leur soutien : Marcela, Silvana, Chiquitina, Alicia, Amélie, Maryam, Loreto, Serena, Marina, Pava, Carlos, Omar, Pancho.... Et a mes deux cousines et amies Clau et Caro !

Un immense merci à ma famille, pour me soutenir dans toutes les choses que j'ai décidées de faire, de croire toujours en moi !!! Mes chers parents « mis gordos », mes deux sœurs Luz et Raque et tous leurs descendants :) Los amo con todo mi corazon, no puede haber caido en una familia mas bonita, gracias !

Gracias « gorditos » por confiar siempre en mi, por todo su amor y su apoyo ...

Gracias por haberme dado alas para aprender a volar ! ... Gracias por estar siempre ahi para levantarme cuando me he caido ...

Y gracias por empujarme a volar otra vez ...

Sigles et Abréviations

3' ou 5'UTR : *3' ou 5' untranslated region*

ADN : Acide désoxyribonucléique

ADNc : ADN complémentaire d'un ARN

ADNg : ADN génomique

AFLP : *amplified fragment length polymorphism*

ARN : Acide ribonucléique

ARNm : ARN messenger

ARNr : ARN ribosomique

ARNt : ARN de transfert

BLAST : *basic local alignment search tool*

CAPS : *cleaved amplified polymorphic sequence*

EST : *expressed sequence tag*

ET : élément transposable

FISH : *Fluorescent in situ hybridization*

GISH : *Genomic in situ hybridization*

Hc-siRNA : *heterochromatic siRNA*

LINE : *long interspersed nuclear element*

LTR : *long terminal repeat*

Ma : Millions d'années

miARN: micro ARN

MITE : *miniature inverted-repeat transposable element*

MPV : *mid parent value*

Nat-siRNA : *natutal antisense small interfering RNA*

NOR : *nucleolus organizer region*

nt : nucléotide

pb : paires de bases

PCR : *polymerase chain reaction*

PCRq : PCR quantitative

rasiRNA : *repeat-associated small interfering RNA*

RFLP : *restriction fragment length polymorphism*

RT : *reverse transcriptase*

SINE : *short interspersed nuclear element*

sRNA: *small RNA*

siRNA : *small interfering RNA*

SSAP : *sequence-specific amplification polymorphism*

tasiRNA : *trans-acting small interfering RNA*

TIR : *terminal inverted repeat*

TRIM : *terminal-repeat retrotransposon in miniature*

WGD : *whole genome duplication*

TABLE DE MATIERES

Table de matières

Remerciements	i
Sigles et abréviations	ii
Introduction Générale	1
Chapitre 1 - L'importance de la polyploïdie dans l'évolution des angiospermes	3
1.1. La polyploïdie	3
1.2. Modes de formation des polyploïdes	4
1.2.1. Formation des polyploïdes par la voie de gamètes non réduits	4
1.2.2. Formation des polyploïdes par la voie de doublement somatique	5
1.3. Récurrence de la polyploïdie chez les plantes	6
1.3.1. La paléopolyploïdie	6
1.3.1. Les polyploïdes actuels	9
1.3.1. Les polyploïdes synthétiques	10
1.4. Conséquences évolutives de la polyploïdie	11
1.5. Polyploïdes versus diploïdes : avantages des polyploïdes sur les diploïdes	12
1.5.1. Hétérosis ou vigueur hybride	12
1.5.2. Redondance génétique : conséquences de la duplication des gènes et devenir des gènes dupliqués	13
1.6. Dynamique des génomes allopolyploïdes	15
1.6.1. Modifications structurales du génome	15
1.6.1.a. Réarrangements chromosomiques : délétions, translocations (réciproques ou non réciproques), recombinaison homéologue	15
1.6.1.b. Activation des éléments transposables	17
1.6.2. Modifications de l'expression des gènes	18
1.6.2.a. Modifications au niveau du transcriptome chez des hybrides et des allopolyploïdes	18
1.6.2.b. Modifications au niveau du protéome chez des hybrides et des allopolyploïdes	21

1.6.3. Modifications épigénétiques	22
1.6.3.a. Modifications au niveau de la méthylation de l'ADN	22
1.6.2.b. Modifications transcriptionnelles et post-transcriptionnelles par l'intermédiaire des petits ARNs non codants.....	23
1.7. Le genre <i>Brassica</i> comme modèle pour l'étude de l'allopolyploïdie	23
1.7.1. Mon modèle d'étude : le colza (<i>Brassica napus</i>)	24
1.7.2. Mon matériel d'étude : des colzas néo-synthétisés	26

Chapitre 2 - L'allopolyploïdie n'a qu'un impact modéré sur l'activation transpositionnelle des éléments transposables..... 29

2.1. Les éléments transposables, un composant majeur du génome végétal	29
2.2. Impact de l'allopolyploïdie sur l'activation des éléments transposables chez le colza : le cas du transposon <i>Bot1</i>	31
2.3. Sarilar et al. (2013), <i>New Phytologist</i> 198: 593-604.....	33

Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids

Chapitre 3 - L'allopolyploïdie entraîne une réponse immédiate et transitoire des populations de petits ARN non codants 75

3.1. La découverte de l'interférence par ARN et du monde des petits ARN non codants	75
3.2. Les petits ARN non codants : définition et classification	79
3.2.1. Les micro ARN (miRNAs) des plantes	81
3.2.1.a. Biogénèse des micro ARN	82
3.2.1.b. Mode d'action des micro ARN	83
3.2.1.c. Caractéristiques évolutives des micro ARN	85
3.2.2. Les autres petits ARN en épingle à cheveux (autres hpRNAs).....	86
3.2.3. Les siRNAs hétérochromatiques	86
3.2.3.a. Biogénèse des siRNAs hétérochromatiques	87
3.2.3.b. Mode d'action des siRNAs hétérochromatiques	88
3.2.3.c. Conservation et évolution des siRNAs hétérochromatiques	89

3.2.4. Les siRNAs secondaires	90
3.2.4.a. Biogénèse des siRNAs secondaires	90
3.2.4.b. Fonction des siRNAs secondaires	91
3.2.4.c. Le cas des siRNAs secondaires liés aux gènes TAS	91
3.2.5. Les siRNAs issus de transcrits natifs antisens (nat-siRNAs)	92
3.3. Les petits ARN non codants d'origine exogène : cas du système de défense antiviral ...	93
3.4. Impact de l'hybridation sur l'expression des petits ARN non codants	94
3.4.1. Mobilisation de populations de petits ARN lors de la formation des gamètes puis au moment de la reproduction	94
3.4.1.a. Les petits ARN du gamétophyte mâle (le grain de pollen)	95
3.4.1.b. Les petits ARN du gamétophyte femelle (le sac embryonnaire)	96
3.4.1.c. Le cas des siRNAs secondaires liés aux gènes TAS	97
3.4.2. Les petits ARN non codants apparaissent impliqués dans le phénomène d'hétérosis	98
3.4.2.a. Une mobilisation de micro ARN et de leurs gènes cibles en réponse à l'hybridation.....	99
3.4.2.b. Les populations de siRNAs de 24-nt présentent des changements en séquence et en abondance après hybridation	100
3.4.3. Les petits ARN non codants à l'origine de l'incompatibilité hybride	100
3.5. Impact de l'allopolyploïdie sur l'expression des petits ARN non codants	102
3.5.1. Des changements d'expression de petits ARN non codants chez les systèmes allopolyploïdes <i>Arabidopsis</i> et blé	102
3.5.2. Martinez Palacios et al. en préparation pour soumission dans <i>Molecular Biology and Evolution</i>	105
<u>Highlighting the immediate but transient response of specific small RNA populations to allopolyploidy using resynthesized <i>Brassica napus</i> allotetraploids</u>	
Conclusions et Perspectives	199
Bibliographie	205

INTRODUCTION GENERALE

I NTRODUCTION

La polyploïdie, ou duplication du génome complet (*whole genome duplication*, WGD) est un processus majeur d'évolution du génome des eucaryotes et en particulier de celui des plantes : toutes les plantes à graines et à fleurs sont polyploïdes ou ont connu au moins un événement de polyploïdisation au cours de leur histoire évolutive (Jiao et al. 2011 ; Figure 0.1.). Ce succès s'explique par le fait qu'un génome polyploïde n'est pas la simple addition des génomes de ses espèces parentales mais que la polyploïdie induit des modifications génomiques majeures, à l'origine d'une nouvelle variabilité génétique utile et nécessaire à toute sélection adaptative et aux étapes de spéciation. Les polyploïdes nouvellement formés, et surtout ceux d'origine hybride – que l'on appelle des allopolyploïdes – peuvent ainsi présenter des valeurs sélectives supérieures à celles de leurs parents, leur conférant une capacité à coloniser de nouvelles niches et à évoluer en tant que nouvelle espèce (Hegarty & Hiscock 2008).



Figure 0.1 : Les angiospermes ont été marquées par de nombreux événements de polyploïdie au cours de leur histoire évolutive. On distingue des espèces dites paléopolyploïdes, où seules des traces des événements passés de polyploïdie subsistent dans leur génome ; c'est le cas par exemple de *Amborella trichopoda* (A), *Arabidopsis thaliana* (B), la vigne (C), le maïs (D), et le riz (E). De nombreuses autres espèces d'angiospermes présentent toujours les différentes copies de génome issues d'événements plus ou moins récents de polyploïdie ; la canne à sucre (8x, F), le blé tendre (6x, H), le blé dur (4x, H), le cotonnier (4x, J), le fraisier (8x, K), le tabac (4x, N), le caféier (4x, O), le bananier (3x, P), et le colza (4x, Q) sont originaires d'événements de polyploïdie plus ou moins anciens, alors que d'autres espèces datent de moins de 150 ans comme *Spartina anglica* (12x, G), *Senecio cambrensis* (4x, L) et *Tragopogon miscellus* (4x, M).

Depuis une dizaine d'années, la polyplœdie connaît un important renouveau auprès de la communauté scientifique qui cherche à identifier les mécanismes moléculaires responsables des changements génomiques accompagnant la formation et l'évolution d'un génome polyplœdie, pour comprendre le succès évolutif de la polyplœdie et plus précisément de l'allopolyplœdie chez les végétaux.

Mon projet de thèse s'inscrit dans ce cadre scientifique. En prenant comme modèle les *Brassica*, dont le génome a connu des cycles récurrents de polyplœdie au cours de son évolution, les plus récents correspondant notamment à la formation du colza (*B. napus*) qui est une espèce allotétraploïde, j'ai contribué à une meilleure connaissance des mécanismes impliqués dans la régulation du génome allopolyplœdie nouvellement formé. Le matériel végétal original que j'ai utilisé pour mes travaux a consisté en des colzas néo-synthétisés (coll. A.-M. Chèvre et collaborateurs, INRA Rennes) : il s'agit de colzas recréés au laboratoire, permettant de suivre les modifications génomiques survenant lors de la formation du nouveau génome allopolyplœdie puis de son évolution au fil des générations, en comparaison des génomes parentaux utilisés comme références. Je me suis tout d'abord intéressée aux éléments transposables, qui constituent une part importante du génome végétal et présentent un pouvoir mutagène majeur en lien avec leur aptitude à se déplacer dans le génome. Mon étude a eu comme objectifs de caractériser l'impact de l'événement d'allopolyplœdie sur *Bot1*, un transposon à ADN spécifique du génome C (Alix et al. 2008), par une approche *transposon display* de type SSAP (*sequence-specific amplification polymorphism*). Puis je me suis attachée à l'étude des petits ARN non codants, agents moléculaires de régulation du génome, impliqués notamment dans le contrôle des éléments transposables. Les petits ARN non codants, découverts au début des années 2000, ont depuis été démontrés comme très diversifiés et impliqués dans des voies de régulation variées, à rôles majeurs dans la régulation de l'expression des gènes et le contrôle de la stabilité du génome. Mes travaux ont ainsi apporté une vision nouvelle quant à l'implication de ces petits ARN non codants dans les mécanismes de mise en place d'un génome allopolyplœdie, où le contrôle de la coexistence de génomes différenciés dans un même noyau apparaît être la première étape clé à surmonter pour assurer le succès futur du génome néo-allopolyplœdie.

CHAPITRE 1

L'importance de la polyploïdie dans l'évolution des angiospermes

CHAPITRE 1

L'importance de la polyploïdie dans l'évolution des angiospermes

Au cours de ce chapitre je vais résumer ce que l'on sait sur les organismes polyploïdes, et plus particulièrement les allopolyploïdes qui sont issus de croisement(s) interspécifique(s). Les organismes polyploïdes semblent avoir des avantages évolutifs par rapport aux espèces diploïdes. Avec le séquençage massif de plusieurs espèces végétales, on s'est aperçu que toutes les angiospermes ont connu au moins un événement de polyploïdisation au cours de leur évolution (Mc Grath & Lynch, 2012 ; Jiao et al. 2011 ; Figure 1.1.). Cela a un grand intérêt et suggère que la polyploïdisation, et plus particulièrement l'allopolyploïdisation, serait un moteur pour l'évolution des espèces et la création de nouvelles espèces qui présentent de nouveaux phénotypes. On sait par exemple que chez un allopolyploïde, la combinaison de deux génomes différenciés confère souvent des caractères supérieurs par rapport à ses parents, comme une augmentation de la taille, du rendement, une meilleure adaptation aux stress environnementaux... Ce phénomène est comparable, et souvent assimilé, au phénomène d'hétérosis ou vigueur hybride qui a d'abord été décrit lors de croisements intraspécifiques entre génotypes différenciés. Cependant, la formation d'un génome allopolyploïde n'est pas sans poser problème. Les événements d'hybridation et de duplication génomique sont ainsi considérés comme des stress ou chocs génomiques, liés à la présence de deux génomes différents dans un même noyau. Il en résulte d'importantes modifications génétiques et épigénétiques durant les premières générations suivant l'hybridation et la duplication du génome, qui peuvent affecter la structure du génome, l'expression des gènes, l'activité des éléments transposables, ou l'expression des petits ARN non codants. Ces derniers sont des molécules qui agissent aussi bien pour réguler l'expression de gènes aux niveaux transcriptionnel et post-transcriptionnel, que pour assurer l'intégrité du génome. La synthèse bibliographique sur les petits ARN non codants, qui est le sujet principale et originale de ma thèse, se trouve développée dans le chapitre 3.

1.1. La polyploïdie

La polyploïdie se définit par la présence de plus de deux lots de chromosomes homologues ou apparentés dans une cellule. Lorsque ces jeux de chromosomes proviennent de la même espèce (= chromosomes homologues), on parle d'espèces autopolyploïdes (Ramsey & Schemske 1998 ; Figure 1.1.). Plusieurs espèces cultivées sont des autopolyploïdes, comme la pomme de terre, la banane et la luzerne. Les espèces contenant au moins deux jeux de

chromosomes provenant d'espèces différentes (= chromosomes homéologues) sont quant à elles connues sous le nom d'allopolyploïdes (Figure 1.1.). Parmi les espèces allopolyploïdes cultivées, on trouve le blé, le coton, le tabac, le caféier ou encore le colza. Dans le détail, on peut aussi trouver des espèces comme la canne à sucre qui résultent d'une hybridation entre des espèces différentes mais très apparentées et qui sont donc « intermédiaires » entre ces deux extrêmes.

Le croisement entre deux espèces différenciées n'aboutit pas toujours à la formation d'un allopolyploïde ; l'hybridation en elle-même peut se heurter à des difficultés lors de la fécondation et du développement de l'embryon. Si ces difficultés sont surmontées, l'obtention d'un hybride interspécifique est possible. Cependant, la plupart du temps, ces hybrides sont stériles car les chromosomes ne se répartissent pas de manière équilibrée au cours de la méiose en raison de l'absence de leur homologue (seuls les homéologues sont présents). Lorsqu'il y a un doublement du stock chromosomique, la fertilité de l'individu est restaurée car les chromosomes peuvent à nouveau s'apparier avec leur chromosome homologue. Les organismes allopolyploïdes, qui présentent donc à la fois des paires de chromosomes homologues mais aussi des jeux de chromosomes homéologues, résultent donc de deux phénomènes distincts, quoique imbriqués : (1) une hybridation interspécifique associée à (2) un doublement chromosomique.

1.2. Modes de formation des polyploïdes (Figure 1.1.)

1.2.1. Formation des polyploïdes par la voie de gamètes non réduits

On considère que la principale voie de formation des polyploïdes implique la formation, chez l'un ou l'autre de ces parents diploïdes (progéniteurs), de gamètes qui ont plus d'un jeu de chromosomes : on parle de gamètes non réduits.

Lors d'une méiose régulière, deux divisions successives font suite à une seule étape de réplication de l'ADN, ce qui conduit à diviser de moitié le nombre de chromosomes ; la méiose produit des cellules haploïdes à n chromosomes (qui donneront naissance aux gamètes) à partir d'un organisme diploïde à $2n$ chromosomes. Lors de la première division (réductionnelle), les chromosomes homologues sont séparés, tandis que dans la seconde division (équationnelle) ce sont les chromatides sœurs produites lors de la réplication qui sont séparées.

Le déroulement de la méiose peut être perturbé au niveau d'une des deux divisions et conduire ainsi à la formation de gamètes « non réduits », possédant $2n$ chromosomes. C'est l'observation de la fréquence non négligeable (~5%) de ces gamètes diploïdes au sein des populations d'espèces diploïdes qui a conduit à penser que ces gamètes joueraient un rôle important dans la formation des polyploïdes (revue Ramsey & Schemske 1998 ; Soltis et al. 2003 ; Comai 2005 ; Chen 2007). Les gamètes diploïdes peuvent être fécondés, ou féconder (1) des gamètes haploïdes et produire des zygotes triploïdes ou (2) d'autres gamètes non réduits, diploïdes, et produire ainsi des zygotes tétraploïdes (Figure 1.1.).

1.2.2. Formation des polyplœides par la voie de doublement somatique

Les organismes polyplœides peuvent aussi être formés sans faire intervenir de gamètes non réduits. Le doublement chromosomique peut se produire de manière spontanée comme conséquence d'une hybridation entre espèces différentes (Ramsey & Schemske 1998). Chez certains hybrides F1 interspécifiques, contenant n chromosomes ($x + x'$), on observe parfois des lignées cellulaires qui ont spontanément $2n$ chromosomes ($2x + 2x'$), alors que le reste de plante est à n chromosomes (chimère polyplœide). Les gamètes formés par la partie à $2n$ chromosomes auront ($x + x'$) chromosomes et, s'ils se fécondent l'un l'autre, ils pourront former un individu S_0 polyplœide ayant $2n$ chromosomes ($2x + 2x'$). *Primula kewensis* constitue un des exemples les plus connus de cette voie de formation des polyplœides (Ramsey & Schemske 1998).

Une deuxième voie de formation possible implique le croisement de deux autopolyploïdes permettant de constituer ainsi un allopolyploïde à $2n$ chromosomes. Le passage par un intermédiaire triploïde est possible, après croisement entre un autotétraploïde (AAAA ou CCCC) et un diploïde (AA ou CC) d'une autre espèce (Comai 2005).

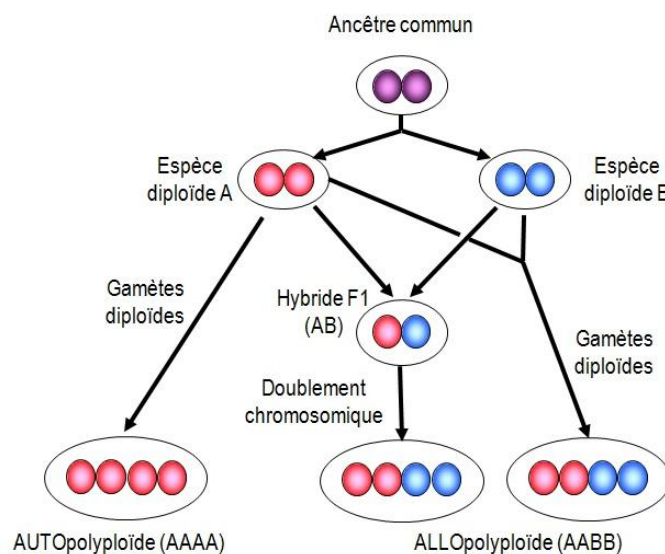


Figure 1.1. Représentation schématique des principaux modes de formation des polyplœides. Plusieurs voies de formation des organismes polyplœides sont possibles : (1) doublement spontané du génome diploïde (par passage de gamètes non réduits diploïdes) conduisant à la formation d'un organisme autotétraploïde ; (2) formation d'un allotétraploïde par hybridation interspécifique suivie du doublement chromosomique ; (3) formation d'un allotétraploïde par l'union de gamètes non réduits diploïdes issus des deux parents diploïdes.

1.3. Récurrence de la polyploïdie chez les plantes

La polyploïdie est peu fréquente chez les Gymnospermes, avec seulement quelques exemples de polyploïdes actuels (Cui et al. 2006). Au contraire, chez les Angiospermes, la polyploïdie est extrêmement fréquente ; on sait actuellement que toutes les Angiospermes ont eu au moins un événement de duplication de leur génome au cours de leur histoire évolutive (Jiao et al. 2011 ; Figure 1.2.). Cette découverte a été réalisée grâce aux énormes avancées des techniques de séquençage haut débit qui ont mis en évidence des traces de duplications au sein des génomes séquencés. Ces duplications sont le produits d'événements anciens de polyploïdie et les organismes qui les contiennent sont communément appelés **paléopolyploïdes**.

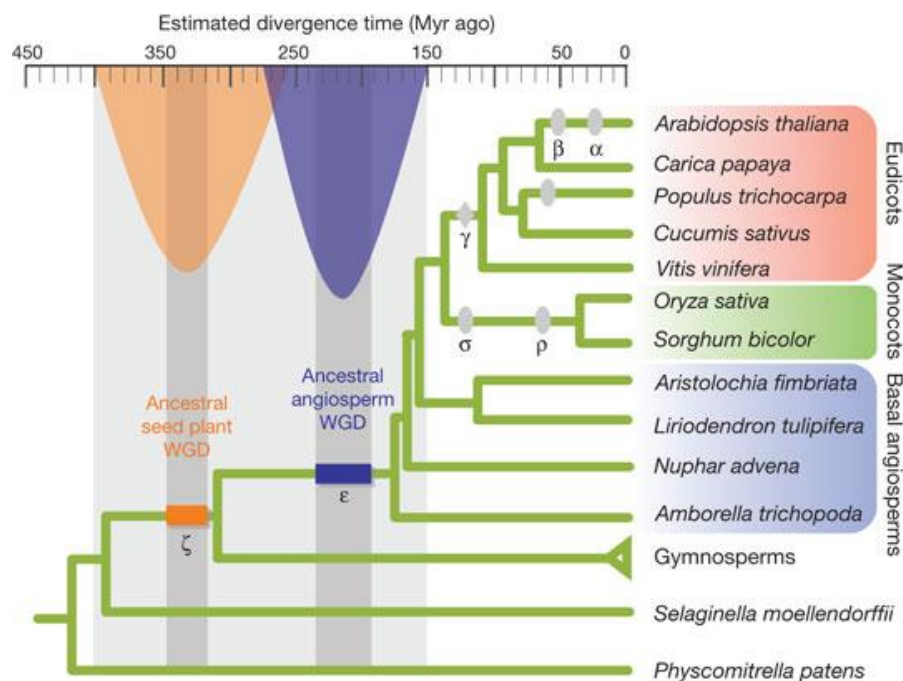


Figure 1.2. Événements de polyploïdisation durant l'évolution des plantes à graines et plantes à fleurs que sont les Angiospermes. La figure montre deux duplications ancestrales (ζ et ε) dans l'évolution des plantes. Les ovales correspondent à des duplications identifiées grâce au séquençage de génomes complets. Le losange (γ) correspond à un événement de triplication (Jiao et al. 2011).

1.3.1. La paléopolyploïdie

Au cours de l'évolution des Angiospermes, plusieurs événements de duplication du génome (WGD – *whole genome duplication*) ont eu lieu (Figure 1.2.). Ces duplications ont été suivies d'une phase de diploïdisation par perte d'un grand nombre (mais pas de tous) de gènes dupliqués. La détection des événements anciens de polyploïdie a été permise par le séquençage de génomes complets d'un nombre croissant d'Angiospermes. Le premier exemple, illustré par la Figure 1.3., concerne le génome de l'espèce modèle *Arabidopsis thaliana*, publié en 2000 (The *Arabidopsis* Genome Initiative 2000).

La Figure 1.3. montre l'existence de nombreux blocs de gènes dupliqués répartis au sein des 5 chromosomes de cette espèce, qui résultent d'une duplication ancestrale de son génome. Ainsi, cette espèce diploïde, choisie au départ pour la « simplicité » de son génome, est en réalité une espèce multiples fois paléopolyploïde. Plusieurs analyses ont en effet montré qu'*Arabidopsis* avait subi cinq événements de polyploïdie successifs : deux duplications communes aux Angiospermes (ζ et ϵ), suivies d'une triplication commune aux dicotylédones (γ), puis de deux duplications plus spécifiques des *Brassicacées* (α et β) (Simillion et al. 2002 ; Jiao et al. 2011 ; Van der Peer et al. 2009 ; Schranz et al. 2006 ; Figure 1.1.).

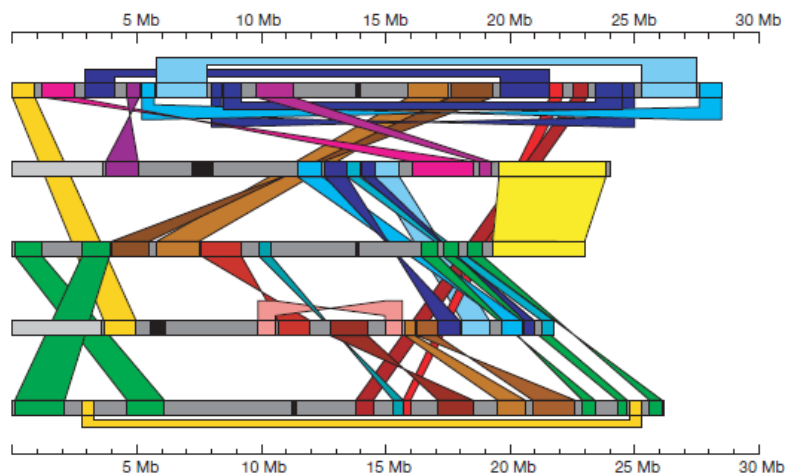


Figure 1.3. Duplications au sein du génome d'*Arabidopsis thaliana*. Les 5 chromosomes sont représentés par des barres horizontales de couleur grise, le chromosome 1 étant en haut et les centromères étant marqués en noir. Les blocs colorés correspondent aux régions dupliquées D'après The *Arabidopsis* Genome Initiative 2000.

Se pose alors la question de la méthodologie à utiliser pour identifier des événements anciens de duplication du génome. Actuellement, il existe plusieurs approches expérimentales pour identifier d'anciens événements de polyploïdisation :

- **Par l'analyse de blocs de synténie** (= conservation du répertoire et de l'ordre des gènes ou des marqueurs) présents en plusieurs copies au sein d'un même génome (Figure 1.4.A). Lorsque la séquence génomique complète de l'espèce considérée n'est pas disponible, ces blocs peuvent être mis en évidence par cartographie ou par hybridation *in situ* en exploitant les ressources génomiques des espèces apparentées (Lysak et al. 2005 ; Salse et al. 2008 ; Thiel et al. 2009) (Figure 1.4.B).

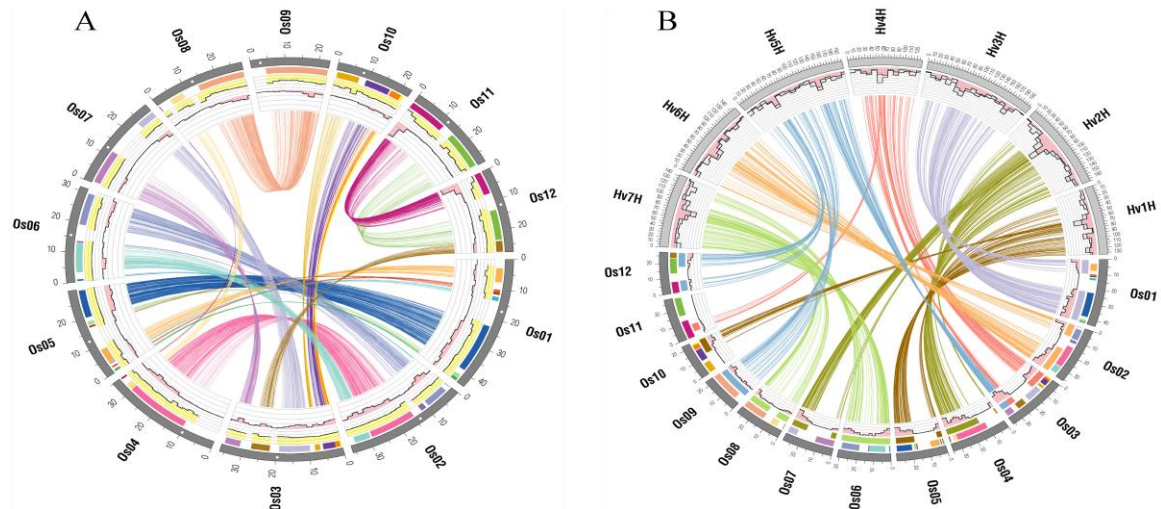


Figure 1.4. Analyse des blocs de duplications au sein d'un même génome (A. le riz, à gauche) et entre espèces proches (B. en comparant l'orge avec le riz, à droite).

A. Duplications dans le génome de riz. Les duplications de gènes dans le génome de riz sont représentées par des lignes connectant les 12 chromosomes du riz (Os01 – Os12). Les gènes dupliqués ont été organisés en 24 paires de segments génomiques situés tout au long de l'ensemble du génome (blocs de différentes couleurs dans le cercle extérieur). Les centromères sont indiqués par des points blancs.

B. Relations de synténie entre le riz et l'orge. Les séquences EST de l'orge sont positionnées sur la carte génétique représentant les 7 chromosomes de cette espèce (Hv1H – Hv7H) en regard de leur homologue sur les 12 chromosomes du riz (Os01 – Os12). Les histogrammes sur les chromosomes de l'orge montrent le nombre total des marqueurs ainsi que le pourcentage de ceux qui présentent une synténie avec le génome du riz. Les barres extérieures sur les chromosomes du riz montrent les segments dupliqués au sein de son génome (Figure 1.4A) tandis que les barres intérieures montrent les régions synténiques avec la carte génétique de l'orge. Les positions pour le génome de riz sont données en Mb et celles de l'orge en cM. D'après Thiel et al. 2009.

- **Par l'analyse du nombre de paires de gènes dupliqués qui présentent un même taux de substitution synonyme** par site où peuvent apparaître des substitutions synonymes (Ks) – et qui sont par conséquent neutres. Cette approche permet non seulement de détecter des événements de polyploïdie anciens (quand tous les gènes dupliqués présentent un même Ks, ce qui montre qu'ils ont divergé les uns des autres, et donc ont été dupliqués tous en même temps) mais aussi de les dater. L'exemple de la Figure 1.5. illustre ces deux points. Si un événement de polyploïdisation survient, un grand nombre de gènes sont dupliqués en même temps, créant un pic dans la distribution des Ks (marqués par les points vert et rouge). Plus les Ks sont élevés, plus anciens sont les événements de duplication qui ont eu lieu (Van der Peer et al. 2009 ; Fawcett et al. 2009), l'événement marqué en vert étant par exemple plus récent (et spécifique du génome d'Arabidopsis) que l'événement marqué en rouge.

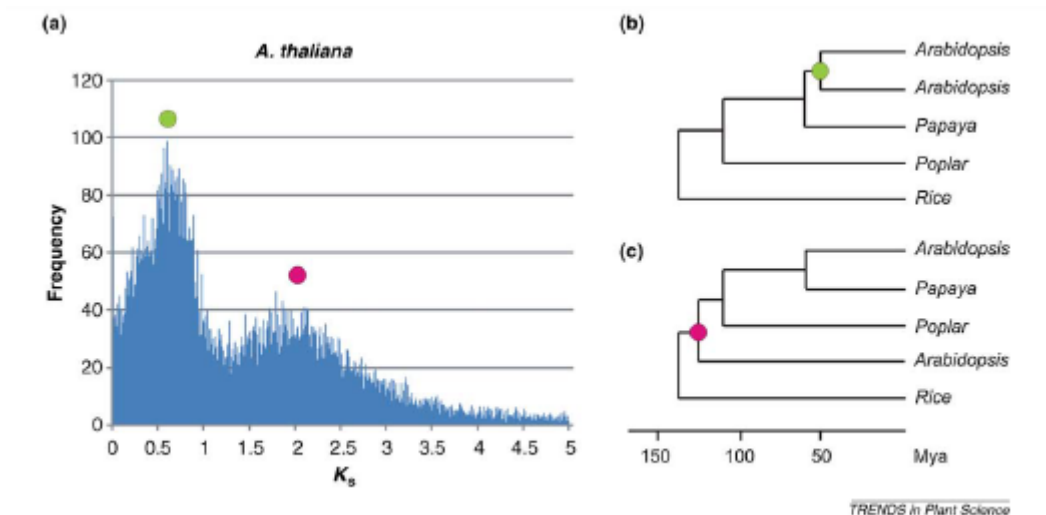


Figure 1.5. Distribution de l'âge des gènes paralogues chez *A. thaliana*. La distribution de l'âge des gènes paralogues basée sur les valeurs de K_s (nombre de substitutions synonymes par site). Chacun de deux pics représente un événement de duplication chez *A. thaliana*, à gauche ($K_s \sim 0.6$) correspond à la plus récente duplication du génome α (point vert), et à droite on retrouve la duplication β (point rouge) (Van der Peer et al. 2009).

Chez les Angiospermes, de traces de polyploïdisation ancestrales on été retrouvées chez de nombreuses espèces considérées aujourd'hui comme diploïdes : la vigne, la papaye, le peuplier, *Medicago truncatula*, le cacaoyer, entre autres. Ces découvertes ont mis en la récurrence de la polyploïdisation au cours de l'évolution des espèces et suggèrent l'importance de ce phénomène pour l'évolution, l'adaptation et la création de nouvelles espèces.

1.3.2. Les polyploïdes actuels

Selon une échelle de temps, on peut distinguer deux groupes de polyploïdes actuels : les polyploïdes jeunes, formés il y a moins de 150 ans et les polyploïdes plus âgés qui ont été formés antérieurement (pour revue Soltis & Soltis, 2009). Selon la date de leur formation, il est possible d'étudier directement l'évolution génomique du jeune allopolyploïde *in situ*, dans les populations naturelles actuelles (par exemple *Tragopogon miscellus/mirus* 4x, *Spartina anglica* 12x, *Senecio cambrensis* 4x). Pour les espèces polyploïdes plus âgées, dont les progéniteurs ne sont pas tous nécessairement connus précisément (par exemple, le génome B des blés), il faut prendre en compte l'évolution des génomes parentaux depuis la formation du polyploïde. Ces espèces étant par ailleurs « plus stables » que les précédentes, il est souvent nécessaire de produire des polyploïdes synthétiques pour analyser l'évolution génomique qui fait suite à la formation du polyploïde considéré. Quelques exemples de polyploïdes plus âgés sont le cotonnier cultivé (*Gossypium hirsutum/barbadense* 4x), le blé dur (*Triticum turgidum durum* 4x), le blé tendre (*Triticum aestivum* 6x), le caféier (*Coffea arabica* 4x), le tabac (*Nicotiana tabacum* 4x) et le colza (*Brassica napus* 4x).

1.3.3. Les polyploïdes synthétiques

En 1937, Blakeslee et Avery ont induit le doublement du stock chromosomique chez des plantes en utilisant la colchicine, un inhibiteur chimique de la division cellulaire mitotique. Cette technique a depuis été largement utilisée pour induire le doublement chromosomique de cellules méristématiques chez des organismes diploïdes et chez des hybrides interspécifiques (Chen & Ni, 2006).

Les polyploïdes synthétiques constituent un excellent matériel génétique pour étudier les changements dans la structure du génome et dans l'expression de gènes qui accompagnent la formation d'un polyploïde, par comparaison avec les génotypes parentaux. Ils présentent plusieurs avantages par rapport aux espèces polyploïdes naturelles :

1. Un des principaux avantages est que les progéniteurs (génotypes) exacts sont connus. Ce n'est pas le cas pour la plupart des allopolyploïdes naturels, sauf pour les polyploïdes les plus récents tels que la spartine, le *tragopogon* ou le *senecio*. De cette façon, la comparaison entre parents et descendants peut se faire plus précisément.
2. On peut mimer la formation d'un polyploïde naturel en suivant les mécanismes qui se mettent en place tout de suite après l'obtention du polyploïde. On peut étudier ce qui se passe dès l'étape d'hybridation (pour les allopolyploïdes) et au cours des différentes générations après la duplication du génome. Cela permet aussi de faire la différence entre les modifications dues à la rencontre des génomes différenciés (hybridation) et celles dues au doublement génomique (allopolyploïdisation).
3. On peut recréer plusieurs événements indépendants pour étudier si les modifications observées sont aléatoires ou au contraire reproductibles.
4. On peut tester le sens des croisements. Il a en effet été montré que la morphologie des polyploïdes peut être très différente selon le sens du croisement, comme par exemple la morphologie florale chez *Tragopogon miscellus* (Figure 1.6.) (Tate et al.2009).

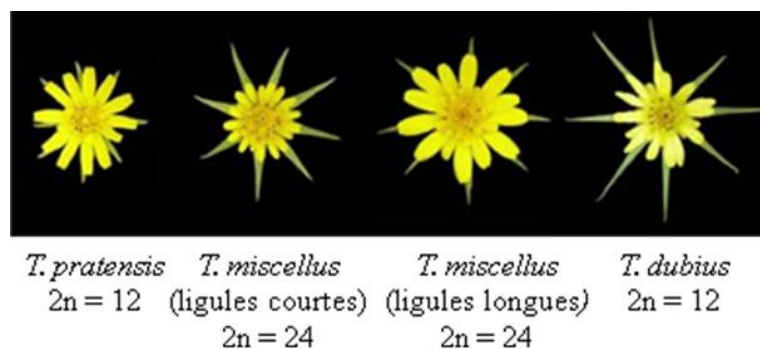


Figure 1.6. Effet du sens du croisement sur la morphologie florale chez *Tragopogon miscellus*. Lorsque *T. pratensis* est le parent maternel, *T. miscellus* a des ligules courtes, alors qu'il présente des ligules longues quand *T. dubius* est le parent maternel. D'après Tate et al.2009.

Compte tenu de tous ces avantages, plusieurs allopolyploïdes synthétiques ont été produits au laboratoire afin d'étudier de près ce qui se passe lors de la formation d'un organisme polyploïde au niveau de la structure du génome, de l'expression de gènes, de l'activation des éléments transposables... Ces études ont permis de démontrer le caractère dynamique des génomes polyploïdes néoformés et ils ont mis en évidence de nombreuses modifications dans l'organisation et la fonction du génome avec des écarts importants par rapport à **l'hypothèse d'additivité des génomes parentaux** (= le polyploïde présente comme caractéristique la moyenne ou l'addition des caractéristiques parentales, selon le trait étudié). Parmi les études menées sur des polyploïdes synthétiques on trouve celles portant sur *Arabidopsis suecica* (Lee & Chen, 2001), le coton (*Gossypium hirsutum* ; Adams et al. 2004 ; Adams & Wendel, 2005), le tabac (*Nicotiana tabacum* ; Petit et al. 2010) et le colza (*Brassica napus* ; Albertin et al. 2006 ; Marmagne et al. 2010 ; Szadkowski et al. 2010, 2011).

1.4. Conséquences évolutives de la polyploïdie

On estime qu'entre 2% et 15% des événements de spéciation chez les Angiospermes seraient causés par un événement de polyploïdisation (Otto & Whitton, 2000 ; Wood et al. 2009). Certains auteurs suggèrent également que la polyploïdie est un facteur clé dans la diversification des espèces végétales (Soltis et al. 2009). Les espèces polyploïdes présentent une grande tolérance aux conditions de stress environnementaux. Ainsi, de nombreuses plantes allopolyploïdes sont capables de coloniser des niches d'où leurs parents diploïdes sont absents (Soltis et al. 2004). D'autres espèces polyploïdes peuvent devenir invasives en raison de leur grande amplitude écologique, notamment dans des environnements stressants ; c'est le cas de la spartine anglaise (*Spartina anglica*) qui s'est formée à partir de l'hybride interspécifique *S. towsendii* lui-même issu du croisement entre *S. alterniflora* et *S. maritima* (Ainouche et al. 2004).

Fawcett et al. (2009) ont constaté qu'un grand nombre de duplications génomiques ancestrales avaient eu lieu au moment de la crise Crétacé-Tertiaire (Figure 1.2.). Au cours de cette période, les conditions climatiques ont été profondément modifiées ; les auteurs proposent donc que les espèces polyploïdes seraient devenues prépondérantes car elles présenteraient alors une meilleure résistance aux changements climatiques et se seraient donc mieux adaptées (Fawcett et al. 2009).

Si l'on considère que la polyploïdie a été très importante dans l'adaptation des espèces (Fawcett et al. 2009 ; Crow & Wagner, 2006), il reste à comprendre quels sont précisément les avantages que confère la polyploïdie par rapport aux organismes diploïdes.

1.5. Polyploïdes versus diploïdes : avantages des polyploïdes sur les diploïdes

La prévalence de la polyploïdie dans le règne végétal suggère que les polyploïdes ont une meilleure valeur sélective (*fitness* en anglais) que leurs progéniteurs. On considère généralement que l'hétérosis et la redondance génétique contribuent à cet avantage (Comai, 2005).

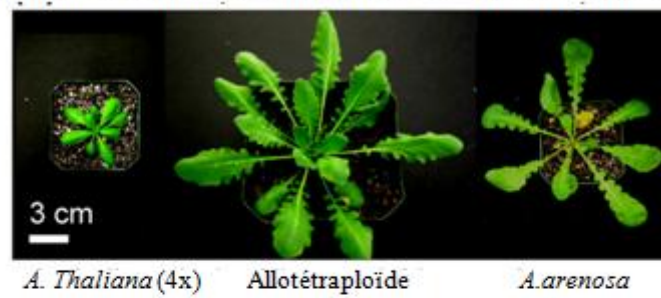
1.5.1. Hétérosis ou vigueur hybride

L'hétérosis ou vigueur hybride est l'augmentation de vigueur observée chez un hybride produit en croisant deux individus génétiquement différents. Ce phénomène a d'abord été observé chez des hybrides F1 de maïs (Birchler et al. 2003) mais il est particulièrement marqué chez les allopolyploïdes par rapport à leurs espèces progénitrices (pour revue Chen 2010). Cette vigueur hybride se manifeste à plusieurs niveaux : augmentation de la taille, de la biomasse, tolérance accrue à des conditions de stress environnementales, etc.

Certains auteurs ont suggéré que ces variations phénotypiques peuvent être en partie dues à l'augmentation de la taille du génome qui conduit « mécaniquement » à une augmentation de la taille des cellules (« gigas effect » : Randolph 1941 ; Stebbins 1971 ; Figure 1.7.). On peut aussi rapprocher l'augmentation des capacités et ou de la vigueur du polyploïde du phénomène d'hétérosis qui est dû en grande partie à l'hétérozygotie de l'hybride. En fait, l'hétérozygotie est fixée chez un allopolyploïde en raison de la présence systématique des copies homéologues (notons que l'hétérozygotie est ici considérée entre génomes, alors qu'elle existe entre allèles chez un hybride diploïde). L'hétérosis observée chez des hybrides intraspécifiques et chez les allopolyploïdes rend ces espèces très intéressantes d'un point de vue agronomique, particulièrement dans l'amélioration variétale.

Par ailleurs, d'importantes modifications génétiques et/ou épigénétiques de la structure et de l'expression des gènes et du génome ont été mises en évidence chez les organismes polyploïdes par rapport aux diploïdes. Ainsi, la régulation non-additive des gènes à l'échelle du génome observée chez plusieurs modèles allotétraploïdes est corrélée avec la divergence de l'expression chez les parents : les hybrides issus de croisements entre espèces parentales plus éloignées présentent un plus grand nombre de gènes avec des niveaux d'expression non-additifs par rapport aux parents, ce qui pourrait fournir des bases moléculaires responsables de la vigueur hybride et de la variation phénotypique observée chez l'allopolyploïde (Chen 2007). Des nombreux gènes du métabolisme, de l'énergie et la photosynthèse, de la biogénèse cellulaire, de la régulation hormonale ou du rythme circadien sont ainsi surexprimés chez les allotétraploïdes d'*A. suecica* (Wang et al. 2006) et pourraient ainsi contribuer à la vigueur hybride observée chez ces derniers (Chen 2007 ; Ni et al. 2008).

(a) Un allotétraploïde formé entre deux espèces d'*A. thaliana*



(b) Fleurs



(c) Graines

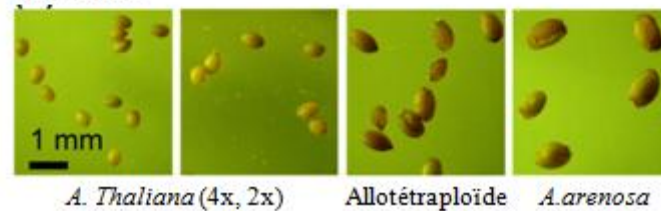


Figure 1.7. Hybrides et allotétraploïdes d'*A. thaliana*. (a) Allotétraploïde stable (génération F8) maintenu par autopollinisation. L'hybride F1 a été produit par la pollinisation d'une plante autotétraploïde d'*A. thaliana* avec le pollen d'*A. arenosa* qui est elle-même autotétraploïde ; (b) fleurs de l'allotétraploïde et ses progéniteurs, *A. thaliana* tétraploïde (en médaillon, diploïde) et *A. arenosa* ; (c) graines de l'allotétraploïde et ses progéniteurs, *A. thaliana* (4x, 2x) et *A. arenosa*. Figure modifiée à partir de Chen 2010.

1.5.2. Redondance génétique : conséquences de la duplication des gènes et devenir des gènes dupliqués

Au-delà du phénomène d'hétérozygotie fixée, la présence de copies dupliquées conduit à de la redondance génétique qui peut s'avérer avantageuse à plusieurs égards.

Le premier avantage conféré par la redondance génétique est la possibilité de masquer des allèles récessifs plus ou moins délétères par les allèles dominants « de type sauvage » présents sur le génome homéologue (allopolyploïde) ou sur les autres chromosomes homologues (autotétraploïde). De plus, le fait d'avoir au moins deux copies de chaque gène donne la possibilité de diversifier leurs fonctions en modifiant une des deux copies même quand il s'agit des gènes essentiels. Chez les organismes diploïdes cette possibilité est très peu fréquente et elle est conditionnée à un événement de duplication segmentaire. Chez les polyploïdes, tous les gènes ont une copie dupliquée qui est disponible pour tester de nouvelles trajectoires évolutives.

Le devenir des gènes dupliqués a été depuis longtemps un sujet de réflexion et d'intérêt majeur sur l'évolution des espèces. Une étape fondamentale a été franchie par Onho, en 1970, qui a proposé que le devenir des gènes dupliqués était différent à court, moyen et long terme. Des études ultérieures portant sur le devenir des gènes dupliqués chez plusieurs espèces ont montré que la plupart des duplications en tandem ou segmentaires aboutissent à la non-fonctionnalisation d'une des copies. Cette **pseudogénéisation** peut se faire par la dégradation d'une copie en gène non fonctionnel accumulant des mutations délétères à sa fonction, ou par des délétions plus ou moins importantes affectant la séquence codante (Lynch & Conery, 2000 ; McGrath & Lynch, 2012). En revanche, un plus grand nombre de gènes dupliqués est généralement conservé chez des espèces polyploïdes (Force et al. 1999 ; Lynch & Conery, 2000) : entre 25 et 75% de rétention des gènes dupliqués selon l'âge de l'événement de polyploïdie (revue Lynch, 2007 ; Otto, 2007 ; McGrath & Lynch, 2012). Comment expliquer des taux de rétention aussi forts ?

1. Il est possible que plusieurs de ces espèces polyploïdes n'aient pas encore atteint « l'équilibre » (une certaine stabilité génomique) et qu'elles soient encore dans une phase de perte des gènes dupliqués ;

2. Il existe des pressions de sélection conduisant à maintenir les copies dupliquées « intactes ». C'est notamment le cas quand les produits de gènes dupliqués sont impliqués dans des interactions multiples ou en amont de réseaux de gènes. Etant donné que la perte d'une copie pourrait altérer la relation stœchiométrique de l'ensemble du complexe ou du réseau protéique, elles sont délétères et contre-sélectionnées, ce qui conduit à une pression de sélection pour maintenir l'équilibre génique et l'ensemble des copies intact (Lynch & Conery, 2000 ; Veitia et al. 2008) ;

3. La duplication des gènes codant plusieurs fonctions peut permettre à chacune de ces fonctions d'évoluer et de se diversifier. Ce phénomène, appelé **sous-fonctionnalisation**, permet à chacune des copies de se spécialiser dans une fonction, ou de s'exprimer dans un organe ou tissu cellulaire particulier, en partageant avec l'autre copie la fonction d'origine du gène ancestral. Un exemple de sous-fonctionnalisation a été mis en évidence chez le coton, où plusieurs gènes dupliqués sont différenciellement exprimés selon les organes (expression tissu-spécifique des loci homéologues ; Adams et al. 2003) ;

4. La rétention des copies dupliquées peut aussi avoir lieu quand une des deux copies acquiert une nouvelle fonction qui serait favorisée par la sélection, ce processus arrivant en général sur le plus long terme et s'appelant **néo-fonctionnalisation**.

L'observation de différences de conservation selon la fonction des gènes, avec une forte rétention constatée pour les gènes impliqués dans la transcription et la transduction de signal (gènes ribosomiques, transférases, kinases, facteurs de transcription), suggère que la principale cause du plus fort niveau de rétention des copies dupliquées par polyploïdie est liée **aux effets de dosage** (point 2 précédent). Tandis que les autopolyploïdes présenteraient l'avantage principale de la redondance génétique et la polysomie (= appariement aléatoire des chromosomes homologues lors de la méiose) qui génèrent de la variabilité par brassage génétique (Parisod et al. 2010).

1.6. Dynamique des génomes allopolyploïdes

Depuis près de 20 ans (Song et al. 1995), des études menées sur des allopolyploïdes synthétiques ont montré que les génomes polyploïdes présentaient des changements dynamiques dans leur séquence d'ADN et dans l'expression de leurs gènes ; cela est probablement dû au « choc génomique » causé par la rencontre de deux génomes différenciés (Chen & Ni, 2006).

1.6.1. Modifications structurales du génome

La taille des génomes polyploïdes, qu'ils soient récents ou anciens, varie significativement par rapport à celle de leurs espèces progénitrices. Ces changements commencent très rapidement après l'hybridation, ce qui suggère qu'ils résultent de l'action de mécanismes se mettant immédiatement à l'œuvre après la formation du polyploïde (Leitch & Leitch, 2008). De nombreux mécanismes conduisant à un réarrangement de la structure des génomes ont été décrits chez les polyploïdes (Chen & Ni, 2006). Ils ont été mis en évidence par PCR, GISH/FISH (*Genomic/Fluorescent in situ hybridization*), AFLP (*Amplified fragment length polymorphism*), par RFLP (*Restriction fragment length polymorphism*) ou par CAPS (*Cleaved amplified polymorphic sequence*).

1.6.1.a. Réarrangements chromosomiques : délétions, translocations (réciproques ou non réciproques), recombinaison homéologue

Le premier exemple de changements structuraux accompagnant la formation d'un polyploïde a été décrit par Song et al. en 1995. En utilisant des marqueurs RFLP, ces auteurs ont mis en évidence des changements rapides dans l'organisation du génome chez des allotétraploïdes synthétiques de *Brassica*. Ils ont détecté un écart à l'additivité des profils parentaux pour certains fragments génomiques : fragments perdus dans la descendance alors qu'ils étaient présents chez les parents ou au contraire, apparition de nouveaux fragments chez les allopolyploïdes par rapport aux parents. Plusieurs de ces changements étaient probablement le résultat de crossovers méiotiques entre chromosomes homéologues ; ils n'avaient été détectés que dans les générations plus avancées d'autofécondation (S3 – S6) et non dans les toutes premières.

Des études plus récentes, portant sur le même matériel végétal que celui que j'ai étudié au cours de ma thèse, ont confirmé le rôle prépondérant des crossovers méiotiques entre homéologues dans l'évolution de la structure des colzas synthétiques (Szadkowski et al. 2010, 2011). Chez ces derniers, si la recombinaison se fait principalement entre régions homologues, des crossovers entre homéologues sont également observables en cytologie chez l'hybride F1 (Szadkowski et al. 2011) ou après doublement chromosomique (dans la

génération S0 ; Szadkowski et al. 2010). Dans la descendance de ces plantes (en particulier celle de l'hybride F1, qui croisé avec un S0 permet d'obtenir des colzas synthétiques issus de gamètes non réduits), les produits de ces crossovers, qui sont des échanges réciproques de matériel génétique entre homéologues, ségrègent et deviennent détectables sous la forme de translocations réciproques ou non-réciproques entre des chromosomes homéologues ; 82% des plantes dans la descendance de l'hybride F1 présentent ainsi au moins une et souvent plusieurs translocations non-réciproques (Szadkowski et al. 2011). Des « réarrangements » comparables s'observent également dans la descendance des allotétraploïdes synthétiques obtenus par doublement à la colchicine (Szadkowski et al. 2010), même s'ils sont alors moins fréquents (Szadkowski et al. 2011).

Contrairement aux observations faites chez les *Brassica*, les allopolyploïdes de blé ont montré entre 10 – 15% des changements génomiques dès la formation de l'hybride F1 (c'est-à-dire avant la première méiose), ces changements étant principalement des délétions spécifiques de certaines régions du génome (Levy & Feldman, 2002). Chez des formes synthétiques d'*Arabidopsis suecica* (produites par croisement entre *A. thaliana* 4x et *A. arenosa* 4x), Pontes et al. (2004) ont observé par FISH d'importants réarrangements structuraux (délétions, duplications, changement de localisation) au niveau des organisateurs nucléolaires (=les régions des chromosomes contenant les gènes d'ARN ribosomique) et des régions 5S.

Les réarrangements par délétion ou duplication de régions génomiques, qu'ils résultent de crossovers méiotiques (translocations non réciproques) ou non, entraînent la présence d'une même région génomique en un nombre variable de copies au sein du génome, ce qui peut conduire à des déséquilibres dans le dosage de certains gènes ; cela peut alors conduire, par effet de cascade, à des déséquilibres plus graves à l'échelle de complexe protéiques ou de réseaux de gènes. Les réarrangements génomiques peuvent donc avoir un effet sur l'expression des gènes qu'ils contiennent, avec soit une surexpression (en cas de duplication du gène ou de ses séquences régulatrices), soit une répression définitive (en cas de délétion ou mutation ponctuelle non synonyme délétère, conduisant par exemple à la modification d'un site d'épissage s'accompagnant de la rétention d'un intron). De même, les remaniements chromosomiques conduisent à faire varier le nombre de copies d'une région génomique d'une plante à l'autre, dans la descendance d'un allotétraploïde : cette source de variabilité doit donc être contrôlée, en utilisant des répliquats biologiques.

Les modifications structurales ne concernent cependant pas toutes les espèces ; chez la Spartine et chez le coton, peu de remaniements sont observés entre génomes homéologues (Ainouche et al. 2004 ; Adams & Wendel, 2004).

1.6.1.b. Activation des éléments transposables

Un autre mécanisme pouvant conduire à un réarrangement de la structure des génomes chez les polyploïdes concerne l'insertion ou l'excision d'éléments transposables (ET) du génome. Les ET sont réputés être activés suite à un stress (Madlung & Comai, 2004). L'hybridation entre deux génomes est considérée comme un stress génomique qui pourrait activer l'expression des éléments transposables car certains des mécanismes contrôlant les ET – comme la méthylation de l'ADN et la compaction de la chromatine –, sont apparemment levés, et de manière *a priori* transitoire, lors de la formation d'un hybride interspécifique (Zhang et al. 2009). Les génomes des espèces polyploïdes plus anciennes montrent parfois d'importantes proliférations d'ET, comme par exemple le coton chez lequel certains ET spécifiques d'un des deux parents (génome A) sont retrouvés sur les deux génomes (A et D) de l'allotétraploïde (Zhao et al. 1998 ; Rong et al. 2010).

La présence d'un ET peut avoir un effet sur l'expression des gènes adjacents. La conséquence de l'insertion des ET peut être un *silencing* définitif si l'élément transposable s'insère dans (1) la séquence promotrice, (2) entre la séquence promotrice et le gène ou (3) dans un exon. Elle peut être réversible pour une insertion d'un ET de classe II (couper-coller), du fait de la possibilité pour ce transposon de s'exciser de son nouveau site d'insertion. De plus, les promoteurs des ET insérés à côté des gènes peuvent avoir un impact sur les niveaux d'expression de ces derniers, comme illustré par l'exemple bien connu du gène *Agouti* chez la souris et de la régulation de son expression (Figure 1.8.). L'insertion d'un ET à proximité d'un gène de la couleur du pelage chez la souris conduit à la surexpression de ce dernier par initiation de la transcription dans le sens 3'→ 5' à partir du promoteur de l'ET. Cette surexpression résulte en une grande variation phénotypique (Morgan et al. 1999).

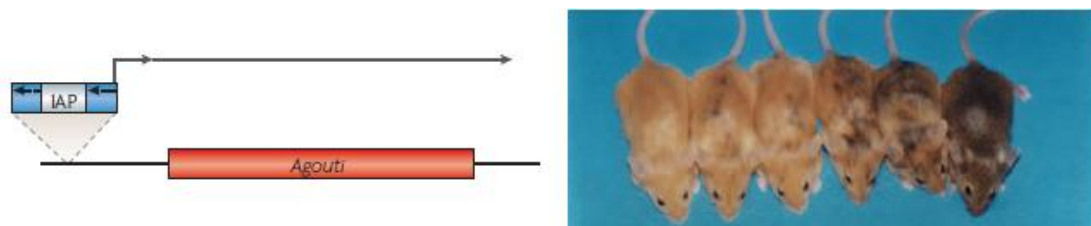


Figure 1.8. Impact des éléments transposables sur l'expression génique : l'exemple du gène *Agouti* chez la souris. Chez la souris, le rétrotransposon IAP (*Intracisternal A-particle*) produit un transcrite qui s'étend jusqu'au gène *Agouti* qui contrôle la couleur du pelage. Le niveau de transcription du gène *Agouti* et donc l'intensité de la couleur du pelage sont soumis à l'état épigénétique de l'ET. Cet état peut être hérité à la descendance. D'après Slotkin & Martienssen, 2007.

Cependant on remarque que l'activation des éléments transposables est loin d'être prévisible et apparaît très dépendante de l'ET que l'on suit et du système polyploïde que l'on étudie (pour revue Parisod et al. 2010).

1.6.2. Modifications de l'expression des gènes

Une autre façon d'appréhender les mécanismes responsables de la plus grande valeur sélective et/ou adaptative des organismes allopolyploïdes est d'analyser si l'expression de certains gènes est modifiée par rapport à celle des progéniteurs. Dans les différents systèmes allopolyploïdes étudiés, la grande majorité des gènes sont transcrits de façon additive. Toutefois, des modifications dans la régulation et l'expression d'une partie des gènes sont souvent observées dès l'étape d'hybridation (revues Adams & Wendel, 2005 ; Chen 2007 ; Doyle et al. 2008).

1.6.2.a. Modifications au niveau du transcriptome chez des hybrides et des allopolyploïdes

Des changements dans l'expression des gènes ont été observés chez des hybrides interspécifiques d'*Arabidopsis* (Shen et al. 2012 ; Fujimoto et al. 2012 ; Andorf et al. 2010), de riz (He et al. 2010), de maïs (Swanson-Wagner, et al. 2006 ; Guo et al. 2006 ; Stupar & Springer, 2006) ou encore de blé (Wang et al. 2006), mais aussi chez des allopolyploïdes d'*Arabidopsis* (Shi et al. 2012 ; Wang et al. 2006), de coton (Flagel et al. 2008), de blé (Pumphrey et al. 2009; Qi et al. 2012), de *Senecio* (Hegarty et al. 2006), de *Tragopogon* (Buggs et al. 2011), et de colza (Marmagne et al. 2010). Ces variations ont été établies par rapport à **l'hypothèse d'additivité**.

L'hypothèse d'additivité : Un génome allopolyploïde provient de l'union des génomes de ses progéniteurs. En théorie, et toutes choses étant égales par ailleurs, on pourrait donc s'attendre à ce qu'il contienne la somme de tous les composants de ses deux génomes parentaux : la quantité d'ADN, la quantité de transcrits, la quantité des protéines produites, ... Cette hypothèse est appelée l'hypothèse d'additivité. Lorsqu'on souhaite tester cette hypothèse de manière quantitative, l'attendu sous l'hypothèse de l'additivité est la valeur moyenne des parents (MPV- *mid-parent value*). On dit que les gènes sont exprimés de façon additive chez un hybride ou un allopolyploïde si leur expression est au niveau de la MPV. Au contraire, si l'expression chez les hybrides ou les allopolyploïdes est supérieure ou inférieure à la MPV, on considère que ces gènes ne sont pas exprimés de façon additive (Figure 1.9). Dans certains cas, les profils d'expression peuvent être supérieurs à la valeur du parent le plus fort ou inférieurs à la valeur du parent le plus faible, se rapprochant respectivement de cas de super-dominance ou sous-dominance.

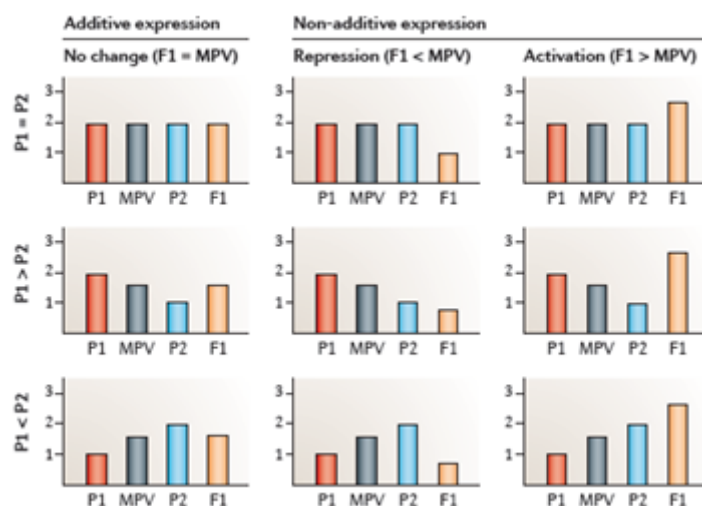


Figure 1.9. Expression additive et non-additive des gènes. La valeur moyenne parentale (MPV) est égale à $(P1+P2) / 2$, où P1 et P2 représentent les niveaux d'expression d'un gène chez les parents 1 et 2. Dans les trois graphes de haut, un gène est exprimé de la même façon chez les deux parents ($P1=P2$), le profil d'expression de l'hybride (F1) peut être additif ($F1=MPV$; à gauche), réprimé ($F1<MPV$; au centre), ou au contraire sur-exprimé ($F1>MPV$; à droite). Les valeurs de l'expression des gènes sont données sur l'axe y. Les graphes des lignes du centre et du bas de la figure montrent que les profils d'expression additif et non-additif chez les hybrides F1 peuvent se produire lorsque le niveau d'expression d'un gène est plus élevé dans le parent P1 (ligne du centre) ou plus élevé dans le parent P2 (ligne du bas). D'après Chen 2013.

Gaeta et collaborateurs (2009) ont testé l'hypothèse d'additivité chez des colzas néo-synthétisés des générations S1 et S6 ; ils ont pour cela mesuré l'expression des gènes en utilisant l'hybridation des transcrits sur une puce à ADN d'*A. thaliana* et l'ont comparé aux profils obtenus chez les parents diploïdes. Cette étude a montré que 1,6 à 32% des gènes présentaient une expression non-additive en génération S6. Cependant les gènes concernés n'étaient pas partagés par tous les individus, traduisant ici l'absence d'une réponse globale et reproductible à l'alloploïdisation (Gaeta et al. 2009).

Les études des différences d'expression entre les alloploïdes et leurs parents peuvent parfois être faites en discriminant l'origine parentale des gènes homéologues. On peut ainsi observer s'il existe un biais lié à cette origine. Dans ce cas, un des deux sous-génomes peut être préférentiellement réprimé (Wang et al. 2006 ; Tate et al. 2009) ou l'alloploïde peut montrer un profil global proche de l'un des deux parents ; ce phénomène est connu sous le nom de **dominance génomique** (Flagel & Wendel, 2010). Ces phénomènes ne sont pas encore expliqués, mais ils montrent l'importance des caractéristiques des progéniteurs dans la mise en place de nouvelles régulations chez l'alloploïde néo-formé.

Malgré la complexité dans l'étude des transcriptomes, certaines tendances ont été constatées :

1. L'expression additive ou non-additive des gènes est en général corrélée avec la distance génétique entre les espèces parentales qui se sont hybridées ; ainsi l'expression non-additive est plus fréquente chez des hybrides interspécifiques que chez des hybrides intraspécifiques (Chen, 2007 ; Jackson & Chen, 2010). Un exemple connu d'expression non-additive est celle des organisateurs nucléolaires pour lesquels une répression épigénétique des gènes d'ARN ribosomiques d'un des parents se met en place chez des hybrides interspécifiques de plantes et animaux (Pikaard 2000). C'est par exemple le cas des gènes ARNr d'*A. thaliana* qui sont réprimés chez l'espèce allotétraploïde *A. suecica* par méthylation de l'ADN, des modifications des histones et par l'expression de petits ARN non codants (Wang et al. 2006 ; Chen et al. 1998 ; Comai et al. 2000 ; Wang et al. 2004 ; Chen & Pikaard, 1997). Le même phénomène de régulation des ARNr a été observé chez les colzas re-synthétisés avec répression des sites NOR provenant du génome C du chou (Ksiazczyk et al. 2011).
2. Les changements dans l'expression de gènes correspondent souvent à des altérations dans des réseaux biologiques donnés. Les gènes qui s'expriment généralement de façon non-additive se trouvent ainsi plus fréquemment dans des catégories fonctionnelles liées à l'énergie, le métabolisme, la réponse aux stress ou la signalisation des phytohormones. Cette tendance a été observée chez des allopolyploïdes d'*Arabidopsis* (Wang et al. 2006), de blé (Pumphrey et al. 2009), de coton (Hovav et al. 2008), et chez des hybrides de maïs (Guo et al. 2004).
3. Les changements dans l'expression des gènes à l'échelle du génome chez des hybrides interspécifiques et des allopolyploïdes peuvent être le résultat de la divergence des mécanismes de régulation en *cis* et en *trans* entre les espèces progénitrices (Wittkopp et al. 2004 ; Tirosh et al. 2009).

La comparaison de l'ampleur des modifications affectant l'expression des gènes chez des hybrides interspécifiques et chez des allopolyploïdes indique que ce serait la rencontre entre deux génomes divergents qui causerait la plupart des variations d'expression, le doublement chromosomique n'ayant apparemment qu'un effet plus secondaire.

1.6.2.b. Modifications au niveau du protéome chez des hybrides et des allopolyploïdes

Différentes études ont étudié les modifications des profils protéomiques intervenant suite à des phénomènes d'hybridation, notamment intraspécifique (chez le maïs – Hoecker et al. 2008 ; Marcon et al. 2010 ; Dahal et al. 2012 ; le riz – Wang et al. 2008) mais relativement peu d'études se sont intéressées aux modifications de l'expression des protéines chez les allopolyploïdes.

De façon pionnière dans le domaine, des études de protéomique comparative ont été menées au laboratoire sur les mêmes colzas synthétiques que ceux que j'ai utilisés au cours de ma thèse. Elles ont démontré que des variations importantes d'expression se faisaient spécifiquement au niveau du protéome (Albertin et al. 2006 ; Marmagne et al. 2010). L'étude comparative des protéomes des premières générations de colzas synthétiques (hybride F1 'AC' et générations S0 et S1 'AACC') avec ceux de leurs parents diploïdes, a montré que plus d'un tiers des protéines exprimées dans la tige et la racine (38% et 25%, respectivement) présentaient une non-additivité significative de leur quantité par rapport à l'attendue à la MPV (Albertin et al. 2006). Ces variations quantitatives apparaissaient ainsi relativement nombreuses, alors que les variations qualitatives (apparition ou disparition de spots protéiques) étaient rares : environ 1% des spots analysés. De plus, il est apparu que c'était essentiellement l'étape d'hybridation interspécifique qui était à l'origine de ces variations, puisque 89% des variations ont été détectées dès l'hybride F1, le reste des écarts à l'hypothèse d'additivité apparaissant au cours des générations S0 et S1. Ceci est en accord avec le fait que très peu de modifications ont été observées chez les autopolyploïdes de *Brassica oleracea* (Albertin et al. 2005), renforçant l'importance de la rencontre entre deux génomes différenciés plutôt que le doublement du génome. L'essentiel des variations observées se sont avérées reproductibles entre quatre lignées de colzas synthétiques obtenues indépendamment à partir des mêmes progéniteurs, traduisant la mise en place d'une régulation de l'expression des gènes dans le génome polyploïde de manière non-stochastique. Par contre, la nature des protéines présentant un écart à l'additivité (identification des protéines obtenue par spectrométrie de masse) n'était liée ni à la catégorie fonctionnelle ni à la localisation subcellulaire (Albertin et al. 2007).

Par la suite, des études de PCR semi-quantitative et quantitative ont été menées sur une centaine de gènes codant les protéines non-additives. Seulement un tiers de ces transcrits montraient une modification dans leur expression dès l'étape de transcription (Marmagne et al. 2010). Ainsi, les régulations post-transcriptionnelles semblaient être responsables de plus des deux tiers de la non-additivité observée au niveau du protéome.

Depuis, d'autres études ont été menées chez des autopolyploïdes et allopolyploïdes d'*Arabidopsis*, par analyse des protéomes des feuilles (Ng et al. 2012), ou encore chez l'espèce de cotonnier *Gossypium barbadense* par étude des protéomes des fibres de coton, en comparant l'espèce sauvage et domestiquée pour caractériser l'impact de la domestication de cette espèce allotétraploïde sur la production de coton (Hu et al. 2013).

1.6.3. Modifications épigénétiques

Des changements épigénétiques suite à un événement d'allopolyploïdisation ont été observés chez plusieurs espèces ; ils concernent principalement la méthylation de l'ADN et les changements de conformation de la chromatine par la mise en place de modifications transcriptionnelles ou post-transcriptionnelles impliquant les petits ARN non codants.

1.6.3.a. Modifications au niveau de la méthylation de l'ADN

Chez la spartine, Salmon et al. (2005) ont utilisé une approche MSAP (*methylation-sensitive amplification polymorphism*) pour analyser les profils de méthylation des deux hybrides indépendants, *Spartina x towsendii* et *Spartina neyrautii*, tous les deux issus du croisement naturel entre *Spartina alterniflora* et *Spartina maritima*. Ces auteurs ont également étudié les profils de méthylation de *Spartina anglica* qui s'est formée à partir du doublement chromosomique de l'hybride *Spartina x towsendii*. Chez les hybrides et l'allopolyploïde, les profils de méthylation sont apparus modifiés d'environ 30% par rapport à ceux observés chez les parents. Ces changements épigénétiques pourraient être à l'origine de la plasticité et l'amplitude écologique de *Spartina anglica* (Ainouche et al. 2004 ; Salmon et al. 2005).

En réalité, l'ampleur de ces modifications de méthylation est très variable selon les espèces. Chez des colzas synthétiques, par exemple, environ 6% des fragments étudiés (par technique RFLP utilisant une enzyme sensible à la méthylation) ont présenté une modification de leur profil de méthylation dès la première génération (Lukens et al. 2006). Par contre, chez des cotons synthétiques, une approche MSAP a mis en évidence une additivité quasi-totale des profils de méthylation (Liu et al. 2001).

Les éléments transposables sont souvent des sites privilégiés d'hyper-méthylation de l'ADN (comme nous le verrons dans le chapitre 3). Lorsqu'ils se trouvent proches des gènes, cette méthylation peut aussi affecter ces derniers, en entraînant leur méthylation (par *spreading*) et en modifiant ainsi leur expression. Un exemple connu est celui observé chez les fleurs belle-de-jour (*Convolvulus tricolor*), où la méthylation d'un élément transposable présente un impact fort sur l'expression du gène *Dfr-B*, qui contrôle la couleur des fleurs, provoquant des panachures au niveau des pétales (Iida et al. 2004 ; Figure 1.10.).

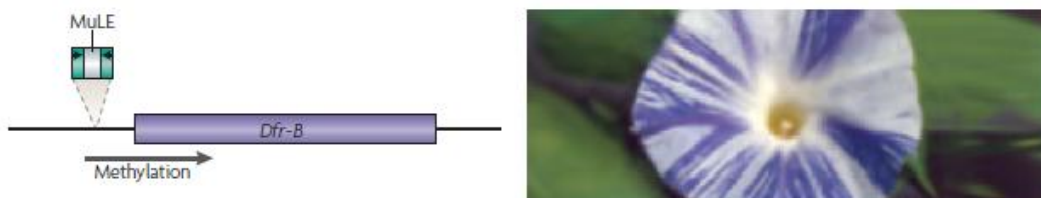


Figure 1.10. Impact de la méthylation d'un élément transposable sur l'expression génique. Chez les fleurs belle-de-jour, la méthylation de l'ADN d'un transposon non-autonome, MuLE, peut se propager jusqu'au promoteur d'un gène contrôlant la couleur des fleurs (*Dfr-B*), et provoquer ainsi des rayures dans la coloration des pétales. D'après Slotkin & Martienssen, 2007.

C'est ainsi que des études ont été menées spécifiquement chez les polyploïdes, visant à caractériser les niveaux de méthylation des éléments transposables suite à un événement d'allopolypléidisation (Kashkush & Khasdan, 2007 ; Parisod et al. 2009). Par exemple, chez la spartine *Spartina anglica*, une approche MSAP a mis en évidence de nombreux changements de méthylation de l'ADN aux sites d'insertion des éléments transposables, surtout en lien avec l'étape de confrontation des génomes différenciés (Parisod et al. 2009). Il est ainsi possible que ces changements de méthylation des ET soient à l'origine d'une partie des modifications observées au niveau de la régulation de l'expression des gènes.

1.6.3.b. Modifications transcriptionnelles et post-transcriptionnelles par l'intermédiaire des petits ARN non codants

En ce qui concerne les petits ARN non codants, seules quelques études relativement récentes ont montré que ces éléments pouvaient s'exprimer de façon non-additive chez des hybrides (comme par exemple de riz – He et al. 2010 ; ou chez la tomate – Shivaprasad et al. 2012) et des allopolypléïdes (chez *A. suecica* – Ha et al. 2009 ; et chez le blé : Kenan-Eichler et al. 2011).

Ces aspects sont spécifiquement développés dans le chapitre 3, entièrement dédié à l'étude des petits ARN non codants.

1.7. Le genre *Brassica* comme modèle pour l'étude de l'allopolypléïdie

Le genre *Brassica* a été marqué par plusieurs événements de polypléïdisation au cours de son évolution. Il a non seulement connu les mêmes événements qu'*Arabidopsis* (α , β , γ ; voir Figure 1.2.) mais aussi des duplications plus récentes qui lui sont spécifiques. Après la séparation entre *Arabidopsis* / *Brassica*, estimée à 25 - 40 millions d'années (Franzke et al. 2011), un événement d'hexaploïdie a affecté le genre *Brassica*, comme en atteste la tripléication complète du génome de l'ensemble de ces espèces (les espèces diploïdes comme *B. rapa* et *B. oleracea* sont ainsi considérées comme des mésopolypléïdes). Cet événement aurait eu lieu il y a environ 13 - 17 millions d'années (Lysak et al. 2005 ; Wang et al. 2011). Ensuite, il y a moins de 10,000 ans, plusieurs événements d'allopolypléïdisation ont donné naissance à trois espèces allotétraploïdes, le colza (*B. napus*, AACCC, $2n=38$), la moutarde brune (*B. juncea*, AABB, $2n=36$) et la moutarde d'Abyssinie (*B. carinata*, BBCC, $2n=34$), à partir du croisement de trois espèces diploïdes, le chou (*Brassica oleracea*, CC, $2n=18$), la navette (*Brassica rapa*, AA, $2n=20$) et la moutarde noire (*Brassica nigra*, BB, $2n=16$). Ces relations ont été décrites en 1935 par U, et on les représente couramment sous forme de triangle, le « triangle de U » (Figure 1.11).

Le genre *Brassica* constitue ainsi un excellent modèle pour l'étude de la polyploïdie et plus précisément de la formation d'un organisme allopolyploïde (pour revue, Jenczewski et al. 2013). Au cours de ma thèse, le colza a ainsi été mon modèle d'étude.

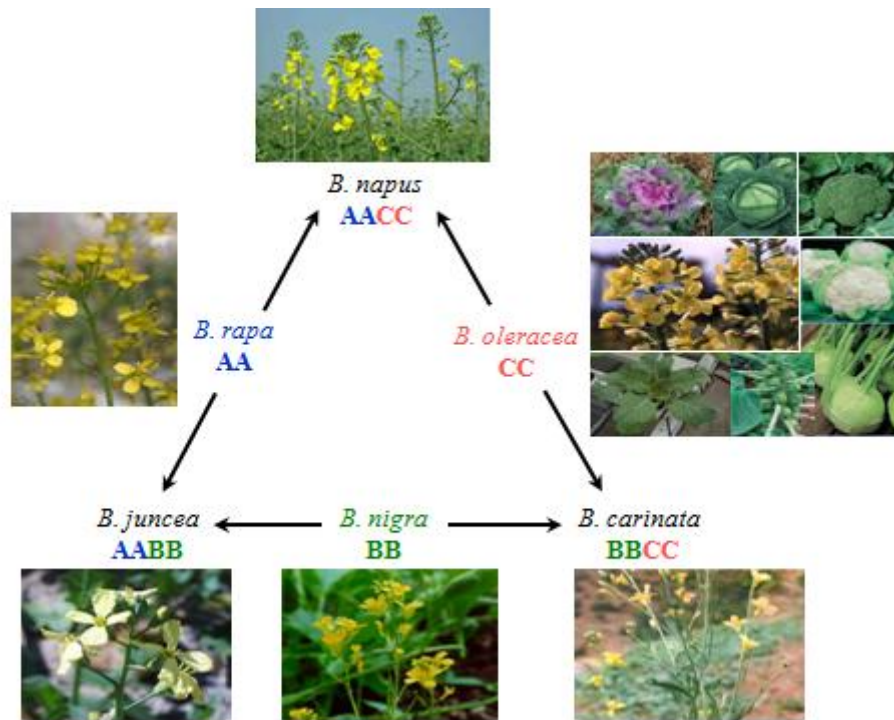


Figure 1.11. Le triangle de U. Représentation de la formation de trois espèces allotétraploïdes (*B. napus*, *B. juncea* et *B. carinata*) à partir de divers croisements entre trois espèces diploïdes (*B. oleracea*, *B. rapa* et *B. nigra*). D'après U, 1935.

1.7.1. Mon modèle d'étude : le colza (*Brassica napus*)

Le colza est une plante annuelle de la famille des Brassicacées, qui est largement cultivée pour la production d'huile alimentaire, pour l'alimentation animale et plus récemment pour la production de biocarburants. On trouve sous la dénomination de *B. napus* deux formes cultivées bien distinctes : (1) les formes oléagineuses ou fourragères appelées colza, et (2) les formes tubéreuses nommées rutabagas. Le colza est une espèce cultivée d'importance économique majeure ; en 2012, plus de 65 millions de tonnes de colza ont été produites dans le monde (<http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E>). Le colza est la seconde espèce oléagineuse cultivée dans le monde, derrière le soja, et la première en Europe.

Le colza (*Brassica napus*, génome AACC, $2n=38$) est une plante allotétraploïde issue de croisements entre la navette (*Brassica rapa*, génome AA, $2n=20$) et le chou (*Brassica oleracea*, génome CC, $2n=18$). Cependant l'origine de cette hybridation, de même que l'identification précise des sous-espèces à l'origine du colza, ne sont pas encore complètement élucidées. L'analyse des génomes chloroplastiques de ces trois espèces a révélé que le colza actuel résulte d'au moins deux (Cifuentes et al. 2010) voire trois événements indépendants d'hybridation interspécifique (Allender & King, 2010) impliquant *B. rapa* comme parent femelle.

Le colza présente plusieurs avantages pour l'étude de la formation d'une espèce allopolyploïde :

- D'abord, sa proximité vis-à-vis de l'espèce modèle *Arabidopsis thaliana* permet de bénéficier de toutes les connaissances acquises chez cette dernière, dont le génome a été le premier à être complètement séquencé et annoté (The *Arabidopsis* Genome Initiative 2000). L'analyse des premières régions séquencées de colza a mis en évidence une très bonne conservation de la colinéarité vis-à-vis d'*Arabidopsis* (Mun et al. 2010). De plus, l'identité nucléotidique entre les séquences codantes est très forte (environ 87%).
- De plus, il est possible de resynthétiser des colzas au laboratoire (on parle de colzas néo-synthétisés ou synthétiques) à partir des espèces diploïdes *B. rapa* et *B. oleracea* afin de mimer la formation du colza, et d'analyser en détail les mécanismes qui accompagnent la rencontre des deux génomes progéniteurs (hybridation interspécifique) puis le doublement chromosomique des génomes parentaux (allopolyploïdisation). On peut d'ailleurs noter que les colzas synthétiques ont constitué l'un des tout premiers modèles mettant en évidence la dynamique des génomes polyploïdes (Song et al. 1992, 1995) et qu'ils restent encore aujourd'hui l'un de ceux pour lesquels on dispose du plus grand nombre de résultats décrivant les conséquences sur le génome, le transcriptome, le protéome ou le phénotype d'un événement de polyploïdie (pour revue, Jenczewski et al. 2013).

A tous ces titres, le colza constituait donc un modèle particulièrement adapté pour aborder la problématique qui m'a intéressée au cours de ma thèse : comprendre comment les composants non codants du génome, et notamment les éléments transposables et les petits ARN, répondent à un événement d'allopolyploïdie.

1.7.2. Mon matériel d'étude : des colzas néo-synthétisés (synthétiques)

Les colzas synthétiques que j'ai étudiés ont été produits par Anne-Marie Chèvre et collaborateurs à l'INRA de Rennes. Ils ont été obtenus à partir de huit croisements indépendants impliquant plusieurs parents diploïdes (Tableaux 1.1. et 1.2.). Les génotypes parentaux *B. oleracea* 'HDEM' et *B. rapa* 'Z1' sont des haploïdes doublés, donc complètement homozygotes. Les génotypes de *B. rapa* C₁₋₃, et de *B. oleracea* C10 et RC ont été produits par des autofécondations successives, conservant donc un taux résiduel d'hétérozygotie.

Tableau 1.1. Parents utilisés pour l'obtention de colzas synthétiques

Espèces	2n	Sous-espèces	Nom commun	Génotype
<i>B.rapa</i> (AA)	20	<i>rapifera</i>	navet	C ₁₋₃
		<i>oleifera</i>	navette d'hiver	Z1
		<i>botrytis italica</i>	brocoli	HDEM
<i>B.oleracea</i> (CC)	18	<i>acephala</i>	chou frisé	C10
		<i>alboglabra</i>	kailan	RC

Tableau 1.2. Croisements réalisés pour l'obtention de colzas synthétiques

Code du croisement*	Parents (maternel-paternel)	Nombre de croisements indépendants
BoEMZ	HDEM x Z1	5
BoRCC	RC x C ₁₋₃	1
BrCRC	C ₁₋₃ x RC	1
BoC10Z1	C10 x Z1	1

*D'après Szadkowski et al. 2010.

La Figure 1.12. montre le schéma d'obtention général des colzas synthétiques. Ces derniers ont été produits, lorsque cela était possible, de deux façons différentes : (1) par doublement du stock chromosomique des hybrides interspécifiques par traitement à la colchicine, et (2) en passant par gamètes non réduits lorsque les hybrides F1 en produisaient. Quand tel était le cas, les gamètes non réduits produits par les hybrides F1 ont été utilisés pour mener des croisements (Figure 1.12., voie de gauche) avec les allotétraploïdes de première génération (S0) obtenus par doublement à la colchicine (Figure 1.12., voie de droite). Disposer de ces deux voies d'obtention a permis de montrer que le mode de formation d'un polyplôïde avait un impact non négligeable sur le comportement des polyplôïdes obtenus (Szadkowski et al. 2011). L'obtention des hybrides a été une tâche particulièrement difficile, avec un taux de réussite de 9 hybrides sur 2556 pollinisations. Ces contraintes expliquent que nous ne disposons pas de croisements réciproques pour l'ensemble des combinaisons.

Les allotétraploïdes de première génération (S0) ont été autofécondés successivement pour produire les générations ultérieures. A chaque génération, un tri a été fait : seules ont été conservés les plantes présentant 38 chromosomes (nombre chromosomique attendu d'un colza) ; toutes les autres plantes ont été éliminées et n'ont donc pas servi à produire la génération suivante. Au final, l'ensemble de ce matériel nous a permis de comparer les parents, les hybrides F1 et les allotétraploïdes synthétiques des générations d'autofécondation successives de S0 à S5.

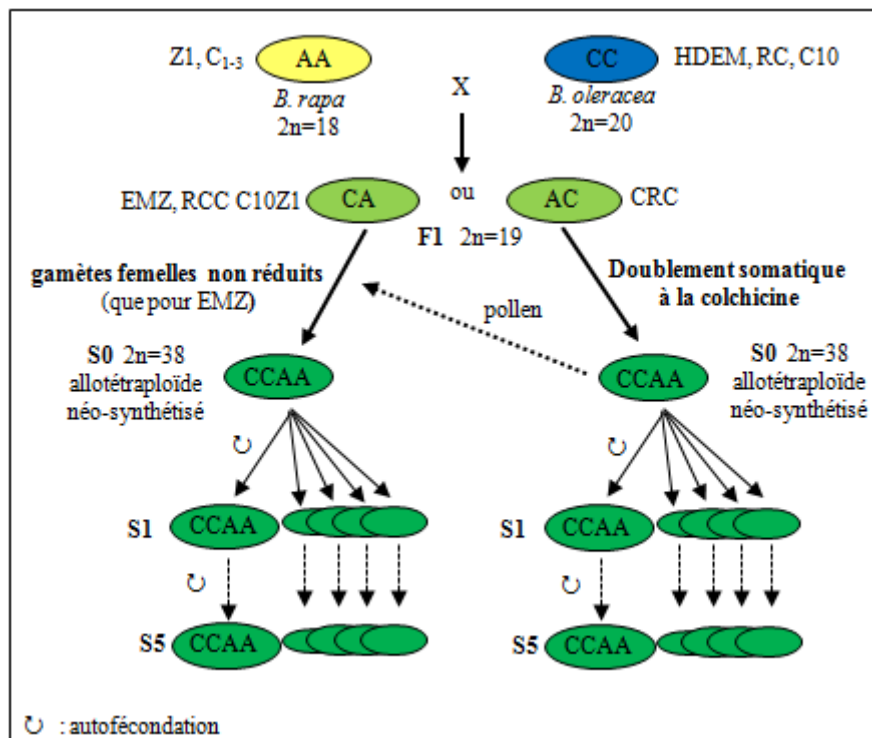


Figure 1.12. Schéma du principe d'obtention de colzas synthétiques. Les différents parents (*B. oleracea* et *B. rapa*) sont utilisés pour créer par croisement des hybrides interspécifiques AC. Ces derniers ont été traités à la colchicine afin de doubler leur stock chromosomique. On obtient ainsi des colzas synthétiques (nommés S0) qui sont autofécondés pendant plusieurs générations. Les hybrides ont également été croisés, comme femelle, avec les colzas synthétiques correspondants, mettant à profit la production (rare) de gamètes non réduits AC par les hybrides interspécifiques AC. D'après K. Alix 2013.

L'ensemble de ces différents colzas synthétiques (jusqu'en génération S2) et leurs progéniteurs diploïdes a servi comme matériel végétal à l'étude décrite au chapitre 2. L'objectif y était d'évaluer l'impact de l'allopolyploïdie sur l'activité de transposition de trois éléments transposables. Je me suis attachée à **caractériser la réactivation transpositionnelle potentielle d'un élément de classe II, le transposon CACTA *Bot1*** (Alix et al. 2008). Ce transposon étant spécifique du génome C du chou, je me suis interrogée quant à la réactivation de *Bot1*, du fait de la confrontation avec un génome différencié, le génome A, qui en est quasiment dépourvu.

Puis seuls les croisements impliquant les deux parents strictement homozygotes HDEM et Z1 ont été utilisés pour l'étude que j'ai menée sur les petits ARN non codants décrite dans le chapitre 3. Le matériel végétal était alors issu de trois croisements indépendants entre *B. oleracea* cv. HDEM et *B.rapa* cv. Z1. La Figure 1.13. présente des photos des parents diploïdes et des allotétraploïdes synthétiques des générations S0 à S5 pour ce croisement nommé (de manière générale) « EMZ ». Ces colzas synthétiques ont ainsi été utilisés pour **évaluer les changements de l'expression des petits ARN, à une échelle globale, suite à l'événement d'allopolyplôidie du colza**. Observer des changements en réponse à l'allopolyplôidie permettrait d'envisager l'implication de ces petits ARN non codants dans les modifications génomiques structurales et fonctionnelles survenant chez les colzas nouvellement créés.

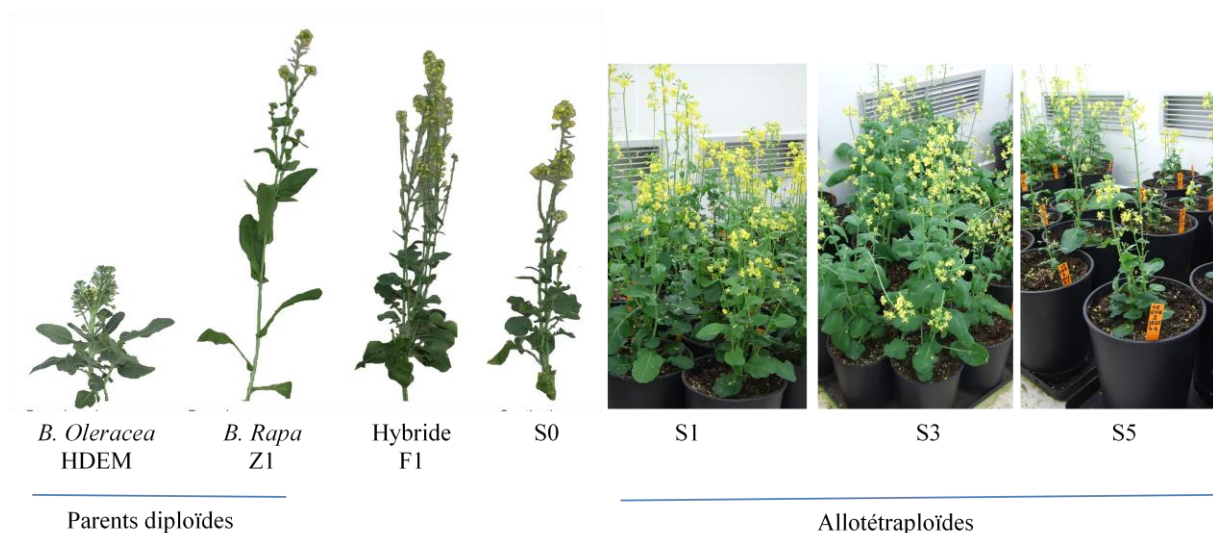


Figure 1.13. Description du matériel végétal *Brassica* utilisé. Photos des parents diploïdes (*B.oleracea* HDEM et *B. rapa* Z1) utilisés pour l'obtention d'un hybride F1, qui a été traité à la colchicine pour l'obtention des allotétraploïdes S0 (Photos de Albertin et al. 2006). Des autofécondations successives ont donné les générations suivantes, de la S1 à la S5 (Photos ©Karine Alix).

CHAPITRE 2

**L'allopolyploïdie n'a qu'un impact modéré
sur l'activation transpositionnelle des
éléments transposables**

L'allopolyploïdie n'a qu'un impact modéré sur l'activation transpositionnelle des éléments transposables

Le « choc génomique » que constituerait la réunion de deux génomes différenciés pour former un organisme allopolyploïde a généralement été associé à une forte activité de transposition des éléments transposables (pour revue, Madlung & Comai 2004). Cette hypothèse avait été proposée par Barbara McClintock dans les années 80s (McClintock 1984). Mes travaux de thèse ont ainsi été l'occasion de contribuer à une étude visant à tester cette hypothèse et à mesurer l'impact de l'allopolyploïdie sur l'activation transpositionnelle des éléments transposables chez les *Brassica* lors de la formation de colzas néo-synthétisés.

2.1. Les éléments transposables, un composant majeur du génome végétal

Les éléments transposables (ET) sont des fragments d'ADN mobiles au sein du génome. Ils sont présents dans tous les organismes vivants, et particulièrement chez les plantes où ils représentent une fraction parfois très grande du génome : au moins 21% du génome d'*Arabidopsis thaliana* (Ahmed et al. 2011) considéré comme le plus petit génome végétal, et plus de 85% du génome de maïs (Schnable et al. 2009) pour atteindre jusqu'à 90% du génome du blé (Yaakov & Kashkush, 2011). Chez *Brassica rapa*, premier génome de *Brassica* entièrement séquencé, il a été estimé que les ET représentaient environ 40% du génome (Wang et al. 2011). Diverses études menées sur l'évolution du contenu en éléments transposables des génomes de *Brassica* ont permis de montrer le rôle majeur des ET dans la dynamique évolutive des *Brassica*. Certains ET ont ainsi accompagné la divergence des espèces et la mise en place de leur spécificité (Alix et al. 2008) et certains semblent même avoir été impliqués dans les mécanismes d'évolution et/ou de régulation des gènes (Alix et al. 2008 ; Sarilar et al. 2011).

On distingue deux classes principales d'ET selon leur mode de transposition (Wicker et al. 2007, Figure 2.1.) :

- Les éléments de classe I, ou rétrotransposons, se déplacent dans le génome selon un mécanisme de « copier-coller » par l'intermédiaire d'une copie ARN, elle-même rétro-transcrite en copie ADN (grâce à la transcriptase inverse) qui est alors insérée en un nouveau site du génome.
- Les éléments de classe II, ou transposons à ADN, se déplacent par l'intermédiaire d'une copie ADN. Cela implique (i) un mécanisme de « couper-coller » avec l'excision d'une copie de l'élément (par une transposase) suivie de l'intégration de cette même copie en un autre site du génome, ou (ii) par répllication directe d'une copie ADN (cas notamment des héliçons, par le biais d'une hélicase).

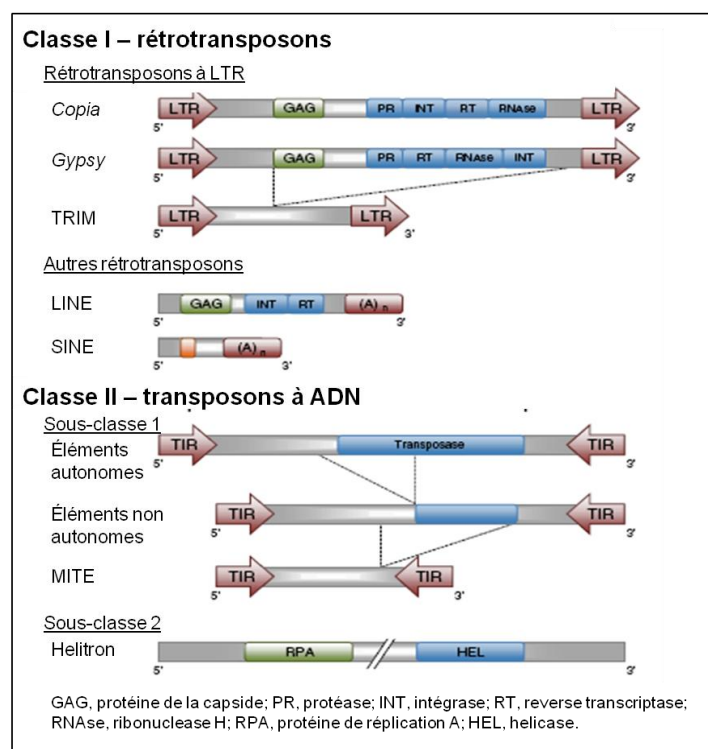


Figure 2.1. Classification des principaux éléments transposables des génomes végétaux. Classe I. Les rétrotransposons à LTR (*Long Terminal Repeats*) sont divisés en *Copia* et *Gypsy* selon l'ordre des gènes qui codent les protéines nécessaires au cycle de transposition. Les rétrotransposons à répétition terminale en miniature (TRIMs) ne possèdent plus de région codante et sont par conséquent non-autonomes. Les rétrotransposons sans LTR regroupent les LINEs et les SINEs (*Long/Short Interspersed Nuclear Elements*). Classe II. La plupart des transposons sont bordés par des répétitions inversées terminales (TIRs, *Terminal Inverted Repeats*) et peuvent être autonomes ou non-autonomes (s'ils contiennent ou non une transposase fonctionnelle). Les éléments transposables à répétitions inversées miniatures (MITEs, *Miniature Inverted-repeat Transposable Elements*) possèdent des caractéristiques des deux classes : ils ont une structure d'éléments de classe II mais défectifs, et peuvent être présents en un grand nombre de copies comme les éléments de classe I. Les héliçons sont des ET qui se répliquent par un mécanisme de « cercle roulant » semblable au mécanisme connu chez les bactéries. Adapté d'après Wicker et al. 2007 et Parisod et al. 2010.

2.2. Impact de l'allopolyploïdie sur l'activation des éléments transposables chez le colza : le cas du transposon *Bot1*

Les éléments transposables, en lien avec leur mobilité dans le génome, représentent une source importante de diversité pour l'expression du génome, mais comme tout agent mutagène, sont responsables de mutations délétères. Ils sont donc maintenus réprimés dans le génome hôte, et leurs activités sont fortement contrôlées par divers mécanismes de régulation épigénétique et d'hétérochromatinisation de l'ADN, i.e. de compactage de la chromatine où sont localisés ces éléments transposables (Slotkin & Martienssen 2007, Lisch 2009). L'hybridation interspécifique et l'allopolyploïdie, à l'origine d'un choc génomique important, s'accompagneraient alors d'une déméthylation plus ou moins globale du génome, qui pourrait entraîner une levée du contrôle des éléments transposables et, par conséquent, une activation transpositionnelle de ces ET maintenus jusque-là réprimés (Madlung & Comai, 2004).

Nous avons mis en place une étude en vue de tester l'hypothèse d'une activation des ET consécutive à l'allopolyploïdisation, en utilisant des colzas néo-synthétisés (encore dits colzas synthétiques) – cf. chapitre 1. L'activation accompagnée de transposition d'un ET doit se traduire par l'apparition de polymorphisme d'insertion entre les colzas synthétiques (*B. napus*, AACC) et leurs parents diploïdes (*B. rapa* AA et *B. oleracea* CC). La technique SSAP (*Sequence-Specific Amplification Polymorphism*) décrite pour la première fois par Waugh et al. (1997) permet de mettre en évidence ce type de variation génétique. Nous avons ainsi choisi d'analyser l'impact immédiat de l'allopolyploïdie sur trois éléments transposables, sélectionnés pour leurs caractéristiques contrastées en termes de mécanisme de transposition, nombre de copies dans le génome et distribution génomique, par comparaison entre les profils SSAP générés pour les colzas synthétiques et ceux des parents diploïdes.

Le travail que j'ai conduit dans le cadre de cette étude a concerné l'analyse complète de l'impact de l'allopolyploïdie sur l'un des trois ET sélectionnés, à savoir l'élément transposable *Bot1* (Alix et al. 2008). Il s'agit d'un ET de la superfamille CACTA ; celle-ci regroupe des transposons à ADN (classe II) de l'ordre des éléments à TIR (*terminal inverted repeats* – Figure 2.1) qui, commençant par le motif « CACTA », sont à l'origine de leur nom. Ce transposon spécifique du génome C de *Brassica*, désigné *Bot1* pour '*Brassica oleracea transposon 1*', a été démontré comme impliqué dans la divergence des génomes A et C des *Brassica*. Il semblerait ainsi qu'à partir d'un élément ancestral commun aux *Brassica*, dont le nombre de copies serait estimé à environ 30 dans le génome de *B. rapa*, il y ait eu divergence de séquences spécifiquement au sein du génome C puis amplification pour atteindre des taux d'au moins 1200 copies chez *B. oleracea* uniquement. L'analyse cytogénétique de la localisation de *Bot1* au sein du génome du colza a mis en évidence un maintien de cette spécificité au sein du génome allotétraploïde.

Mon étude avait comme objectif d'évaluer si la spécificité génomique de *Bot1* pour le génome C pouvait favoriser l'activation de ce transposon lors de l'allopolypléidisation du colza, en l'absence de mécanismes capables de contrôler son activation dans le génome A de *B. rapa*. Même si la majorité des profils SSAP sont apparus additifs entre les parents diploïdes et les colzas synthétiques correspondants (pas de différence au niveau des bandes amplifiées), je me suis intéressée à l'étude détaillée des cas de non-additivité où les bandes SSAP correspondantes pouvaient marquer des sites candidats de nouvelles insertions de *Bot1*. De façon intéressante, les cas de non-additivité détectés ont été les plus nombreux en génération S2 des colzas synthétiques, soit après deux générations d'autofécondation. J'ai procédé à la caractérisation par clonage et séquençage de trente bandes SSAP non-additives, afin de déterminer l'origine du polymorphisme observé par SSAP. La plupart des bandes candidates (67%) ont ainsi montré des profils additifs suite aux vérifications réalisées par PCR. Néanmoins, j'ai pu mettre en évidence quelques réarrangements structuraux et un total de trois événements de transposition (1 nouvelle insertion et 2 excisions) de *Bot1* qui ne concernaient d'ailleurs que des individus de génération S2. Il est à noter que ces transpositions ont été les seules mises en évidence chez les colzas synthétiques, et donc associées au plus jeune des ET étudiés (*Bot1* étant spécifique du génome C).

Mon travail a pu mettre en évidence les capacités de *Bot1* à transposer au sein du génome de *Brassica*. Ce résultat nous amène à nous interroger quant à l'absence d'envahissement du génome A par *Bot1*, qui aurait pu survenir lors de la formation du colza naturel (phénomène rencontré chez le coton pour un rétrotransposon spécifique d'un des génomes parentaux, Zhao et al. 1998). Une des explications à ce phénomène serait l'existence de mécanismes de répression spécifiques de *Bot1* qui se mettrait en place au cours de l'événement d'allopolypléidie, empêchant une activation anarchique de ce transposon et permettant ainsi la formation d'un nouveau génome allopolypléide stable. Reste à identifier la nature de ces mécanismes de régulation, où les petits ARN, connus comme étant des acteurs clés du contrôle de la transposition des ET, pourraient jouer un rôle important.

L'ensemble de l'étude, correspondant en majorité au travail de thèse de Véronique Sarilar qui a analysé l'impact de l'allopolypléidie sur les deux autres ET sélectionnés (un rétrotransposon de type *Athila* – Alix et al. 2005, et le MITE *BraSto* – Sarilar et al. 2011) a été valorisé par publication d'un article où j'apparais comme deuxième auteur. La fiche-résumé de l'article et l'article lui-même sont donnés ci-après.

2.3. Sarilar et al. (2013), *New Phytologist* 198: 593-604

Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids

Véronique SARILAR, Paulina MARTINEZ PALACIOS, Agnès ROUSSELET, Céline RIDEL, Matthieu FALQUE, Frédérique EBER, Anne-Marie CHEVRE, Johann JOETS, Philippe BRABANT and Karine ALIX

Contexte et objectifs : L'hybridation interspécifique et en particulier l'allopolyploïdie jouent un rôle important dans la formation et l'adaptation de nouvelles espèces. Cependant l'hybridation entre deux génomes différenciés est depuis longtemps considérée comme un « choc génomique » qui pourrait entraîner l'activation des éléments transposables. Même si cette activation présenterait des conséquences importantes pour le génome, relativement peu d'études ont été réalisées chez les espèces polyploïdes afin de mesurer l'impact réel de l'allopolyploïdie sur l'activité des ET. Nous avons donc entrepris de tester l'activation transpositionnelle des ET chez les *Brassica*, suite à l'événement d'allopolyploïdie du colza par analyse de colzas synthétiques.

Méthodes : Nous avons mesuré l'impact de l'allopolyploïdie sur l'activation de trois ET choisis pour leurs caractéristiques contrastées en termes de mode de transposition, nombre de copies et localisation dans le génome, par utilisation de la technique SSAP (*Sequence-Specific Amplification Polymorphism*). Les profils SSAP générés ont été comparés entre les colzas synthétiques et leurs progéniteurs diploïdes, permettant d'étudier les modifications survenant au niveau des sites d'insertion des ET sélectionnés. Les bandes différemment amplifiées (non-additives) ont été caractérisées pour déterminer l'importance des restructurations génomiques liées aux ET ciblés.

Principaux résultats: Les profils SSAP obtenus sont apparus majoritairement additifs entre colzas synthétiques et parents diploïdes. Nous n'avons donc pas détecté de changements structuraux massifs associés aux trois éléments étudiés. Nous avons pu mettre en évidence quelques réarrangements génomiques et assez peu de transpositions, ces dernières étant spécifiques du transposon *Bot1*, le plus jeune des ET étudiés.

Conclusion : Notre étude a montré que l'activation des éléments transposables suite à un événement d'allopolyplœdie était une réponse hautement spécifique de certains ET uniquement. L'absence de polymorphisme au niveau de la structure de l'ADN, malgré l'observation de profils SSAP polymorphes entre parents et descendances (>60% des cas étudiés par clonage et séquençage), nous a amenés à faire l'hypothèse qu'une partie des différences dans les profils SSAP pouvait être liée à des changements dans les niveaux de méthylation de l'ADN. Nos résultats posent donc maintenant la question du rôle des mécanismes épigénétiques dans le contrôle des ET lors de la formation d'un nouveau génome allopolyplœide.

Mots clés

Brassica, éléments transposables, hybridation interspécifique, MITE, polyploïdie, rétrotransposon à LTR, technique SSAP, transposon CACTA.

1 Original research article / full paper

2 **Allopolyploidy has a moderate impact on restructuring at three contrasting**
3 **transposable element insertion sites in resynthesized *Brassica napus***
4 **allotetraploids**

5

6 Véronique SARILAR^{1,2}, Paulina MARTINEZ PALACIOS³, Agnès ROUSSELET⁴, Céline RIDEL⁴,
7 Matthieu FALQUE⁴, Frédérique EBER⁵, Anne-Marie CHEVRE⁵, Johann JOETS⁴, Philippe
8 BRABANT^{1,6} and Karine ALIX^{1,6*}

9

10 ¹AgroParisTech, 16 rue Claude Bernard, F-75231 Paris cedex 05, France

11 ²CNRS, ³Univ Paris-Sud, ⁴INRA: ⁶UMR 0320 / 8120 Génétique Végétale, Ferme du Moulon, F-91190
12 Gif-sur-Yvette, France

13 ⁵INRA, UMR 1349 IGEPP, BP 35327, F-35653 Le Rheu cedex, France.

14

15 *author to whom correspondence should be addressed

16 e-mail: alix@moulon.inra.fr, phone: +33/(0)1 69 33 23 72, fax: +33/(0)1 69 33 23 40

17

18 word count / main text: **6456** < 6500 words

19 word counts introduction: 982

20 materials and methods: 1502

21 results: 2003

22 discussion: 1836 (<1937, 30% of the MS)

23 acknowledgments: 133

24 Number of tables: 3

25 Number of figures: 2

26 Supporting information: 3

27

28 Figure 1 and Table 1 should be reproduced in colour

29

30 *Manuscript accepted for publication in New Phytologist*

1 **Summary**

- 2 • The role whole genome duplication (WGD) plays in evolution and adaptation is
3 particularly well illustrated in allopolyploids, where WGD is concomitant with
4 interspecific hybridisation. This “Genome Shock”, usually accompanied by structural
5 and functional modifications, has been associated with the activation of transposable
6 elements (TEs). However, the impact of allopolyploidy on TEs has been studied in
7 only a few polyploid species, and not in *Brassica* which has been marked by recurrent
8 polyploidy events.
- 9 • We developed sequence-specific amplification polymorphism (SSAP) markers for
10 three contrasting TEs, and compared profiles between resynthesized *B. napus*
11 allotetraploids and their diploid *Brassica* progenitors. To evaluate restructuring at TE
12 insertion sites we scored changes in SSAP profiles and analyzed a large set of
13 differentially amplified SSAP bands.
- 14 • No massive structural changes associated with the three TEs surveyed were detected.
15 However, several transposition events specific to the youngest TE originating from the
16 *B. oleracea* genome were identified.
- 17 • Our study supports the hypothesis that TE responses to allopolyploidy are highly
18 specific. The changes observed in SSAP profiles lead to hypothesize that they may
19 partly result from changes in DNA methylation, asking the question of the role of
20 epigenetics during the formation of a new allopolyploid genome.

21

22 **Key words**

23 *Brassica*, CACTA transposon, interspecific hybridisation, LTR-retrotransposon, MITE,
24 polyploidy, SSAP

1 Introduction

2 Polyploidy, or whole-genome duplication (WGD), is now recognised as an important
3 evolutionary force in plants, particularly in angiosperms, with all flowering plant lineages
4 having experienced one or more rounds of ancient polyploidy (Cui *et al.*, 2006; Jiao *et al.*,
5 2011). The recurrent polyploidisation events which have marked angiosperm evolutionary
6 history occurred prior to or simultaneously with major evolutionary transitions and adaptive
7 radiation of species. This striking coincidence suggests that WGD played a predominant part
8 at the origin of adaptive speciation (De Bodt *et al.*, 2005), as illustrated by the key role WGD
9 appears to have played in the survival and proliferation of plant lineages during the
10 Cretaceous-Tertiary mass extinction event (Fawcett *et al.*, 2009). Indeed, polyploidy is one of
11 the main mechanisms (with unequal crossing over and retroposition) that results in genetic
12 redundancy, which is then acted upon by selection, drift and mutation, and represents, with
13 the emergence of new gene functions, an important source of novelty (Ohno, 1970; Zhang,
14 2003). Polyploids are usually divided into two categories: autopolyploids arising within a
15 single species by chromosome doubling, and allopolyploids resulting from genome
16 duplication implying interspecific hybridisation between two or more distinct species. It is
17 currently estimated that, in natural populations, allopolyploids are more prevalent than
18 autopolyploids, even if the rate of autopolyploid formation appears higher than that of
19 allopolyploids (Ramsey & Schemske, 1998). Thus the evolutionary success of allopolyploids
20 seems to be due to the interspecific hybridisation resulting in immediate evolutionary
21 advantages (even if the question of the evolutionary advantages of genome multiplication *per*
22 *se* still remains unanswered, Parisod *et al.*, 2010b).

23 One of the main features of allopolyploids is their ‘doubled interspecific hybrid’
24 composition which fixes the heterozygosity at the origin of the well-known phenomenon of
25 heterosis, or hybrid vigour, responsible for increased growth and robustness in hybrids

1 compared to their parents (Chen, 2010). In addition, this hybridisation step - the merger of
2 differentiated complete genomes in one single cell – may constitute a major “Genome
3 Shock”, as first hypothesized by McClintock (1984). This is accompanied by rapid and
4 profound structural and functional modifications (Chen & Ni, 2006; Doyle *et al.*, 2008;
5 Hegarty & Hiscock, 2008; Soltis & Soltis, 2009; Gaeta & Pires, 2010) representing an
6 important source of novelty. Such an immediate response from neo-allopolyploid genomes
7 explains their ability to overcome competition with their parental species and colonize new
8 habitats (Hegarty & Hiscock, 2008), as well as a large extent of their adaptability and indeed
9 their evolutionary success.

10 Hybridisation and allopolyploidy have often been associated with the transposition of
11 transposable elements (TEs). TEs are a major component of plant genomes (e.g. 21% of the
12 smallest plant genome from *Arabidopsis thaliana* (Ahmed *et al.*, 2011), and over 85% of the
13 reference maize genome (Schnable *et al.*, 2009)). They are divided into two main classes
14 depending on their mode of transposition (Wicker *et al.*, 2007): class I elements or
15 retrotransposons move via reverse-transcribed RNA intermediates, and class II elements or
16 DNA transposons move via DNA intermediates. As a consequence of their repetitiveness in
17 the genome, TEs have played a major role in shaping the structure of plant genomes through
18 transposition but also recombination (Vitte & Panaud, 2005). TEs are also involved in
19 unequal and illegitimate recombination events and can therefore generate a large variety of
20 structural mutations (Devos *et al.*, 2002; Vitte *et al.*, 2011). When TEs are activated,
21 associated with their amplification and mobility across the genome, they can be a major
22 source of diversity for gene expression, acting as continuous mutagenic agents, but can also
23 be responsible for deleterious effects. Thus TEs are mainly silenced and their capacity to
24 recombine as well as their transcriptional /transpositional activity is usually controlled by
25 heterochromatinisation and epigenetic host regulatory machineries (Slotkin & Martienssen

1 2007, Lisch 2009). Two major natural phenomena are assumed to reactivate TEs: (i) biotic
2 and abiotic stresses (Grandbastien 1998, Kalendar *et al.*, 2000), and (ii) interspecific
3 hybridisation and allopolyploidy in response to the “Genome Shock” accompanied by
4 extensive demethylation which could potentially lead to derepression and activation of
5 previously silenced TEs (Madlung & Comai, 2004). Nevertheless, the effective impact of
6 allopolyploidy on TEs has been studied in only a few specific polyploid systems (reviewed in
7 Parisod *et al.*, 2010a), and surprisingly has not yet been surveyed in any polyploid *Brassica*
8 species.

9 The evolutionary history of the *Brassica* genus has been particularly marked by
10 polyploidy. In addition to the α (3R) polyploidy event which is unique to the Brassicaceae
11 family (Barker *et al.*, 2009), an ancient triplication was shown to be unique to the Brassiceae
12 tribe, with strong evidence from various approaches for genome triplication in diploid
13 *Brassica* species (Schranz *et al.*, 2006). Finally, additional rounds of allopolyploidisation
14 occurred recently (less than 0.01 million years ago), with *B. napus* (AACC, $2n=4\times=38$) in
15 particular originating from natural interspecific hybridisation between the diploid species *B.*
16 *rapa* (AA, $2n = 2\times = 20$) and *B. oleracea* (CC, $2n = 2\times = 18$). Consequently, this
17 allotetraploid *Brassica* species and its progenitors represent a good model for investigating
18 the impact of allopolyploidy on TEs, in order to identify potential structural modifications
19 linked to TE derepression. Resynthesized *B. napus* allotetraploids have been successfully
20 used to characterize the regulation of gene expression following allopolyploidy by
21 comparative proteomics (Albertin *et al.*, 2006) and transcriptomics (Gaeta *et al.*, 2009;
22 Marmagne *et al.*, 2010). The extent of structural changes occurring during the formation of
23 the allopolyploid genome has also been investigated (Gaeta *et al.*, 2007; Szadkowski *et al.*,
24 2010, 2011). In the present study, we analyzed the same plant material to characterize the
25 impact of allopolyploidy on three contrasting TEs (considering their genomic localisation,

1 copy number in the genome, and mechanism of transposition). We developed a transposon
2 display approach to identify the extent and nature (transposition vs. genomic rearrangement)
3 of the structural modifications which can be triggered by these TEs in the early generations of
4 a newly formed *Brassica* allopolyploid genome.

5

6 **Materials and Methods**

7 **Plant material and DNA extraction**

8 The resynthesized *B. napus* allotetraploids we analyzed were produced by A.-M. Chèvre *et al.*
9 and used in previous studies (as specified below) – *nb*: all the resynthesized *B. napus* lines
10 and the different diploid *Brassica* genotypes (except Z1) are the property of INRA Le Rheu,
11 France (contact A.-M. Chèvre). Three homozygous doubled haploid lines, ‘HDEM’ *B.*
12 *oleracea* var. *botrytis italica*, ‘RC34’ *B. oleracea* var. *alboglabra*, ‘Z1’ *B. rapa* ssp. *oleifera*
13 (with Z1 provided by K.C. Falk, Agriculture and Agri-Food Canada), and two selfed
14 progenies of ‘C_{1.3}’ *B. rapa* ssp. *rapifera* and ‘C10’ *B. oleracea* var. *acephala* were used as
15 diploid progenitors. Reciprocal crosses between RC34 and C_{1.3} led to the synthetic
16 allotetraploids BoRCC (with RC34 as the maternal parent) and BrCRC (with C_{1.3} as the
17 maternal parent) (Szadkowski *et al.*, 2010). Another cross between C10 (maternal parent) and
18 Z1 resulted in a biannual synthetic allotetraploid BoC10Z1 (specifically analysed in the
19 present study). Five independent crosses between HDEM and Z1 led to the synthetic
20 allotetraploids BoEMZ (Albertin *et al.*, 2006; Marmagne *et al.*, 2010; Szadkowski *et al.*,
21 2010). Three of the five S0 synthetic lines (obtained from colchicine treatment of the first
22 hybrids F1), namely BoEMZ1₃₈, BoEMZ2₃₈ and BoEMZ3₃₈, were then used each as paternal
23 parent in crosses with the corresponding hybrid F1 to produce three other synthetic
24 allotetraploids obtained from female unreduced gametes, designated BoEMZ1_{19,38},

1 BoEMZ2_{19,38} and BoEMZ3_{19,38} (Szadkowski *et al.*, 2011). Therefore, our plant material
2 consisted of 11 independently resynthesized *B. napus* allotetraploids and their progenies,
3 represented by a total of 109 individual plants comprising the first hybrids F1 (CA or AC), the
4 corresponding doubled hybrids S0 (CCAA or AACC), and the following generations of
5 selfing S1 and S2. All S0, S1 and S2 plants were verified for chromosome numbers $2n=38$
6 using flow cytometry and classical cytogenetics. Details of the plant material used are given
7 in Table 1. Three cuttings per plant, randomized in three blocks, were grown under controlled
8 conditions (18°C during 8h night and 21°C during 16h). Samples were collected and bulked
9 from the three cuttings from 10 to 12 am. Total genomic DNA was extracted from young
10 leaves following a standard CTAB protocol.

11 **The three transposable elements surveyed**

12 We chose to survey three contrasting *Brassica* TEs, considering their genomic localisation,
13 copy number in the genome, and mechanism of transposition. To represent class I TEs we
14 chose a 10kb-long *Athila*-like retrotransposon (Alix & Heslop-Harrison, 2004). This TE is
15 represented by approximately 400 copies in the *B. oleracea* genome, mainly concentrated in
16 the pericentromeric regions of the chromosomes (Alix *et al.*, 2005). This retrotransposon was
17 previously used to develop PCR-based markers which preferentially target centromeres
18 (Pouilly *et al.*, 2008). This element should be efficient to identify new TE-related genomic
19 rearrangements occurring in the usually hyper-methylated pericentromeric regions, in
20 response to an eventual demethylation following allopolyploidy. We also targeted the
21 *Brassica* MITE *BraSto* which is 250bp-long with approximately 125 and 310 copies in the *B.*
22 *oleracea* and *B. napus* genomes, respectively. This element has been demonstrated to be
23 located preferentially in the gene space (Sarilar *et al.*, 2011). As we observed signs of recent
24 proliferation in the diploid and allotetraploid *Brassica* species (Sarilar *et al.*, 2011), it
25 appeared worthwhile to examine whether this transposon is mobilised immediately after the

1 polyploidisation event. Finally, the class II CACTA transposon *Bot1* was chosen for its
2 *Brassica* C genome-specificity, genome A is nearly devoid of any *Bot1* copies (Alix *et al.*,
3 2008). This 10kb-long transposon has gone through several rounds of amplification in the C-
4 genome, while only a few ancestral copies are still present in *B. rapa*. *Bot1* is present at
5 ~1500 copies in the *B. oleracea* genome and is highly dispersed along the full length of all *B.*
6 *oleracea* chromosomes (as observed using FISH - Alix *et al.*, 2008).

7 **Sequence-Specific Amplification Polymorphisms - SSAPs**

8 Our SSAP strategy followed the protocol optimized for the *Athila*-like retrotransposon and
9 described by Pouilly *et al.* (2008), which was based on previously described protocols with
10 some modifications (Waugh *et al.*, 1997; Ellis *et al.*, 1998). Additional primers were
11 specifically designed for *BraSto* and *Bot1*. The endonuclease *EcoRI* was used to restrict
12 genomic DNA, which was used, after ligation to adaptors, as the template for a two step-
13 SSAP reaction including pre-amplification prior to selective amplification. See Table 2 for
14 sequences and designation of adaptors and primers, as well as the list of the SSAP primer
15 combinations used. Final amplified products were separated by electrophoresis in denaturing
16 5.5% acrylamide sequencing gels using a LI-COR® DNA analyzer (LI-COR Biosciences,
17 USA) (Pouilly *et al.*, 2008). All the SSAP fingerprints were analyzed manually, and any
18 ambiguous SSAP bands were removed from the analysis. First, we compared the parental
19 SSAP multiband profiles to each other for each cross, and then we compared the SSAP
20 profiles of the resynthesized *B. napus* allotetraploids to both parental SSAP profiles, again for
21 each cross. SSAP bands were scored as present (amplified) or absent (not amplified) for each
22 line of the resynthesized *B. napus* allotetraploids surveyed: (i) SSAP bands amplified in
23 synthetic lines and present in at least one of the corresponding diploid parents were
24 considered as additive for these lines (add); (ii) extra SSAP bands amplified in synthetic lines
25 only, were scored as non-additive (N₊) for these lines; (iii) SSAP bands that were missing in

1 synthetic lines compared to diploid parents were scored as non-additive (N.) for these lines.
2 Proportion of non-additivity (i.e. global non-additivity) was estimated for any given cross,
3 and for each primer \times generation combination. It corresponds to the ratio calculated as the
4 division between the total number of non-additive bands and the total number of amplified
5 bands (parental SSAP bands + SSAP bands newly amplified in progenies): $N_{nad} = [(N. + N_+) / (n \times N_{diploid\ parents} + N_+)] \times 100$, with 'n' the number of lines per generation, and $N_{diploid\ parents}$
6 the total number of bands amplified in the diploid progenitors ($N_{diploid\ parents} = N_{specific\ to\ parent\ 1} + N_{specific\ to\ parent\ 2} + N_{parent\ 1\ and\ parent\ 2}$). We classified the global non-additivity observed into four
7 main classes: (i) strict additivity ($N_{nad}=0$), (ii) $N_{nad}<5\%$, (iii) $5\% \leq N_{nad} \leq 10\%$, (iv)
8 $N_{nad}>10\%$, taking into account each 'cross \times primer \times generation' combination as a single
9 case.
10
11

12 **Statistical analyses**

13 We looked for differences in the total number of SSAP bands between the two parental
14 species (*B. oleracea* vs. *B. rapa*), the five parental genotypes (three *B. oleracea* genotypes
15 and two *B. rapa* genotypes) and the different primers. To analyze the non-additivity of the
16 SSAP profiles among the different synthetic lines, we looked for differences in the number of
17 non-additive SSAP bands in relation to the crosses, generations and primers. The χ^2 tests
18 for contingency tables were all performed under the following H0 hypothesis: "the number of
19 non-additive SSAP bands observed for a specific condition is proportional to the number of
20 individuals tested for this condition". As an example, for the parental SSAP profiles and the
21 variable 'number of SSAP bands generated', the 'species' effect was tested against a 3/5-2/5
22 theoretical distribution, as we studied three different *B. oleracea* genotypes and two different
23 *B. rapa* genotypes (each genotype being represented by the same number of individuals). In
24 the results section we reported the P_{value} (p) associated with each χ^2 test performed for the
25 different comparisons we made. The Wilcoxon signed-rank test was also applied to test the

1 significance of more non-additive SSAP bands appearing or missing in one given
2 combination, with $W > 0$ indicating $N_+ > N_-$. (with $H_0: E(N_+) = E(N_-)$).

3 **Molecular characterization of non-additive SSAP bands**

4 For each TE surveyed, approximately 30 non-additive SSAP bands were cloned and
5 sequenced in order to identify the molecular polymorphism responsible for the non-additive
6 profile observed. SSAP bands were excised from dried silver stained polyacrylamide gels (our
7 protocol was modified slightly from that described in the Promega technical manual, with
8 fixation in 10% ethanol for 5 min, and oxidation in 1% nitric acid for 3 min), suspended in 40
9 μ l sterile ultrapure water, submitted to three freezing/defrosting cycles, then boiled for 5 min,
10 and finally re-amplified by PCR with the unlabelled primers used for the selective
11 amplification. PCR products were cloned (pGEM-T vector, Promega) and transformed in
12 *Escherichia coli* strain DH5 α . PCR was used to check that the DNA inserts were of the
13 correct size and at least five clones per amplified SSAP band were sequenced (GenoScreen,
14 France).

15 Nucleotide sequences from clones of the same SSAP band were aligned using
16 ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>); the predominant consensus
17 sequences were used as queries for similarity searches against non-redundant nucleotide
18 databases at NCBI (<http://www.ncbi.nlm.gov>) with the programme BLASTN and against
19 *Arabidopsis* genes at TAIR (<http://www.arabidopsis.org/Blast>) using WU-BLAST2. Primers
20 compatible with the corresponding TE selective SSAP primer were designed against genomic
21 regions flanking putative insertions of the TE surveyed. To identify the molecular nature of
22 the differentially amplified SSAP bands, new primer pairs were used for direct PCR
23 amplification on genomic DNA from individuals showing non-additive patterns, controls
24 represented by the diploid parents and synthetic lines displaying additive patterns. Amplified

1 products were separated by electrophoresis in 2.5% agarose gels, and visualized after staining
2 with ethidium bromide.

3 Nucleotide sequences of the SSAP bands and primer sequences with detailed results
4 are provided as Supporting Information Tables S1, S2 and S3.

5

6 **Results**

7 **SSAP profiling and summary of SSAP bands differentially amplified between *B. rapa*** 8 **and *B. oleracea* diploid parents**

9 Structural changes in copies of the retrotransposon family were detected using *Athila* primers,
10 which were designed in the nucleotide sequence of reverse transcriptase (Pouilly *et al.*, 2008).
11 Seven primer combinations were used (Table 2). For further statistical analyses, non-
12 ambiguous SSAP bands were scored; number of the scored bands varied from 12 to 52,
13 depending on the primer pair and cross considered. When SSAP patterns between the two
14 diploid parental species were compared (*B. oleracea* and *B. rapa* represented by three and two
15 different genotypes, respectively), *B. oleracea* displayed more SSAP bands than *B. rapa*
16 ($p \leq 0.05$). Statistical analysis showed a genotype effect ($p \leq 0.05$) with the highest number of
17 *Athila*-SSAP bands amplifying from HDEM (199 – sum of the bands amplified with the seven
18 SSAP combinations) and the lowest number from C_{1,3} (153). A primer effect was also
19 detected ($p \leq 0.001$) and was due to a deficit of bands amplified with the Athila40/18-E55
20 primer pair.

21 The *BraSto* SSAP primer (BrMiselect), designed in the terminal inverted repeat (TIR)
22 of the MITE, was used in combination with three different selective *Eco* primers (Table 2). It
23 amplified between 20 and 38 SSAP bands. We analyzed the SSAP patterns obtained for the
24 five parental genotypes and observed that *B. oleracea* displayed more *BraSto* SSAP bands

1 than *B. rapa* ($p \leq 0.001$). The primer combination had an effect on the total number of SSAP
2 bands amplified ($p \leq 0.05$), with significantly more bands amplified with the selective primer
3 E55 (Eco-CGA). The E80 (TAC)-BrMiselect primer pair amplified less SSAP bands than the
4 other two primer pairs.

5 The *Bot1* SSAP primer (3Botsselect) was defined upstream the CACTA motif of the
6 transposon and used in combination with three selective *Eco* primers (Table 2). Consistent
7 with the C genome-specificity of *Bot1*, 32 to 58 SSAP band levels were amplified, with more
8 bands in *B. oleracea* than *B. rapa* ($p \leq 0.001$). The selective primer, E55 (Eco-CGA), amplified
9 more SSAP bands than E33 (AAG) which also amplified more bands than E35 (ACA). Thus
10 once again a primer effect was observed ($p \leq 0.01$).

11 **The *Athila*-like SSAP profiles were mainly additive in the resynthesized *B. napus*** 12 **allotetraploids compared with their diploid parents**

13 The *Athila*-like SSAP multiband profiles in *B. napus* lines were compared with the parental
14 SSAP patterns, enabling the additivity hypothesis to be tested (see materials and methods).
15 However, when the entire dataset as a whole was considered, with no distinction between
16 crosses, generations or primers, the SSAP bands were mainly additive, with 52% showing
17 additivity (142 cases among the 273 cases studied corresponding to the different cross \times
18 generation \times primer combinations; Table 3). For the remaining cases, 32% showed levels of
19 non-additivity between 1% and 5%; 13% between 5% and 10%, and only 3% showed a level
20 of non-additivity above 10% (with a maximum of 17% for two combinations involving two
21 different F1 hybrids). When the total number of non-additive SSAP bands for every cross was
22 considered, we detected a cross effect ($p \leq 0.001$), with cross BrCRC showing a higher number
23 of non-additive SSAP bands (+190%) than expected (i.e. as estimated under the hypothesis
24 that the number of non-additive SSAP bands is proportional to the number of individuals
25 tested, equivalent here to the mean number of SSAP bands per line multiplied by the number

1 of lines originating from the cross BrCRC). Interestingly, the reciprocal cross BoRCC did not
2 show any particular deviation from the expected number of non-additive SSAP bands. The
3 three crosses BoEMZ1_{19,38}, BoEMZ3₃₈ and BoEMZ5₃₈ generated fewer than expected non-
4 additive SSAP bands (-74%, -73% and -75%, respectively – Table 3). The generation (F1, S0,
5 S1 or S2) did not have an influence on the number of non-additive bands; nevertheless, the
6 appearance of newly amplified SSAP bands was more frequent than the disappearance of
7 parental SSAP bands ($W > 0$, $N_+ > N_-$) in the generations F1 ($p \leq 0.05$), S0 ($p \leq 0.001$) and S1
8 ($p \leq 0.001$). In addition, we found a significant cross effect ($p \leq 0.001$) at each generation (with
9 F1 and S0 considered together). Interestingly, BoEMZ2_{19,38} was the only cross to display a
10 deficit of non-additive SSAP bands in S1 and a high excess of non-additive bands in S2; the
11 general rule was that crosses displayed comparable levels of non-additivity in both
12 generations. Finally, we estimated the impact of each primer pair on the number of non-
13 additive SSAP bands amplified. We found that this parameter had a significant effect
14 ($p \leq 0.001$): more non-additive SSAP bands were obtained with the primer pair Athila40/18-
15 E55 (+69%), and to a lesser extent with Athila40/18-E35 and Athila100/18-E84, than
16 theoretically expected, while the primer pair Athila100/18-E80 generated less non-additive
17 SSAP bands (-68%).

18 **The *BraSto*-SSAP profiles of the resynthesized *B. napus* allotetraploids showed the**
19 **highest levels of non-additivity compared to their diploid parents**

20 Next we quantified global non-additivity of the *BraSto*-SSAP profiles across the different *B.*
21 *napus* synthetic lines. Most of the SSAP bands were additive but in a lower proportion than
22 observed with the *Athila*-like element or *Bot1* (see below and Table 3): for 51% of the
23 different cases studied (60 among the 119 cases corresponding to the different cross \times
24 generation \times primer combinations) less than 5% of the SSAP bands (out of the total number
25 of SSAP bands amplified) were non-additive, with 22% showing strict additivity. For the

1 remaining cases, 20% displayed a level of non-additivity greater than 10% (with one
2 combination reaching 19% in S2). Taking into consideration the different crosses available,
3 we observed significant differences ($p \leq 0.001$) between the 11 crosses which generated non-
4 additive SSAP bands. Of note, more non-additive SSAP bands were amplified from crosses
5 BrCRC and BoEMZ2_{19,38} than theoretically expected (+43% and +40%, respectively), while
6 the cross BoEMZ3₃₈ generated less non-additive SSAP bands than expected (-52%). We also
7 observed that the reciprocal cross BoRCC displayed the expected number of non-additive
8 SSAP bands, and that the different BoEMZ crosses did not show homogeneous behavior,
9 even when considering the method used to produce them (colchicine treatment or female
10 unreduced gametes – Table 1). Such differences for non-additivity between crosses were
11 significant only for the S1 and S2 generations ($p \leq 0.05$ and $p \leq 0.001$, respectively). We found
12 that generation (F1, S0, S1 or S2) had an impact on the proportion of non-additivity ($p \leq 0.05$)
13 so that the number of non-additive SSAP bands was, at first stable, but then increased
14 significantly in the S2 generation. Interestingly, this non-additivity in S2 was characterized by
15 a higher number of lost (N_-) parental SSAP bands rather than a gain (N_+) of novel bands
16 ($W < 0$, $N_- > N_+$). The appearance of newly amplified SSAP bands was relatively more frequent
17 ($W \geq 0$, $N_- < N_+$) in the F1 hybrids and S0 generation ($p \leq 0.05$), whereas it was more balanced in
18 the S1 (not significant). We found an effect ($p \leq 0.001$) of the selective *Eco*- primer used for
19 SSAP amplification: the BrMiselect - E55 primer pair amplified +88% of non-additive SSAP
20 bands than theoretically expected, whereas discrepancies of -28% and -60% were observed
21 for E36 and E80 respectively.

22 ***Bot1*-SSAP profiles included significantly more non-additive bands in the S2 generation**
23 **of *B. napus* synthetic lines**

24 When the *Bot1*-SSAP profiles of the different individuals from all the different *B. napus*
25 synthetic lines were compared with the profiles of the corresponding diploid parents, again

1 the majority of the SSAP bands surveyed were additive. Indeed, 71% of the different cases
2 studied (86 of the 114 cases corresponding to the different cross \times generation \times primer
3 combinations) displayed less than 5% non-additivity, with 26% showing additivity (Table 3).
4 The remaining 22% displayed between 5% and 10% non-additivity, and 7% at levels greater
5 than 10% (with one cross \times primer combination showing 17% non-additive bands in S2).
6 When the non-additivity of the SSAP profiles for each cross was analyzed, we observed
7 differences ($p \leq 0.001$) between crosses for the number of non-additive SSAP bands. The
8 relative proportion of non-additive SSAP bands for cross BoRCC was consistent with
9 expected numbers. Its reciprocal cross BrCRC, however, displayed more non-additive SSAP
10 bands than theoretically expected (+51%). Among the eight different BoEMZ synthetic
11 progenies, five showed the expected number of non-additive SSAP bands. A greater
12 proportion of non-additivity (+39%) was found in cross BoEMZ_{319,38}, whereas fewer non-
13 additive SSAP bands than expected were observed in the crosses BoEMZ₁₃₈ (-31%) and
14 BoEMZ₃₃₈ (-37%). Thus the method used to generate the BoEMZ crosses (colchicine
15 treatment or female unreduced gametes) did not have an effect. Interestingly, we estimated
16 that the differences observed between crosses for the general non-additivity of the SSAP
17 profiles analyzed were only statistically significant in the S2 generation ($p \leq 0.001$). To
18 complete this analysis, we examined the impact of generation. While a significant effect was
19 found ($p \leq 0.001$) by testing all four generations (F1, S0, S1 and S2) for homogeneity in the
20 proportion of non-additivity, no conclusion could be drawn when the three generations F1, S0
21 and S1 ($p = 0.1$) were tested. This indicated that relatively more non-additivity was observed
22 for the S2. When the number of disappearing (N_-) vs. appearing (N_+) SSAP bands was
23 evaluated, for each generation, most of the non-additive SSAP bands corresponded to newly
24 amplified SSAP bands ($W > 0$, $N_+ < N_-$; $p \leq 0.001$ for each generation). Of note, the non-
25 additivity estimated here depended on the primer used to amplify the SSAP ($p \leq 0.001$), with

1 the selective primer E55 producing more non-additive bands in all four generations (+41% in
2 total), and E35 producing fewer non-additive bands (-39%).

3 Table 3 provides a summary of the non-additivity represented by each TE-SSAP profile, the
4 parameters influencing non-additivity and their respective effects.

5 **Some TE-mediated structural changes did occur but overall very few transposition**
6 **events occurred in the early generations of resynthesized *B. napus* allotetraploids**

7 A total of 21 ($N_+=17$, $N_-=4$), 27 ($N_+=21$, $N_-=6$) and 30 ($N_+=18$, $N_-=12$) non-additive bands
8 which amplified with the *Athila*-, *BraSto*- and *BotI*-SSAPs respectively, were cloned and
9 sequenced. See supplemental information Tables S1A, S2A and S3A for the description and
10 full molecular characterization of the non-additive SSAP bands analyzed. For 17 SSAP bands,
11 more than one major sequence was obtained and thus analyzed, illustrating the complexity of
12 the detailed analysis of multiband profiles. For the three TE-derived SSAP profiling
13 approaches, not all the SSAP bands could be confidently related to the targeted TE. Only 35%
14 of the *Athila*-SSAP sequences contained *Athila*-like or any reverse transcriptase fragments,
15 and only 38% of the *BraSto*-SSAP sequences contained the *BraSto*-specific TIRs (Tables
16 S1B, S2B). Nevertheless, more than half of the *BotI*-SSAP sequences did contain the typical
17 CACTA ending motif (Table S3B).

18 PCR tests were then developed by designing PCR primers in all the cloned sequences;
19 primers were used in combination with the corresponding TE-specific primer (*Athila*40/18 or
20 *Athila*100/18; *BrMiselect*, *3Botselect*) in order to verify the presence/absence of the
21 corresponding non-additive SSAP band in the two diploid parents and synthetic lines of
22 interest (those showing the non-additive SSAP profile, and others used as references). Figure
23 1 gives a summary of the information obtained from the detailed analysis of each SSAP
24 sequence. Firstly, for several SSAP bands (25% of all the analyzed SSAP sequences) we
25 could not draw any conclusions about the nature of the polymorphism revealed by SSAP due

1 to invalid PCR test results. This was particularly marked for the *Athila*-SSAP profiling, with
2 PCR tests of 39% of the *Athila*-SSAP sequences failing to give conclusive results (multiband
3 amplicons or smears, data not shown). For a total of 61% of all the sequences analyzed, we
4 showed that these were not accompanied by any structural modifications, as the resulting PCR
5 test profiles were additive in all the synthetic lines tested compared with the diploid parents.
6 We could identify structural changes associated with the 14% remaining sequences. These
7 represented 9% and 19% of the sequences analyzed for *Athila*- and *BraSto*-SSAP,
8 respectively. According to sequence analysis, it appears that partial genomic rearrangements
9 in the vicinity of the insertions of the TEs surveyed have occurred, rather than new
10 transposition events. Interestingly, while 7% of the *BotI*-SSAP sequences investigated were
11 of this kind, three *BotI*-SSAP sequences could be associated with putative transposition
12 events, which corresponded to one new insertion and two excision events (one excision event
13 is illustrated in Figure 2). These polymorphic sites were all identified in the S2 generation of
14 the synthetic lines.

15

16 **Discussion**

17 **Allopolyploidy was not accompanied by massive reshuffling of TEs in early generations** 18 **of resynthesized *B. napus* allotetraploids**

19 Until recently Barbara McClintock's Genome Shock hypothesis (1984), which proposed that
20 a burst of transposition occurs immediately after allopolyploidisation, had been rarely
21 investigated in detail. During the last decade, efforts have been made to examine the impact of
22 allopolyploidy on the mobilisation of TEs across the genome. The mobilisation of various TE
23 families has been studied in different polyploid plant species at different evolutionary time
24 scales (e.g. *Gossypium* – Zhao *et al.*, 1998, Hu *et al.*, 2010; *Triticum* – Kashkush *et al.*, 2002,

1 Charles *et al.*, 2008; *Arabidopsis* – Madlung *et al.*, 2005; *Nicotiana* – Petit *et al.*, 2007;
2 *Spartina* – Parisod *et al.*, 2009). This led to the conclusion that “the effect of allopolyploidy
3 on TE genome fractions may be more complex than generally assumed” (Parisod *et al.*,
4 2010a). In the present study, we analyzed early generations of different resynthesized *B.*
5 *napus* allotetraploids for changes in the insertion profiles of three contrasting TEs. By
6 developing a SSAP profiling approach specific to each of the three TEs surveyed, we could
7 observe a general trend for additivity of the insertion profiles between the synthetic lines and
8 their corresponding diploid parents. In other words, we could not detect any massive
9 structural changes associated with the three contrasting TEs surveyed, immediately after
10 allopolyploidisation. As in previously published studies on the topic, our results reinforce the
11 concept that highly specific TE responses occur during a particular allopolyploidy event. In
12 *Arabidopsis* polyploids, Madlung *et al.* (2005) demonstrated that *Sunfish* transposons are
13 mobilised, while no evidence for transposition was reported from a global survey of potential
14 TE re-activation (Beaulieu *et al.* 2009). In synthetic *Nicotiana tabacum* allopolyploids, it was
15 demonstrated that only a few specific young populations of the *Tnt1* retrotransposon were
16 affected and mobilized by the polyploidisation event (Petit *et al.*, 2010). These observations
17 underline the lack of any general rules concerning the global impact of allopolyploidy on TE
18 mobilisation.

19 The very few transposition events we could identify were related specifically to the
20 putative activity of the *Bot1* CACTA transposon, activity which was already suspected
21 because of the existence of *Bot1* transcripts in the *Brassica* EST database (Alix *et al.*, 2008).
22 This *Brassica* C genome specific transposon appears to be the youngest TE surveyed in the
23 present study. Indeed, the low level of sequence divergence between *Bot1* and its C genome
24 specificity have suggested that this element was amplified after divergence from the A
25 genome (Alix *et al.*, 2008). For comparison, *BraSto* is present in both the A and C genomes,

1 while the *Athila*-like element is closely related to the *Athila* retrotransposon of *Arabidopsis*.
2 Therefore, it appears that the age of a TE could be indicative of its potential transposition
3 activity, as was previously observed for *Tnt1* in *Nicotiana* (Petit *et al.*, 2010). With regards to
4 the suspected transposition activity of *Bot1*, which is highly supported by our current data,
5 one evolutionary question still remains: why did *Bot1* not spread to the A subgenome during
6 the recent formation of the natural allotetraploid *B. napus* (Alix *et al.*, 2008)? Indeed, this
7 phenomenon was demonstrated for *Gossypium* allopolyploids with a *copia*-like
8 retrotransposon specific to one parental genome (Zhao *et al.*, 1998). One possibility is that
9 specific *Bot1* repression mechanisms set in immediately after allopolyploidisation, avoiding
10 anarchic transposition across the genome and ensuring stable establishment of the newly
11 formed allopolyploid species (even if very few copies may escape this control). In addition,
12 even if the hypothesis that the few transpositions depicted here may reflect the stochastic
13 standard activity of *Bot1* in the diploid parental genome (which would thus reinforce our
14 conclusion of the low impact of allopolyploidy on *Bot1* activation) cannot be completely
15 ruled out, our data still ask the question of the mechanisms controlling *Bot1* activity. The
16 actors involved and the extent of this regulatory mechanism remain to be elucidated,
17 especially considering that TE transcription was also demonstrated as a major target for
18 control by small interfering RNAs (Mirouze *et al.*, 2009).

19 **Meiotic recombination appears to be involved in the TE-mediated structural changes** 20 **observed**

21 We identified several genomic rearrangements which resulted in part of the non-additivity
22 depicted (at least 11% of all the non-additive SSAP bands analyzed, however some of the
23 undetermined non-additive SSAP bands may also correspond to other genomic
24 rearrangements). It was striking that rearrangements were significantly more numerous at the
25 MITE *BraSto* insertion sites, and that they were all observed in the S1 generation synthetic

1 lines (Table S2A). This makes sense in that *BraSto* copies were found to be associated with
2 the *Brassica* gene space, and were usually in more interstitial positions than the other two TEs
3 surveyed here (Sarilar *et al.*, 2011). This data echoes major results reported previously from
4 the analysis of the same resynthesized *B. napus* allotetraploids: it was demonstrated that
5 structural changes were generated by the first meiosis – which took place in S0 to produce the
6 S1 lines – and were significantly more abundant in interstitial and distal locations along the
7 chromosomes (Szadkowski *et al.*, 2010). It can thus be hypothesized that restructuring at TE
8 insertion sites is strongly influenced by meiotic recombination. This is also supported by the
9 identification of a pronounced generation effect on SSAP non-additivity for the two TEs
10 *BraSto* and *Bot1*, with significantly more non-additive SSAP bands detected at the second
11 selfed generation S2 (particularly marked with *Bot1*-SSAP). This effect was not significant in
12 the *Athila*-SSAP profiles and *Athila*-like copies are concentrated in pericentromeric regions
13 (Alix *et al.*, 2005) where the recombination rate is low due to the hypermethylation of this
14 region. We also found that the *Eco* primer used had a significant effect on generating non-
15 additive SSAP bands (with the selective *Eco* primer E55 generating significantly more non-
16 additive SSAP bands than the others). This finding underlines the importance of the location
17 of the genomic regions targeted by the different primer pairs used on the extent of the changes
18 affecting the TE insertion sites analyzed by SSAPs. Here again, our observations highlight the
19 putative key role of the recombination rate in generating non-additive SSAP profiles. Indeed,
20 reactivation of *Tnt1* retrotransposons in resynthesized *Nicotiana tabacum* allotetraploids was
21 effective only in the S4 generation, after four rounds of meiosis, and was accompanied by
22 major restructuring at the *Tnt1* insertion sites (Petit *et al.*, 2010).

23 Among the additional parameters examined for their involvement in SSAP non-
24 additivity, we paid attention to the way each allotetraploid was formed. Szadkowski *et al.*
25 (2011) demonstrated that the type of allotetraploid formation pathway involved had a

1 significant effect on the generation of genomic rearrangements. They reported that
2 significantly more rearrangements occurred in synthetic lines originated from unreduced
3 gametes. In our present study however we could not validate this effect. Similar non-
4 additivity proportions were found in the synthetic BoEMZ_{19,38} allotetraploids produced from
5 unreduced gametes as those of the synthetic BoEMZ₃₈ allotetraploids obtained following
6 colchicine treatment of F1 hybrids. In our study, in all the different generation × primer
7 combinations analyzed and for the three different TEs surveyed, the cross effect (i.e. the union
8 at random of two samples of gametes) was the major effect, suggesting a statistically
9 significant proportion of stochastic SSAP non-additivity.

10 **Towards a hypothesis that some non-additive SSAP bands may result from changes in** 11 **DNA methylation**

12 A relatively high number of non-additive SSAP bands was found to be additive at the
13 genomic structure level, according to the PCR tests performed (corresponding to 61% of all
14 the non-additive SSAP bands analyzed in detail). Non-additivity results, which could not be
15 associated with transposition events or structural rearrangements (indicated as ‘no structural
16 modification’) could be explained by competition between multiple target sites during PCR
17 amplification, but also polymorphisms at the *EcoRI* endonuclease restriction sites we used to
18 develop the SSAP markers. The *EcoRI* recognition site is GAATTC, and it has been
19 demonstrated that *EcoRI* is moderately sensitive to cytosine methylation in eukaryotic cells,
20 as in bacteria, with a reduced rate of cleavage at hemi-methylated GAATT^{m5}C and no
21 cleavage with DNA containing GAATT^{m5}C on both strands (Brennan *et al.*, 1986; Nelson &
22 McClelland, 1991). It is thus tempting to hypothesize that some non-additive SSAP bands
23 observed may correspond to methylation changes at the restriction sites. Epigenetic responses
24 of the plant genome to allopolyploidy have been previously documented. Remodelling of
25 DNA methylation was frequently observed in various polyploid plant species, corresponding

1 mainly to rapid alterations occurring immediately after the allopolyploidisation event (e.g.
2 *Arabidopsis* – Madlung *et al.*, 2005) or already in the F1 hybrids (e.g. *Triticum* – Shaked *et*
3 *al.*, 2001; *Spartina* – Parisod *et al.*, 2009; *Senecio* – Hegarty *et al.*, 2011). In *Brassica*,
4 changes to the extent of DNA methylation were described in newly resynthesized *B. napus*
5 allotetraploids (Lukens *et al.*, 2006) and also in the same plant material as that surveyed in the
6 work reported here (A. Salmon and A.-M. Chèvre, unpublished). In addition, different studies
7 on *Arabidopsis* hybrids have revealed the influence of the maternal cytoplasm on epigenetic
8 regulation and control of transposons and gene expression in reciprocal crosses through the
9 production of specific small RNAs (Martienssen 2010; Lu *et al.*, 2012). With the involvement
10 of small RNAs in mediating DNA methylation (Simon & Meyers, 2011), we can thus
11 hypothesize that cytotype (cross direction) would also influence changes in DNA methylation
12 in response to interspecific hybridization. Notably, when testing the cross effect for the
13 reciprocal crosses BoRCC and BrCRC, we identified significantly more non-additive SSAP
14 bands in the progeny carrying the A cytoplasm, for each of the three TE-SSAP profiles
15 (BrCRC progeny – Table 3); however we did not observe any deviations from the expected
16 proportions of non-additive SSAP bands in the corresponding progeny carrying the C
17 cytoplasm (BoRCC). This can be compared to the previously demonstrated significant
18 cytoplasmic effect on the transmission of meiosis-related genomic rearrangements to
19 progenies (Szadkowski *et al.*, 2010). While a deficit in DNA methylation was recently
20 demonstrated to boost recombination at least in euchromatic regions (Melamed-Bessudo &
21 Levy, 2012), the relationship between changes in DNA methylation patterns and the structural
22 rearrangements which occur following interspecific hybridisation and allopolyploidisation
23 (Hegarty *et al.*, 2011) should also be investigated in newly resynthesized *B. napus*
24 allotetraploids.

1 In conclusion, our study of the impact of allopolyploidy on the restructuring of three
2 contrasting TEs in resynthesized *B. napus* allotetraploids demonstrated the lack of an anarchic
3 burst of transposition, as generally assumed until recently. The non-additivity of TE-anchored
4 SSAP profiles between synthetic *B. napus* allotetraploids and their diploid progenitors,
5 accompanied or not by structural rearrangements, asks also the question of some potential
6 concordance with DNA methylation changes. These observations imply that the tight control
7 of TEs at their insertion sites, by heterochromatinisation and other hypermethylation
8 processes, was partly relaxed in response to the recent allopolyploidisation event. The very
9 few *Bot1* transposition events identified may be thus related to a higher activation of TEs
10 primarily at the level of transcription, but this still needs to be confirmed. Future studies are
11 needed to ascertain whether structural and functional changes induced by allopolyploidy do
12 not reflect the upstream disruption of the epigenetic regulatory machinery of the neo-
13 polyploid genome.

14

15 **Acknowledgements**

16 We thank Jean-Claude Letanneur and graduate students from AgroParisTech (Wenxuan
17 Zhang, Thomas Murarasu, Jordan Peltier, Florent Guinot) for technical assistance. We also
18 particularly thank Clémentine Vitte and Adrienne Ressayre for valuable discussions, and
19 Leigh Gebbie for corrections to the English. We also thank the three anonymous referees for
20 providing us with valuable comments and suggestions. This work was supported by the
21 project “Effect of polyploidy on plant genome biodiversity and evolution” funded by the
22 French *Agence Nationale de la Recherche* (Biodiversity programme #ANR-05-BDIV-015).
23 VS was supported by a PhD fellowship from the French *Direction Générale de*
24 *l'Enseignement et de la Recherche* (DGER) via AgroParisTech and the French *Centre*

1 *National de la Recherche Scientifique (CNRS)*. PMP was supported by a PhD fellowship
2 from the French *Ministère de l'Enseignement Supérieur et de la Recherche (MESR)*.

3

4 **References**

- 5 **Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H. 2011.** Genome-wide evidence for
6 local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*.
7 *Nucleic Acids Research* **39**(16):6919–6931.
- 8 **Albertin W, Balliau T, Brabant P, Chèvre A, Eber F, Malosse C, Thiellement H. 2006.**
9 Numerous and rapid nonstochastic modifications of gene products in newly synthesized
10 *Brassica napus* allotetraploids. *Genetics* **173**: 1101-13.
- 11 **Alix K, Heslop-Harrison. 2004.** The diversity of retroelements in diploid and allotetraploid
12 *Brassica* species. *Plant Molecular Biology* **54**: 895-909.
- 13 **Alix K, Joets J, Ryder CD, Moore J, Barker GC, Bailey JP, King GJ, Heslop-Harrison**
14 **JS. 2008.** The CACTA transposon *Bot1* played a major role in *Brassica* genome
15 divergence and gene proliferation. *The Plant Journal* **56**: 1030-1044.
- 16 **Barker MS, Vogel H, Schranz ME. 2009.** Paleopolyploidy in the Brassicales: analyses of
17 the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis*
18 and other Brassicales. *Genome Biology and Evolution* **1**: 391-399.
- 19 **Alix K, Ryder CD, Moore J, King GJ, Heslop-Harrison JS. 2005.** The genomic
20 organization of retrotransposons in *Brassica oleracea*. *Plant Molecular Biology* **59**:
21 839-851.
- 22 **Beaulieu J, Jean M, Belzile F. 2009.** The allotetraploid *Arabidopsis thaliana*-*Arabidopsis*
23 *lyrata* subsp. *petraea* as an alternative model system for the study of polyploidy in
24 plants. *Molecular and General Genetics* **281**: 421-435.

1 **Brennan CA, Van Cleve MD and Gumpert RI. 1986.** The effects of base analogue
2 substitutions on the cleavage by the *EcoRI* restriction endonuclease of
3 octadeoxyribonucleotides containing modified *EcoRI* recognition sequences. *The*
4 *Journal of Biological Chemistry* **261**: 7270–7278.

5 **Charles M, Belcram H, Just J, Huneau C, Viollet A, Couloux A, Segurens B, Carter M,**
6 **Huteau V, Coriton O. et al. 2008.** Dynamics and differential proliferation of
7 transposable elements during the evolution of the B and A genomes of wheat. *Genetics*
8 **180**: 1071-1086.

9 **Chen ZJ, Ni Z. 2006.** Mechanisms of genomic rearrangements and gene expression changes
10 in plant polyploids. *BioEssays* **28**: 240-252.

11 **Chen ZJ. 2010.** Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant*
12 *Science* **15**: 57-71.

13 **Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson**
14 **JE, Arumuganathan K, Barakat A et al. 2006.** Widespread genome duplications
15 throughout the history of flower plants. *Genome Research* **16**: 738-749.

16 **De Bodt S, Maere S, Van de Peer Y. 2005.** Genome duplication and the origin of
17 angiosperms. *Trends in Ecology and Evolution* **20**: 591-597.

18 **Devos KM, Brown JKM, Bennetzen JL. 2002.** Genome size reduction through illegitimate
19 recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* **12**:
20 1075-1079.

21 **Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008.**
22 Evolutionary genetics of genome merger and doubling in plants. *Annual Review of*
23 *Genetics* **42**: 443-461.

- 1 **Ellis TH, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ. 1998.** Polymorphism of
2 insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity
3 analysis in pea. *Molecular and General Genetics* **260**: 9-19.
- 4 **Fawcett JA, Maere S, Van de Peer Y. 2009.** Plants with double genomes might have had a
5 better chance to survive the cretaceous-Tertiary extinction event. *Proceedings of the*
6 *National Academy of Sciences of the United States of America* **106**: 5737-5742.
- 7 **Gaeta RT, Pires JC. 2010.** Homoeologous recombination in allopolyploids: the polyploidy
8 ratchet. *New Phytologist* **186**:18-28.
- 9 **Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007.** Genomic changes in
10 resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The*
11 *Plant Cell* **19**: 3403-3417.
- 12 **Gaeta RT, Yoo SY, Pires JC, Doerge RW, Chen ZJ, Osborn TC. 2009.** Analysis of gene
13 expression in resynthesized *Brassica napus* allopolyploids using Arabidopsis 70mer oligo
14 microarrays. *PLoS One* **4**: e4760.
- 15 **Grandbastien MA. 1998.** Activation of plant retrotransposons under stress conditions.
16 *Trends in Plant Science* **3**: 181-187.
- 17 **Hegarty MJ, Batstone T, Barker GL, Edwards KJ, Abbott RJ, Hiscock SJ. 2011.**
18 Nonadditive changes to cytosine methylation as a consequence of hybridization and
19 genome duplication in *Senecio* (Asteraceae). *Molecular Ecology* **20**: 105-113.
- 20 **Hegarty MJ, Hiscock SJ. 2008.** Genomic clues to the evolutionary success of polyploid
21 plants. *Current Biology* **18**: R435-R444.
- 22 **Hu G, Hawkins JS, Grover CE, Wendel JF. 2010.** The history and disposition of
23 transposable elements in polyploid *Gossypium*. *Genome* **53**: 599-607.

- 1 **Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho**
2 **LP, Hu Y, Liang H, Soltis PS et al. 2011.** Ancestral polyploidy in seed plants and
3 angiosperms. *Nature* **473**: 97-102.
- 4 **Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. 2000.** Genome evolution
5 of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in
6 response to sharp microclimatic divergence. *Proceedings of the National Academy of*
7 *Sciences of the United States of America* **97**: 6603-6607.
- 8 **Kashkush K, Feldman M, Levy AA. 2002.** Gene loss, silencing and activation in a newly
9 synthesized wheat allotetraploid. *Genetics* **160**: 1651–1659.
- 10 **Lisch D. 2009.** Epigenetic regulation of transposable elements in plants. *Annual Review in*
11 *Plant Biology* **60**: 43-66.
- 12 **Lu J, Zhang C, Baulcombe DC, Chen ZJ. 2012.** Maternal siRNAs as regulators of parental
13 genome imbalance and gene expression in endosperm of *Arabidopsis* seeds.
14 *Proceedings of the National Academy of Sciences of the United States of America* **109**:
15 5529-5534.
- 16 **Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. 2006.** Patterns of
17 sequence loss and cytosine methylation within a population of newly resynthesized
18 *Brassica napus* allopolyploids. *Plant Physiology* **140**: 336-348.
- 19 **Madlung A, Comai L. 2004.** The effect of stress on genome regulation and structure. *Annals*
20 *of Botany* **94**: 481-495.
- 21 **Madlung A, Tyagi AP, Watson B, Jiang HM, Kagochi T, Doerge RW, Martienssen R,**
22 **Comai L. 2005.** Genomic changes in synthetic *Arabidopsis* polyploids. *Plant Journal*
23 **41**: 221-230.

- 1 **Marmagne A, Brabant P, Thiellement H, Alix K. 2010.** Analysis of gene expression in
2 resynthesized *Brassica napus* allotetraploids: transcriptional changes do not explain
3 differential protein regulation. *New Phytologist* **186**: 216-227.
- 4 **Martienssen RA. 2010.** Heterochromatin, small RNA and post-fertilization dysgenesis in
5 allopolyploid and interploid hybrids of *Arabidopsis*. *New Phytologist* **186**: 46-53.
- 6 **McClintock B. 1984.** The significance of responses of the genome to challenge. *Science* **226**:
7 792-801.
- 8 **Melamed-Bessudo C, Levy AA. 2012.** Deficiency in DNA methylation increases meiotic
9 crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*.
10 *Proceedings of the National Academy of Sciences of the United States of America* **109**:
11 E981-E988; doi:10.1073/pnas.1120742109.
- 12 **Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J,
13 Weigel D, Paszkowski J, Mathieu O. 2009.** Selective epigenetic control of
14 retrotransposition in *Arabidopsis*. *Nature* **461**: 427-431.
- 15 **Nelson M, McClelland M. 1991.** Site-specific methylation: effect on DNA modification
16 methyltransferases and restriction endonucleases. *Nucleic Acids Research* **19**: 2045-
17 2071.
- 18 **Ohno S. 1970.** Evolution by gene duplication. Springer-Verlag, Heidelberg, Germany. 150pp.
- 19 **Parisod C, Salmon A, Zerjal T, Tenailon M, Grandbastien M-A, Ainouche ML. 2009.**
20 Rapid structural and epigenetic reorganization near transposable elements in hybrid
21 and allopolyploid genomes in *Spartina*. *New Phytologist* **184**: 1003-1015.
- 22 **Parisod C, Alix K, Just J, Petit M, Sarilar S, Mhiri C, Ainouche M, Chalhoub B,
23 Grandbastien M-A. 2010a.** Impact of transposable elements on the organization and
24 function of allopolyploid genomes. *New Phytologist* **186**: 37-45.

- 1 **Parisod C, Holderegger R, Brochmann C. 2010b.** Evolutionary consequences of
2 autopolyploidy. *New Phytologist* **186**: 5-17.
- 3 **Petit M, Lim KY, Julio E, Poncet C, de Borne FD, Kovarik A, Leitch AR, Grandbastien**
4 **M-A, Mhiri C. 2007.** Differential impact of retrotransposon populations on the
5 genome of allotetraploid tobacco (*Nicotiana tabacum*). *Molecular Genetics and*
6 *Genomics* **278**: 1-15.
- 7 **Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Lim KY, Kovarik A, Leitch AR,**
8 **Grandbastien M-A, Mhiri C. 2010.** Mobilization of retrotransposons in synthetic
9 allotetraploid tobacco. *New Phytologist* **186**: 135-147.
- 10 **Pouilly N, Delourme R, Alix K, Jenczewski E. 2008.** Repetitive sequence-derived markers
11 tag centromeres and telomeres and provide insights into chromosome evolution in
12 *Brassica napus*. *Chromosome Research* **16**: 683-700.
- 13 **Ramsey J, Schemske DW. 1998.** Pathways, mechanisms, and rates of polyploidy formation
14 in flowering plants. *Annual Review Ecology and Systematic* **33**: 589-639.
- 15 **Sarilar V, Marmagne A, Brabant P, Joets J, Alix K. 2011.** ‘BraSto’, a *Stowaway* MITE
16 from *Brassica*: recently active copies preferentially accumulate in the gene space.
17 *Plant Molecular Biology* **77**: 59-75.
- 18 **Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001.** Sequence elimination and
19 cytosine methylation are rapid and reproducible responses of the genome to wide
20 hybridization and allopolyploidy in wheat. *The Plant Cell* **13**: 1749-1759.
- 21 **Schnable PS, Ware D, Fulton RS, et al. 2009.** The B73 maize genome: complexity,
22 diversity, and dynamics. *Science* **326**: 1112-1115.
- 23 **Schranz ME, Lysak MA, Mitchell-Olds T. 2006.** The ABC's of comparative genomics in
24 the Brassicaceae: building blocks of crucifer genomes. *Trends in Plant Science* **11**:
25 535-542.

- 1 **Simon SA, Meyers BC. 2011.** Small RNA-mediated epigenetic modifications in plants.
2 *Current Opinion in Plant Biology* **2**: 148-55.
- 3 **Slotkin RK, Martienssen R. 2007.** Transposable elements and the epigenetic regulation of
4 the genome. *Nature Reviews Genetics* **8**: 272-285.
- 5 **Soltis PS, Soltis DE. 2009.** The role of hybridization in plant speciation. *Annual Review of*
6 *Plant Biology* **60**: 561-588.
- 7 **Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre A-M. 2011.**
8 Polyploid formation pathways have an impact on genetic rearrangements in
9 resynthesized *Brassica napus*. *New Phytologist* **191**: 884-894.
- 10 **Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O,**
11 **Manzanares-Dauleux MJ, Delourme R, King GJ et al. 2010.** The first meiosis of
12 resynthesized *Brassica napus*, a genome blender. *New Phytologist* **186**: 102-112.
- 13 **Vitte C, Panaud O. 2005.** LTR retrotransposons and flowering plant genome size:
14 emergence of the increase/decrease model. *Cytogenetic and Genome Research* **110**:
15 91-107.
- 16 **Vitte C, Peterson T, Bennetzen JL. 2011.** Transposable elements and the evolving structure
17 of the maize genome. *In: Advances in Maize* (Prioul J-L, Thévenot C, Molnar T)
18 (vol.3) pp.1-36.
- 19 **Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W. 1997.**
20 Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome
21 revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular and*
22 *General Genetics* **253**: 687-694.
- 23 **Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P,**
24 **Morgante M, Panaud O et al. 2007.** A unified classification system for eukaryotic
25 transposable elements. *Nature Reviews Genetics* **8**: 973-982.

1 **Zhang J. 2003.** Evolution by gene duplication: an update. *Trends in Ecology and Evolution*
2 **18:** 292-298.

3 **Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH.**
4 **1998.** Dispersed repetitive DNA has spread to new genomes since polyploidy
5 formation in cotton. *Genome Research* **8:** 479-492.

6

7

1 **Figure legends**

2 **Figure 1.** Characterization of the polymorphisms identified by SSAP profiling of three
3 diverse transposable elements in newly resynthesized *Brassica napus* allotetraploids.

4 (a) Overall findings from a total of 97 sequences analyzed, with no distinction made between
5 the three TEs studied. Seventy eight non-additive SSAP bands were cloned and sequenced
6 then analyzed to determine whether a transposition event or structural rearrangement may
7 have occurred. (b) Characterization of the polymorphisms revealed by SSAP for the three TEs
8 surveyed. Twenty three, 32 and 42 sequences obtained from non-additive SSAP bands related
9 to the *Athila*-like element, the MITE *BraSto* and the transposon *Bot1*, respectively, were
10 analyzed as above.

11 **Figure 2.** PCR validation of a putative excision event involving a *Bot1* copy which occurred
12 in synthetic line #93 (S2 generation) of the resynthesized *B. napus* allotetraploid BoEMZ3_{19,38}

13 (a) A partial profile of the polymorphisms observed among the BoEMZ3₃₈ and BoEMZ3_{19,38}
14 progenies which identified the non-additive *Bot1*-SSAP band #17. (b) Ethidium bromide-
15 stained agarose gel showing results of the PCR validation of the SSAP polymorphism shown
16 in (a).

17 The star indicates (a) the missing non-additive SSAP band in synthetic line #93 for which the
18 parental band (from the *B. oleracea* parent 'HDEM') was cut out of the polyacrylamide gel,
19 cloned and sequenced, and (b) the corresponding absence of amplicon after PCR validation
20 while the band was amplified as expected in the diploid parent *B. oleracea* (HDEM) and the
21 synthetic line #76 used as a reference.

22 Faint amplification in the *B. rapa* parent 'Z1' may correspond to aspecific amplification (as
23 verified for some PCR products, data not shown) regarding the design of the PCR test
24 performed here. M=size marker.

1 **Supporting Information**

2 **Table S1.** (S1A) Molecular characterization of the 21 cloned and sequenced non-additive
3 *Athila*-SSAP bands. (S1B). Summary of the *Athila*-SSAP polymorphisms revealed by 23
4 sequences analyzed from the 21 different non-additive SSAP bands cloned.

5 **Table S2.** (S2A) Molecular characterization of the 27 cloned and sequenced non-additive
6 *BraSto*-SSAP bands. (S2B). Summary of the *BraSto*-SSAP polymorphisms revealed by 32
7 sequences analyzed from the 27 different non-additive SSAP bands cloned

8 **Table S3.** (S3A) Molecular characterization of the 30 cloned and sequenced non-additive
9 *Bot1*-SSAP bands. (S3B). Summary of the *Bot1*-SSAP polymorphisms revealed by 42
10 sequences analyzed from the 30 different non-additive SSAP bands cloned

11

1 **Table 1.** Description of the plant material surveyed in the present study: scheme showing the
 2 generation of the resynthesized *B. napus* allotetraploids, sample sizes and details of the
 3 synthetic progenies

4

Parental genotypes and crosses (♀×♂)	C _{1.3} × RC (AA×CC)	RC × C _{1.3} (CC×AA)	C10 × Z1 (CC×AA)	HDEM × Z1 (CC×AA)	
Hybrid F1	1 (AC)	1 (CA)	1 (CA)	5 (CA)	
<i>somatic chromosome doubling</i>					<i>female unreduced gametes</i>
Synthetic offspring	BrCRC	BoRCC	BoC10Z1	BoEMZi ₃₈	BoEMZi _{19,38}
S0	1 (AACC)	1 (CCAA)	1 (CCAA)	5 (CCAA)	3 (CCAA)
<i>selfing</i>					
S1	5	5	5	22	14
<i>selfing</i>					
S2	5	5	-	23	15

5

6

1 **Table 2.** List of primer sequences and primer pairs used for SSAPs

2

	<i>Primer name</i>	<i>Sequence (5' to 3')</i>	<i>Dye</i>
Common primers	Adaptor <i>EcoRI</i> (1)	CTCGTAGACTGCGTACC	
	Adaptor <i>EcoRI</i> (2)	AATTGGTACGCAGTCTAC	
	Eco+0	GACTGCGTACCAATTC	
	E33	Eco+AAG	
	E35	Eco+ACA	
	E36	Eco+ACC	
	E55	Eco+CGA	
	E80	Eco+TAC	
	E84	Eco+TCC	
<i>Athila</i>-SSAP*	Athila157/17 (preamplification)	TTCTCAATCAGCTTAAA	
	Athila40/18	GTAGTCATCATCAAATCC	5' IRD 700
	Athila100/18	GGCAGAGTTCATAGCACT	5' IRD 800
	<i>Selective primer pairs used:</i>	Athila40/18+E35; Athila40/18+E55; Athila40/18+E84; Athila100/18+E35; Athila100/18+E36; Athila100/18+E80; Athila100/18+E84	
<i>BraSto</i>-SSAP	BrMipreamp (preamplification)	GRAAACTGTTCAAYYAATTAR	
	BrMiselect	CAAACATCAAGTATTATG	5' IRD 700
	<i>Selective primer pairs used:</i>	BrMiselect+E36; BrMiselect+E55; BrMiselect+E803	
<i>Bot1</i>-SSAP	3Botpreamp (preamplification)	ATTATTTCCGACGACT	
	3Botsselect	TCCTTGCTGTTTTCTTGT	5' IRD 700
	<i>Selective primer pairs used:</i>	Botsselect+E33; 3Botsselect+E35; 3Botsselect+E55	

3 * from Pouilly *et al.*, 2008

4

1 **Table 3.** Characterization of the SSAP profiling approach developed for the three contrasting
2 TEs *Athila*-like, *BraSto* and *Bot1*. The proportion of non-additive SSAP bands identified and
3 the main parameters affecting this non-additivity are shown in detail.

4

SSAP approach:	<i>Athila</i> -like SSAP	<i>BraSto</i> -SSAP	<i>Bot1</i> -SSAP
Number of SSAP band levels scored:	12 to 52	20 to 38	32 to 58
Percentage of cases* with:			
Strict additivity	52%	22%	26%
<5% non-additivity	32%	29%	45%
5-10% non-additivity	13%	29%	22%
>10% non-additivity	3%	20%	7%
(maximal non-additivity detected, in generation x)	(17% in F1)	(19% in S2)	(17% in S2)
Cross effect	Significant BrCRC +190% [#] BoEMZ1 _{19,38} -74% BoEMZ3 ₃₈ -73% BoEMZ5 ₃₈ -75%	Significant BrCRC +43% BoEMZ2 _{19,38} +40% BoEMZ3 ₃₈ -52%	Significant BrCRC +51% BoEMZ1 ₃₈ -31% BoEMZ3 ₃₈ -37% BoEMZ3 _{19,38} +39%
Cross × Generation effect	Cross effect detected at each generation	Differences between crosses significant in S1 and S2	Differences between crosses significant in S2
Generation effect	Not significant N ₊ >N ₋ in F1, S0, S1 N ₊ =N ₋ in S2	Significant More non-additive bands in S2 N ₊ >N ₋ in F1 and S0 N ₊ =N ₋ in S1 N ₊ <N ₋ in S2	Significant More non-additive bands in S2 N ₊ >N ₋ in F1, S0, S1, S2
Primer effect (primer indicated if it generated more non-additive bands)	Significant Athila40/18E55 +69% Athila40/18-E35 +29% Athila100/18-E84 +28%	Significant E55 +88%	Significant E55 +41% E33 (in S2 only) +10%

5 * a single case is defined as a cross × generation × primer combination

1 # all the percentages were calculated from the results of χ^2 contingency tests as follows: deviation =
2 $((N_{obs}-N_{theor})/N_{theor})\times 100$, N_{obs} : number of non-additive SSAP bands observed, N_{theor} : number
3 of non-additive SSAP bands theoretically expected according to the contingency table.
4 N_+ = number of newly amplified SSAP bands not present in the parental SSAP profiles; N_- : number of
5 SSAP bands present in at least one of the diploid parents but absent in one synthetic line.

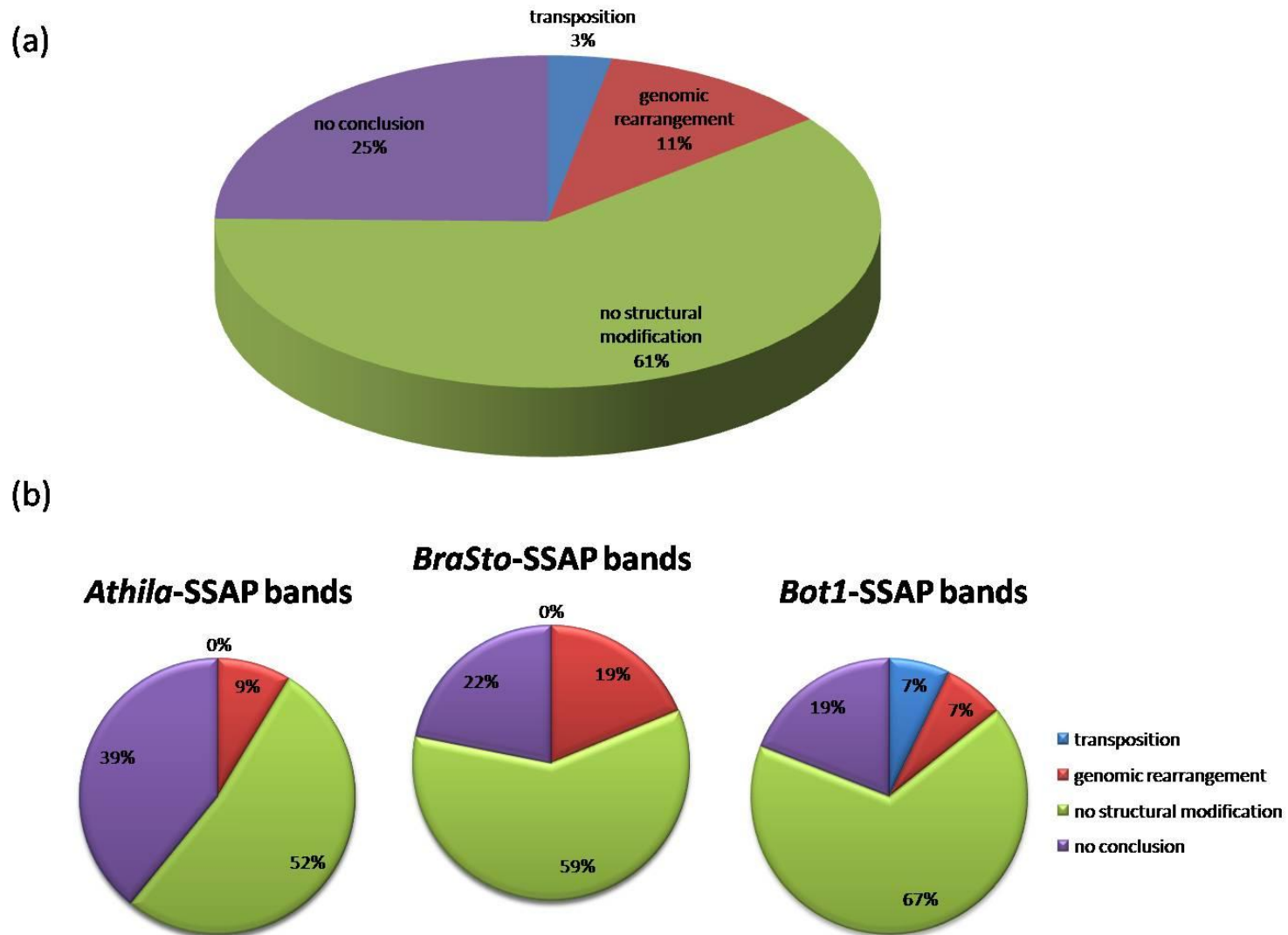


Figure 1

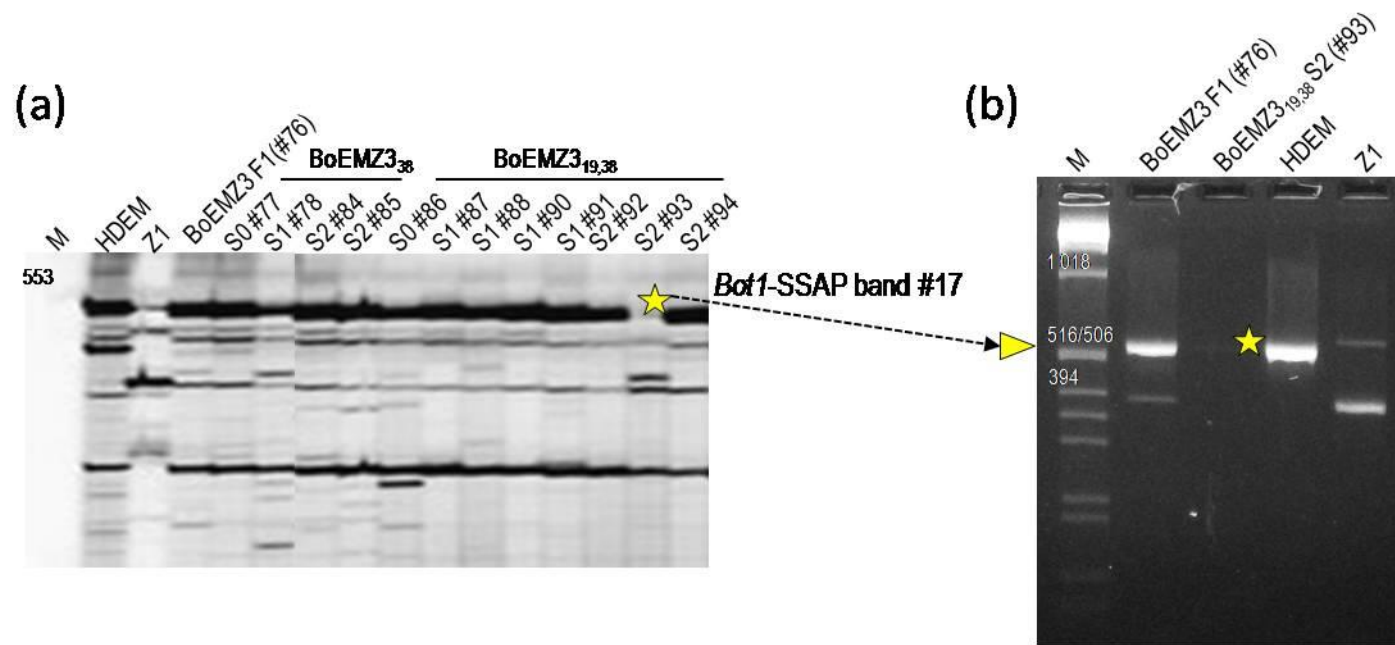


Figure 2

Les résultats obtenus chez *Brassica* et présentés dans le chapitre 2 tendent à montrer que l'activation des ET suite à l'allopolyploïdisation n'est pas une généralité mais plutôt un phénomène spécifique de chaque ET. Ils remettent donc quelque peu en cause l'hypothèse de l'explosion de transposition des ET suite à l'allopolyploïdie, jusque-là couramment admise (McClintock 1984). L'analyse de différents systèmes allopolyploïdes (*Arabidopsis*, *Gossypium*, *Nicotiana*, *Spartina*, *Triticum*...), pour la transposition de différentes familles d'ET, a conduit également à mettre en évidence que l'activité de transposition entraînée par l'allopolyploïdie ne concernait qu'une très faible fraction des ET du génome concerné (pour revue, Parisod et al. 2010). Ainsi l'activation des ET suite à l'allopolyploïdie ne semble pas être un phénomène massif comme on le croyait, mais plutôt un mécanisme restreint à quelques familles d'ET, spécifiques de l'espèce polyploïde étudiée.

Notre étude suggère que les ET resteraient en grande partie réprimés pendant l'alloploïdisation, même si certains sont capables de transposer (comme illustré avec *Bot1*). Elle nous amène donc à nous interroger quant aux mécanismes responsables de ce contrôle des ET. Les petits ARN figurent parmi les acteurs épigénétiques majeurs de la répression des ET, et ils agiraient notamment en contrôlant leur transcription (Mirouze et al. 2009). Dans la suite de mon projet de thèse, je me suis donc attachée à étudier si ces petits ARN pouvaient effectivement jouer un rôle dans le contrôle des ET pendant l'allopolyploïdisation, et plus généralement s'ils possédaient un rôle prépondérant lors de la formation d'une nouvelle espèce allopolyploïde.

CHAPITRE 3

**L'allopolyploïdie entraîne une réponse immédiate
et transitoire des populations de petits ARN
non codants**

L'allopolyplœdie entraîne une réponse immédiate et transitoire des populations de petits ARN non codants

Les petits ARN non codants sont des acteurs majeurs de la régulation de l'expression des gènes et de la stabilité structurale du génome. Ils sont impliqués dans la répression des gènes au niveau post-transcriptionnel, par une action directe sur les transcrits en conduisant à leur dégradation ou au blocage de la traduction, mais aussi transcriptionnel, à l'aide du maintien (ou de la mise en place) de régions hétérochromatiques dans le génome, en contrôlant la méthylation de l'ADN et le remodelage de la chromatine. L'étude de l'évolution des mécanismes de régulation par petits ARN a mis en évidence une origine ancestrale de ces mécanismes de régulation, avec un rôle initial de défense contre l'invasion du génome hôte par des éléments endogènes ou exogènes, que sont les éléments transposables et les séquences de type viral (Vaucheret & Fagard, 2001).

Dans le cadre de l'allopolyplœdie du colza, nous nous sommes intéressés à évaluer l'impact que pourrait avoir l'allopolyplœdisation sur l'expression des petits ARN non codants, et à identifier éventuellement des populations de petits ARN répondant spécifiquement à l'allopolyplœdie. De tels petits ARN joueraient un rôle majeur dans la formation du génome néo-allopolyplœide, notamment par le contrôle des éléments transposables de chacun des génomes parentaux, sujets à la réactivation suite au choc génomique engendré par la rencontre de ces génomes différenciés dans une même cellule.

3.1. La découverte de l'interférence par ARN et du monde des petits ARN non codants

Les organismes procaryotes et eucaryotes emploient diverses catégories d'acides ribonucléiques, ARN, pour établir et maintenir leurs fonctions cellulaires de base. Les rôles classiquement étudiés de ces ARN concernent la biosynthèse des protéines, où ils servent (i) d'intermédiaires entre la séquence d'acide désoxyribonucléique (ADN) et les protéines, en tant qu'ARN messagers (ARNm), mais aussi (ii) d'éléments structuraux dans la constitution des ribosomes, et ce sont alors les ARN ribosomiques (ARNr), ou encore (iii) d'adaptateurs pour les acides aminés au cours de la traduction protéique en ce qui concerne les ARN de transfert (ARNt). La découverte relativement récente d'un nombre toujours croissant de petits ARN dits non codants, c'est-à-dire n'ayant pas la finalité de coder une quelconque séquence protéique, a complètement changé la vision biologique que l'on avait des ARN et de leur rôle dans la régulation de l'expression des gènes (Ossowski et al. 2008 ; Vaucheret 2008).

En 1990, l'équipe de Richard Jorgensen (*DNA plant technology corporation, Oakland, California, USA*) décrit chez les plantes un mécanisme moléculaire où la surexpression d'un transgène en orientation sens (i.e. correspondant à la version de l'ARN traduit en protéine) résultait, de façon inattendue, en une forte réduction (voire suppression complète) de l'expression du dit transgène mais aussi du gène endogène homologue – amenant à nommer ce phénomène de **co-suppression**. Les objectifs initiaux de son étude étaient de surexprimer un gène codant une *chalcone synthase* (CHS), enzyme impliquée dans la biosynthèse des anthocyanes, pour intensifier la pigmentation violette des pétales d'une fleur de pétunia, et chercher à déterminer ainsi si cette enzyme était un facteur limitant majeur dans la biosynthèse des anthocyanes (Napoli et al. 1990). Au lieu d'obtenir des fleurs à coloration des pétales plus intense, ils ont obtenu une grande gamme de variations au niveau du phénotype des plantes transgéniques, allant de fleurs complètement blanches à des fleurs à coloration en mosaïque présentant de grands secteurs blancs (Figure 3.1.). Sans que le mécanisme moléculaire soit encore découvert, R. Jorgensen et ses collaborateurs mettaient en évidence l'existence d'interactions entre molécules d'ARN conduisant à une extinction de l'expression du gène concerné. C'est avec *Caenorhabditis elegans* que le mode d'interaction entre molécules d'ARN allait être caractérisé, mettant par ailleurs en évidence la généralité biologique du mécanisme de régulation de l'expression génique nouvellement découvert.



Figure 3.1. Mise en évidence du phénomène de co-suppression chez le pétunia, par les travaux de Jorgensen. Phénotypes floraux obtenus pour deux plantes transgéniques obtenues après transformation d'un génotype de pétunia à fleur violette (A) à l'aide d'une construction 'promoteur 35S CMV – gène CHS': aucune des fleurs n'est plus sombre que les fleurs parentales, mais toutes montrent des absences de coloration, illustrant le phénomène de co-suppression de l'expression du transgène et du gène endogène homologue. Photos publiées dans Napoli et al. 1990.

La première caractérisation d'une interaction ARN-ARN impliquant un ARN antisens (i.e. ARN complémentaire de l'ARN messager cible, dont l'association bloque la traduction de ce dernier) chez *C. elegans* remonte aux travaux de Victor Ambros (*Dartmouth Medical School, Hanover, New Hampshire, USA*) et collaborateurs. Ils procèdent en effet au clonage et à la caractérisation du gène *lin-4*, un agent régulateur majeur du développement larvaire de *C. elegans* par son rôle de répression de l'expression de deux autres gènes *lin-14* et *lin-28*, du fait de l'association entre les transcrits de *lin-4* qui sont complémentaires des ARN messagers de *lin-14* et *lin-28* (Lee et al. 1993). La baisse correspondante de la synthèse des protéines LIN-14 et LIN-28 marque l'entrée dans un nouveau stade du développement. *Lin-4* est un gène anormalement court, ne codant pour aucune protéine et présentant des transcrits très courts de 22-nt provenant de précurseurs à structure tige-boucle ; il s'agit du **premier micro ARN identifié** (pour revue, Ambros 2001).

Puis en 1998, toujours chez *C. elegans*, les équipes d'Andrew Fire (*Carnegie Institution, Baltimore, Maryland, USA*) et de Craig Mello (*University of Massachusetts Medical School, Worcester, Massachusetts, USA*) démontrent, au cours de travaux en collaboration, que la présence d'ARN double brin portant une séquence identique à celle d'un ARN messenger est capable d'induire une diminution de la quantité du transcrit ciblé et de réprimer ainsi l'expression du gène correspondant, et ce de manière beaucoup plus efficace que lors de l'introduction d'une molécule ARN antisens simple brin. Un aspect particulier à ce mécanisme de répression génique est qu'il peut se mettre en place à partir de l'introduction d'un ARN double brin (sens+antisens) mais aussi d'un ARN simple brin antisens ou même sens, impliquant des mécanismes moléculaires dédiés à la synthèse d'ARN double brin. Ils nomment alors ce mécanisme de **gene silencing** (similaire à celui de la co-suppression identifié chez le pétunia) **d'interférence par ARN** (*RNA interference, RNAi*) (Fire et al. 1998 ; pour revue Mello & Conte, 2004).

Cette découverte leur vaut d'obtenir le prix Nobel en 2006 :

The Nobel Prize in Physiology or Medicine 2006: Andrew Z. Fire, Craig C. Mello

Prize motivation: "for their discovery of RNA interference - gene silencing by double-stranded RNA"

Durant la même période, David Baulcombe (*Sainsbury Laboratory, Norwich, UK*) mène ses recherches sur la compréhension des mécanismes de défense mis en place par les plantes contre les virus, connus pour utiliser des ARN doubles brins comme moyen de multiplication dans la cellule. Avec comme modèles la tomate et le tabac, et en étudiant les mécanismes de répression génique survenant suite à l'introduction d'un transgène ou d'un virus, il identifie ainsi des populations de brins d'ARN antisens beaucoup plus courts, de l'ordre de 25-nt, accumulés dans la cellule végétale, et identifiés comme les intermédiaires actifs de l'ARN interférence : on vient de découvrir les **petits ARN non codants**, et en particulier les micro ARN des plantes, agents moléculaires indispensables à la mise en place de la répression post-transcriptionnelle de l'expression des gènes - *post-transcriptional gene silencing, PTGS* (Hamilton & Baulcombe, 1999).

Cette régulation post-transcriptionnelle de l'expression de gènes par les petits ARN a été rapidement retrouvée chez la plupart des organismes eucaryotes (pour revue, Hammond et al. 2001), incluant la drosophile (Hammond et al. 2000) et les mammifères (Elbashir et al. 2001). Au cours des années suivantes, les activités de recherche menées sur la compréhension mécanistique de l'interférence par ARN ont permis l'identification des protéines et complexes protéiques spécifiquement impliqués : c'est notamment chez la drosophile que l'identification du complexe RISC, *RNA-induced silencing complex* (Hammond et al. 2000) puis de l'endoribonucléase DICER (Berstein et al. 2001) a été faite pour la première fois. Aujourd'hui, le RNAi est devenu un outil efficace pour étudier la fonction biologique des gènes, par annulation contrôlée et ciblée de l'expression de ces derniers.

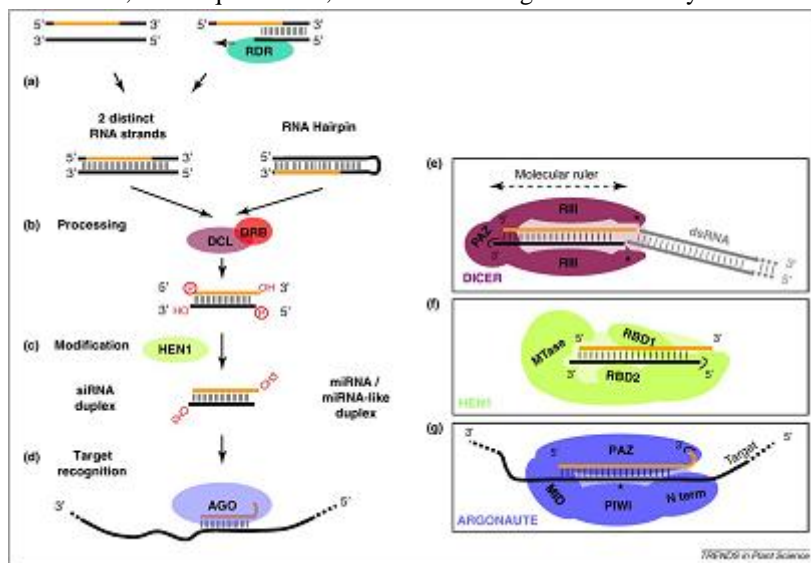
Description des protéines impliquées dans les voies de biogénèse et de régulation par petits ARN non codants

Les voies de biogénèse puis d'action des petits ARN non codants font appel à des protéines spécifiques qui sont des membres de grandes familles de protéines (Vaucheret 2008 ; Wassenegger et al. 2006), impliquant des phénomènes de redondance et de compensation fonctionnelle, même si ces relations sont loin d'être clairement élucidées.

- ◆ Les protéines de type **Dicer**, **DCL** (*Dicer-like protein*), sont à l'origine de la coupure de l'ARN double brin en fragments double brin (duplex) de 21-nt à 25-nt. Chez *Arabidopsis* on connaît l'existence de quatre protéines de type Dicer, DCL1, -2, -3 et 4. L'action de ces protéines DCL est souvent associée à celle d'une protéine DRB (*dsRNA-binding protein*), à l'origine de la sélection du substrat ARN à découper ; chez *Arabidopsis*, DCL2 et DCL3 ne requièrent pas cette protéine pour fonctionner.

- ◆ La protéine **HEN1** (*Hua Enhancer 1*) agit pour méthylter les extrémités 3' des petits ARN, notamment des micro ARN, assurant leur protection contre la dégradation.

- ◆ Les petits ARN non codants sont ensuite chargés dans les complexes protéiques **RISC** (*RNA-induced silencing complex*). La définition d'un complexe RISC repose sur la présence *a minima* d'une protéine **Argonaute**, **AGO**. Un seul brin du duplex résultant du découpage par DCL va être pris en charge par le complexe RISC : c'est le petit ARN actif, et on parle alors de *guide strand* (e.g. miRNA) alors que l'autre brin, le *passenger strand* ou encore *star strand* (e.g. miRNA*) va être dégradé. Ce complexe va reconnaître les ARN messagers cibles, sur la base de complémentarité de séquence avec le petit ARN chargé. Dix protéines Argonaute ont été identifiées chez *Arabidopsis*, dont les activités biologiques de six d'entre-elles sont connues. Les protéines AGO1, AGO4 et AGO7 mènent au clivage de leurs ARNm cibles, AGO1 et AGO10 ont la capacité de bloquer la traduction de leurs cibles, tandis que AGO4, AGO6 et AGO9 guident la méthylation de sites ADN cibles (Vaucheret, 2008).



Domaine PAZ (Piwi/Ago/Zwille). Il s'agit d'un domaine d'association à l'ARN (*RNA-binding domain, RBD*) très conservé, que l'on trouve au niveau des protéines Dicer et Argonaute, et qui maintient stable l'interaction entre la protéine et la molécule d'ARN, que ce soit pour découper le dsRNA initial (Dicer) ou pour guider le blocage de l'ARN messager cible afin d'en assurer la dégradation ou d'en empêcher la traduction (Argonaute).

Domaines RIII. Au niveau des protéines Dicer, deux domaines de RNase de type III interviennent dans la coupure de chacun des brins de l'ARN double brin. La distance entre le domaine PAZ et les sites opérationnels de coupure détermine la taille des duplex générés.

Représentation simplifiée du réseau moléculaire de biogénèse et d'action des petits ARN non codants chez les plantes, identifiant les principales protéines impliquées. D'après Vazquez et al. 2010.

D'autres protéines impliquées dans les voies de régulation par petits ARN jouent des rôles clés chez les plantes :

- ◆ Les protéines **RDR** sont des polymérases d'ARN dépendantes d'ARN qui permettent, à partir d'un brin ARN, la synthèse d'ARN double brin correspondant au précurseur *dsRNA* à l'origine des petits ARN non codants. Ces rôles ont été particulièrement décrits chez les plantes pour les protéines RDR2 et RDR6.

- ◆ Les ARN polymérases **Pol IV** et **Pol V**, spécifiques des plantes, sont à l'origine de la transcription des ARN messagers impliqués dans les processus de répression des gènes au niveau transcriptionnel, par guidage des modifications des marques épigénétiques de la chromatine (méthylation de l'ADN, modification post-traductionnelle des histones).

Les détails d'intervention de ces différentes protéines en fonction de la classe de petits ARN non codants considérée sont donnés dans le texte principal.

Depuis la découverte des premières séquences de petits ARN impliquées dans ce mécanisme d'interférence par ARN, de nombreuses études ont permis de compléter l'identification de ces petits ARN de 21 à 26-nt, mettant en évidence la complexité et le nombre important des acteurs moléculaires de la régulation de l'expression des gènes.

3.2. Les petits ARN non codants : définition et classification

Chez les plantes, les petits ARN non codants sont impliqués dans des processus biologiques très divers tels que le développement de la plante, les réponses aux stress, ou les mécanismes de défense contre les virus et autres agents pathogènes, impliquant des rôles spécifiques au niveau de la régulation de l'expression génique et du maintien de régions hétérochromatiques dans le génome. Du fait de la diversité de fonctions biologiques où l'on trouve impliqués les petits ARN, leur classification a généralement pris en compte comme critère premier leur mode de biogénèse.

De manière générale, les petits ARN non codants se forment à partir de précurseurs ARN simple brin (ssRNA) qui adoptent une structure particulière en tige-boucle (= épingle à cheveux, *hairpin*), ou de précurseurs ARN double brin (dsRNA). Ces derniers sont produits à partir (1) de gènes endogènes, comme dans le cas des gènes codant les micro ARN et que l'on appelle les gènes *MIR* (*MIR genes*), (2) de répétitions inversées (IR) ou de transcrits antisens qui se chevauchent au moins partiellement (cas des NAT-siRNAs), (3) d'ARN viraux, ou encore (4) à partir de produits issus de l'action d'ARN polymérases dépendantes d'ARN (RDRP). Ces structures d'ARN double brin sont reconnues spécifiquement par des endonucléases de type Dicer (DCL) qui vont les couper en petits fragments, le plus souvent de 21-nt à 25-nt, et donc générer les petits ARN non codants correspondants.

Initialement (et permettant une classification relativement claire et simple à appréhender), les petits ARN non codants ont été classifiés en fonction de leur origine / voie de biogénèse et non en fonction de leur rôle dans la cellule (Voinnet 2009). (1) Les petits ARN produits à partir du découpage de précurseurs ARN double brin sont connus comme étant **les petits ARN interférents, siRNAs** – *small interfering RNAs*, tandis que (2) ceux qui présentent une origine endogène et possèdent une structure tige-boucle une fois transcrits ont été nommés **micro ARN, miRNAs**. Cette classification n'était cependant pas totalement satisfaisante, car trop simplifiée, notamment du fait de l'existence de précurseurs à structure en tige-boucle mais qui produisaient des siRNAs dépourvus des principaux critères des micro ARN (en termes de *MIR genes*, Meyers et al. 2008), nommés alors siRNAs de type micro ARN, *miRNA-like siRNAs*, sur proposition de Zhu et collaborateurs en 2008 (Vazquez et al. 2010).

La classification la plus récente des petits ARN chez les plantes a été proposée par Axtell (2013). Dans cette classification, les petits ARN ont été divisés en deux classes principales selon la nature des précurseurs à l'origine des petits ARN : (1) **les ARN en épingle à cheveux, hpRNAs – hairpinRNAs**, issus de précurseurs simple brin qui possèdent une structure complémentaire en tige-boucle et (2) **les petits ARN interférents, siRNAs – small interfering RNAs** toujours originaires de précurseurs double brin formés par l'hybridation intermoléculaire de deux brins d'ARN complémentaires (Figure 3.2.). A cette classification primaire, assez similaire à celle décrite précédemment (Voinnet 2009) se superposent les classifications secondaires et tertiaires prenant en compte le rôle biologique du petit ARN considéré (Figure 3.2. et description plus détaillée dans la suite du manuscrit).

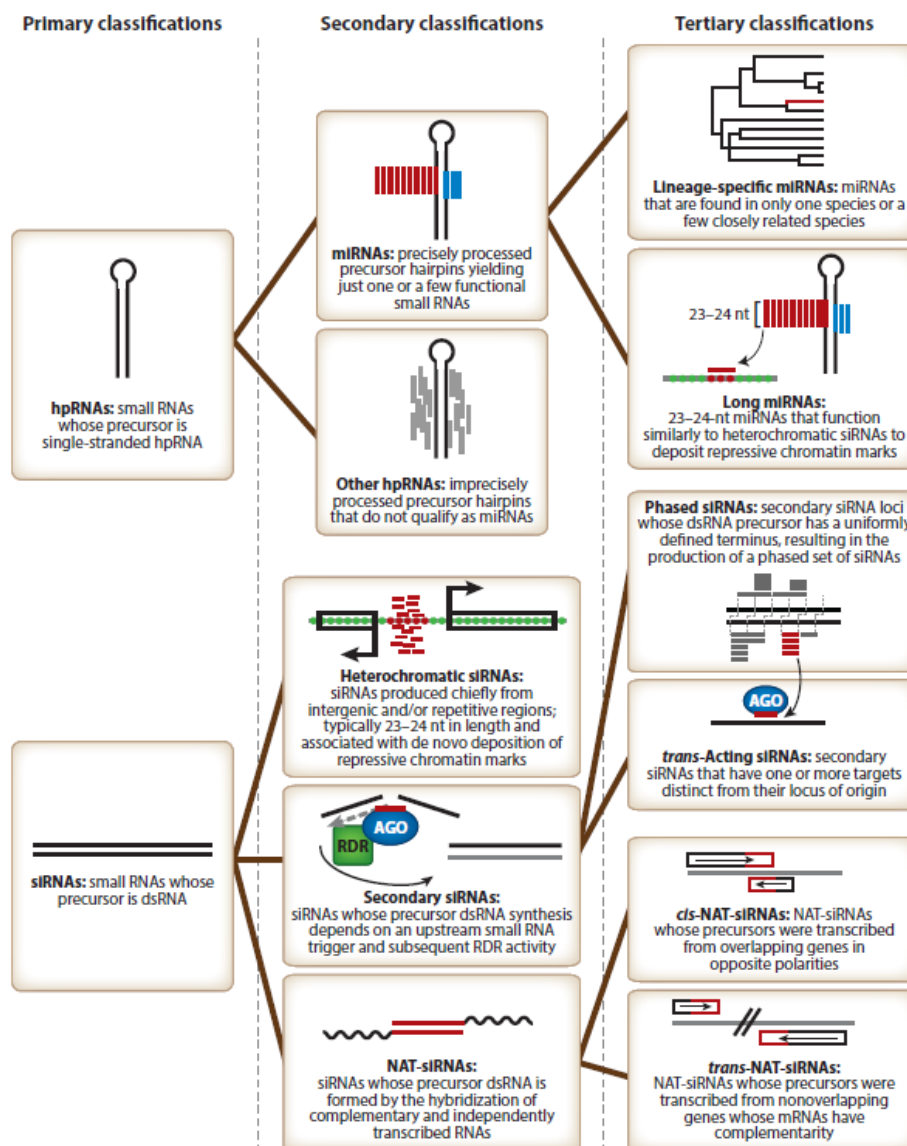


Figure 3.2. Système de classification hiérarchique des petits ARN non codants chez les plantes. Les lignes épaisses en noir indiquent les relations hiérarchiques. Abréviations : *dsRNA*, *doubled-stranded RNA*, ARN double brin ; *hpRNA*, *hairpin RNA*, ARN en épingle à cheveux (à structure tige-boucle) ; *miRNA*, *micro RNA*, micro ARN ; *NAT-siRNA*, *natural antisens transcript small interfering RNA*, petit ARN interférent issu de transcrits natifs antisens ; *siRNA*, petit ARN interférent. D'après Axtell 2013.

L'utilité de ces classifications doit être évaluée à la fois par le degré selon lequel les regroupements de petits ARN reflètent une réalité biologique, et la mesure dans laquelle ces regroupements facilitent la discussion sur le sujet. La classification primaire semble assez simple et robuste, car chez les plantes, les petits ARN proviennent nécessairement de l'un des deux types de précurseur décrits. Certains regroupements au niveau de la classification secondaire apparaissent également très robustes: au sein de chaque classe des miRNAs, siRNAs hétérochromatiques et siRNAs secondaires, les petits ARN concernés ont les mêmes protéines impliquées au niveau de leur biogenèse, les mécanismes d'action ont été démontrés expérimentalement, et les distributions de taille de ces petits ARN clairement identifiées. En outre, les caractéristiques qui définissent ces trois groupes de petits ARN sont conservées entre espèces végétales. En revanche, les hpRNAs (autres que des miRNAs) ainsi que les NAT-siRNAs pourraient être encore scindés en différentes classes: pour chaque catégorie, les exigences en protéines impliquées dans leur biogénèse sont variables, et les distributions de taille de ces petits ARN également. Le système de classification doit donc être considéré comme un travail en cours et surtout comme étant susceptible d'évoluer et de s'adapter aux connaissances que l'on ne cesse encore d'acquérir sur le monde des petits ARN non codants (Axtell 2013).

3.2.1. Les micro ARN (miRNAs) des plantes

Les micro ARN (miRNAs) sont représentés chez l'ensemble des eucaryotes (He et al. 2008; Siomi et Siomi 2010), ainsi que certains virus à ADN. Une particularité de la classe des micro ARN des plantes est qu'elle représente la plus petite classe de petits ARN, en termes d'effectifs, par rapport à l'ensemble des populations de petits ARN, *a contrario* des animaux (ou encore de la drosophile) où il s'agit de la classe la plus représentée. C'est d'ailleurs pour cette raison que des efforts de méthodologie dans l'annotation des micro ARN des plantes ont particulièrement été fournis par la communauté végétaliste (Meyers et al. 2008).

Bien qu'ils n'aient été découverts que depuis environ une décennie, les miRNAs sont connus comme les régulateurs de l'expression des gènes les plus importants (Sun 2012). Ils régulent négativement l'expression des gènes **au niveau post-transcriptionnel** (on parle de *post-transcriptional gene silencing, PTGS*), soit par clivage et dégradation de l'ARN messager cible, soit par inhibition de la traduction (Bartel 2007 ; Brodersen et al. 2008). Ils sont impliqués dans la régulation de presque tous les processus biologiques et métaboliques, tels que le développement des organes, les voies de signalisation, et les réponses aux stress environnementaux (Jones-Rhoades et al. 2006 ; Voinnet 2009).

3.2.1.a. Biogénèse des micro ARN

La plupart des miRNAs des plantes présentent une taille de 21 nucléotides (certains pouvant être de 20 ou 22-nt). Ils font intervenir spécifiquement la protéine DCL1 pour leur biogénèse, et la protéine AGO1 pour leur fonction (même si quelques exceptions ont été décrites). Un schéma général de la biogénèse des micro ARN est donné en Figure 3.3.

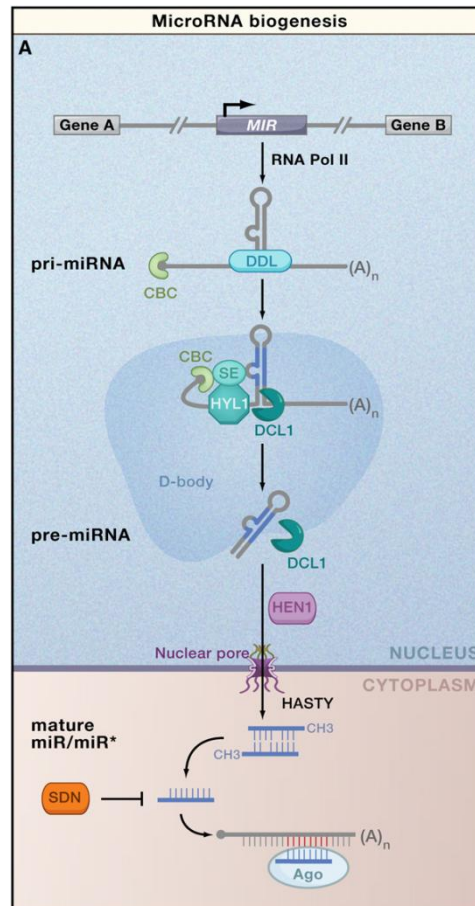


Figure 3.3. Schéma général de l'organisation moléculaire de la biogénèse des micro ARN chez les plantes. D'après Voinnet 2009.

Dans le noyau de la cellule. Les micro ARN sont codés par des gènes endogènes appelés gènes *MIR* (*MIR genes*). Chez les plantes, ils se retrouvent principalement localisés au niveau des régions intergéniques. Comme les gènes codants pour les protéines, les gènes de micro ARN sont transcrits par l'ARN polymérase II (Pol II) (Jones-Rhoades et al 2006). Les transcrits des micro ARN sont appelés des micro ARN primaires (pri-miRNAs). L'ARN polymérase Pol II génère des pri-miRNAs capés (avec coiffe) et polyadénylés (Lee et al. 2004). Une fois que les pri-miRNAs ont été transcrits, la protéine de liaison à l'ARN DAWDLE (DDL) se lie aux pri-miRNAs, pour éviter leur dégradation et augmenter ainsi leur taux de conversion en précurseurs de miRNA (pre-miRNAs) et favoriser la formation des micro ARN matures. L'enzyme de type Dicer DCL1 coupe alors les extrémités du pri-miRNA pour générer les pre-miRNAs qui présentent des structures secondaires en tige-boucle

(Meyers et al. 2008). Ce processus nécessite l'action de plusieurs protéines : une protéine de liaison à l'ARN double brin appelée HYPONASTIC LEAVES1 (HYL1), la protéine SERRATE (SE), le complexe nucléaire de liaison à la coiffe, (CBC) et la protéine effective Dicer DCL1 (Voinnet 2009 ; Sun 2012). Les pre-miRNA sont ensuite clivés par DCL1 et HYL1 en un duplex miRNA: miRNA*, toujours dans le noyau de la cellule (Song et al 2007). Ces duplex présentent 2 nucléotides débordant au niveau de l'extrémité 3' de chaque brin, et peuvent être facilement dégradés par une classe d'exonucléases, les SMALL RNA DEGRADING NUCLEASE (SDN) (Ramachandran et Chen 2008). Afin de stabiliser ces duplex miRNA:miRNA* et empêcher leur dégradation, une méthyltransférase dédiée aux petits ARN, appelée Hua Enhancer 1 (HEN1), méthyle immédiatement les extrémités 3' de chaque brin (Yu et al 2005).

Dans le cytoplasme de la cellule. Le duplex miRNA:miRNA*, aux extrémités méthylées et donc protégées, est ensuite transporté du noyau vers le cytoplasme par action de la protéine HASTY (Park et al. 2005). Dans le cytoplasme, le duplex est séparé en deux brins, le miRNA guide (*guide miRNA*) et le miRNA*.

3.2.1.b. Mode d'action des micro ARN

Après exportation, l'action des miRNAs se passe donc dans le cytoplasme de la cellule. Le miRNA guide est alors chargé dans la protéine AGO1 (dans la majorité des cas, une moindre proportion étant reliée à l'activité d'AGO10 ; Voinnet 2009) et incorporé dans le complexe inducteur de silencing, RISC (*RNA-induced silencing complex*), tandis que le miRNA* est majoritairement dégradé, par un mécanisme encore méconnu. Le micro ARN lié au complexe RISC va réguler l'expression de gènes cibles, de manière spécifique, par reconnaissance de similarité de séquence des ARN messagers correspondants, initiant leur dégradation ou le blocage de leur traduction.

Tous les ARN messagers cibles de micro ARN de plantes vérifiées expérimentalement à ce jour, présentent un degré de complémentarité de séquence important avec les miRNAs concernés. La complémentarité de séquence parfaite est rare, et des exemples de régulation par miRNA avec plus de cinq mésappariements entre le miRNA et sa cible ont été répertoriés (Vaucheret com. pers.). Cela conduit ainsi à ce qu'un unique miRNA puisse avoir plusieurs dizaines à une centaine de cibles différentes, notamment quant le mécanisme de répression est le blocage de la traduction, la complémentarité de séquence étant plus relâchée (Pratt & MacRae, 2009). Les mésappariements ont cependant tendance à n'être autorisés qu'en des sites précis (pour revue Axtell 2013) : soit au niveau du premier nucléotide de la région complémentaire, correspondant au premier nucléotide à l'extrémité 5' du miRNA, soit au niveau de l'extrémité 3' du miRNA, sachant que la tolérance des mésappariement dans la moitié côté 5' est globalement plus faible (Figure 3.4., Mallory et al. 2004).



Figure 3.4. Schéma général de mise en place de l'appariement entre le micro ARN et sa cible ARNm au niveau de la région de complémentarité de séquence, chez les plantes. A noter que chez les animaux, une région dite graine (*seed region*) s'étend des nucléotides 2 à 7 (ou 1 à 8) où un appariement exact est requis. Figure adaptée d'après Axtell 2013.

A noter que certains sites en position interne peuvent aussi autoriser des mésappariements, entraînant la formation de renflements dans la région concernée ; ces cas ont été rarement observés chez les plantes, et ne permettent alors pas le rôle classique de clivage ou blocage de la traduction de la cible par le micro ARN (Axtell 2013).

Une fois que la cible d'un micro RNA est identifiée, on observe une baisse du niveau d'accumulation de l'ARNm cible sous forme pleine longueur, en lien avec l'activité endonucléolytique de la protéine AGO1 associée au micro RNA qui amène au clivage de l'ARNm cible, généralement entre les positions 10 et 11 de la région d'appariement. C'est ainsi que la validation expérimentale de l'identification des cibles de micro ARN se fait par une approche de séquençage haut débit du 'dégradome', qui permet d'identifier les résultats directs de clivage par miRNA en séquençant les extrémités 5' des populations d'ARNm non capés (e.g. *PARE, parallel analysis of RNA ends* – German et al. 2009) à confronter aux sites prédits de clivage des miRNAs étudiés, par utilisation d'algorithmes dédiés (CleaveLand – Addo-Quaye et al. 2009). Cependant, l'accumulation des connaissances sur le mode d'action des micro ARN montre que dans de nombreux cas, leur impact de répression serait plus marqué au niveau de la synthèse de la protéine concernée que sur le niveau d'accumulation de l'ARNm, mettant en évidence l'importance, si ce n'est la dominance, du mécanisme de répression par blocage de la traduction. En conclusion, d'après les données expérimentales acquises sur l'action des micro ARN, il semblerait que la plupart des miRNAs de plantes répriment leurs cibles géniques par un mécanisme impliquant une combinaison moléculaire (1) de déstabilisation de l'ARNm cible (souvent par clivage) et (2) de répression traductionnelle (Voinnet 2009).

Notons enfin, qu'en prenant en compte le mode de biogénèse des micro ARN, certains micro ARN de 24-nt ont pu être identifiés : essentiellement chez *Arabidopsis* (Vazquez et al. 2008) et le riz (Wu et al. 2010), où l'on parle alors de micro ARN longs (*long miRNAs*, Figure 3.2.). La génération de ces miRNAs de 24-nt à partir de structures en tige-boucle codées par des gènes endogènes fait cependant intervenir la protéine Dicer DCL3. D'ailleurs, la fonction de ces miRNAs de 24-nt apparaît atypique et plus proche de celle des siRNAs (voir ci-dessous).

En effet, ils partagent la voie de régulation des siRNAs hétérochromatiques et conduisent à des modifications des marques chromatinienne dans les régions où sont localisés leurs gènes cibles, agissant alors comme répresseurs potentiels au niveau transcriptionnel de l'expression des gènes. Il s'agirait de « proto-miRNAs », originaires de gènes *MIR* d'évolution récente, à l'inverse des miRNAs classiques provenant de gènes *MIR* plus anciens (Vazquez et al. 2008).

3.2.1.c. Caractéristiques évolutives des micro ARN

Les micro ARN possèdent un ensemble d'ARN messagers cibles définis, sur la base du principe de complémentarité de séquence (3.2.1.b), et les familles de micro ARN peuvent être conservées sur de longues distances évolutives (Jones-Rhoades et al. 2006 ; pour revue, Cuperus et al. 2011), avec certains micro ARN conservés par exemple entre les angiospermes et la mousse *Physcomitrella patens* – sur 91 familles de micro ARN identifiées chez *Arabidopsis*, 22 d'entre elles sont retrouvées chez les monocotylédones, et 9 chez *P. patens* (Fahlgren et al. 2010). Les micro ARN conservés ciblent des ARNm homologues entre plusieurs espèces (Jones-Rhoades 2012), ce qui démontre que les relations micro ARN / ARNm cible ont souvent été stables pendant de longues périodes dans l'évolution des plantes.

Cependant, chez les plantes, on a pu remarquer que relativement peu de micro ARN étaient conservés entre espèces, avec la majorité d'entre eux apparaissant spécifiques d'une espèce ou d'un groupe d'espèces très apparentées (pour revue, Cuperus et al. 2011). Généralement, ces micro ARN présentent des niveaux d'expression très faibles et ne possèdent que peu ou pas de cibles géniques à réguler (Fahlgren et al. 2010). Ce sont les *lineage-specific miRNAs* de la classification tertiaire donnée en Figure 3.2. Ces observations impliquent que les gènes *MIR* à l'origine de ces miRNAs présentent un turn-over évolutif important, et que la majorité d'entre eux sont donc d'évolution très récente au sein d'un génome donné. C'est ainsi qu'il a été estimé chez *Arabidopsis* que 18% et 22% des loci de miRNAs identifiés respectivement chez *A. lyrata* et *A. thaliana* apparaissaient uniques ou ayant fortement divergé au sein de l'espèce considérée, la divergence des génomes de ces deux espèces datant d'il y a environ 10 millions d'années (Fahlgren et al, 2010). Il a par ailleurs été montré que ces jeunes gènes *MIR* évoluaient principalement sous neutralité, au contraire des gènes *MIR* plus anciens sous sélection purifiante, impliquant un rôle fonctionnel mineur des miRNAs correspondants, soumis majoritairement à l'effet de la dérive génétique (Fahlgren et al. 2010). Cette évolution rapide des gènes à l'origine de petits ARN de type micro ARN, toujours par phénomène de duplication, pourrait représenter un moyen de maintenir latent, en attente, un mécanisme de régulation efficace au cours de l'évolution, où ces agents de régulation de l'expression de gènes doivent notamment adapter des mécanismes de réponse aux stress, dont la mise en place se doit d'être rapide et efficace.

3.2.2. Les autres petits ARN en épingle à cheveux (autres hpRNAs)

Il s'est avéré que certains loci dans le génome des plantes produisaient des petits ARN non codant présentant, comme les micro ARN classiques, des précurseurs à ARN simple brin à structure tige-boucle. Cependant ils ne présentent pas l'ensemble des critères inclus dans l'annotation des micro ARN, dont celui primordial de l'existence d'un site précis d'excision du duplex miRNA:miRNA* de la tige formée dans la structure tige-boucle (considéré comme le premier critère d'annotation – Meyers et al. 2008), conduisant d'ailleurs à des micro ARN dits phasés. De tels petits ARN ont été observés chez *Arabidopsis* suite à leur formation à partir de répétitions inversées (*inverted-repeats, IR*), mais les précurseurs présentaient alors des structures tige-boucle plus longues que celles correspondant à des micro ARN classiques, et leurs tailles variaient de 21 à 24-nt ; de plus, ils feraient intervenir les autres protéines Dicer DCL2, -3 et -4 dans leur biogénèse (Dunoyer et al. 2010). On parle aussi d'*endoIR-siRNAs*.

Le rôle de ces hp-RNAs dans les mécanismes de régulation de l'expression des gènes et de contrôle de la structure du génome demeure encore flou, de même que leur évolution et conservation chez les plantes. On a cependant pu mettre en évidence leur mobilité à travers les tissus vasculaires de la plante, s'accompagnant d'une répression **au niveau transcriptionnel** (on parle de *transcriptional gene silencing, TGS*) par méthylation *de novo* de sites génomiques distants. On a d'ailleurs émis l'hypothèse de leur rôle éventuel dans l'adaptation aux conditions environnementales, et dans la mémoire épigénétique transgénérationnelle (Dunoyer et al. 2010). Des efforts dans les méthodes d'annotation de ces petits ARN au niveau des génomes végétaux aujourd'hui séquencés permettraient de mieux cerner leur éventuelle signification biologique et évolutive.

3.2.3. Les siRNAs hétérochromatiques

La plupart des siRNAs hétérochromatiques présentent une taille de 24 nucléotides, certains de 23-nt, ce qui les distingue facilement des miRNAs. Les siRNAs hétérochromatiques ont des exigences bien spécifiques en ce qui concerne les protéines des familles RDR, DCL, et AGO : la plupart dépendent spécifiquement de RDR2 et DCL3 pour leur biogénèse, et au niveau de leur fonction, ils interagissent avec les membres du clade AGO4 des protéines Argonaute. Selon un processus spécifique aux plantes, deux ARN polymérases (que l'on ne trouve que chez les végétaux) sont impliquées dans la régulation épigénétique dirigée par ces petits ARN : les ARN polymérases Pol IV et Pol V, qui ont évolué à partir de l'ARN polymérase Pol II (Haag & Pikaard, 2011).

3.2.3.a. Biogénèse des siRNAs hétérochromatiques

Dans la classification donnée en Figure 3.2., ces siRNAs hétérochromatiques sont endogènes, c'est-à-dire originaires directement du génome considéré. Le modèle établi pour la biogénèse des siRNAs hétérochromatiques (Wierzbicki et al. 2012, dont le schéma est donné en Figure 3.5.) commence par la transcription, à l'aide de l'ARN polymérase Pol IV, d'un ARN simple brin (ssRNA) à partir d'un site correspondant essentiellement à une séquence d'ADN répétée ou à un élément transposable (d'après *Arabidopsis thaliana* – Kasschau et al. 2007).

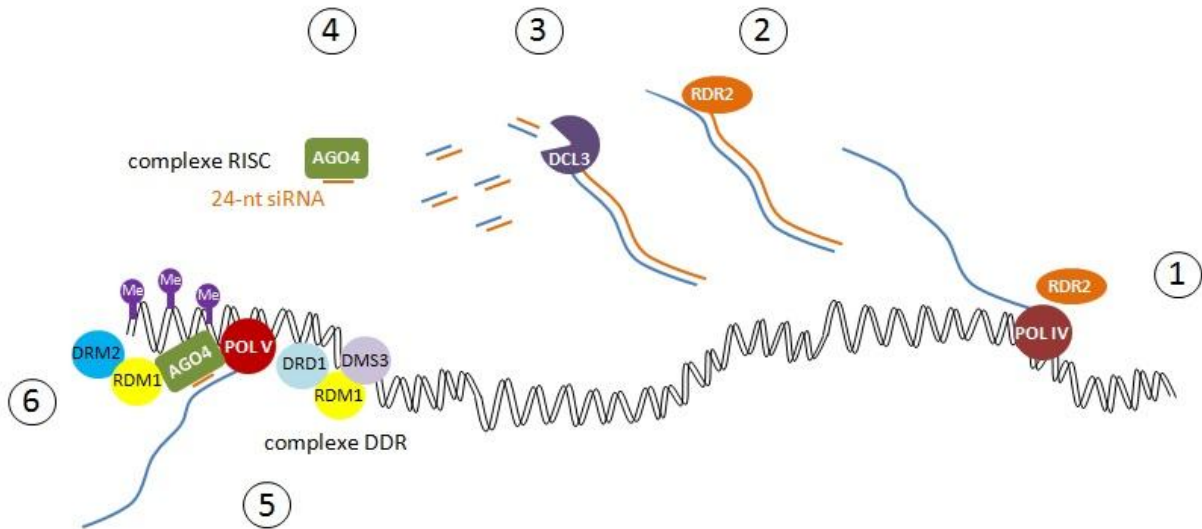


Figure 3.5. Représentation du modèle de méthylation de l'ADN guidée par ARN chez *Arabidopsis thaliana*, impliquant la biogénèse de petits ARN interférents de 24-nt dits hétérochromatiques. (1) La polymérase IV initie la transcription d'un ARN (à partir d'un élément transposable, par exemple) qui (2) par action de la protéine RDR2, sera à l'origine d'un ARN double brin, (3) précurseur des duplex de petits ARN de 24-nt produits par la protéine Dicer DCL3. (4) Les petits ARN de 24-nt, après désappariement, sont chargés dans la protéine Argonaute AGO4 (ou également AGO6, AGO9) pour former le complexe RISC qui (5) ciblera un ARN en cours de transcription par la polymérase Pol V, dont l'activité aura été initiée par le complexe DDR (DRD1, DMS3, RDM1). AGO4 (/6/9) du complexe RISC contribue alors (6) au recrutement de la protéine RDM1 qui va interagir avec la méthyltransférase DRM2 à l'origine de la méthylation *de novo* des cytosines de l'ADN. Figure adaptée d'après Wierzbicki et al. 2012 et Zaratiegui & Martienssen, 2012.

Puis, il y a synthèse de l'ARN en double brin (dsRNA) médiée par l'ARN polymérase dépendante d'ARN, RDR2 (pour *RNA-dependent RNA polymerase 2*). L'ARN double brin qui en résulte est alors coupé par la protéine Dicer DCL3 en produits de 24-nt qui constituent la population de petits ARN interférents hétérochromatiques appariés en duplex. Après désappariement de ces duplex, un brin de 24-nt est chargé dans le complexe RISC sous contrôle d'une protéine Argonaute du clade AGO4 (soit AGO4, -6, et -9, d'après la nomenclature définie chez *Arabidopsis*). Notons que la régulation de l'activité de transcription de Pol IV (de même que Pol V, ci-dessous) demeure encore inconnue, et que les agents moléculaires à l'origine de l'initiation de la transcription de ces sites endogènes spécifiques restent à identifier.

3.2.3.b. Mode d'action des siRNAs hétérochromatiques

Les siRNAs hétérochromatiques dirigent des modifications épigénétiques de régions génomiques ciblées. La régulation épigénétique fait référence à des modifications dans le fonctionnement des gènes, qui sont héritées au moment de la mitose et/ou de la méiose, et qui n'impliquent pas de changements dans la séquence de l'ADN (Bird 2007). Cette régulation est réalisée par marques épigénétiques comprenant (1) la méthylation des cytosines de l'ADN (5-méthylcytosine) au niveau des sites en contexte CG, CHG et CHH (H ≠ G) chez les plantes (alors que ce sont les sites en contexte CG qui sont majoritairement ciblés chez les mammifères), et (2) des modifications post-traductionnelles des histones (Matzke et al. 2009 ; Law & Jacobsen, 2010 ; Saze et al. 2012). Ce type de modification, mené par les siRNAs, consiste en l'hétérochromatinisation des loci ciblés par méthylation à l'aide d'une méthyltransférase (DRM2, *Domains rearranged methyltransferase 2*) dans un processus appelé méthylation de l'ADN guidée par ARN (*RdDM – RNA-directed DNA methylation*), s'accompagnant de la méthylation des histones au niveau de ces mêmes loci (avec notamment les phénomènes de méthylation en position H3K9 (lysine 9 de l'histone H3) responsables de la répression du locus concerné, pour revue Saze et al. 2012). L'action de ces siRNAs est donc essentiellement **au niveau transcriptionnel (TGS)**.

De façon indépendante de la biogénèse des siRNAs hétérochromatiques, le complexe protéique DDR (constitué des protéines DRD1, DMS3 et RDM1) permet la transcription d'ARN par la polymérase Pol V (Figure 3.5.). On a pu observer que les sites préférentiels de Pol V correspondaient à des promoteurs proches d'éléments transposables relativement jeunes (du moins chez *Arabidopsis*, Zhong et al. 2012) – sans doute pour en contrôler l'activité transcriptionnelle et donc transpositionnelle. Les transcrits nouvellement synthétisés vont alors servir de base au recrutement du complexe RISC portant la protéine AGO4 (ou AGO6, AGO9) chargée du siRNA, qui est responsable de la reconnaissance du transcrit par complémentarité de séquence. A ce niveau, AGO4 interagit à la fois avec le domaine C-terminal de la grande sous-unité de Pol V et RDM1, cette dernière interagissant avec la méthyltransférase DRM2. DRM2 permet la méthylation *de novo* des cytosines de l'ADN, spécifiquement aux sites de recrutement du complexe protéique médié par le siRNA hétérochromatique associé à AGO4.

De façon intéressante, les loci qui ont ainsi été hétérochromatinisés par la voie impliquant les petits ARN interférents hétérochromatiques sont, dans une majorité des cas et de façon assez paradoxale, ceux qui ont tout d'abord été transcrits par Pol IV, et qui sont donc à l'origine des siRNAs impliqués, mais également au final de leur propre répression épigénétique par action *cis* (Wierzbicki et al. 2012).

Cependant, diverses études ont montré que ces siRNAs hétérochromatiques pouvaient diriger des modifications épigénétiques sur d'autres régions génomiques (action *trans*), voire d'autres molécules d'ADN, que celles dont ils étaient originaires (pour revue, Mirouze 2012), du moment que le siRNA en question trouve un site d'interaction à séquence complémentaire. De par leur mobilité dans l'organisme (préférentiellement des parties végétatives de la plante vers les racines – pour revue, Melnyk et al. 2011), ils peuvent ainsi propager des modifications au niveau des marques épigénétiques entre cellules, tissus, ou même entre individus voire entre espèces (ou peut-être même entre règnes...). Tel est aussi le cas de leur rôle dans les phénomènes d'hybridation, intra- ou interspécifique, qui sera développé ci-dessous (3.4)

3.2.3.c. Conservation et évolution des siRNAs hétérochromatiques

Les siRNAs hétérochromatiques constituent une classe de petits ARN endogènes que l'on retrouve pour l'ensemble des angiospermes étudiées : ils correspondent alors à la classe des petits ARN non codants la plus représentée, d'après les données issues de séquençages haut débit de banques de petits ARN. Cependant, au-delà des plantes à fleurs, lorsque l'on s'intéresse aux représentants des spermatophytes dont on possède aujourd'hui des données génomiques, il apparaît que ces derniers sont dépourvus de ces populations de siRNAs de 24-nt de même que de protéine similaire à DCL3, suggérant une absence de ces modes de régulation du génome ou du moins l'existence de mécanismes alternatifs. Le mécanisme étant par ailleurs globalement conservé au niveau de *Physcomitrella* ou de *Selaginella* (étant là, spécifique de certains tissus), il s'agirait donc d'une perte de ce mécanisme de régulation épigénétique chez les conifères étudiés (pour revue, Axtell 2013).

Au contraire de l'existence d'une conservation de certaines familles de micro ARN entre familles de plantes (3.2.1.c), il apparaît que les sites à l'origine de siRNAs hétérochromatiques, au niveau individuel, ne soient que peu ou pas conservés, même entre espèces proches. La classe de ces petits ARN non codants est ainsi marquée par l'existence d'une très grande diversité de siRNAs (Lelandais-Brière et al. 2009). De nombreuses régions génomiques impliquées dans la biogénèse de tels siRNAs correspondent à des régions riches en éléments transposables (complets ou dégénérés). Il est donc envisageable que ces régions soient soumises à une forte évolution, en lien avec la dynamique d'évolution de ces éléments transposables connue pour être rapide au niveau de l'évolution structurale du génome des plantes (en termes de position ou d'augmentation / réduction du nombre de copies dans le génome).

3.2.4. Les siRNAs secondaires

Les siRNAs secondaires regroupent les petits ARN interférents qui sont originaires de précurseurs d'ARN double brin et dont la production est initiée par un (ou plusieurs) autre petit ARN, synthétisé en amont – d'où le terme de secondaire. Généralement, le petit ARN à l'origine de ces siRNAs secondaires est soit un micro ARN, soit déjà un siRNA secondaire. Apparemment, ce système de production de siRNAs secondaires n'existerait que chez les plantes, les nématodes et les champignons, où ils seraient impliqués dans l'amplification du mécanisme de régulation par ARN interférence dans l'organisme, par la mise en place de cascades de régulation (Vazquez & Hohn, 2013). Ils agissent essentiellement **au niveau post-transcriptionnel (PTGS)**.

3.2.4.a. Biogénèse des siRNAs secondaires

Dans la majorité des cas, la biogénèse des petits ARN interférents secondaires endogènes requiert (1) le recrutement – par la protéine AGO associée au petit ARN initiateur de la production des siRNAs secondaires – de la polymérase dépendante d'ARN, RDR6, à l'origine de la synthèse du brin d'ARN complémentaire, et en présence du cofacteur SGS3 (*Suppressor of gene silencing 3*) puis (2) de la protéine Dicer DCL4 qui va couper le précurseur ARN double brin obtenu en siRNAs secondaires de 21-nt, ou éventuellement de 22-nt (Vazquez & Hohn, 2013).

De nombreux (mais pas tous) siRNAs secondaires sont phasés, du fait qu'ils proviennent alors du découpage successif, par la protéine Dicer, des précurseurs d'ARN double brin présentant une terminaison clairement définie, car issus du clivage de l'ARNm primaire par la protéine Argonaute dont l'activité a été médiée par le petit ARN non codant initiateur. On parle de *phased siRNAs*. Par ailleurs, certains de ces siRNAs secondaires peuvent agir à distance, au niveau de cibles différentes de celles dont ils sont originaires : ce sont alors des *trans-acting siRNAs*, *tasiRNAs*. La distinction faite entre ces deux populations de petits ARN non codants dans la classification tertiaire donnée en Figure 3.2. n'est donc pas aussi tranchée et encore assez floue.

Toutes les interactions 'petit ARN non codant / ARN messager cible' ne sont pas capables d'induire la production de petits ARN interférents secondaires. Celle-ci va être privilégiée dans différentes conditions (pour revue, Axtell 2013) : (1) quand les ARNm cibles présentent des extrémités 5' ou 3' aberrantes après une mauvaise maturation de l'ARNm, (2) ou s'ils ont plusieurs sites ciblés par différents petits ARN ; (3) quand la complémentarité de séquence entre le petit ARN et l'ARNm cible est très élevée (peu de mésappariement voire aucun) et associée à un fort taux d'expression de l'un des deux acteurs de l'interaction ; (4) quand la protéine Argonaute chargée du petit ARN initiateur est soit AGO2 soit AGO7 ; (5) enfin, quand la protéine AGO1 à l'origine du clivage est chargée d'un micro ARN de 22-nt.

3.2.4.b. Fonction des siRNAs secondaires

Les siRNAs secondaires endogènes correspondent à une classe de petits ARN bien conservée chez les végétaux : on les trouve aussi bien chez les angiospermes que chez des espèces végétales plus ancestrales comme la mousse *Physcomitrella patens*. De plus, des siRNAs secondaires particuliers ont été identifiés comme conservés entre différentes espèces végétales plus ou moins apparentées. Ce maintien au cours de l'évolution renforce l'idée qu'ils représentent une classe de petits ARN spécifique, à fonction biologique déterminée.

Alors que les ta-siRNAs produits à partir de loci bien identifiés, appelés les gènes *TAS*, ont été les premiers siRNAs secondaires à être caractérisés (3.2.4.c), la question se posait de savoir si des ARN messagers pouvaient conduire au même type de production de siRNAs secondaires. Par bioinformatique et biologie moléculaire, différents rôles biologiques ont ainsi pu leur être attribués, concernant différents mécanismes biologiques (pour revue, Vazquez & Hohn, 2013). Chez *Arabidopsis*, de nombreux cas de régulation des ARN codant les protéines PPR (*pentatricopeptide proteins* – des centaines de gènes codant ces protéines sont présents dans le génome des plantes et contrôlent le bon déroulement de l'expression des gènes mitochondriaux et chloroplastiques) ont pu être mis en évidence, avec les micro ARN miR161/miR400 comme initiateurs du processus (Howell et al. 2007 ; Chen et al. 2007). De même, des populations de siRNAs secondaires ont été identifiées comme impliquées dans la régulation de la voie de signalisation des auxines, avec les ARN codant les récepteurs à l'auxine de type TIR/AFB2 (TAAR) à l'origine de la production de siRNAs secondaires de 22-nt, impliquant en amont le micro ARN miR393 (Si-Ammour et al. 2011). Enfin, divers travaux ont démontré la production de siRNAs secondaires, correspondant à des pha-siRNAs, à partir des ARNm de gènes de résistance aux pathogènes et surtout de gènes de type NBS-LRR, chez *Arabidopsis* (Howell et al. 2007), mais aussi la tomate (Shivaprasad et al. 2012), le tabac (Li et al. 2012) et *Medicago* (Zhai et al. 2011), montrant la généralité du processus biologique.

3.2.4.c. Le cas des siRNAs secondaires liés aux gènes TAS

Ces ta-siRNAs sont les siRNAs secondaires endogènes les mieux caractérisés. Il s'agit du cas où un micro ARN initie leur production à partir d'un ARN *TAS*, correspondant à la transcription de loci précis dans le génome, les gènes *TAS* (*TAS genes*). La transcription de ces gènes, par la polymérase Pol II, entraîne la production de longs ARN primaires (pri-*TAS*) qui ne codent aucune protéine ; ils ne serviraient que de précurseurs ARN dans la production ciblée de ta-siRNAs. La biogénèse de ces derniers suit les principales étapes décrites précédemment (3.2.4.a), l'étape de médiation étant assurée, une fois le pri-*TAS* exporté dans le cytoplasme, par un micro ARN spécifique chargée dans AGO1 qui entraîne le clivage de l'ARN. Une fois les ta-siRNAs produits, ils agissent en réprimant leurs populations d'ARN messagers cibles uniquement par clivage, après chargement dans la protéine Argonaute AGO1 ou AGO7, selon la nature du gène *TAS* dont ils sont originaires (Vaucheret 2006).

A ce jour, quatre groupes d'ARN TAS ont été identifiés chez *Arabidopsis* : TAS1, TAS2, TAS3 et TAS4. Les ta-siRNAs originaires des TAS1 et TAS2 ont été montrés comme impliqués dans la régulation des PPR (avec miR173, donc complémentaires de ce qui a été décrit précédemment), et ceux issus de TAS4 (impliquant miR828) comme ciblant les facteurs de transcription de type MYB, et par là-même affectant en particulier la voie de biosynthèse des anthocyanes. Pour TAS3, son rôle a été largement documenté, avec la production de ta-siRNAs médiée par miR390 et qui ciblent les facteurs de réponse à l'auxine *ARF3* et *ARF4* (Allen et al. 2005), jouant un rôle essentiel dans l'organisation des organes de la plante, l'identité des méristèmes et le cycle de développement de la plante.

Une des particularités de ces ta-siRNAs est leur intervention au niveau de la régulation en cascade de réseaux complexes de gènes, par la production de nouvelles populations de siRNAs secondaires, allant jusqu'à quatre cycles de production successifs de siRNAs secondaires depuis l'initiation par un micro ARN, selon des études menées chez *Arabidopsis* (cas de miR173 et de TAS1, TAS2 – pour revue, Vazquez & Hohn, 2013).

3.2.5. Les siRNAs issus de transcrits natifs antisens (nat-siRNAs)

Contrairement aux types de petits ARN interférents (siRNAs) décrits précédemment, l'origine de l'ARN double brin donnant naissance à ces nat-siRNAs n'implique pas l'intervention d'une polymérase RDR. Le précurseur ARN double brin est ici issu de la simple hybridation de transcrits indépendants, en raison de leurs séquences ARN complémentaires (même partiellement). Ces NAT (*natural antisens transcripts*) proviennent (1) de la transcription de gènes chevauchants, s'effectuant en sens opposés – ils génèrent alors des *cis*-nat-siRNAs, ou potentiellement (2) de gènes indépendants physiquement mais présentant une région de séquence complémentaire autorisant leur appariement – produisant des *trans*-nat-siRNAs. Actuellement, seuls des cas de *cis*-nat-siRNAs ont été décrits chez les plantes, les *trans*-nat-siRNAs étant plutôt du domaine de l'hypothèse, et répondraient spécifiquement à des phases de développement ou des conditions de stress (Wang et al. 2006 ; Zhou et al. 2009). L'action de ces nat-siRNAs a pour l'instant été révélée uniquement **au niveau post-transcriptionnel (PTGS)** par clivage d'un ARN messager cible.

Relativement peu de cas de production de *cis*-nat-siRNAs ont été répertoriés et étudiés en détails ; ils concernent essentiellement le génome d'*Arabidopsis*. La première paire de gènes NAT étudiée donne naissance à un seul type de siRNAs de 24-nt, générés par DCL2 (entraînant aussi la production de siRNAs secondaires de 21-nt par l'action de DCL1), et impliqués dans la tolérance au stress salin (Borsani et al. 2005). La deuxième paire de gènes NAT caractérisée est impliquée dans la résistance à la bactérie *Pseudomonas syringae* et génère des siRNA de 22-nt obtenus par action de DCL1 (Katiyar-Agarwal et al. 2006). Enfin, un troisième cas de paire de gènes NAT apparaît aussi impliquée dans la résistance aux pathogènes, mais elle produit alors des siRNA plus longs, de 30 à 40-nt (lsiRNAs) par l'action de DCL1 et DCL4, impliquant une voie de régulation par AGO7 (Katiyar-Agarwal et al. 2007).

D'après ces différentes études, il apparaît que les *cis*-nat-siRNAs ne présentent pas de mode de biogénèse homogène (par exemple, soit DCL1, DCL2 ou DCL4 est impliquée), ni de taille précise (22-nt ou 24-nt), amenant à se poser la question de la justification de cette classe de petits ARN non codants. Par ailleurs, l'ensemble des locus à l'origine éventuelle d'ARN double brin de type NAT ne semble pas produire d'excès particulier de petits ARN non codants, au contraire : seulement 6% de ces sites chez *Arabidopsis* (sur 1439 paires identifiées) et 16% chez le riz (sur 767) ont pu être associés à la production de siRNAs à au moins 10 RPM (Zhang et al. 2012), et plus récemment ils ont été identifiés à un taux de 18% (sur 994 sites prédits) chez le soja (Zheng et al. 2013). Cependant, des efforts de recherche par prédiction des paires de gènes NAT à l'origine possible de ces nat-siRNAs ont été faits pour un ensemble de plus de 65 espèces végétales, prenant en considération l'impact fonctionnel éventuellement important de ces siRNAs en interaction avec les voies majeures de régulation par petits ARN non codants, que nous avons décrites précédemment (PlantNATsDB <http://bis.zju.edu.cn/pnatdb/> Chen et al. 2011).

3.3. Les petits ARN non codants d'origine exogène : cas du système de défense antiviral

Les mécanismes d'ARN interférence et de répression de gènes constituent par ailleurs le principal système immunitaire des plantes dirigé contre les virus. La mise en place de ce processus est initiée par la réplication du génome viral, aboutissant dans tous les cas à la production d'ARN double brin (que ce soit par repliement de l'ARN simple brin viral, ou du fait de la réplication du génome viral). Cet ARN double brin permet l'entrée dans le système de défense de la plante par recrutement des voies de régulation par petits ARN non codants (pour revue, Incarbone & Dunoyer, 2013).

Ce sont les protéines Dicer DCL4 et DCL2 (à rôle de « suppléante » de DCL4 et donc de garde-fou – Ruiz-Ferrer & Voinnet, 2009) qui interviennent dans la production des petits ARN dérivés de virus (*virus-derived small RNAs*, *vsRNAs*) de 21-nt et 22-nt respectivement, en coupant l'ARN double brin viral présent dans la cellule. Les petits ARN viraux ainsi produits sont chargés au niveau des complexes RISC, ici de type antiviral, impliquant les protéines Argonaute AGO1 (pour les 21-nt) et AGO2 (pour les 22-nt) pour permettre la **répression post-transcriptionnelle** de l'ARN viral. Lorsqu'il s'agit de virus à ADN, la protéine DCL3 est également recrutée dans le processus de défense et produit alors des *vsRNAs* de 24-nt, qui comme les siRNAs de 24-nt, conduisent à des modifications des marques épigénétiques du génome viral (méthylation des cytosines de l'ADN) par l'intermédiaire d'une protéine AGO4, agissant ainsi **au niveau transcriptionnel** de l'expression du virus par la voie de type *RdDM* (décrite en Figure 3.5.).

De plus, afin de pouvoir faire face au rythme élevé de réplication et au mouvement des virus dans la plante et arrêter l'invasion virale, des mécanismes d'amplification de la réponse antivirale ont été sélectionnés, faisant intervenir les petits ARN secondaires. Comme décrit précédemment (3.2.4.a.), ces vsRNAs secondaires requièrent l'activité de polymérase dépendantes d'ARN (surtout RDR1 et RDR6 et plus rarement RDR2) à l'origine de la production de nouveaux précurseurs dsRNA, pris en charge comme nouveaux substrats par les protéines DCL. Souvent, ces vsRNAs secondaires représentent les populations de petits ARN les plus nombreux impliqués dans la réponse antivirale. Par ailleurs, par leur capacité de mouvement entre cellules, les vsRNAs secondaires sont à l'origine de l'expansion de la réponse antivirale à l'ensemble de la plante entière.

Ce cas de mécanisme moléculaire généralisé de réponse aux attaques virales met en évidence la forte interconnexion qui existe entre les différentes voies de régulation par petits ARN (impliquant des protéines et des populations de petits ARN clairement identifiées). Notons cependant que, malgré la complexité des niveaux d'intervention de ce système immunitaire développé chez les plantes, les virus ont su contourner ces mécanismes de régulation en produisant des suppresseurs viraux du mécanisme de RNA silencing (*viral suppressors of RNA silencing*, VSRs – pour revue, Ruiz-Ferrer & Voinnet, 2009). Encore une illustration de l'inépuisable course aux armements entre hôtes et pathogènes...

3.4. Impact de l'hybridation sur l'expression des petits ARN non codants

Diverses études ont pu démontrer une mobilisation des populations de petits ARN non codants lors de croisements entre plantes, que ce soit en contexte intraspécifique ou interspécifique. Je vais m'attacher à présenter les principales connaissances sur ce sujet issues de travaux récents menés chez *Arabidopsis* mais aussi quelques espèces cultivées.

3.4.1. Mobilisation de populations de petits ARN lors de la formation des gamètes puis au moment de la reproduction

Initialement, les connaissances sur l'implication des petits ARN dans la régulation du génome et de l'expression des gènes ont été obtenues à partir de l'étude de la plante, en tant que sporophyte diploïde. Ce n'est que plus tard que l'on s'est intéressé aux gamétophytes, mâle et femelle, et que l'on a pu mettre en évidence un rôle des petits ARN dans la conservation de l'intégrité du génome et le développement du gamète, en préparation de la phase de reproduction et de fécondation qui se poursuit par l'étape majeure (et primordiale en agronomie pour ce qui est des espèces cultivées) du développement de la graine contenant le sporophyte diploïde de la nouvelle génération.

Chez les plantes, les cellules germinales se différencient tardivement au cours du développement du sporophyte (et non durant l'embryogénèse, comme chez les animaux), à partir de cellules souches somatiques qui entrent progressivement en division méiotique. Et, comme nous l'avons vu précédemment (3.2.3.), l'ensemble des cellules de la plante présente des marques épigénétiques d'hétérochromatinisation de l'ADN, en rapport avec leur contenu en éléments transposables largement réprimé par les mécanismes de la voie de méthylation de l'ADN médiée par ARN, *RdDM*, associée à l'activité d'ADN-méthyltransférases. Or, il apparaît souvent nécessaire d'avoir une élimination d'un ensemble de marques épigénétiques (si ce n'est de toutes) au niveau des cellules germinales – on parle de reprogrammation du génome – lors de la formation des gamètes. S'intéressant à ce processus, il a ainsi été observé des changements importants de ces marques épigénétiques dans le grain de pollen tout comme dans le sac embryonnaire (pour revues, Le Trionnaire & Twell, 2010 ; Calarco & Martienssen, 2011 ; Castel & Martienssen, 2013).

3.4.1.a. Les petits ARN du gamétophyte mâle (le grain de pollen)

En ce qui concerne la méthylation de l'ADN, chez la plante modèle *Arabidopsis thaliana*, contrairement aux mammifères pour lesquels on observe une déméthylation globale du génome au cours de la spermatogénèse (Popp et al. 2010), les cellules de la lignée germinale mâle conservent une grande partie de la méthylation symétrique, à des niveaux similaires à ceux des cellules somatiques (Calarco et al. 2012; Ibarra et al. 2012), mais montrent une baisse des niveaux de méthylation en des sites asymétriques pour les éléments transposables et notamment les rétrotransposons (qui se retrouvent par la suite méthylés chez l'embryon en développement ; Jullien et al. 2012).

Spécifiquement au niveau de l'ADN du noyau végétatif du grain de pollen, on observe une très forte décondensation de la chromatine, associée à une absence de la protéine DDM1, *Decreased DNA Methylation* (nom donné par rapport au phénotype du mutant correspondant *ddm1*), qui est impliquée dans la conformation de la chromatine par contrôle de la méthylation de l'ADN (Slotkin et al. 2009). Cela entraîne la réactivation massive de certaines familles d'éléments transposables (ce sont essentiellement des rétrotransposons qui ont été identifiés, notamment *Athila*) dont le contrôle transcriptionnel (TGS) ne peut être assuré du fait de l'absence de siRNAs de 24-nt spécifiques à ces ET, dans le noyau végétatif. Les nouveaux transcrits d'ET ainsi synthétisés sont alors pris en charge dans les voies de biogénèse des petits ARN et conduisent à la production d'une nouvelle classe spécifique de petits ARN de 21-nt, antisense par rapport aux ET transcrits ; ces nouveaux siRNAs sont alors transportés jusqu'aux cellules spermatiques, où ils peuvent y assurer le contrôle transcriptionnel de l'expression des ET correspondants (Slotkin et al. 2009, Figure 3.6.A).

3.4.1.b. Les petits ARN du gaméophyte femelle (le sac embryonnaire)

Le gaméophyte femelle a été soumis à moins d'études que le gaméophyte mâle, du fait de son accès plus difficile en lien avec sa position renforcée dans les tissus sporophytiques. Il a ainsi été mis en évidence que, comme dans le noyau végétatif du grain de pollen, les deux noyaux de la cellule centrale présentaient un ADN sous une forme d'hétérochromatine différente de celle des cellules somatiques, présentant notamment de très faibles niveaux de méthylation. Cette hypométhylation, liée à la surexpression de la protéine DME1 (DEMETER), une ADN glycosylase qui enlève les groupes méthyl des cytosine (Baroux et al. 2007,) et à la sous-expression de l'ADN méthyltransférase MET1 (ou DMT1) de maintien de la méthylation de l'ADN (Jullien et al. 2008), entraîne alors une forte activité transcriptionnelle. Par contre, au niveau de l'oosphère, l'ADN est sous forme chromatinienne condensée, avec des marques épigénétiques répressives importantes correspondant notamment à un enrichissement en méthylation des sites de la lysine de l'histone H3, H3K9me2 (Baroux et al. 2011). Cette particularité amène l'ADN de l'oosphère a présenté un très faible niveau de transcription.

Une conséquence directe de l'hypométhylation de la cellule centrale est la transcription d'éléments transposables qui ne sont alors plus soumis à aucun contrôle de répression d'ordre transcriptionnel (Figure 3.6.B). Pris en charge dans une voie de biogénèse des petits ARN, ils aboutissent à la production de petits ARN de 24-nt (et non de 21-nt comme dans le noyau végétatif du grain de pollen) sous contrôle de la polymérase Pol IV – on parle d'ailleurs de populations de *p4-siRNAs*. Ces petits ARN de 24-nt conduisent à l'hyperméthylation de sites de l'ADN autres que CG, correspondant surtout à des rétroéléments (Mosher et al. 2009 ; Hsieh et al. 2009). Par la suite, ces petits ARN de 24-nt migreraient vers l'oosphère où, pris en charge par la protéine AGO9, ils assureraient un renforcement du contrôle des ET dans la cellule germinale du gamète femelle.

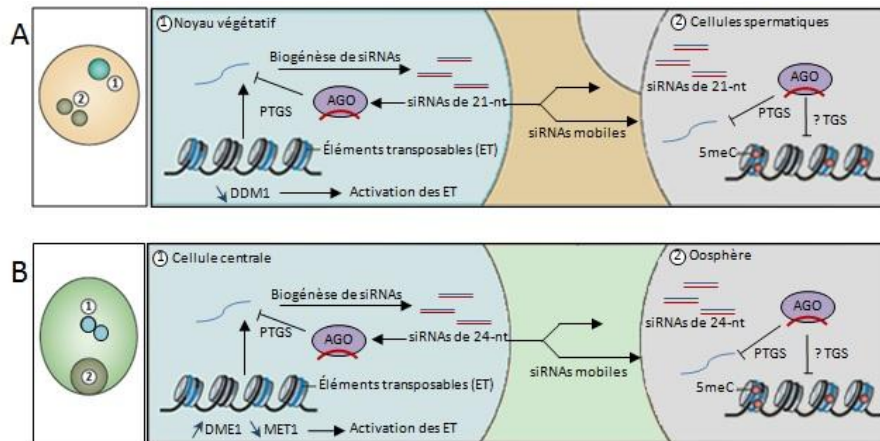


Figure 3.6. Répression des éléments transposables induite par ARN interférence dans les cellules germinales – modèle développé à partir des connaissances acquises chez *A. thaliana*. **A.** Dans ① le noyau végétatif du gamétophyte mâle, la protéine DDM1 n'est pas synthétisée, entraînant la possibilité aux éléments transposables (ET) d'être exprimés. Les transcrits des ET sont alors processés en petits ARN interférents de 21-nt (siRNAs) qui, par leur mobilité, peuvent entraîner un contrôle du type post-transcriptionnel (PTGS) de l'expression des ET correspondants dans ② les cellules spermatiques. **B.** Dans ① la cellule centrale binucléée du gamétophyte femelle, la protéine MET1 est réprimée, alors que l'ADN glycosylase DEMETER (DME1) est exprimée, entraînant une perte de la méthylation des cytosines et permettant la transcription des ET. Ceci active la voie de méthylation de l'ADN guidée par ARN (*RdDM*) impliquant Pol IV et produit des siRNAs de 24-nt qui peuvent être acheminés jusqu'à ② l'oosphère pour renforcer la répression des ET, par l'action de AGO9. Figure traduite de Castel & Martienssen, 2013.

3.4.1.c. Les petits ARN de la fécondation puis du développement de la graine

D'après ce que nous venons de décrire sur les gamétophytes mâle et femelle, il s'avère qu'au cours de la reproduction, alors que l'oosphère hérite de l'information génétique portée par les parents mâle et femelle à l'issue de la rencontre des gamètes pour donner le zygote, elle hérite aussi de populations de petits ARN produites par les cellules végétatives des gamétophytes puis transportées dans les cellules germinales, et qui se trouvent être impliquées spécifiquement dans le contrôle des éléments transposables (Mirouze 2012). Une hypothèse est donc la sélection d'un mécanisme moléculaire permettant aux éléments transposables de s'exprimer dans les cellules végétatives des gamétophytes, afin qu'ils soient utilisés comme matrices dans la production de petits ARN à rôle majeur dans le renforcement de la répression de ces mêmes ET dans les cellules germinales (Martienssen 2010) mais aussi au cours du développement ultérieur de la graine (Mosher et al. 2009 ; Slotkin et al. 2009). Ce mouvement de petits ARN entre cellules des gamétophytes présente des implications importantes quant à la transmission de l'information épigénétique à la génération suivante.

Ce mécanisme d'héritage de populations de petits ARN, les 21-nt à action PTGS provenant du parent mâle et les 24-nt à action TGS du parent femelle, ne permettrait pas uniquement la simple répression des ET. Ils interviendraient en synergie lors de la fécondation puis au cours de la formation du zygote, apportant un mécanisme efficace de contrôle des ET, notamment vis-à-vis de ceux provenant de 'l'autre' génome.

3.4.2. Les petits ARN non codants apparaissent impliqués dans le phénomène d'hétérosis

L'évolution par hybridation entre espèces est très fréquente chez les eucaryotes, mais c'est chez les espèces végétales qu'elle est la plus répandue (notamment en lien avec les processus de polypléidie ; on parle alors d'allopolypléidie, cf. Chapitre 1) – chez les animaux, l'hybridation interspécifique conduit dans de nombreux cas à la stérilité des individus hybrides ou du moins à une baisse importante de leur fertilité. Le succès de l'hybridation interspécifique est lié à ce que celle-ci génère des niveaux accrus d'hétérozygotie, qui sont responsables du phénomène d'hétérosis, encore appelé vigueur hybride, phénomène décrit et étudié à l'origine chez des hybrides intraspécifiques.

L'hétérosis fait communément référence à des niveaux supérieurs dans la production de biomasse, la taille, le taux de croissance et la fertilité (par exemple, au niveau de la production en graines), parfois aussi la valeur sélective, dans une descendance d'individus hybrides par rapport aux parents. Cette vigueur hybride a été très étudiée en sélection variétale, et correspond généralement au résultat d'un croisement réalisé au sein d'une même espèce, comme dans le cas des variétés hybrides F1 de maïs obtenues par croisement entre lignées homozygotes (Birchler et al. 2003 ; Figure 3.7.). La base de ce phénomène réside en la complémentarité et/ou l'interaction entre allèles différents, s'appuyant donc sur un niveau d'hétérozygotie élevé, conduisant à ce que les espèces présentant une forte vigueur hybride soient plutôt allogames.

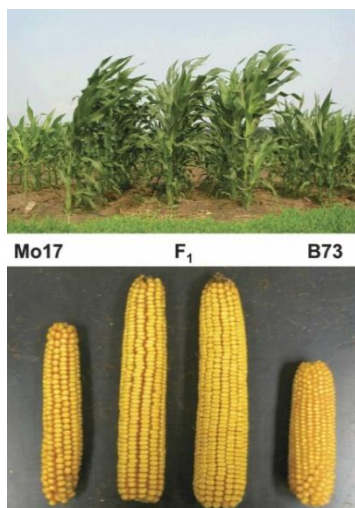


Figure 3.7. Illustration de la vigueur hybride, appelée hétérosis, chez le maïs. On remarque facilement la taille beaucoup plus grande des hybrides F1 au niveau du développement de la plante (les 3 lignes centrales plantées au champ) et des épis de maïs (les 2 au centre de la photo) en comparaison des lignées parentales (Mo17 à gauche et B73 à droite). C'est ainsi que le maïs est cultivé quasi-exclusivement sous forme de variétés hybrides simples (un croisement de deux lignées) depuis les années 1960s, en raison des forts rendements associés. D'après Springer & Stupar, 2007.

L'étape proprement dite d'hybridation interspécifique, qui correspond à la rencontre de génomes différenciés dans une même cellule, entraîne ainsi des interactions génomiques collectivement connues sous le terme de « choc génomique » (McClintock, 1984). Ces interactions peuvent alors être bénéfiques (vigueur hybride ; nouvelle diversité génétique) ou parfois désavantageuses (instabilité génomique ; réactivation d'éléments transposables, au moins au niveau transcriptionnel).

Diverses études ont documenté des changements génétiques et épigénétiques rapides, survenant en réponse à ce choc génomique ; récemment, des travaux menés sur différentes espèces végétales ont suggéré un rôle important des petits ARN non codants en réponse au choc génomique, par leur implication dans la stabilité du génome hybride nouvellement créé et dans les changements au niveau de la régulation de l'expression des gènes qui ont pu être observés (correspondant à des écarts par rapport à l'attendu de la moyenne des profils parentaux). Ces observations ont ainsi abouti à l'hypothèse que les petits ARN étaient directement impliqués dans la viabilité des hybrides et l'expression de l'hétérosis.

3.4.2.a. Une mobilisation de micro ARN et de leurs gènes cibles en réponse à l'hybridation

La biogénèse des micro ARN peut être modulée au niveau post-transcriptionnel (après transcription du gène *MIR*) aux étapes de clivage par DCL1 des miRNA primaires (pri-miRNA) ou des précurseurs de miRNA (pre-miARN) (Cuperus et al. 2011 ; Jones - Rhoades et al. 2006). De ce fait, des divergences de séquence au niveau des précurseurs de miRNA et/ou la régulation différentielle des facteurs de production des miRNA (comme les gènes de leur biogénèse, par exemple) pourraient conduire à un gain de nouveaux profils d'expression comme à une baisse de l'expression (voire une perte) de certains micro ARN, entre génotypes au sein de la même espèce ou entre espèces différentes. L'expression différentielle de certaines miRNAs à l'origine de profils d'expression de gènes spécifiques de l'hybride par rapport à ses parents a ainsi été démontrée chez différentes espèces comme *Arabidopsis* (Shen et al. 2012), le riz (He et al. 2010), la canne à sucre (Zanca et al. 2010), ou le blé (mais là au niveau de l'allopolyploïde uniquement, Kenan-Eichler et al. 2011).

En particulier, un travail original a été mené sur des lignées d'introgession de tomate issues d'un croisement entre la tomate cultivée, *Solanum lycopersicum*, et une espèce sauvage apparentée, *Solanum pennellii* (Eshed & Zamir, 1995). Les auteurs démontrent que les caractères transgressifs présentés par certaines lignées d'introgession en comparaison des parents sont corrélées à des niveaux de sur-expression de certains petits ARN (Shivaprasad et al. 2012). Par exemple, il apparaît que le micro ARN miR395, pourtant additif au niveau de l'hybride F1, est surexprimé dans la génération F2 mais aussi dans une lignée d'introgession présentant une plus grande tolérance au stress salin. La gestion de ces populations de petits ARN et notamment des micro ARN pour l'expression transgressive de caractères d'intérêt agronomique semble donc une nouvelle piste pour l'amélioration variétale basée sur l'exploitation des hybrides.

3.4.2.b. Les populations de siRNAs de 24-nt présentent des changements en séquence et en abondance après hybridation

Le choc génomique entraîné par l'événement d'hybridation peut être à l'origine d'une réactivation, du moins au niveau de leur transcription, des éléments transposables (McClintock, 1984 ; pour revue Madlung & Comai, 2004). En accord avec cette hypothèse, il s'avère que l'on observe généralement une baisse des niveaux d'expression des populations de petits ARN de 24-nt, originaires de loci riches en éléments transposables ou autres éléments répétés, immédiatement après l'événement d'hybridation (chez *Arabidopsis*, Groszmann et al. 2011 ; le riz, He et al. 2010 ; le maïs, Barber et al. 2012). On peut remarquer que cet impact de l'hybridation sur les petits ARN de 24-nt apparaît toujours beaucoup plus important que celui observé sur les micro ARN (très clairement illustré pour des hybrides intra-spécifiques d'*Arabidopsis thaliana*, Groszmann et al. 2011). Par ailleurs, plus les petits ARN de 24-nt présentent des différences d'expression entre les deux parents, plus ils semblent soumis à une réduction de leur accumulation dans l'hybride correspondant. Enfin, ces siRNAs de 24-nt affectés par l'événement d'hybridation sont souvent associés à des éléments transposables localisés près de gènes. Des changements dans le contrôle transcriptionnel de ces ET permettent donc d'envisager une répercussion directe sur l'activité transcriptionnelle des gènes associés, amenant l'hybride à présenter des profils d'expression différents de ceux observés chez les parents, et potentiellement responsables de l'hétérosis.

Néanmoins, il faut garder à l'esprit que bon nombre d'hybridations interspécifiques ne sont pas viables, avec parfois un avortement précoce des graines d'origine hybride. Par ailleurs, bien que le phénomène d'hétérosis soit généralement observé chez les hybrides qui ont surmonté les barrières de l'hybridation, certains hybrides s'avèrent par la suite stériles, un phénomène connu sous le nom de létalité hybride (Dobzhansky 1936). Là aussi, il a été récemment suggéré que les petits ARN non codants pouvaient agir comme une barrière à l'hybridation entre espèces trop éloignées génétiquement (Martienssen 2010).

3.4.3. Les petits ARN non codants à l'origine de l'incompatibilité hybride

Le succès des hybrides interspécifiques dépend en grande partie du bon développement de la graine. Une graine contient un embryon diploïde et un albumen triploïde en raison de la double fécondation chez les angiospermes (Haig & Westoby, 1989 ; 3.4.1.c.). Chez les plantes, la létalité post-zygotique ou l'avortement de la graine est souvent accompagnée de l'arrêt du développement de l'embryon et d'un retardement de la cellularisation de l'albumen dans des croisements interploïdes (Scott et al. 1998), ainsi que dans des graines d'hybrides interspécifiques (Bushell et al. 2003). Cette incompatibilité hybride pourrait être attribuée à la réactivation d'éléments transposables et à la perturbation du mécanisme d'empreinte et de reprogrammation génomique. Par exemple, dans des croisements interspécifiques entre *A. thaliana* diploïde et *A. arenosa* tétraploïde, le rétrotransposon *Athila*, présent dans l'hétérochromatine et normalement réprimé, se retrouve exprimé à partir de l'allèle paternel, entraînant un arrêt du développement de la graine (Josefsson et al. 2006). Or chez *A. thaliana*, au niveau du gamétophyte femelle (comme nous l'avons décrit précédemment, 3.4.1.b.), des

siRNAs de 24-nt passent de l'albumen (cellule centrale) vers l'embryon (oosphère), renforçant le contrôle des éléments transposables dans l'embryon. La réactivation de l'élément *Athila* dans les graines hybrides serait donc le résultat de la divergence de séquence de l'élément *Athila* entre les deux espèces parentales ou d'une production insuffisante des *p4-siRNAs* maternels, conduisant à un contrôle inefficace de ce rétrotransposon et à une instabilité génomique à l'origine de l'avortement de la graine (Martienssen 2010 ; Figure 3.8.).

A cela s'ajoute le fait que ces *p4-siRNAs* maternels ont récemment été démontrés comme impliqués dans le maintien du dosage approprié des génomes parentaux et la régulation de l'expression des gènes soumis à empreinte (Lu et al. 2012), le développement normal de la graine nécessitant un déséquilibre des génomes parentaux au niveau de l'albumen triploïde avec un ratio de 2m:1p – 2 doses du génome maternel : 1 dose du génome paternel (Haig & Westoby, 1989 ; Scott et al. 1998).

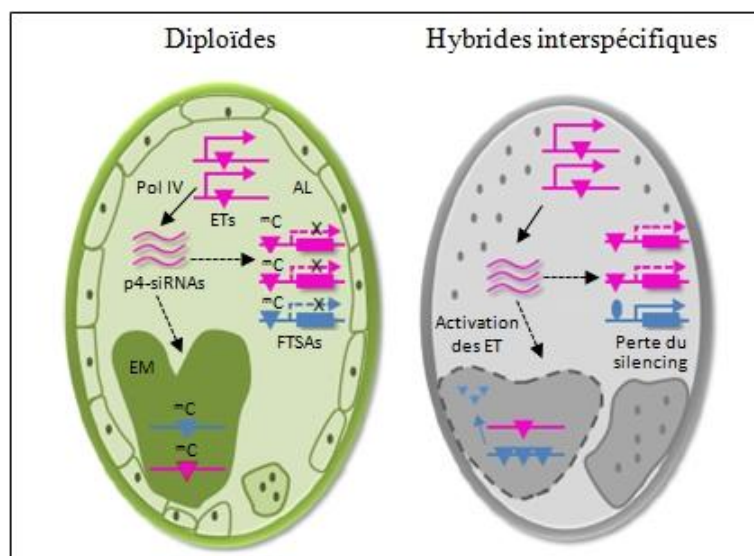


Figure 3.8. Représentation schématique d'un mécanisme potentiellement impliqué dans l'incompatibilité hybride. La production d'une classe particulière de *siRNAs* de 24-nt dépend de l'ARN polymérase IV Pol IV (ce sont des *siRNAs* hétérochromatiques (3.2.3.), mais on peut aussi parler de *p4-siRNAs*) et de loci maternels d'éléments transposables (ET) (triangles roses) dans l'albumen (AL). Ces *siRNAs* maternels sont prédits comme étant mobilisés (ligne pointillée) dans l'embryon (EM) pour réprimer les ET correspondants par méthylation des cytosine (^mC) et assurer ainsi le maintien de l'intégrité du génome (à gauche, chez les diploïdes). Des facteurs de transcription spécifiques de l'albumen (FTSAs) associés à ces ETs peuvent aussi être des cibles potentielles des *p4-siRNAs*, et leur expression ou *silencing* (marqué ×) pourrait être critique pour le développement de l'albumen. Chez des hybrides interspécifiques à avortement précoce (à droite), la répression des ET et des FTSAs peut être compromise par la variation du nombre de copies et/ou la divergence de séquence des loci d'ET du génome paternel (ovales bleus). Le dosage et la divergence des *p4-siRNAs* maternels conduisent alors à la réactivation des ET dans l'embryon et à l'arrêt du développement de l'albumen. Figure traduite de Ng et al. 2012.

En conclusion, les populations de petits ARN non codants présenteraient des implications différentes dans le devenir de l'individu hybride nouvellement formé, selon leur origine : les populations de petits ARN de 24-nt d'origine maternelle interviendraient plutôt dans le développement de graines viables, par maintien de l'intégrité du génome hybride, alors que la reprogrammation de certains micro ARN dans les hybrides contribueraient à des phénotypes nouveaux, spécifiques de l'hybride et traduisant le phénomène d'hétérosis.

Cette reprogrammation de petits ARN non codants chez les hybrides pourrait correspondre à un mécanisme majeur pour l'acquisition de nouveaux profils d'expression des gènes, rendant possible une plus grande capacité à répondre aux stimuli environnementaux mais aussi aux différents signaux du développement, assurant ainsi une plus grande adaptabilité des génomes hybrides.

3.5. Impact de l'allopolyploïdie sur l'expression des petits ARN non codants

Chez les allopolyploïdes, l'hybride issu d'un croisement interspécifique, possède l'ensemble de ses chromosomes à l'état dupliqué. Une question qui se pose généralement est de comprendre ce qu'entraîne cette duplication, par rapport à l'hybridation initiale. Il s'agit d'évaluer si, en plus de fixer l'état hétérozygote et de restaurer un niveau satisfaisant de fertilité par reformation des paires de chromosomes homologues, la polyploïdie n'entraînerait pas d'innovations génomiques particulières. Cette question n'a encore été abordée que rarement au niveau de la mobilisation des petits ARN non codants.

3.5.1. Des changements d'expression de petits ARN non codants chez les systèmes allopolyploïdes *Arabidopsis* et blé

L'équipe de Z. Jeffrey Chen a séquencé des banques de petits ARN non codants obtenus à partir de feuilles et de fleurs d'un allotétraploïde synthétique d'*Arabidopsis* pris aux générations F1 et F7, et de leurs espèces parentales *A. thaliana* et *A. arenosa* (Figure 3.9.), ainsi que de l'espèce allotétraploïde naturelle correspondante *A. suecica* (Ha et al. 2009). Notons ici que la génération F1 correspond déjà à un allotétraploïde, le croisement initial étant réalisé entre génotypes autotétraploïdes. Leur analyse a porté sur un ensemble de 818 000 séquences de 20 à 25 nucléotides de long obtenues pour les 10 banques réalisées, ce qui représente aujourd'hui une faible profondeur de séquençage. Ils ont analysé les différentes classes de petits ARN non codants en comparant les populations de petits ARN chez les allotétraploïdes avec celles des parents autotétraploïdes.



Figure 3.9. Allotétraploïdes synthétiques d'*Arabidopsis*. L'allotétraploïde a été obtenu par croisement entre *Arabidopsis thaliana* Columbia (doublé, 4x) × *A. arenosa* (4x) C24. Figure modifiée à partir de Chen, 2010.

Les siRNAs de 24-nt hétérochromatiques (ou ici, rasiRNAs pour *repeat-associated siRNAs*) sont principalement issus de transposons et de séquences répétées qui sont très présents dans les centromères et dans les régions hétérochromatiques (Lippman & Martienssen, 2004). Ils divergent rapidement, même au niveau d'espèces étroitement apparentées (Chen 2010), avec ici moins de 20% des rasiRNAs qui ont été trouvées communes aux deux parents des allopolyploïdes *A. thaliana* et *A. arenosa*. La population de rasiRNA est apparue relativement faible dans la génération F1, avec de nombreux siRNAs de 24-nt correspondant à des éléments transposables qui ont été perdus à cette génération (environ 5.5%), permettant potentiellement la réactivation des éléments transposables. Les auteurs ont interprété cela comme étant une conséquence du « choc génomique » provoqué lors de la formation des hybrides. Au fil des générations, **la stabilité du génome allopolyploïde se rétablit, grâce à la régénération de rasiRNAs dans les allotétraploïdes de 7^{ème} génération** : on retrouve dans la génération F7, et de façon similaire chez l'allotétraploïde naturel *A. suecica*, les siRNAs de 24-nt qui étaient absents à la génération F1 (Ha et al. 2009). Cela indique qu'il faut plusieurs générations pour rétablir des profils d'expression de siRNAs de 24-nt impliqués dans l'intégrité du génome (Jackson & Chen, 2010). Néanmoins, il semblerait que ces changements de populations de siRNAs n'aient pas de corrélation avec les changements d'expression des gènes.

Par contre, alors que les populations de micro ARN (et ta-siRNAs) apparaissent très conservées en séquence au niveau des parents autotétraploïdes, elles montrent majoritairement des niveaux d'expression non-additifs entre la valeur moyenne parentale attendue et ce qui est observé chez l'allotétraploïde en génération F7 et *A. suecica*. Ces niveaux d'expression ayant ensuite été démontrés comme corrélés négativement aux niveaux d'expression des gènes cibles potentiels, il semblerait que ce soit **ces micro ARN qui entraînent une régulation différentielle de l'expression des gènes chez l'allopolyploïde**, mécanismes potentiellement impliqués dans le phénomène d'hétérosis et la meilleure capacité d'adaptation du génome allopolyploïde, une fois stabilisé.

Une autre étude a été menée chez l'espèce allohexaploïde blé, en comparant les populations de petits ARN non codants au sein de l'espèce diploïde *Aegilops tauschii* (génome DD), l'espèce tétraploïde *Triticum turgidum* ssp. *durum* (AABB), leur hybride synthétique triploïde (ABD), et l'hexaploïde dérivé (AABBDD) (Figure 3.10.), similaire en structure à *T. aestivum* ssp. *aestivum* (Kenan-Eichler et al. 2011). Un total de 6 millions de séquences comprises entre 18 et 35 nucléotides de long ont été analysées (incluant aussi les séquences des ARNr et ARNt) pour les 4 banques qui ont été construites, avec une banque pour chaque structure génomique. La plupart des **miRNAs étudiés (21 au total) se sont révélés être exprimés au niveau de la valeur moyenne parentale attendue, chez l'hybride triploïde et l'allohexaploïde**, même si quelques cas de sur-expression ont pu être observés au niveau de l'allohexaploïde.

Par contre, les niveaux d'expression des **siRNAs de 24-nt correspondant à des éléments transposables ont présenté une baisse significative chez l'allopolyploïde, mais pas l'hybride**. Les travaux menés chez le blé ont ainsi mis en évidence un impact différent de l'hybridation et de l'allopolyploïdisation sur l'expression des petits ARN, la baisse significative des populations de siRNAs de 24-nt apparaissant similaire à ce qui a été observé chez des allopolyploïdes d'*Arabidopsis*, en lien avec un réactivation des éléments transposables (démonstré chez le blé pour le rétrotransposon *Veju*). Ces résultats mettent en évidence une étape de déstabilisation du génome néo-allopolyploïde dans les premières phases de spéciation par allopolyploïdie.

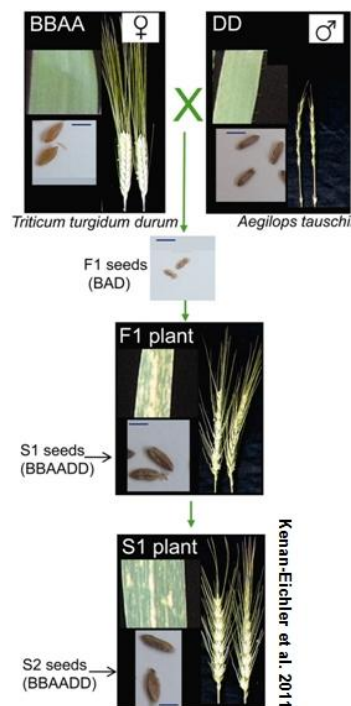


Figure 3.10. Description du matériel végétal utilisé pour obtenir des allopolyploïdes de blé synthétiques. *Triticum turgidum durum* (parent femelle, génome BBAA) et *Aegilops tauschii* (parent mâle, génome DD), ont été croisés pour obtenir des plantes hybrides F1 (génome BAD). Des graines allopolyploïdes S1 ont été obtenues ; par autofécondation elles ont donné des graines S2, deuxième génération d'allopolyploïde synthétique de blé. D'après Kenan-Eichler et al. 2011.

Dans le cadre de mon projet de thèse portant sur l'étude de la mise en place du génome allotétraploïde du colza, nous avons entrepris d'étudier l'impact global de l'allopolyploïdie sur les petits ARN non codants, dans un contexte de dynamique du génome allopolyploïde nouvellement formé. Il s'agissait de voir (1) si ces petits ARN réagissaient spécifiquement à l'allopolyploïdisation, et à quel moment, présentant alors des niveaux d'expression différents selon les générations étudiées, ou (2) si au contraire ils n'étaient pas affectés par l'événement de création du génome allopolyploïde avec une expression stable, 'additive', par rapport à ce que l'on pouvait observer chez les parents diploïdes.

3.5.2. Martinez Palacios et al. en préparation pour soumission dans *Molecular Biology and Evolution*

Highlighting the immediate but transient response of specific small RNA populations to allopolyploidy using resynthesized *Brassica napus* allotetraploids

Paulina MARTINEZ PALACIOS, Marie-Pierre JACQUEMOT, Marion TAPIE, Agnès ROUSSELET, Matthieu FALQUE, Andrew LLOYD, Eric JENCZEWSKI, Gilles LASSALLE, Anne-Marie CHÈVRE, Christine LELANDAIS, Martin CRESPI, Philippe BRABANT, Johann JOETS and Karine ALIX

Contexte et objectifs : L'hybridation interspécifique s'avère être un processus majeur d'évolution des espèces, où son succès et son omniprésence chez les plantes apparaissent directement liés au phénomène de polyploïdie. Cette duplication complète du génome, quand elle est associée à un événement de croisement interspécifique – on parle alors d'allopolyploïdie – a notamment permis aux végétaux d'assurer un bon niveau de fertilité (par stabilité méiotique) en association avec le phénomène bien connu d'hétérosis, alors fixé chez ces néo-allopolyploïdes, à l'origine de nouvelles capacités d'adaptation et notamment du succès évolutif de ces nouveaux génotypes hybrides.

Les études menées ces dix dernières années sur différents systèmes allopolyploïdes ont permis de mettre en évidence la mise en place de nombreuses régulations tant au niveau de l'expression que de l'architecture structurale du génome néo-allopolyploïde, de suite après l'événement d'allopolyploïdisation. En particulier, on a pu observer des modifications importantes au niveau des marques épigénétiques du génome. Alors que l'hétérosis et les phénomènes de vigueur hybride associés sembleraient trouver des explications mécanistiques dans les voies de régulation par petits ARN non codants, ces derniers apparaissent être de bons candidats en tant qu'acteurs moléculaires de la régulation globale du génome néo-allopolyploïde.

En prenant comme modèle le colza (*Brassica napus*, AACC), espèce allotétraploïde issue de l'hybridation entre *B. rapa* (AA, 2n=20) et *B. oleracea* (CC, 2n=18), nous avons évalué l'impact global de l'allopolyploïdisation sur les petits ARN non codants, par étude de colzas néo-synthétisés.

Méthodes : A partir d'extraits d'ARN totaux de tige, nous avons procédé à la construction et au séquençage par technique Illumina de onze banques de petits ARN pour trois colzas néo-synthétisés indépendants, pris à deux générations d'autofécondation différentes (S1 comme génération très précoce, et S5 comme génération plus avancée) et pour leurs progéniteurs diploïdes. Après curation des résultats de séquençage de ces banques et élimination des séquences identiques aux ARN ribosomiques et de transfert, c'est un total de 14 917 073 séquences de 18 à 26 nucléotides (nt) qui ont été analysées.

Par bioinformatique, nous avons mené l'étude comparative de la distribution des séquences obtenues, au sein des différentes générations (parents, S1, S5). La caractérisation de populations de petits ARN montrant des réponses spécifiques à l'événement

d'allopolyplœdie a été réalisée par une annotation bioinformatique détaillée, et la validation de leur dynamique d'expression s'est faite par hybridation de type *Northern* et/ou RT-PCR quantitative, en incluant alors les générations de création des hybrides interspécifiques F1 et des colzas néo-synthétisés S0 (provenant du doublement somatique des F1) et S3. Nous avons pu identifier l'origine de certains petits ARN d'intérêt par exploitation de données de transcriptome haut débit (*RNAseq*) obtenues pour des méiocytes de variétés de colza prises à des niveaux de ploïdie différents.

Principaux résultats: L'analyse de la distribution des populations de petits ARN séquencés pour les deux générations des trois colzas néo-synthétisés étudiés en comparaison de leurs parents diploïdes, a mis en évidence une hausse significative des petits ARN de 21-nt s'accompagnant d'une baisse des populations de 24-nt, de suite après allopolyplœdisation, en génération S1, par rapport aux générations parentales et S5.

Alors que les micro ARN ne semblaient pas être particulièrement affectés par l'événement d'allopolyplœdie, les mêmes familles de micro ARN se retrouvant dans les trois générations étudiées et à des niveaux d'expression similaires, les petits ARN de type siRNAs spécifiques des éléments transposables se sont révélés particulièrement touchés. Ils sont apparus responsables de la baisse globale de la classe des 24-nt en génération S1, cependant le détail de l'annotation des ces petits ARN a montré des réponses différentes selon la famille d'ET qu'ils ciblaient. Ainsi, les petits ARN de 24-nt ciblant les rétrotransposons à LTR étaient significativement plus nombreux chez les colzas néo-synthétisés, et dans les deux générations S1 et S5, que chez les parents diploïdes.

Nous avons également identifié des populations spécifiques de petits ARN de 21-nt responsables de la hausse de la classe des 21-nt en génération S1, et démontré qu'elles étaient surexprimées dès la création de l'hybride F1 et en génération S0, avec des niveaux d'expression similaires à ceux des parents au moins à partir de la génération S3. De façon intéressante, nous avons montré que des populations de petits ARN de 24-nt identiques en séquence à ces petits ARN de 21-nt étaient préférentiellement exprimées dans la génération plus avancée S5 des colzas néo-synthétisés. Pour certains de ces petits ARN de 21-nt, nous avons caractérisé une origine de type viral, du fait de leur similarité de séquence avec des transcrits de virus à ARN, notamment présents de manière spécifique dans le transcriptome de méiocytes de colzas haploïdes.

Conclusion : Notre étude a mis en évidence une réponse immédiate mais transitoire des petits ARN non codants à l'événement d'allopolyplœdie du colza.

Nous avons pu démontrer que les micro ARN n'étaient que globalement peu affectés par l'allopolyplœdie. Quant aux siRNAs de 24-nt ciblant spécifiquement les éléments transposables, ils ont montré des changements d'expression spécifiques des familles d'ET qu'ils pouvaient cibler : ces résultats sont en adéquation avec les changements des niveaux de transcrits de ces ET, que nous avons démontrés, dans le cadre d'une autre étude, comme étant également spécifiques de chaque famille. Notre étude confirme ainsi la réponse complexe et non prévisible des éléments transposables à l'allopolyplœdie. Il semblerait d'ailleurs que les rétrotransposons à LTR soient particulièrement la cible d'une régulation par petits ARN, suggérant un contrôle efficace de leur transposition.

La mise en évidence de populations de petits ARN de 21-nt surexprimés en génération S1, et de populations de petits ARN de 24-nt parfaitement similaires en séquence, suggère la mise en place de voies complémentaires de régulation par petits ARN. Elles consisteraient tout d'abord en l'assurance d'un contrôle précoce, 'd'urgence', au niveau post-transcriptionnel (PTGS) suivi progressivement d'une régulation plus stable, sur le long terme, qui s'établirait au niveau transcriptionnel (TGS). L'identification parmi ces petits ARN de 21-nt, de petits ARN spécifiques de séquences de virus à ARN, dont les transcrits sont exprimés dès la création des hybrides F1, nous amène à concevoir l'idée d'une réactivation d'éléments du génome de type endovirus en réponse au choc génomique engendré par l'hybridation interspécifique de l'événement d'allopolyploïdisation. Ces résultats originaux nous amènent à considérer ces endovirus comme des acteurs moléculaires de réponse à l'hybridation à l'origine du phénomène de nécrose hybride, et dont le contrôle immédiat et actif par petits ARN de 21-nt et de 24-nt serait nécessaire au succès de l'hybride et à la viabilité du néo-allopolyploïde. Notre étude démontre l'importance des petits ARN non codants comme gardiens de l'intégrité du génome suite au choc génomique dont est responsable un événement d'allopolyploïdie.

Mots clés

allopolyploïdie, *Brassica*, éléments transposables, endovirus, hybridation interspécifique, micro ARN (miRNAs), petits ARN interférents (siRNAs), RT-PCR quantitative, séquençage NGS

Highlighting the immediate but transient response of specific small RNA populations to allopolyploidy using resynthesized *Brassica napus* allotetraploids

Paulina MARTINEZ PALACIOS¹, Marie-Pierre JACQUEMOT², Marion TAPIE^{3,4}, Agnès ROUSSELET², Matthieu FALQUE², Andrew LLOYD⁵, Eric JENCZEWSKI⁵, Gilles LASSALLE⁶, Anne-Marie CHEVRE⁶, Christine LELANDAIS^{4,7}, Martin CRESPI⁷, Philippe BRABANT^{3,8}, Johann JOETS² and Karine ALIX^{3,8*}

¹Université Paris-Sud, ²INRA: ³UMR 0320 / 8120 Génétique Végétale, Ferme du Moulon, F-91190 Gif-sur-Yvette, France

⁴Université Paris Diderot Paris 7, 5 rue Thomas Mann, F-75205 Paris cedex 13, France

⁵INRA, Institut Jean-Pierre Bourgin, UMR 1318 INRA-AgroParisTech, F-78026 Versailles cedex, France

⁶INRA, UMR 1349 IGEPP, BP 35327, F-35653 Le Rheu cedex, France

⁷Institut des Sciences du Végétal, CNRS, F-91198 Gif-sur-Yvette cedex, France

⁸AgroParisTech, 16 rue Claude Bernard, F-75231 Paris cedex 05, France

*author to whom correspondence should be addressed

e-mail: alix@moulon.inra.fr, phone: +33/(0)1 69 33 23 72, fax: +33/(0)1 69 33 23 40

Author contributions: K.A. designed research; P.M.P., A.L., G.L., P.B. and J.J. performed bioinformatic and statistical analyses; P.M.P., M.-P.J., M.T., A.R., M.F. and C.L. performed molecular experiments; P.M.P., C.L., M.C., E.J., P.B., J.J., and K.A. analyzed data; A.-M.C. created the plant material; and P.M.P., J.J. and K.A. discussed the results and wrote the paper.

*Manuscript **in preparation** for submission to *Molecular Biology and Evolution* (10.35)
In response to the call for papers in *Experimental Evolution**

Introduction

Considering interspecific hybridization as a major evolutionary process of speciation in eukaryotes has long been a question under debate (Mable 2013). Indeed, hybrids were long assumed as evolutionary dead ends notably with regards to their often low fertility associated with reduced fitness. This view of interspecific hybrids as “reproductive accidents” was particularly common in animal genetics. Thanks to the technical improvements made in genomics, which have facilitated the access to complete genome sequences, there has been increasing evidence to support that more and more eukaryote species, including animals, are of hybrid origins (reviewed by Seehausen 2004 and Mallet 2007; e.g. *Heliconius* butterflies – Mavárez et al. 2006, Dasmahapatra et al. 2012; modern humans – Green et al. 2010, Abi-Rached et al. 2011). Nevertheless, it is in plants that interspecific hybridization has played a major role in species evolution.

The prevalence of interspecific hybridization in plants has been usually associated with the evolutionary success of polyploidy (whole genome duplication, WGD), all seed and flowering plants having experienced at least one round of WGD during the course of their evolution (Jiao et al. 2011). Indeed, allopolyploidy (the merger of two or more divergent genomes in a single nucleus following interspecific hybridization associated with genome doubling) has been recognized as one of the main factors of the higher adaptive potential of polyploid plants with respect to the increased genetic diversity allopolyploidy may result in. The first advantage of allopolyploids in comparison to homoploids is the immediate recovering of homologous chromosome pairs as a prerequisite to regular meiosis (which usually also requires a tight genetic control of crossing-over formation and distribution; Cifuentes et al. 2010, Grandont et al. 2013) and restoring fertility in the resulting hybrids, thanks to chromosome set doubling (which can be directly correlated to the rarity of

polyploidy in animals, notably with sexual chromosomes, where managing fertility and sex determination with genome duplications remains challenging; Collares-Pereira et al. 2013). Then the major advantage of allopolyploids is their ‘doubled interspecific hybrid’ composition that directly results in (i) gene redundancy accompanied by an increased number of alleles and the availability to acquire modified or new functions from duplicated genes (i.e. respectively sub- and neofunctionalization), and in (ii) fixing heterozygosity at the origin of the well-known phenomenon of heterosis, or hybrid vigour (Comai 2005). In addition to the rapid and profound structural and functional modifications that accompany the formation of a neo-allopolyploid and its further stabilisation as a new allopolyploid species (Madlung and Wendel, 2013), these different advantages of being allopolyploid explain the ability of allopolyploids to overcome competition with their parental species and colonize new habitats (Hegarty and Hiscock, 2008), as well as the large extent of their adaptability and thus evolutionary success. While epigenetic changes during neo-allopolyploid formation have been evidenced for different allopolyploid systems (e.g. *Arabidopsis* – Madlung et al. 2002, 2005; Ni et al. 2009; *Brassica* - Lukens et al. 2006, Książczyk et al. 2011; *Senecio* – Hegarty et al. 2011; *Spartina* – Salmon et al. 2005, Parisod et al. 2009; *Triticum* – Shaked et al. 2001), and may thus be involved in expression variation in allopolyploids, heterosis has also been recently suggested to originate from the confrontation of epigenetic patterns and populations of small non coding RNAs (sRNAs) inherited from each progenitor (Chen 2010, Barber et al. 2012, Ng et al. 2012). Moreover, at the onset of the neo-allopolyploid formation is the necessity to overcome hybrid seed failure, which ranges from early seed inviability to hybrid sterility. As demonstrated in animals, genome reprogramming has been shown to occur in plants, but specifically in vegetative cells, with the loss of DNA methylation at the origin of transposable element reactivation, resulting in the production of novel TE-derived sRNAs (Slotkin et al. 2009). If parental TE sequences differ substantially, as it should be the case

during interspecific hybridization, the TE-derived sRNAs inherited from each parent represent key factors to control the reactivated TEs, and their abundance directly correlated to their efficiency in repressing TEs may determine the outcome of the interspecific cross (Martienssen 2010). Such an epigenetic determinism of both the viability and the higher fitness of hybrids opens new avenues of investigation, including that of the evolutionary significance of epigenetic regulations in the success of interspecific hybridization and allopolyploidy.

Small non coding RNAs from plants have been subjected to various classifications (e.g. Axtell 2013; Martinez de Alba et al. 2013), which are in permanent revision because of the still continuous advances made in the knowledge and understanding of the “small RNA world” (Finnegan and Matzke, 2003). Based on their modes of biogenesis, sRNAs may be divided into two major classes (Martinez de Alba et al. 2013; Mirouze 2012): (i) the microRNAs (miRNAs), generally 21 nucleotide (nt) -long, which are produced from endogenous genes, the *MIR* genes (Voinnet 2009), and (ii) the small (or short) interfering RNAs (siRNAs) that represent a highly diverse class of 21- to 24-nt sRNAs (and thus subjected to numerous sub-classifications). These siRNAs are all produced from double-stranded RNA (dsRNA) precursors of various origins, such as inverted repeats, natural antisens overlapping transcripts, or dsRNAs synthesized from single strand RNAs (ssRNAs) by RNA-dependent RNA polymerases (RDRs) with primary ssRNAs resulting from the transcription of transposable elements (TEs) or other repetitive sequences (Bond and Baulcombe, 2014). Regarding the modes of action of these various types of sRNAs, miRNAs and 21-nt siRNAs have been demonstrated to mostly mediate post-transcriptional gene silencing (PTGS) by endonucleolytic cleavage of messenger RNA targets or repression of translation, while 24-nt siRNAs usually repress gene activity by transcriptional gene silencing (TGS) through modifications of epigenetic marks of chromatin (DNA methylation, histone

modifications) at specific genomic loci. For every sRNA, the highly specific action of gene repression is based on base-pairing (Mallory and Vaucheret, 2006), with selectivity being mediated by sequence similarity with the sRNA which has been loaded into one Argonaute (AGO) protein as the major component of the effector silencing protein complex RISC (RNA-induced silencing complex). Only two studies have been specifically dedicated to the analysis of the mobilization of sRNAs during allopolyploidization events, demonstrating immediate changes in expression levels of sRNAs at the onset of the allopolyploid formation in both the allopolyploid *Arabidopsis* and wheat models (Ha et al. 2009; Kenan-Eichler et al. 2011). To complement these two studies, evaluating the global dynamics of expression of sRNAs following allopolyploidisation, with the possibility to distinguish the hybridization step from the polyploidisation one, should add to our knowledge of the response of the different sRNA pathways to allopolyploidy, which represents one of the major evolutionary processes of plant speciation. This is what we aimed to study in *Brassica*.

The *Brassica* genus represents a good model to study speciation by allopolyploidy, with the allotetraploid species *B. napus* (oilseed rape, AACC, $2n=38$) originating from interspecific hybridization between the diploid species *B. rapa* (AA, $2n=20$) and *B. oleracea* (CC, $2n=18$) that occurred less than 10 000 years ago. As the current diploid *Brassica* species are still closely related to the true diploid progenitors of *B. napus*, the allotetraploid can be resynthesized artificially, and therefore represents an excellent plant material to analyze the genomic consequences of polyploidy over different timescales (Jenczewski et al. 2013). This has been particularly well demonstrated by previously published studies dealing with the impact of allopolyploidy on genome rearrangement (Gaeta et al., 2007; Szadkowski et al., 2010, 2011), gene expression (Albertin et al., 2006; Gaeta et al., 2009; Marmagne et al., 2010) or transposable element reactivation (Sarilar et al. 2013). In the present study, we characterized the global response of sRNA populations to allopolyploidy in *Brassica*, by

comparing distributions of sRNAs in three resynthesized *Brassica napus* allotetraploids surveyed at two different generations and in their diploid progenitors, using high-throughput sequencing of sRNA populations. Access to the first hybrids resulting directly from crosses between the diploid progenitors indicated that hybridization triggers part of the immediate changes in sRNA expression levels depicted in the early generations of the neo-allopolyploids, while a progressive parental-like inheritance of sRNAs was observed across the more advanced generations of allopolyploidy. We notably discuss the involvement of small non coding RNAs in the success of allopolyploid establishment, thanks to the control of genome integrity by silencing of transposable elements and other endogenous viral-like elements that appear particularly subject to instability during the collision of the two differentiate genomes responsible for the new speciation event.

Results

To study the global impact of allopolyploidy on the expression of small RNAs, 23,102,628 sequence reads were generated by Illumina sequencing of eleven libraries, constructed from stems of resynthesized *B. napus* allotetraploids at two different generations of selfing and their diploid progenitors *B. oleracea* and *B. rapa* (Fig. S1). After the removal of adaptors, low quality reads and sequences that matched perfectly ribosomal and transfer RNAs, we focused our analysis on the 14,917,073 sequence reads of 18-26-nt in size. To allow further comparisons, read counts were normalized against the total number of reads analyzed for each library.

Expression levels of small RNAs experienced rapid changes following allopolyploidy

Size distribution of small RNAs showed that the 24-nt class was the most abundant class in all libraries, accounting on average for 45.54 % of the total number of redundant reads per

library, followed by the 21-, 23- and 22-nt classes which represented on average 20.63 %, 14.03 % and 8.02 % of the reads for each library, respectively (Table S1). The pattern of size distribution of sRNAs we observed for both diploid and allotetraploid *Brassica* genotypes was thus similar to patterns of sRNA distribution which were previously depicted for various plants (as reported by Romanel et al. 2012, even if exceptions can be reported such as in grapevine, Pantaleo et al. 2010), indicating a good representation of sRNA expression in the eleven libraries surveyed. Interestingly, as shown in Fig. 1A, the 21-nt sRNA class was significantly higher in the S1 generation than in the parental generation by comparison to the MPV ($P<0.005$), but also higher than in the S5 generation ($P<0.05$), whereas no significant differences were observed when comparing the abundance of the 21-nt sRNAs in the S5 generation to the MPV ($P=0.22$). At the S1 generation, the increase of the 21-nt sRNA class was accompanied simultaneously by a significant decrease of the 24-nt class by comparison to the MPV ($P<0.05$). In the same manner, the 24-nt class was depicted as lower in the S1 than in the S5 generation, and this difference was significant ($P<0.01$) when considering the S5 lines from the two allotetraploids BoEMZ2₃₈ and BoEMZ5₃₈. As for the 21-nt class, no significant differences were observed between the relative abundances of the 24-nt sRNAs in the S5 generation and the parental generation (tested using the MPV, $P=0.22$). Interestingly, we could observe the same dynamics for the 22-nt class as for the 21-nt class, with significant increases in the S1 generation compared to the MPV ($P<0.01$) and the S5 generation ($P<0.01$), while no significant differences could be made between the S5 lines and the parents (still tested using the MPV, $P=0.73$). The 23-nt class was similar to the 24-nt class: its abundance was lower in the S1 generation than the MPV ($P<0.05$) and the relative abundance estimated for the S5 generation ($P<0.05$ when considering BoEMZ2₃₈ and BoEMZ5₃₈), while no significant differences could be made between the relative abundance of the 23-nt reads in the S5 generation and the MPV ($P=0.70$).

We then focused on the two most abundant classes of reads, and tested if the variations observed for the 21-nt and 24-nt classes across the generations were mainly (i) quantitative, with the same populations of sRNAs showing different levels of expression across the generations, or (ii) qualitative, with populations of sRNAs being expressed specifically in each generation. After removing sequence redundancy, the 24-nt class was still the most abundant class in all the libraries, i.e. the most diverse, and accounted on average for 55.36% of the total number of unique reads per library, followed by the 23-, 21- and 22-nt classes representing on average 19.65%, 9.41% and 7.38% of the unique reads for each library, respectively (Table S1). We observed a general drastic reduction of the 21-nt class, suggesting high levels of expression for a major part of the 21-nt sRNAs in the three generations surveyed (Fig. 1B); the still highly abundant 24-nt class could be related to the presence of numerous unique 24-nt sRNAs but expressed at relatively low levels (in accordance with previously published studies that demonstrated the usual high sRNA diversity of the 24-nt class, e.g. Lelandais-Brière 2009). The only significant differences we detected across the generations for the relative amounts of both 21-nt and 24-nt unique sRNAs corresponded to a slight increase in the level of the 21-nt sRNA diversity estimated for the S1 generation in comparison to the ones estimated for the parental generation ($P < 0.05$ against the MPV) and the S5 generation ($P < 0.05$, taking into consideration the two allotetraploids BoEMZ₂₃₈ and BoEMZ₅₃₈). It can be noticed that similar differences were observed only for the 22-nt unique sRNAs, with a higher diversity in the S1 generation compared to the diploid parents (tested using the corresponding MPV, $P < 0.05$) and the S5 lines ($P < 0.05$). To complete our analysis, we determined the qualitative composition of the sRNA diversity for the 21-nt and 24-nt classes; to take into account background noise of sRNA expression, we focused the analysis on different sets of expressed sRNAs according to conditions of expression levels. Only sRNAs with normalized expression levels ≥ 35 reads per million (RPM) within each library

were analysed, taking into consideration the work from Fahlgren et al. (2009) who demonstrated highly reproducible data for the more abundant sRNAs with >30 RPM ($\log_2(\text{RPM}) > 5$) when considering individual sRNA sequences. The results shown in Figure 1C indicated a higher diversity within the S1 generation (50% and 41% of the 21- and 24-nt sRNAs analysed respectively, were S1-specific) suggesting the expression of specific sRNA populations in response to allopolyploidy.

Altogether, the results obtained from the analysis of the size-specific distribution of small RNAs across the three generations surveyed, clearly showed that early generations (S1) of the *B. napus* neo-allopolyploids exhibited an over-expression of 21-nt sRNAs accompanied by an under-expression of 24-nt sRNAs in comparison to the MPV (used as the reference for testing additivity), while the more advanced generations (S5) tended to restore expected mid-parent expression levels for both sRNA classes. We thus carried out the detailed annotation of these two classes of reads by bioinformatics (*SI materials and methods*); our major goal was to identify small RNA populations that were responsible for the deviation of the sRNA expression patterns observed in the S1 generation (i) from the expected additivity of the parental patterns, as well as (ii) from the expression patterns observed in the S5 generation which were highly similar to the mid-parental ones.

The global expression of microRNAs was not significantly affected by allopolyploidy

We first identified conserved microRNAs (miRNAs) by comparing the entire set of unique reads for each of the two 21-nt and 24-nt classes to the plant miRBase database (release 19). Among the 24-nt reads, it can be said that no conserved miRNAs were identified (one microRNA with only 1 raw read count was retrieved from the sequencing data obtained for the BoEMZ2₃₈ S5 line) in accordance with the rarity of 24-nt miRNAs in the plant miRBase. On the contrary, among the 21-nt class, more than one third of the sRNA reads could be

identified as already known microRNAs, representing on average 33.93%, 30.91% and 42.80% of the 21-nt reads from the parental, the S1 and the S5 generations, respectively. Indeed, 127, 107 and 101 unique reads of 21 nucleotides that perfectly matched registered miRNAs (with no mismatch), representing a total of 414 368, 550 949 and 282 423 raw read counts were identified for the respective parental, S1 and S5 generations. These miRNAs corresponded to a total of 34 microRNA families (Fig. 2A, and detailed in Table S2), including 24 miRNA families which were notably already identified in *Brassica*. In terms of diversity, nearly all the unique retrieved miRNA sequences were found in the three generations analysed, with only a few reads that were specific of one generation, and usually of one replicate within the concerned generation, and always displaying very low levels of expression (6 unique miRNAs, represented each by 1 to 4 raw read counts only, Fig. 2A).

To complete our analysis, we performed microRNA prediction on the still non-annotated 21-nt and 24-nt fractions of the libraries with miRCat from the UEA plant sRNA toolkit (<http://srna-tools.cmp.uea.ac.uk/>) that identifies mature miRNAs as well as their precursors (Moxon et al. 2008). Using the complete genome sequences of *Brassica rapa* and *Arabidopsis thaliana* as references, we could not predict any miRNA candidate among the 24-nt reads, while 198, 129 and 113 unique reads (Fig. 2B), representing a total of 64 160, 17 014 and 42 910 raw read counts, were predicted as miRNA candidates from the 21-nt class for the respective parental, S1 and S5 generations, according to hairpin prediction. To provide here a new set of fully described new putative *Brassica* miRNAs, details for locations along the *B. rapa* genome of the putative *MIR* genes and characterization of the corresponding miRNA precursors are given in Table S3. It has to be noticed that the identification of the miRNA* strands in addition to the guide miRNA strands for 92 out of the collection of the 274 newly predicted *Brassica* microRNAs (Table S3) adds robustness to our current *Brassica* miRNA prediction which may be useful for further investigation. Therefore, within the 21-nt class,

predicted miRNAs were estimated as representing on average 5.88%, 1.44% and 8.08% of the 21-nt reads from the parental, the S1 and the S5 generations, respectively. The diversity of the *Brassica* predicted miRNAs appeared slightly higher in the diploid parents in comparison to the neo-allopolyploids, even if caution has to be taken as no gathering in larger microRNA families can be done, like for already identified miRNAs.

While we observed similar populations of known miRNA families between the diploid progenitors and the two generations of resynthesized *B. napus* allotetraploids, indicating equivalent levels of miRNA diversity across the generations, we tested for their differential expression levels between the generations. We could not find any significant differences when comparing the relative abundance of the 21-nt fraction annotated as known microRNAs in the S1 generation to the MPV ($P=0.22$) or to the one estimated in the S5 generation ($P=0.64$), this latest one being neither significantly different from the MPV ($P=0.44$). Similarly, no significant differences were observed when considering the relative abundance of the whole fraction of sRNAs annotated as miRNAs (known and predicted) – S1 vs. MPV, $P=0.23$; S1 vs. S5, $P=0.75$. In summary, the results we obtained from the bioinformatic analysis of the dynamic of expression of miRNAs following allopolyploidy indicated that the microRNAs were not responsible for the immediate changes we observed in the S1 generation. Quantification of expression by real-time quantitative RT-PCR (qRT-PCR) of a set of five different known microRNAs were in agreement with the sequencing data; accordingly, the three microRNAs miR159a, miR160a and miR166a were validated as displaying stable expression levels across our experimental design (Fig. S2) and were thus further used as references for validation of the dynamics of expression of other sRNAs of interest across the generations by qRT-PCR experiments.

The expression changes of small RNAs derived from transposable elements in response to allopolyploidy were highly diverse

For each of the two 21-nt and 24-nt classes, we identified small RNAs derived from transposable elements among the reads which were not previously categorized as miRNAs or putative miRNAs. According to recent classifications of sRNAs, we further denominated these reads as siRNAs (corresponding to the primary classification from Axtell et al. 2013). The sequence reads were subjected to local BLAST analyses (word size = 21 or 24, no mismatch) using Repbase and a locally developed *Brassica* transposable element database (unpublished data; Sarilar, Martinez Palacios, Joets, Alix) as references for transposable element sequences. Among the 21-nt class, relatively few reads were similar to TE sequences corresponding on average to 8.04% (representing 71,454 raw read counts), 3.79% (52,207) and 7.16% (41,299) of the 21-nt reads from the parental, the S1 and the S5 generations, respectively. No significant differences could be found when comparing the relative abundance of the 21-nt fraction annotated as TE-derived siRNAs in the S1 generation to the MPV ($P=0.30$) or to the one estimated in the S5 generation ($P=0.82$), which was neither different from the MPV ($P=0.21$). Within the 24-nt class, TE-siRNAs were estimated as representing on average 14% (597,714 raw read counts), 12% (189,912) and 14% (236,162) of the 24-nt reads from the parental, the S1 and the S5 generations, respectively. We observed a significant decrease of the abundance of the 24-nt TE-siRNAs in the S1 generation in comparison to the MPV ($P<0.05$) as well as to the abundance observed in the S5 generation ($P<0.01$, when considering the allotetraploids BoEMZ2₃₈ and BoEMZ5₃₈). Here again, no significant differences could be observed when comparing the abundance of the 24-nt TE-siRNAs in the S5 generation to the MPV ($P=0.28$). Therefore, the fraction of sRNAs annotated as TE-siRNAs was demonstrated to contribute, at least partly, to the overall decrease of the 24-nt class observed in the S1 generation we depicted previously (Fig. 1). We thus undertook the annotation of the TE-siRNA fraction, for each of the 21-nt and 24-nt classes, to describe in more detail the changes encountered by populations of TE-siRNAs

according to the TE class / order they originated from (following the TE classification from Wicker et al. 2007). From the results obtained after TE-annotation, we could observe different dynamics of expression of TE-siRNAs in relation with the TEs they derived from (Fig. 3). For instance, SINE-related siRNAs appeared significantly less abundant in the S1 generation of the allotetraploids in comparison with the parental generation, for the corresponding 21-nt and 24-nt siRNAs ($P < 0.05$ in comparison to MPV), and the same dynamics were observed for siRNAs derived from LINEs, with the S5 generation displaying additive expression profiles in comparison with the parental ones in both cases. For siRNAs derived from Helitrons and transposons of the TIR order, no significant differences were observed for the relative abundances of the 21-nt siRNAs across the generations, while a significant decrease of the corresponding 24-nt siRNAs was identified in the generation S1 in comparison with the diploid parents ($P < 0.05$ in comparison to MPV); the S5 generation was still closely similar to the parental generation when compared to the mid-parent values of the relative siRNA abundances surveyed. The most important changes in TE-siRNA abundances following allopolyploidisation were observed for the siRNAs originating from LTR-retrotransposons, with opposite directions of variation for the 21- and the 24-nt siRNAs: (i) the 21-nt siRNAs derived from LTR-retrotransposons were significantly more abundant in both the S1 and S5 generations of the allotetraploids in comparison to the parental generation ($P < 0.01$ in comparison to MPV), with no differences between these two generations of selfing, (ii) while the amounts of the corresponding 24-nt siRNAs were similar in the parental and S1 generations (S1 vs. MPV, $P = 0.89$), increasing significantly in the S5 generation for the two allotetraploids BoEMZ2₃₈ and BoEMZ5₃₈ (S5 vs. MPV and S5 vs. S1, $P < 0.02$).

Quantification of expression by qRT-PCR was performed on a set of eleven different 24-nt TE-siRNAs, but usable results were obtained for only four of them; three were derived from TIR transposons and one from LTR-retrotransposons. For the TIR-siRNA24_37, the

results were in accordance with the sequencing data and the general dynamics of expression observed for the 24-nt TIR-derived siRNAs (depicted in Fig. 3): as shown in Figure 4, we observed a significant decrease of its expression in the S1 generation compared to the parental one (S1 vs. MPV, $P=0.0005$), while its level of expression in the S5 generation was increasing and more similar to the one estimated for the mid-parent value ($P\sim 0.08$). Interestingly, qRT-PCR performed on the first hybrids F1 and the generations S0 and S3 led to identify an immediate response of this population of 24-nt TIR-siRNA24_37, with a significant decrease observed in the F1 (F1 vs. MPV, $P<0.02$) which became highly significant in the somatic-doubling generation S0 (S0 vs. MPV, $P<0.001$). Then, its levels of expression estimated in the S3 generation were not significantly different from those obtained for the S5 generation ($P=0.54$). For the three other 24-nt TE-siRNAs analysed (data not shown), the analysis of variance (ANOVA) within the generations was not significant; comparing means using a Student's t-test identified tendencies for the 24-nt LTR-siRNA24_16 to be more expressed in the first generation S0 of the allopolyploids (S0 vs. MPV, $P<0.05$), and for the TIR-siRNA24_36 to be more abundant in the S5 generation compared to the S1 and S3 generations ($P\sim 0.05$), while the 24-nt TIR-siRNA24_22 displayed relatively stable expression levels in the different generations of the allopolyploids compared to the diploid parents with a slight decrease in the S1 generation (S1 vs. MPV, $P=0.07$). For these three populations of siRNAs, expression levels of the TEs they putatively originated from were also assayed by qRT-PCR; for the TIR-transposon on which the 24-nt TIR-siRNA24_22 mapped, we identified higher expression levels in the S5 generation compared to the S1 and S3 generations ($P<0.05$ in both cases), and for the two other TEs analysed, they displayed the same dynamics of expression as the ones depicted for the corresponding 24-nt siRNA populations (data not shown).

Specific populations of small non-coding RNAs were demonstrated as responsible for the relative increase of expression of the 21-nt class in the S1 generation

In order to identify sRNA populations responsible for the global over-expression of 21-nt sRNAs observed in the early generations (S1) of the resynthesized *B. napus* allotetraploids, we categorized the remaining reads to be annotated, as (i) reads mapping *Brassica* genomic sequences (taking into account all the released *Brassica* genomic resources at the time of our study) and (ii) reads of undetermined origins, as being not present in the genomic dataset surveyed. While the reads mapping *Brassica* genomic sequences were equally abundant in the two generations of the neo-allotetraploids, with no significant deviation from the expected mid-parent value (ANOVA $P_v=0.55$), the 21-nt reads of still undetermined origins were significantly over-represented in the S1 generation (ANOVA, $P_v<0.001$). Indeed, a total of 167 263 (representing 2.3% of all the reads sequenced for the concerned libraries), 709 359 (16.6%), and 92 828 (2.7%) raw read counts were identified for the respective parental, S1 and S5 generations. These proportions represented a highly significant increase of the abundance of the 21-nt sRNAs in the S1 generation in comparison to the MPV ($P<10^{-4}$) as well as to the abundance estimated in the S5 generation ($P<0.001$), while no significant differences were observed when comparing the abundance of these 21-nt sRNA populations in the S5 generation to the MPV ($P=0.45$). We sorted the unique 21-nt reads based on their levels of expression, taking into consideration that a 21-nt sRNA to be further analysed has to be expressed at least 500 times greater in the S1 generation than in both the parental and the S5 generations (according to the crude bioinformatic data). A total of 68 sRNA sequences were thus retrieved from the undetermined sRNA reads, with expression levels ranging from 501 to 7192 raw read counts in the S1 generation, and only 0 to 6 reads in each of the parental and S5 generations; they represented a total of 109 548 reads from the S1 generation and only 81 and 66 reads from the parents and the S5 generation, respectively (Table S4).

We then performed the validation of the expression levels for the 21-nt sRNA populations specific of the S1 generation, by qRT-PCR: from the 68 sequences, a total of 18 sRNAs were tested for qRT-PCR analyses; results could be used for 10 of them (Fig. S3). The first hybrids F1 as well as the S0 and S3 generations of the three resynthesized *B. napus* allotetraploids were included in the qRT-PCR experiments. For all the ten siRNAs we assayed the relative abundance across the different generations by qRT-PCR, we clearly demonstrated a significant increase of the siRNA reads in the S1 generation in comparison with the diploid progenitors and the S5 generation, producing results which were in agreement with bioinformatic predictions. Indeed, while all sRNAs were detected in the S1 generation with PCR products obtained in all the cases surveyed, no products were amplified in most of the cases surveyed for the parental and S5 generations or amplicons were obtained with high Cq values (Fig. S3); it was common to identify ΔCq of more than 10 between the parents/S5 and the S1 generations. Interestingly, qRT-PCR analyses performed on the F1 hybrids and the S0 generation led to very similar results as those obtained for the S1 generation, while the S3 generation led to results identical as those generated from the S5 generation. For instance, the siRNA21_21 (Fig. S3, p.5/16) was detected in both the F1 hybrids and the S1 generation with Cq=20 and in the S0 generation with Cq=23, while it was detectable at Cq=33 and Cq=35 for the S5 generation and the diploid parents, respectively and not detectable in the S3 generation. In addition to qRT-PCR, expression levels were also assayed by Northern blot hybridization for four different sRNAs. While the S1-specific expression of the different 21-nt siRNAs was indeed validated, it was striking to observe additional hybridization signals at the size of 24-nt in the two generations S1 and S5 (Fig. 5), with co-existence of distinct populations of siRNAs putatively involved in the silencing of the same targets in S1. The presence of the 24-nt siRNAs only in the S5 generation suggested a progressive shift in the sRNA regulation pathways mobilized during the formation of the neo-allopolyploids.

We hypothesized that the absence of these specific 21-nt sRNAs from the *Brassica* genomic resources was related to their potential origins from genome components located in heterochromatic regions (supported by the expression of specific 24-nt siRNA populations involved in silencing by RdDM; Fig. 5); indeed, these regions, enriched in transposable elements and other repetitive sequences, are usually not sequenced during complete genome sequencing projects. We undertook the identification of the molecular origins of these specific 21-nt sRNA populations by exploiting mRNAseq data generated from high throughput sequencing of RNAs extracted from meiocytes of two different oilseed rape varieties; these varieties were both analysed at two different ploidy levels, haploids vs. diploids (unpublished data; Lloyd, Delannoy, Jenczewski). Interestingly, twenty-nine of the sixty-eight 21-nt sRNAs matched perfectly with mRNAs that were only transcribed in the haploid forms of one of the two varieties surveyed. In particular, the two 21-nt siRNAs, siRNA21_26 and siRNA21_39, corresponded to transcripts showing high similarities with viruses (E-values of respectively $1e-165$ with 99% identity covering 319bp of the first viral sequence (accession #X13062) and $4e-76$ with 92% identity covering 210bp of the other virus sequence (accession #HQ388350) retrieved from the mRNAseq data). Validation of the expression of the corresponding viruses by qRT-PCR revealed high levels of expression in the F1 hybrids, the S0 and S1 generations, but no expression in both diploid progenitors as well as in the S3 and S5 generations (Fig. S3). In silico mapping of the full set of the sRNA reads, excluding the miRNAs (known and predicted) on the two viral sequences, revealed an over-representation of virus-derived siRNAs of both 21- and 24-nt long in the S1 generation, with only a very few corresponding reads in the parental and S5 generations (Fig. 6). Such patterns indicate again the co-existence of distinct populations of siRNAs involved in two different regulation pathways that were recruited immediately after the allopolyploidisation event to efficiently silence reactivated endovirus-like sequences in response to the allopolyploidy-induced genomic shock.

Discussion

Allopolyploidy, combining interspecific hybridization with whole genome duplication, has played a major role in plant evolution, and has been the subject of important research activity to decipher the reasons of such an evolutionary success (Ainouche and Jenczewski, 2010). It has been thus demonstrated that the neo-allopolyploid genome encounters numerous changes, at both the structural and functional levels, which may be mainly triggered by the merger of divergent genomes in the same nucleus – the so-called “genomic shock”. The *Brassica* polyploid system, with the ability to resynthesize allotetraploids artificially, represents one of the excellent models to study allopolyploidy. Notably, *B. napus* and the resynthesized allotetraploid forms of this species have been deeply studied for the changes occurring at the onset of its formation, by analysing the genomic, meiotic, transcriptomic and proteomic levels (reviewed by Jenczewski et al. 2013). The originality of the present work is to complement all the previous studies performed on the same material of resynthesized *B. napus* allotetraploids by providing the first step towards the integration of the different changes observed during the early stages of the allopolyploid formation with the major molecular mechanisms of maintenance of genome integrity and regulation of gene expression that involve small non coding RNAs.

Small RNAs experience non-stochastic and immediate expression changes during allopolyploidisation

In the present study, we used high-throughput sequencing of small RNAs from a total of eleven libraries, in order to analyse the global changes of sRNA expression during the formation of three independent resynthesized *B. napus* allotetraploids (surveyed at two different generations, S1 and S5) in comparison with their diploid progenitors. Analysis of the size-specific distribution of the ~15 million sequence reads of 18-26-nt in size revealed that small RNA populations experienced rapid changes in the early generations of the neo-allopolyploids: while the S5 generation displayed global sRNA expression levels similar to the parental ones, the S1 generation underwent drastic modifications in the relative abundances of sRNA populations. Comparing relative abundances of sRNA reads and levels of diversity in sRNA sequences across the three generations surveyed, we could establish that the differences observed in size-distribution were mainly related to differential expression of sRNA populations: 21-nt sRNA populations were over-expressed in the S1 generation while 24-nt sRNAs were under-expressed, in comparison with the expression levels estimated for both the parental and S5 generations. One of the most interesting characteristics of our experimental design was the possibility to compare sRNA mobilization between three independent resynthesized *B. napus* allotetraploids. The two previous studies dealing with the mobilization of sRNAs during allopolyploidisation were performed on only one resynthesized *A. suecica* allopolyploid (surveyed at the two generations F1 and F7, Ha et al. 2009) or by the analysis of sRNA sequencing data of one single library constructed using RNA extracts from a pool of synthetic allopolyploid wheats with no access to the variability between resynthesized allopolyploids (Kenan-Eichler et al. 2011). We thus had the opportunity to generate innovative and highly interesting results, as we detected variations in sRNA expressions that were reproducible in the three independently created *B. napus* allotetraploids,

indicating non-stochastic regulation of sRNA expression in response to the allopolyploidisation event. Such immediate and directed changes in sRNA expressions, with a rapid return to the parental state, may suggest an immediate and necessary remobilisation of sRNAs to face the genomic shock that accompany the early stages of allopolyploidisation. The annotation of the 21- and 24-nt sRNA fractions allowed the identification of specific sRNA populations involved in this early response to allopolyploidy.

Severe changes in LTR-derived siRNAs indicate a PTGS-to-TGS shift in silencing of LTR retrotransposons at the first stages of the neo-allopolyploid formation

In the global analysis of the changes encountered by TE-derived siRNAs during allopolyploidisation, we observed a drastic decrease of the corresponding 24-nt fraction in the S1 generation only. This echoes the general observations made by previous studies on the immediate impact of hybridization (such as in Arabidopsis, Groszmann et al. 2011 ; rice, He et al. 2010 ; or maize, Barber et al. 2012) and allopolyploidisation (as observed for the first allopolyploids F1 of Arabidopsis, Ha et al. 2009; and S1 of wheat, Kenan-Eichler et al. 2011). However, detailed annotation of the TE-derived sRNAs could identify different trends of regulation of TEs depending on their nature, with differential involvements of 21- and 24-nt TE-derived sRNAs across the generations. Changes in the S1 generation only could be observed for SINEs and LINEs with both the corresponding 21-nt and 24-nt siRNAs being under-expressed, as well as for Helitrons and TIR, for which only a significant decrease for the matching 24-nt siRNAs was identified. But for LTR-retrotransposons, which represent the most abundant order of TEs in plant genomes, the scenario was quite different with the 21-nt LTR-siRNAs being over-expressed in both the S1 and S5 generations, while the 24-nt LTR-siRNAs were over-expressed in the S5 generation only. This reminds the recent work published by Marí-Ordóñez et al. (2013) that aimed at reconstructing *de novo* silencing of an

active and transcribed plant LTR-retrotransposon, *EVD*, in *Arabidopsis*; they demonstrated an immediate PTGS control of the reactivated retrotransposon by 21-nt sRNAs, progressively replaced by TGS mediated by 24-nt siRNAs for a longer established silencing of the active retrotransposon. Accordingly, it can be assumed that LTR-retrotransposons became transcriptionally active upon hybridization and/or polyploidisation in response to the genomic shock at the origins of the RdDM-based TGS deregulation, with an immediate mobilization of 21-nt siRNAs produced from the transcribed retrotransposons to enhance PTGS and jugulate the proliferation of the LTR-retrotransposons. Then, 24-nt siRNAs were produced to set up TGS which progressively succeeded PTGS (with the distinct 21- and 24-nt siRNA populations coexisting in the S5 generation). Such dynamics of TE repression is also in agreement with the occurrence of both 21- and 24-nt siRNAs in the S1 generation (here, taking place earlier for the concerned targets) as depicted by Northern blot hybridization (Fig.5) for elements we supposed to be located in usually highly heterochromatic regions. Populations of 21- and 24-nt siRNAs can be thus regarded as the guardians of genome integrity in the neo-allopolyploid, ensuring its maintenance in response to the allopolyploidisation-induced genomic shock.

Specific 21-nt and 24-nt sRNA populations may be involved in hybrid viability by controlling reactivated endogenous viral-like elements

Another particularly original and interesting result was the identification of populations of 21-nt small RNAs corresponding to virus-like sequences that were highly expressed in the S1 generation of the three resynthesized *B. napus* allotetraploids under study and demonstrated to be also highly expressed in the F1 hybrids and the first generation of the allotetraploids S0. In addition, we demonstrated the transcription from the corresponding virus-like sequences in response to the merger of the two differentiated genomes in the hybrid F1, as well as in the

early generations of the neo-allopolyploids (S0 and S1), having thus the same dynamics of expression as the 21-nt virus-derived siRNAs. Additionally, analysis of sequencing data revealed the expression of populations of 24-nt siRNAs targeting the same virus-like sequences, which were also specifically expressed in the S1 generation. Such observations led us to the hypothesis that endogenous viral elements (EVEs) of the *Brassica* genome may be reactivated in response to the stress caused by interspecific hybridization at the origins of the new allopolyploid genome. The EVEs were recently demonstrated to be present in all eukaryotic genomes, and may correspond to cases of exaptation and domestication of viruses (such as domestication of retrotransposons), with all types of viruses being concerned (Feschotte and Gilbert, 2012). Becoming endogenous, they can express themselves in response to stresses (as already depicted in *Musa*, Côte et al. 2010). We can thus formulate an original hypothesis, namely that autoimmune reactions with necrosis responsible for hybrid incompatibility between species (Bomblies and Weigel, 2007) could result from the reactivation of some EVEs, which are expressed with no tight control by small RNAs; these EVEs lead the hybrid plant to respond to this “viral invasion” by a generalized apoptosis reaction. In the case of the resynthesized *B. napus* allotetraploids we studied, the expression of 21-nt and 24-nt small RNAs specific of the reactivated EVEs would avoid any proliferation of these endoviruses, thus ensuring the viability of the new allopolyploids.

While the present study gains insight into the involvement of specific populations of small non coding RNAs in ensuring genome integrity and thus hybrid/allopolyploid viability during the early stages of the formation of the neo-allopolyploids, additional work has to be done to get a more precise picture of the role of sRNAs in regulation of gene expression following allopolyploidy. However, one of the strengths of our study is to provide an important collection of small RNAs that can be expressed in *Brassica*. For instance, we took advantage

of our work to predict a total of 274 new microRNAs (with more than one third being supported by sequencing of the corresponding miRNA*) representing a nice set of sequences to consider further when working on the regulation of gene expression in *Brassica*. This work can also be considered as an essential starting point to refine the understanding of specific regulations that have been demonstrated previously to occur during the allopolyploidisation event of *B. napus*.

Materials and Methods

Plant material and RNA extraction

Our plant material consisted of three independently resynthesized *Brassica napus* allotetraploids (i.e. resulting from three independent crosses), namely BoEMZ1₃₈, BoEMZ2₃₈ and BoEMZ5₃₈ and their diploid progenitors *B. oleracea* ('HDEM', female donor) and *B. rapa* ('Z1'). It comprised the first hybrids F1 (CA), the corresponding doubled hybrids S0 (CCAA), and the following generations of selfing S1, S3 and S5 (Fig. S1). All the plant material was produced by A.-M. Chèvre and collaborators, and was already used in several studies in which it was fully described (e.g. Szadkowski et al., 2010; Sarilar et al. 2013). At least two cuttings per plant were grown under controlled conditions (18°C during 8h night and 21°C during 16h), and samples were collected individually from each single cutting, at once. Total RNAs were extracted from stems using the TRIzol® reagent following the manufacturer's protocol (Invitrogen), with addition of 0.2 µl/ml beta-mercapto-ethanol to TRIzol® extemporarily.

Small RNA library construction and sequencing

Small RNA library construction and sequencing using the Illumina GA II sequencing system were performed by ImaGenes GmbH (Berlin, Germany). A total of eleven small RNA libraries were constructed and sequenced, corresponding to RNA extracts from the diploid progenitors (with two biological replicates for each progenitor) as well as a 1:1 parental RNA mix used as a technical replicate, and RNA extracts from the three resynthesized *B. napus* allotetraploids surveyed at two different generations of selfing S1 and S5 (Fig. S1).

Bioinformatic analysis

Adaptors, low-quality reads and reads smaller than 18-nt and longer than 26-nt were removed from the sRNA sequence data prior to further analysis. Each small RNA library was analyzed individually, with small RNA reads being normalized by library size for further comparisons (see below, statistical analysis). Size distribution of sRNAs was first computed for comparisons between generations. Then, full annotation of the sRNA classes of interest was performed to tentatively identify their genomic origins. The detailed bioinformatic analysis of sRNA sequences is provided in *S1 Materials and Methods*, with the corresponding bioinformatic flowchart. The sRNA sequence data have been deposited in the Sequence Read Archive (SRA) database at the NCBI; accession numbers [XXXX](#).

Statistical analysis of the differential sRNA expressions supported by NGS data

Read counts were normalized to adjust for differences in library size and coverage, allowing direct comparisons between sRNA data from different libraries; each raw read count was multiplied by 10^6 and then divided by the total read count of the whole library (the unit becoming 'reads per million', RPM). To identify differences in expression for specific sRNA populations (by size or by category determined after annotation within each size class) we

performed a one-way analysis of variance. The classification variable was the generation with five different levels HDEM, Z1, 1:1 parental RNA extract mix, S1 and S5. As the differences between the profiles obtained from the 1:1 parental RNA extract mix and the mean of the two parents were never found significant (see below), we calculated the mean parent value (MPV) as the weighted mean of HDEM, Z and of the 1:1 RNA extract mix from the two diploid parents. We then performed the three following contrasts: MPV vs. S1, MPV vs. S1, and S1 vs. S5 (using a Student's t-test procedure).

When estimating significant allopolyploidy-related differences in expression of sRNAs between the diploid progenitors and the different resynthesized *B. napus* allotetraploid lines, sRNA expression data in the synthetic lines were compared to the mid-parent expression value (MPV). When analyzing size distribution of sRNAs across the generations, the calculated mid-parent value (MPV_c), corresponding to the average expression data estimated from the four individual parental sRNA libraries, was compared to sRNA data obtained for the library constructed from the 1:1 parental RNA mix which represented a technical replicate (MPV_t). As no significant differences were found (with P -values such as $0.18_{26-nt} \leq P \leq 0.96_{20-nt}$), the MPV taken as the reference for all the comparisons made further corresponded to the mean value of sRNA expression data obtained from the five parental libraries.

Real time quantitative reverse transcription PCR (qRT-PCR) analysis

RNA extracts were treated with DNase I (DNA-Free Ambion #1906). First-strand cDNA was synthesized from 500 ng of total RNA using the miScript II RT kit (Qiagen), using the miScript HiFlex Buffer in order to allow quantification of both sRNAs and mRNAs. Real-time quantitative RT-PCR (qRT-PCR) for sRNAs was performed using miScript SYBR Green PCR kit (Qiagen) on a 7500 Real-Time PCR System (Applied Biosystems). Reactions

were performed following manufacturer's instructions, by adding to the mix 0.5 μ M of the oligonucleotide corresponding to the direct sequence of the targeted sRNA and 5 μ l of the synthesized cDNA diluted 300 times, and subjected to the following programme: 95°C for 15 min, 40 cycles of 94°C for 15 s, Ta for 30 s and 70°C for 45 s; Ta ranged from 55°C to 65°C and was optimized for each sRNA primer. Dissociation curves were used to detect non-specific amplification immediately after the real-time PCR run. Quantitative RT-PCR assays were performed on the diploid parents, the three F1 hybrids, and three lines of the corresponding S0 doubled hybrids and the following generations of selfing S1, S3 and S5; two biological replicates were tested (except for two F1 hybrids and two S0 lines for which only one plant was available), and at least two technical replicates were performed.

Quantitative RT-PCR data were analysed using the LinRegPCR software v.2012.x (Ruijter et al. 2009). For each sample, baseline correction was performed on the raw fluorescence data, allowing the determination of a window-of-linearity in order to use linear regression analysis to determine the PCR efficiency per sample from the slope of the regression line. The mean PCR efficiency per reaction and the Cq value were then used to calculate a starting concentration N0 per sample. The relative expression levels were obtained from the ratio between the N0 of the sRNA under survey and the normalization factor. In our study, the normalization factor corresponded to the mean N0 value of three additively expressed micro RNAs (namely miR159a, miR160a, miR166a) which were first verified for their highly stable expression in our samples using GeNORM (Vandesompele et al. 2002), as shown in Fig. S2.

Following the same procedure and on the same dilute cDNA templates, we also assayed transcript amounts for (i) mRNAs specific of transposable elements matching perfectly the 24-nt sRNAs under study and for (ii) transcripts retrieved by similarity search against *B. napus* meiosis RNAseq data (E. Jenczewski, unpublished), which matched

perfectly some of the 21-nt sRNAs of interest. A list of the different primers used for both sRNA and mRNA qRT-PCR experiments is given in Table S5 (only those producing exploitable results are provided). Only qRT-PCR data points for sRNAs and mRNAs with significant expression levels ($C_q \leq 35$) were retained for further statistical analyses when required.

Small RNA blot analysis

Small RNA blot analysis was performed as described in Lelandais-Brière et al. (2009) using oligonucleotide probes complementary to the sRNA sequences to be analysed, with an oligonucleotide probe complementary to U6 snRNA used to normalize RNA concentrations.

References

- Abi-Rached L., Jobin M.J., Kulkarni S., McWhinnie A., Dalva K. *et al.* 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334: 89-94.
- Ainouche M.L., and Jenczewski E. 2010. Focus on polyploidy. *New Phytologist* 186: 1-4.
- Albertin W, Balliau T, Brabant P, Chèvre A, Eber F, Malosse C, Thiellement H. 2006. Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics* 173: 1101-13.
- Axtell M.J. 2013. Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64: 137–59.
- Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP. 2012. Repeat associated small RNAs vary among parents and following hybridization in maize. *Proc Natl Acad Sci USA* 109: 10444-10449.

- Bomblies K. and Weigel D. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* 8: 382-393.
- Chen Z.J. 2010. Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science* 15: 57-71.
- Cifuentes M, Grandont L, Moore G, Chèvre A-M, Jenczewski E. 2010. Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol* 186: 29-36.
- Collares-Pereira M.J., Matos I., Morgado-Santos M., Coelho M.M. 2013. Natural pathways towards polyploidy in animals: the *Squalius alburnoides* fish complex as a model system to study genome size and genome reorganization in polyploids. *Cytogenetic and Genome Research* 140: 97-116.
- Comai 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6: 836-846.
- Côte FX, Galzi S, Folliot M, Lamagnère Y, Teycheney P-Y and Iskra-Caruana M-L. 2010. Micropropagation by tissue culture triggers differential expression of infectious endogenous *Banana streak virus* sequences (eBSV) present in the B genome of natural and synthetic interspecific banana plantains. *Molecular Plant Pathology* 11: 137-144.
- Dasmahapatra K.K. *et al.*, The *Heliconius* Genome Consortium, 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94-98.
- Fahlgren N., Sullivan C.M., Kasschau K.D., Chapman E.J., Cumbie J.S., Montgomery T.A., Gilbert S.D., Dasenko M., Backman T.W.H., Givan S.A., and Carrington J.C. 2009. Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA* 15: 992-1002.
- Feschotte C., and Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics* 13: 283-296.

- Finnegan E.J. and Matzke M.A. 2003. The small RNA world. *Journal of Cell Science* 116: 4689-4693.
- Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant Cell* 19: 3403-3417.
- Gaeta RT, Yoo SY, Pires JC, Doerge RW, Chen ZJ, Osborn TC. 2009. Analysis of gene expression in resynthesized *Brassica napus* allopolyploids using Arabidopsis 70mer oligo microarrays. *PLoS One* 4: e4760.
- Grandont L., Jenczewski E. and Lloyd A. 2013. Meiosis and its deviations in plants. *Cytogenetic and Genome Research* 140: 171-184.
- Green R.E. *et al.* 2010. A draft sequence of the Neandertal genome. *Science* 328: 710-722.
- Groszmann M., Greaves I.K., Albertyn Z.I., Scofield G.N., Peacock W.J., and Dennis E.S. 2011. Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci. USA* 108: 2617–2622.
- Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ. 2009. Small RNAs serve as genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences, USA* 106: 17835-17840.
- He G., Zhu X., Elling A.A., Chen L., Wang X., Guo L., Liang M., He H., Zhang H., Chen F., Qi Y., Chen R., and Deng X.-W. 2010. Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *Plant Cell* 22: 17-33.
- Hegarty MJ, Batstone T, Barker GL, Edwards KJ, Abbott RJ, Hiscock SJ. 2011. Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in *Senecio* (Asteraceae). *Molecular Ecology* 20: 105-113.

- Hegarty MJ, Hiscock SJ. 2008. Genomic clues to the evolutionary success of polyploid plants. *Current Biology* 18: R435-R444.
- Jenczewski E., Chèvre A.-M., and Alix K. 2013. Chromosomal and gene expression changes in *Brassica* allopolyploids. In: Polyploid and Hybrid Genomics. Z.J. Chen and J.A. Birchler (Eds). Wiley-Blackwell, USA, 384 pages, pp 171-186.
- Kenan-Eichler M., Leshkowitz D., Tal L., Noor E., Melamed-Bessudo C., Feldman M. and Levy A.A. 2011. *Genetics* 188:263-272.
- Książczyk T., Kovarik A., Eber F., Huteau V., Khaitova L. et al. 2011. Immediate unidirectional epigenetic reprogramming of NORs occurs independently of rDNA rearrangements in synthetic and natural forms of a polyploid species *Brassica napus*. *Chromosoma* 120: 557-571.
- Lelandais-Brière C., Naya L., Sallet E., Calenge F., Frugier F., Hartmann C., Gouzy J., Crespi M. 2009. Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* 21: 2780-2796.
- Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. 2006. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiology* 140: 336-348.
- Mable B. 2013. Polyploids and hybrids in changing environments: winners or losers in the struggle for adaptation? *Heredity* 110:95-96.
- Madlung A, Masuelli R.W., Watson B., Reynolds S.H., Davison J., Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* 129: 733-746.
- Madlung A, Tyagi AP, Watson B, Jiang HM, Kagochi T, Doerge RW, Martienssen R, Comai L. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant Journal* 41: 221-230.

- Madlung A. and Wendel J.F. 2013. Genetic and epigenetic aspects of polyploidy evolution in plants. *Cytogenetic and Genome Research* 140: 270-285.
- Mallet J. 2007. Hybrid speciation. *Nature* 446: 279-283.
- Mallory A.C., and Vaucheret H. 2006. Functions of microRNAs and related small RNAs in plants. *Nature Genetics* 38: S31-S36.
- Marí-Ordóñez A., Marchais A., Etcheverry M., Martin A., Colot V. and Voinnet O. 2013. Reconstructing *de novo* silencing of an active plant retrotransposon. *Nature Genetics* 45: 1029-1039.
- Marmagne A, Brabant P, Thiellement H, Alix K. 2010. Analysis of gene expression in resynthesized *Brassica napus* allotetraploids: transcriptional changes do not explain differential protein regulation. *New Phytologist* 186: 216-227.
- Martienssen RA. 2010. Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. *New Phytologist* 186: 46-53.
- Martinez de Alba A.E., Elvira-Matelo E., and Vaucheret H. 2013. Gene silencing in plants: a diversity of pathways. *Biochimica et Biophysica Acta* 1829: 1300-1308.
- Mirouze M. 2012. The small RNA-based odyssey of epigenetic information in plants: from cells to species. *DNA and Cell Biology* 31: 1650-1656.
- Moxon S., Schwach F., MacLean D., Dalmay T., Studholme D.J., Moulton V. 2008. A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 24: 2252–2253.
- Ng DW-K, Lu J, Chen ZJ. 2012. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr Opin Plant Biol* 15: 154-161.
- Ni Z., Kim E.D., Ha M., Lackey E., Liu J. et al. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457: 327-331.

- Pantaleo V., Szittyá G., Moxon S., Miozzi L., Moulton V., Dalmay T., Burgyan J. 2010. Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *The Plant Journal* 62: 960-976.
- Parisod C, Salmon A, Zerjal T, Tenailon M, Grandbastien M-A, Ainouche ML. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist* 184: 1003-1015.
- Romanel E., Silva T.F., Corrêa R.L., Farinelli L., Hawkins J.S., Schrago C.E.G., Vaslin M.F.S. 2012. Global alteration of microRNAs and transposon-derived small RNAs in cotton (*Gossypium hirsutum*) during Cotton leafroll dwarf polerovirus (CLRDV) infection. *Plant Molecular Biology* 80: 443-460.
- Ruijter J.M., Ramaker C., Hoogaars W.M.H., Karlen Y., Bakker O., van den Hoff M.J.B., Moorman A.F.M. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research* 37: e45.
- Salmon A., Ainouche M.L. and Wendel J.F. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* 14: 1163-1175.
- Sarilar V., Martinez Palacios P., Rousselet A., Ridel C., Falque M., Eber F., Chèvre A.-M., Joets J., Brabant P., and Alix K. 2013. Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. *New Phytologist* 198: 593-604.
- Seehausen O. 2004. Hybridization and adaptive radiation. *Trends in Ecology and Evolution*. 19: 198-207.
- Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *The Plant Cell* 13: 1749-1759.

- Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136: 461-472.
- Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre A-M. 2011. Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytologist* **191**: 884-894.
- Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ et al. 2010. The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytologist* **186**: 102-112.
- Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3(7): research0034.1-0034.11.
- Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* 136: 669–687.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O et al. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973-982.

Figures

Figure 1. Size distribution of small RNAs in three resynthesized *B. napus* allotetraploids (BoEMZ1₃₈, BoEMZ2₃₈, BoEMZ5₃₈ – in order) surveyed at two different generations of selfing (S1 and S5) in comparison with the diploid progenitors *B. oleracea* (HDEM) and *B. rapa* (Z1). (A) Size distribution of redundant sRNAs is provided with (B) size distribution of unique reads (illustrating the diversity of sRNAs which are different in sequences). (C) The diversity for reads with more than 35 RPM was analysed across the generations, for both the 21-nt (left) and the 24-nt (right) classes; the results indicated a higher diversity within the S1 generation with 50% and 41% of the 21- and 24-nt reads analysed respectively, which were S1-specific, suggesting the expression of specific sRNA populations in response to allopolyploidy. Only 8 % of the significantly expressed sRNAs were common to the three generations surveyed.

Figure 2. Distribution of the sRNAs annotated as microRNAs in the resynthesized *B. napus* allotetraploids surveyed at the two generations S1 and S5 in comparison with the diploid parents. Venn diagrams showing distributions of (A) the 34 known microRNA families as registered in miRBase v19 present in the sequencing data, and (B) the 274 putative individual microRNAs we identified according to hairpin prediction. For the know microRNA families, nearly all the miRNAs were found in the different sRNA libraries; only 4 unique reads specific of the diploid parents and 2 unique reads specific of the S5 generation were identified as generation-specific but they all displayed very low levels (concerned miRNAs indicated on the figure with corresponding raw read counts in brackets). For these analyses, the sequencing data obtained from each library were pooled by generation: diploid parents (pool of 5 libraries), S1 generation (3), S5 generation (3).

Figure 3. Normalized abundances of the sRNAs corresponding to sequences of transposable elements in the three resynthesized *B. napus* allotetraploids surveyed at the two generations S1 and S5 in comparison with the diploid progenitors, considering the TE class and order the TE-siRNAs originated from. Abundances of TE-siRNAs were estimated for each of the (A) 21-nt and (B) 24-nt class of sRNAs. Annotation of TE-siRNAs was performed according to the TE classification published in Wicker et al. 2007.

Figure 4. Analysis of the expression of 24-nt TIR-siRNA24_37 in the resynthesized *B. napus* allotetraploids in comparison with the diploid parents and the first hybrids F1. (A) The main features of this 24-nt siRNA putatively derived from a TIR transposon are provided, including the Illumina sequencing data that suggested a decrease of its abundance in the S1 generation. (B) The results obtained by qRT-PCR were subjected to an ANOVA and a Student's t-test analysis for comparison of means: the TIR-siRNA24_37 was significantly less expressed in the generations S0 and S1 in comparison to the parental and the S5 generations ($P < 0.001$), the S5 generation showing similar expression profiles as the expected mid-parent profile ($P \sim 0.08$). Relative expressions (N_0/N_{ref}) were estimated with N_{ref} corresponding to the mean expression level of miR159a, miR160a and miR166a (previously tested as shown in Fig. S2).

Figure 5. Validation of the expression levels estimated for the 21-nt siRNA21_04, by Northern blot analysis.

Bioinformatic estimates from the analysis of sequencing data predicted high levels of expression in the S1 generation only (767 RPM against no reads in the parental and the S5 generations). Such expression levels were validated by Northern blot analysis with hybridization signals at the size of 21-nt only in the three different S1 genotypes analysed (★). Additional signals at the size of 24-nt were identified in the two generations S1 and S5,

indicating co-existence (in S1) of distinct populations of siRNAs involved in the silencing of the same targets.

Figure 6. Mapping of all the collection of sRNAs excluding miRNAs on the two virus-like transcripts identified as corresponding to specific 21-nt siRNA populations over-represented in the S1 generation of the resynthesized *B. napus* allotetraploids.

Results of the mapping using all the (A) 21-nt sRNA reads and (B) the 24-nt reads. Numbers of reads mapping each of the viral sequences are given as reads per million for each generation (parents, S1, S5). The virus-derived siRNAs are mainly expressed in the S1 generation; the 21-nt virus-derived siRNAs potentially involved in PTGS of the reactivated endoviruses are significantly more abundant than the 24-nt virus-derived siRNAs that may mediate TGS. The two sets of sRNAs mapping the viral sequences were identified as totally independent.

Mapping was performed using Bowtie 2; results of alignments were viewed using Integrative Genomics Viewer, IGV.

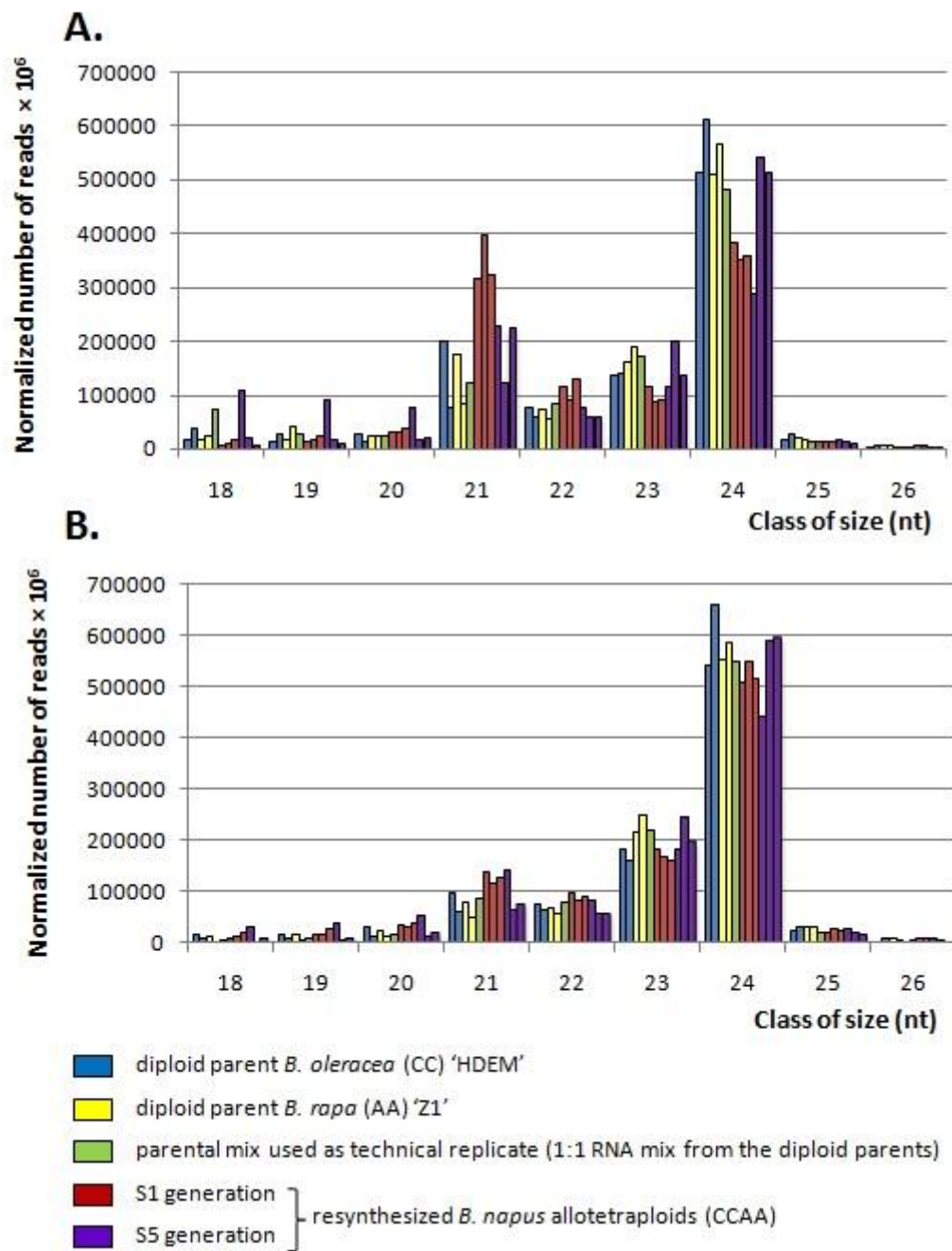


Figure 1 (1/2)

C.

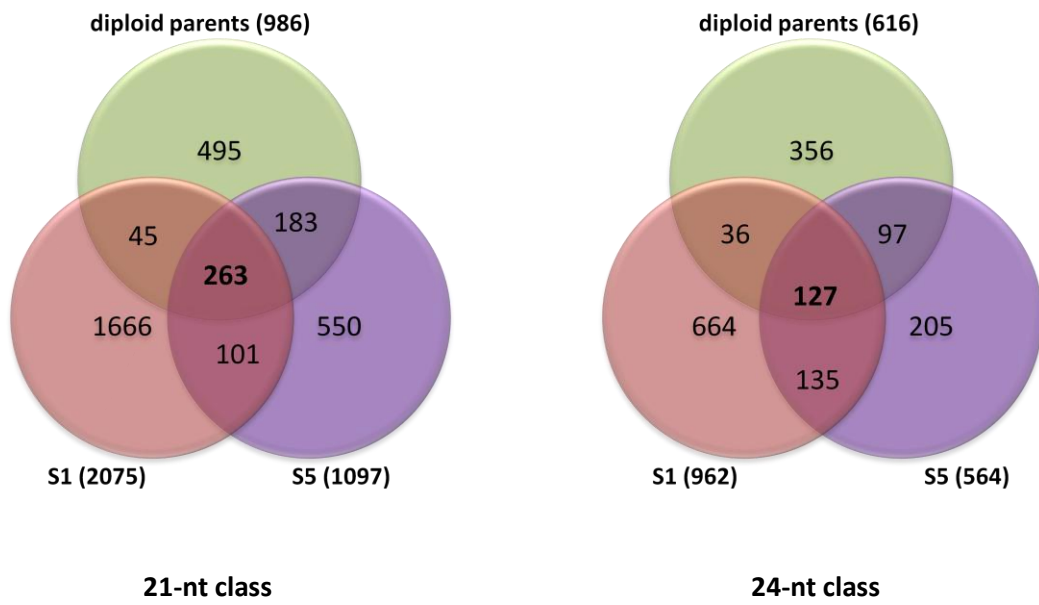
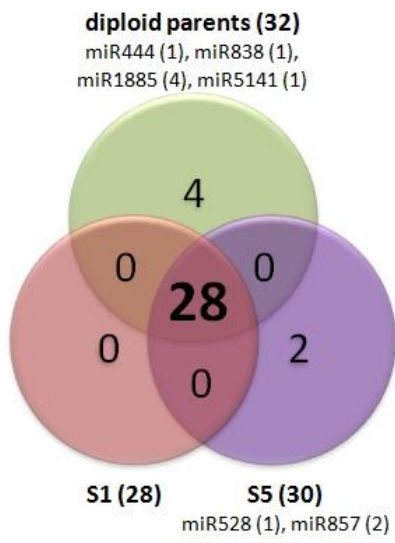


Figure 1 (2/2)

A. Known microRNA families



B. Predicted microRNAs

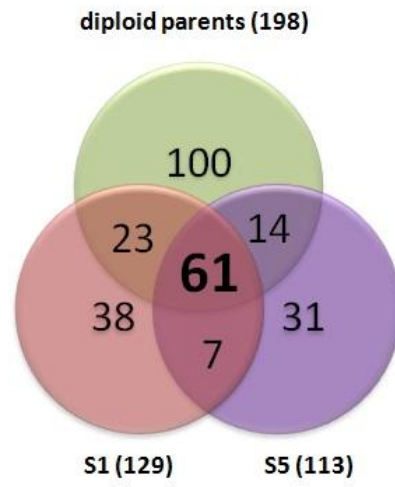


Figure 2

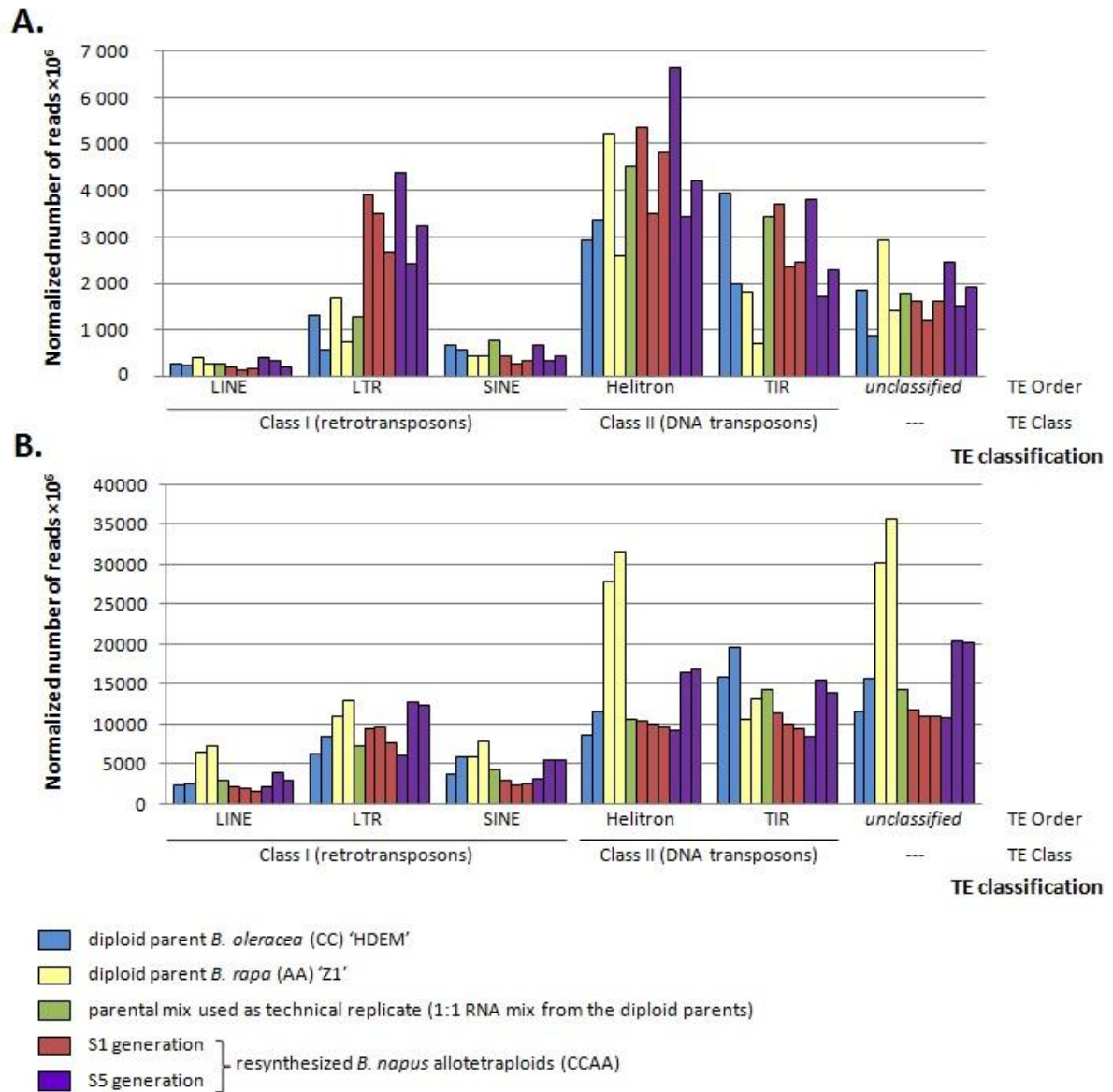


Figure 3

A.

24-nt TIR-siRNA24_37

Nucleic sequence	AGAGTTGAGAAGTGAGGTTTTGGG		
Sequencing data	Parents	S1	S5
(normalised number of reads $\times 10^6$)	182	47	88

B.

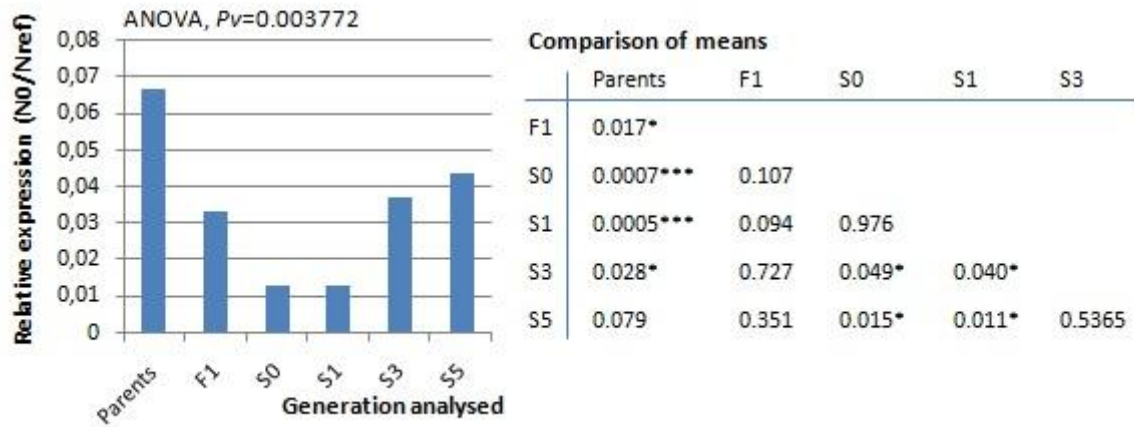


Figure 4

Bioinformatic sequencing data for siRNA21_04 (RPM)

Parents	S1	S5
0	767	0

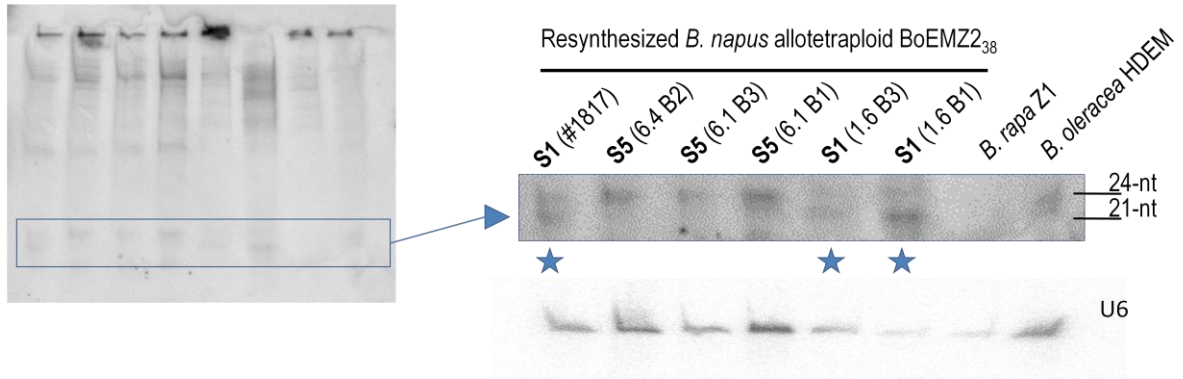


Figure 5

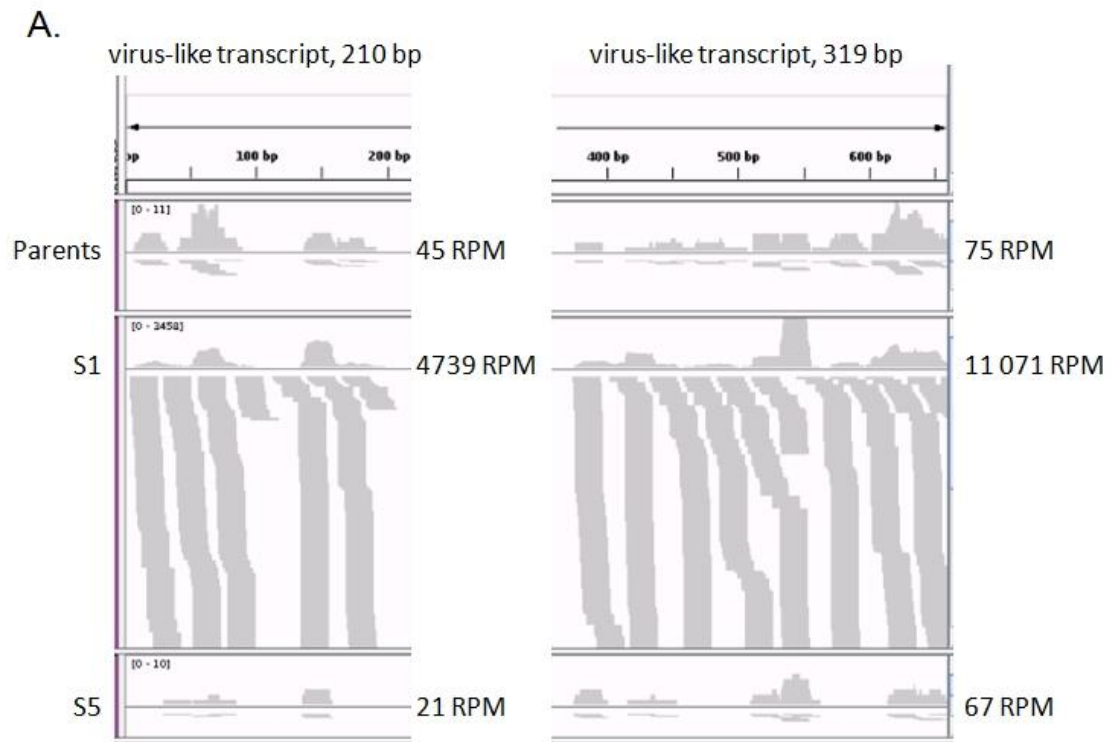


Figure 6

Supplemental Information

S1 Materials and Methods: Detailed bioinformatic analysis

A total of 20,397,093 sequence reads were obtained after Illumina sequencing of the eleven small RNA libraries which were specifically constructed for our study (Fig. S1).

- *Short-read quality control and pre-processing*

We analyzed the quality of the Illumina short-reads using the quality control tool for high throughput sequence data FastQC (Version 0.10.1). We removed the adaptors sequences (Sequence-File Pre-processing tool from the UEA sRNA toolkit; Moxon et al. 2008), and we removed the empty reads, low complexity reads and reads smaller than 18-nt and bigger than 26-nt by using the Filter tools from the UEA sRNA toolkit. In a subsequent step, reads similar to tRNAs and rRNAs were filtered out using the same Filter Tool (Moxon et al. 2008). To increase the sensitivity of this step, we have build our own *Brassica* rRNA and tRNA databases from de novo prediction using the RNAmmer 1.2 software (Lagesen et al. 2007) and the tRNAscan-SE 1.21 software (Lowe et Eddy 1997) (Schattner et al. 2005) on the genome sequence of *Brassica rapa* (*Brassica rapa* Genome Sequencing Project Consortium, 2011). We predicted 79 rRNA unique sequences and 1138 tRNAs sequences. These sequences have been added to the set of Rfam 10.1 sequence database (Griffiths-Jones et al. 2003) and matched against all the small RNAs libraries with no mismatches allowed. The relative amount of reads corresponding to rRNAs was larger for the 21nt read fraction compared to the 24nt read fraction. Only a low fraction of reads match to the tRNA sequences whatever the library considered.

- *Identification of microRNAs*

The short-read libraries were searched for known miRNA with the miRProf tool against the mirBase database (Version 19, database plant_mature, the options group-organisms, group-

family and collapse-match-groups were enabled). We performed de novo prediction of new potential miRNA using the miRCat tool (options were set accordingly to the documentation: -genomehits 16 --hit_dist 200 --maxgaps 3 --max_overlap_length 70 --max_percent_unpaired 60 --max_unique_hits 3 --maxsize 24 --min_abundance 5 --min_energy -25.0 --min_gc 10 --min_hairpin_len 75 --min_paired 17 --minsize 20 --percent_orientation 80 --pval 0.1 --window_length 100). Two predictions were performed on pre-processed short-reads by using the *B. rapa* and the *A. Thaliana* genome sequences.

The quantification of smallRNA was conducted as described in methods.

- *Identification of smallRNA similar to transposable element*

Sequences of small RNA remaining unannotated were compared to a home-made *Brassica* transposable element database (Sarilar et al., personal communication) using Blast (word size = 21 or 24) and only perfect matches were considered. A second comparison was performed against the Repbase database using the same Blast approach.

- *Alignment of the unidentified short-read on the B rapa, B napus and B. oleracea genome sequences*

Short-read which were not identified as miRNA or similar to TE sequences were aligned to the genome of 1) *B rapa*, 2) *B Oleracea* (GSS sequences) and 3) *B napus* assembly (under MTA; B Chalhoub, INRA, URGV Evry, France). The steps 2 and 3 were performed successively on short-reads that fail to map in the previous steps. Alignments were done using the Bowtie 2 software with no or one mismatch allowed (setting k = 10).

Table S1

TableS1. Size distribution of the redundant reads of the small RNA libraries without reads corresponding to rRNAs and tRNAs**RAW READ COUNTS**

nt	HDEM-1	HDEM-2	Z1-1	Z1-2	HDEM+Z1(1:1)	S1	S1-EMZ2	S1-EMZ5	S5	S5-EMZ2	S5-EMZ5
18	24113	2919	32234	55193	72673	5375	16203	29076	54585	34132	10375
19	21040	1946	30693	99660	26697	10350	26655	43407	46491	28206	12866
20	44901	1075	49513	55767	23214	22931	55455	67976	38578	30289	29308
21	316868	5633	355260	208399	121318	224295	694320	587791	116216	205930	285277
22	123939	4417	150840	135279	81960	81938	160431	235259	39279	100006	75714
23	219535	10298	328846	473390	169422	82501	150646	165459	58050	332906	175173
24	818685	45497	1043044	1420723	476931	270808	614231	649693	146370	904432	646344
25	25457	2017	43059	45469	14474	8893	21337	25029	9004	24712	14825
26	4268	488	12275	12647	2362	2134	7052	10668	2870	6350	4434
	1 598 806	74 290	2 045 764	2 506 527	989 051	709 225	1 746 330	1 814 358	511 443	1 666 963	1 254 316

14 917 073

NORMALIZATION PER MILLION READS (RPM)

nt	HDEM-1	HDEM-2	Z1-1	Z1-2	HDEM+Z1(1:1)	S1			S5		
18	15081,88	39291,96	15756,46	22019,71	73477,51	7578,70	9278,32	16025,50	106727,44	20475,56	8271,44
19	13159,82	26194,64	15003,20	39760,19	26992,54	14593,39	15263,44	23924,16	90901,63	16920,59	10257,38
20	28084,08	14470,32	24202,69	22248,71	23470,98	32332,48	31755,17	37465,59	75429,72	18170,17	23365,72
21	198190,40	75824,47	173656,39	83142,53	122661,02	316253,66	397588,08	323966,38	227231,58	123536,04	227436,31
22	77519,72	59456,19	73732,85	53970,69	82867,31	115531,74	91867,52	129665,15	76800,35	59992,93	60362,78
23	137311,84	138618,93	160744,84	188862,92	171297,54	116325,57	86264,34	91194,24	113502,38	199708,09	139656,20
24	512060,25	612424,28	509855,49	566809,37	482210,73	381836,51	351726,76	358084,24	286190,25	542562,73	515295,99
25	15922,51	27150,36	21047,88	18140,24	14634,23	12539,04	12218,19	13794,96	17605,09	14824,56	11819,19
26	2669,49	6568,85	6000,20	5045,63	2388,15	3008,92	4038,18	5879,77	5611,57	3809,32	3534,99
	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00	1000000,00

Table S1

PERCENTAGE
(%)

nt	HDEM-1	HDEM-2	Z1-1	Z1-2	HDEM+Z1(1:1)	S1-EMZ1	S1-EMZ2	S1-EMZ5	S5-EMZ1	S5-EMZ2	S5-EMZ5	mean fraction of the full library per size class (%)
18	1,5082	3,9292	1,5756	2,2020	7,3478	0,7579	0,9278	1,6026	10,6727	2,0476	0,8271	3,04
19	1,3160	2,6195	1,5003	3,9760	2,6993	1,4593	1,5263	2,3924	9,0902	1,6921	1,0257	2,66
20	2,8084	1,4470	2,4203	2,2249	2,3471	3,2332	3,1755	3,7466	7,5430	1,8170	2,3366	3,01
21	19,8190	7,5824	17,3656	8,3143	12,2661	31,6254	39,7588	32,3966	22,7232	12,3536	22,7436	20,63
22	7,7520	5,9456	7,3733	5,3971	8,2867	11,5532	9,1868	12,9665	7,6800	5,9993	6,0363	8,02
23	13,7312	13,8619	16,0745	18,8863	17,1298	11,6326	8,6264	9,1194	11,3502	19,9708	13,9656	14,03
24	51,2060	61,2424	50,9855	56,6809	48,2211	38,1837	35,1727	35,8084	28,6190	54,2563	51,5296	46,54
25	1,5923	2,7150	2,1048	1,8140	1,4634	1,2539	1,2218	1,3795	1,7605	1,4825	1,1819	1,63
26	0,2669	0,6569	0,6000	0,5046	0,2388	0,3009	0,4038	0,5880	0,5612	0,3809	0,3535	0,44
	100,0000	100,0000	100,0000	100,0000	100,0000	100,0000	100,0000	100,0000	100,0000	100,0000	100,0000	100,00

Table S2

Table S2. List of the different microRNA families registered in miRBase and retrieved from our sequencing data.

One colour corresponds to one known microRNA family; a total of 34 microRNA families were thus retrieved from our database.

Results obtained with miRProf				Results obtained with the locally developed script					
matching miRNAs	PARENTS	S1	S5	vs	sequences	Parents (5)	S1 (3)	S5 (3)	ID-miR
miR1140(0)	149	27	8		ACAGCCTAAACCAATCGGAGC	149	27	8	miR1140
miR156(0)	158	73	151		TGACAGAAGAGAGTGAGCACA	104	33	103	miR156a
miR156(0); miR156-5p(0); miR157(0)	17816	5127	35841		CTGACAGAAGATAGAGAGCAC	1	1	8	miR156b
miR156*(0)	62	39	18		TTGACAGAAGAAAGAGAGCAC	50	36	38	miR156c
miR157*(0)	15	19	13		GCTTACTCTCTCTGTCCACC	42	4	1	miR156e*
	18051	5258	36023		GCTCACTCTCTATCTGTCCACC	10	6	0	miR156f*
	23+	42+	28+		TTGACAGAAGATAGAGGGCAC	3	3	2	miR156g
					TTGACAGAAGAGAGTGAGCAC	4219	956	1315	miR156k
					TTGACAGAAGATAGAGAGCAC	17816	5127	35841	miR157a
					GCTCTCTATACTTCTGTCCACC	2	6	2	miR157c*
						22247	6172	37310	
miR159(0)	734	1555	603		TTTGGACTGAAGGGAGCTCTA	60	147	53	miR159
miR159(0); miR159.1(0)	221865	452424	162920		TTTGGATTGAAGGGAGCTCTA	221782	452281	162869	miR159a
miR319(0)	15169	17413	5451		TTTGGATTGAAGGGAGCTCTG	83	143	51	miR159a.1
miR319*(0)	119	46	9		TTTGGATTGAAGGGAGCTCTT	446	976	354	miR159b
	237887	471438	168983		TTTGGATTGAAGGGAGCTCCT	2	1	2	miR159c
					CTTGGATTGAAGGGAGCTCTA	226	431	194	miR159f
					TTGGACTIONGAAGGGAGCTCCCT	14949	17190	5353	miR319a
					AGAGCTTCCTTGAGTCCATTC	113	43	9	miR319a*
					CTTGGACTGAAGGGAGCTCCT	2	1	0	miR319b
					TTGGACTIONGAAGGGTGTCCCT	6	4	3	miR319b
					AGAGCTTTCTTCGGTCCACTC	6	1	0	miR319b*
					CTTGGACTGAAGGGAGCTCCC	79	70	19	miR319c
					TTGGACTIONGAAGGGAGCTCCTT	90	93	52	miR319c
					GGAGATTCTTTCAGTCCAGTC	0	2	0	miR319c*

Table S2

				TTTGGACTGAAGGGAGCTCCT	40	53	24	miR319e
				TTGGACTGAAGGGAGCTCCCA	5	3	0	miR319g
					237889	471439	168983	
miR160(0)	1306	147	461	TGCCTGGCTCCCTGTATGCCA	1305	147	461	miR160a
miR160*(0)	489	57	115	GCGTATGAGGAGCCATGCATA	486	29	81	miR160a*
	1795	204	576	TGCCTGGCTCCCTGTATGCCG	1	0	0	miR160e
	3+	28+	34+		1792	176	542	
miR161(0)	27	2	17	TCAATGCACTGAAAGTGACTA	27	2	17	miR161
miR162(0)	1809	966	710	TCGATAAACCTCTGCATCCAG	1809	965	710	miR162a
miR162*(0)	16	34	20	GGAGGCAGCGGTTTCATCGATC	15	30	14	miR162a*
	1825	1000	730	TCGATAAGCCTCTGCATCCAG	0	1	0	miR162b
	1+	4+	6+		1824	996	724	
miR164(0)	18948	6037	15700	TGGAGAAGCAGGGCACGTGCA	17885	5134	13290	miR164a
				TGGAGAAGCAGGGCACGTGCG	904	830	2109	miR164c
				TGGAGAAGCAGGGTACGTGCA	17	1	1	miR164c
				TGGAGAAGCAGGGCACGTGCT	141	71	300	miR164d
				TGGAGAAGCAGGGCACGTGTG	1	1	0	miR164h
					18948	6037	15700	
miR165(0)	234	146	126	TCGGACCAGGCTTCATCCCC	234	146	126	miR165a
miR166(0)	49509	15552	24034	TCGGACCAGGCTTCATCCCC	49469	15538	24020	miR166a
miR166*(0)	3543	4351	2735	GGAATGTTGTCTGGCTCGGGG	12	5	7	miR166a*
	53286	20049	26895	TCGGACCAGGCTTCATTCTA	1	0	0	miR166b
				GGAATGTTGTCTGGCTCGAGG	34	22	0	miR166b*
				GGAATGTTGTCTGGCTCGAGG	3476	4288	2720	miR166c*
				CTCGACCAGGCTTCATTCCC	4	0	0	miR166e
				TCGAACCAGGCTTCATCCCC	10	3	2	miR166e
				GGAATGTTGTCTGGCACGAGG	19	33	7	miR166e*
				TCGGACCAGGCTTCATTCTC	9	3	4	miR166g
				GGAATGTTGTTGGCTCGAGG	2	3	1	miR166g*

Table S2

				TCCGACCAGGCTTCATTCCC	1	0	0	miR166j
				TCGGACCAGGCTTCATTCCCT	15	6	8	miR166m
				TCGGACCAGGCTTCATTCCCT	0	2	0	miR166n
					53286	20049	26895	
miR167(0)	116	73	91	TGAAGCTGCCAGCATGATCTA	104	72	91	miR167a
miR167(0); miR167-5p(0)	33	6	2	TAAGTGCCAGCATGATCTTG	8	1	0	miR167c
	149	79	93	TGAAGCTGCCAGCATGATCTC	1	0	0	miR167c
				TGAAGCTGCCAGCATGATCTG	3	0	0	miR167d
				TGAAGCTGCCAGCATGATCTT	33	6	2	miR167f
					149	79	93	
miR168(0)	22397	7941	16500	TCGCTTGGTGCAGATCGGGAC	1	3	2	miR168a
miR168(0); miR168-5p(0)	1	3	2	TCGCTTGGTGCAGGTGCGGAA	22397	7941	16500	miR168a
miR168*(0)	205	158	185	CCCGCCTTGCATCAACTGAAT	205	158	185	miR168a*
	22603	8102	16687		22603	8102	16687	
miR169(0)	3782	446	439	CAGCCAAGGATGACTTGCCGA	343	12	18	miR169a
miR169*(0)	87	25	11	CAGCCAAGGATGACTTGCCGG	1265	290	101	miR169b
	3869	471	450	TGAGCCAAGGATGACTTGCCG	205	3	18	miR169d
				GCAAGTTGACCTTGCTCTGT	74	5	9	miR169d*
				GCAAGTTGACCTTGCTCTGC	2	0	0	miR169f*
	15+	20+	2+	TAGCCAAGGATGACTTGCCCTA	134	2	1	miR169f.1
				TAGCCAAGGATGACTTGCCCTG	1045	32	18	miR169h
				TGAGCCAAGGATGACTTGCCG	782	107	283	miR169i
				GAGCCAAGGATGACTTGCCGT	1	0	0	miR169l
				GAGCCAAGGATGACTTGCCGG	2	0	0	miR169m
				TCAGCCAAGGATGACTTGCCG	1	0	0	miR169s
					3854	451	448	
miR170*(0); miR171*(0)	38	139	23	TATTGGCCTGGTTCACTCAGA	37	139	23	miR170*
miR171(0)	1291	481	398	TGATTGAGCCGCGCCAATATC	250	40	164	miR171a
miR171(0); miR171*(0)	0	2	0	TTGAGCCGTGCCAATATCACT	1	0	0	miR171a

Table S2

miR171*(0)	41	5	51	TGATTGAGCCGCGTCAATATC	38	25	33	miR171b
	1370	627	472	TGATTGAGCCGTGCCAATATC	0	0	2	miR171b
				TTGAGCCGTGCCAATATCACG	1002	416	198	miR171b
				AGATATTAGTGCGGTTCAATC	41	5	51	miR171b*
				AGATATTGGTGCGGTTCAATC	1	2	0	miR171c*
				TTGAGCCGTGCCAATATCACA	0	0	1	miR171f
					1370	627	472	
miR172(0)	24285	29226	5984	AGAATCCTGATGATGCTGCAG	8	13	0	miR172
miR172*(0)	4110	1607	191	AGAATCTTGATGATGCTGCAT	1643	1206	1969	miR172a
	28395	30833	6175	GTGGCATCATCAAGATTCACA	1	1	0	miR172a*
				AGAATCTTGATGATGCGGCAA	1	0	0	miR172b
				GCAGCACCATCAAGATTCACA	3047	1181	75	miR172b*
	19+	19+	8+	GCAGCACCATTAAGATTCACA	1057	335	108	miR172b*
				AGAATCTTGATGATGCTGCAG	22456	27737	3339	miR172c
				TGAATCTTGATGATGCTGCAC	1	0	0	miR172c
				GGAGCATCATCAAGATTCACA	5	90	8	miR172c*
				GGAATCTTGATGATGCTGCAT	153	246	666	miR172e
				GGAATCTTGATGATGCTGCAG	4	5	2	miR172g
					28376	30814	6167	
miR2111(0)	0	0	15	TAATCTGCATCCTGAGGTTTA	0	0	15	miR2111a
miR2111*(0)	10	2	9	GTCTCGGGATGCGGATTACC	0	1	1	miR2111a*
	10	2	24	ATCCTCGGGATACAGATTACC	2	1	7	miR2111b*
				ATCCTCGGGATACGGATTACC	8	0	1	miR2111b*
					10	2	24	
miR390(0)	1	0	1	AAGCTCAGGAGGGATAGCGCC	3317	458	792	miR390a
miR390(0); miR390-5p(0)	3317	458	792	AAGCTCAGGAGGGATAGCACC	1	0	1	miR390b
	3318	458	793		3318	458	793	
miR391(0)	239	551	3865	TTCGAGGAGAGATAGCGCCA	239	551	3865	miR391
miR391*(0)	383	15	38	ACGGTATCTCTCCTACGTAGC	383	15	38	miR391*

Table S2

	622	566	3903		622	566	3903	
miR393(0)	137	7	3	TCCAAAGGGATCGCATTGATC	137	7	3	miR393
miR395(0)	3020	335	448	CTGAAGTGTGGGGGAACTC	312	39	12	miR395a
miR395*(0)	13	5	0	CTGAAGTGTGGGGGGACTC	2682	295	432	miR395b
	3033	340	448	GTTCCCTTTAACGCTTCATTG	9	4	0	miR395b*
				CTGAAGTGTGGGGGGACTT	2	0	2	miR395c
				GTTCTCTGAGCACTTCATTG	4	1	0	miR395e*
				CTGAAGTGTGGAGGAACTC	24	1	2	miR395i
					3033	340	448	
miR396(0)	1134	645	266	TTCCACAGCTTTCTTGAACTA	1	3	0	miR396a
miR396*(0)	2460	1381	468	TTCCACAGCTTTCTTGAAGT	900	394	166	miR396a
	3594	2026	734	GTTCAATAAAGCTGTGGGAAG	1200	1297	244	miR396a*
				TTCCACAGCTTTCTTGAAGT	233	248	100	miR396b
				GCTCAAGAAAGCTGTGGGAAA	1258	84	224	miR396b*
				GTTCAATAAAGCTGTGGGAAA	2	0	0	miR396b*
					3594	2026	734	
miR397(0)	38	7	4	TCATTGAGTGCAGCGTTGATG	38	7	3	miR397a
miR398(0)	2834	43	23	TGTGTTCTCAGGTCACCCCTT	25	0	0	miR398a
miR398*(0)	32	0	1	GGAGTGCATGAGAACACGGA	31	0	1	miR398a*
	2904	50	28	TGTGTTCTCAGGTCACCCCTG	2803	43	23	miR398b
				TGTGTTCTCAGGTCGCCCTG	6	0	0	miR398b
			1+	GGGTCGACATGAGAACATATG	1	0	0	miR398c*
					2904	50	27	
miR399(0)	197	30	106	TGCCAAAGGAGAGTTGCCCTG	195	22	100	miR399b
miR399*(0)	0	0	1+	TGCCAAAGGAGATTTGCCCGG	2	8	6	miR399f
miR400(0)	322	49	22	TATGAGAGTATTATAAGTCAC	322	49	22	miR400
miR403(0)	1638	1510	990	TTAGATTCACGCACAAACTCG	1638	1510	990	miR403
				TTAGATTCACGCACAAACTTG	0	0	1	miR403a
miR408(0)	158	18	26	ATGCACTGCCTCTCCCTGGC	4483	521	529	miR408

Table S2

miR408(0); miR408-3p(0)	4483	521	529	CTGCACTGCCTCTCCCTGGC	41	1	1	miR408
miR408*(0)	765	130	464	ACAGGGAACATGCAGAGCATG	4	0	3	miR408*
	5406	669	1019	CAGGGAACAAGCAGAGCATGG	765	130	464	miR408*
				TGCACTGCCTCTCCCTGGCT	117	17	25	miR408b
					5410	669	1022	
miR824(0)	140	84	117	TAGACCATTTGTGAGAAGGGA	140	84	117	miR824
miR824*(0)	229	33	100	CCTTCTCATCGATGGTCTAGA	229	33	100	miR824*
	369	117	217		369	117	217	
miR845(0)	220	21	3	CGGCTCTGATACCAATTGATG	220	21	3	miR845a
miR858(0)	5	2	0	TTTCGTTGTCTGTTGACCTT	5	2	0	miR858
				TTCGTTGTCTGTTGACCTTG	7	18	1	miR858b
					12	20	1	
microRNAs avec une très faible representation dans les banques (moins de 5 reads au moins dans un banque)								
miR444(0); miR444.2(0)	1	0	0	TGCAGTTGTTGTCTCAAGCTT	1	0	0	miR444b.2
miR528(0)	0	0	1	TGGAAGGGGCATGCAGAGGAG	0	0	1	miR528
miR838(0)	1	0	0	TTTTCTTCTTCTTGCACA	1	0	0	miR838
miR857(0)	0	0	2	TTTTGTATGTTGAAGGTGTAT	0	0	2	miR857
miR1885(0)	4	0	0	non trouvé	0	0	0	
miR5141(0)	1	0	0	non trouvé	0	0	0	
Reads totaux identifiés comme étant des microRNAs connus	410136	549974	281103		414302	550836	282343	
	4166-	862-	1240-		4166+	862+	1240+	6268

Table S3A

Table S3. List of the microRNAs predicted from analysis of the *B. rapa* genome.

Location on *B. rapa* chromosomes of the predicted *MIR* genes are indicated, as well as identification of putative miRNA* from our sequencing sRNA data.

Sequence	MIR genes	Chromosome	miRNA*
AAAGCTCAGGAGGGATAGCGC	6	A02 : A03 : A03** : A04** : A05** : A10	CGCTGTCCATCCTGAGTTTCA GCTGTCCATCCTGAGTTTCAT CGCTATCCATCCTGAGTTCTA
AAATTATGGAGTGGGGATGGA	2	A02 : A03	NO
AACAGCCTAAACCAATCGGAG	1	A03	GCTCCGATTGGCTTTAGGCTG
AACAGCCTAAACCAATCGGAG	1	A03	GCTCCGATTGGCTTTAGGCTG
AAGAGGTTCTGGACACGTAGG	1	A05	NO
AAGCGAGGGGTTGGACATCTG	1	A09	NO
AAGCGGTGCGGGACAAGCGGT	1	A06	NO
ACAGAAGATAGAGAGCACAGA	1	A07	NO
ACAGGGAACAAGCAGAGCATG	1	A05	NO
ACAGTATAGAACCAGAGAGAC	1	A09	NO
ACATCTGTTAGGGTTTCTTTG	1	A06	NO
ACCAGAGGTCATTAGCTCAAC	1	A08	NO
ACGGTCTTCGGTCTCGGGCGC	1	A01	GTGGTCCAGACTAGAGACCGA
ACGTGATAAGCCTCTGAAGAA	1	A08	NO
ACGTTGAAGGGTCGTCGAGAG	1	A02	NO
ACTAGTCTGTTGGGCTCGGGC	1	A06	NO
ACTCACGGGTCGACTCGGCC	3	A03* : A08* : Scaffold000111*	CGAGTTGGCGAGTCGCCGGCG
ACTCGAACTCGTATGGCGGA	2	A03 : A08	NO
ACTCGTTACGGATCGGTGCT	2	Scaffold000155 : Scaffold013212	NO
ACTGTAACTGCGGAGCGGGG	1	A07	TACGGTGCGGTTTTAACAGTT
ACTGTAACTGCGGACCGGGG	1	A08	NO
ACTGTAACTGCGGGGCGGGG	4	A01 : A01 : A03 : A05*	CTGCGGTGCGGTTTTAACAGT
ACTGTAACTGTGGGACGGGG	1	A06	NO
AGAGGGAGATCGTCGAGGTGA	1	Scaffold000419	NO
AGAGGGATCAGTAAAAGATGA	1	A09	ATCTTTAGCTGGTCTCTCTCC TCTTTAGCTGGTCTCTCTCC
AGAGGTAGAAATAGAGGTGGG	1	A05	NO
AGAGTTGTTAAAGGGTGAGGT	1	Scaffold000378	NO
AGAGTTTCCTTAAGTCCATTC	1	A08	NO
AGATGGATCATCCGGATGACT	2	A06 : Scaffold000167	NO
AGATTCGGAAGCCTTTGA	1	A03	NO
AGCCAAGGATGACTTGCCGGA	3	A02 : A06** : A09	CTGGCAAGTTGTCTTCGGCT GGCAAGTTGTCTTCGGCTAC
AGCTGCTAAGCTATGGATCCC	2	A02 : A07*	CTTGCATATCTTAGGAGCTTT
AGCTTCGATAAGAGACGGTT	1	A09	NO
AGGGAGCTTTCTTCGGTCCAC	1	A04	NO
AGGGGAATGTTGTCTGGCACG	1	A02	GTCGGACCAGGCTTCATTCCC GGACCAGGCTTCATTCCCCTC GACCAGGCTTCATTCCCCTCA
AGGGTTGGGTTAAACAGGGC	1	A10	NO
AGGTGGTCATACTGTCAACAG	1	A07	TTTGGCATTCTGTCCACCTCC TTGGCATTCTGTCCACCTCT
AGTCTGTTCCAAGCTCTGGC	1	A01	NO
AGTGAAGGATCGGTTTGGTGA	1	A10	NO
AGTGATCCGTTTGTGTTAGGC	1	A02	NO
AGTGGTGTCTTGAGGAGTTGG	1	A03	NO
AGTGTTGGGTTGGAGAAGGTT	2	A03 : A05	NO
ATAAATCCCAAGCATCATCCA	1	A09	ATGATTCTGGGATTTATCTT

Table S3A

ATAAATTTCCGGGTTTTCCGGGC	1	A09	NO
ATAACTTGGTTTTGCTCCTAC	1	A01	NO
ATAGAGTGAACCATTAGAGCA	3	A01 : A03 : A07	NO
ATAGAGTGACGTTGGAGATGC	1	A05	NO
ATAGGGTTAGGGCTGGTACTA	2	A07 : A09	NO
ATCAAAGGACGAAGGACGAGC	1	A05	NO
ATCAACGTTGGCTCAATTATG	1	A10	NO
ATCAGAGAAGGGTCATTGTGG	3	A01: Scaffold000104 : Scaffold000155	NO
ATCATGCGATCTCATTGGATT	1	A04	NO
ATCATGCGATCTCTTCGGATT	1	A07	NO
ATCATGGATCTGGACCGCGGG	1	A09	NO
ATCCTCGGGACACAGATTACC	1	A10	NO
ATCTTCGGCTAAGAGACGGTT	1	A05	NO
ATGAAGAGCGGCATGAACGAA	5	A04 : A07 : A10 : Scaffold000381 : Scaffold000607	NO
ATGAAGGTGTAAGAAGGTAGC	1	A04	NO
ATGAAGTGAGGATGATAGGGA	1	A01	NO
ATGATCAAGGAAGGGGGCACG	1	A01	NO
ATGATTTGTAICTTTCCATA	1	A05	NO
ATGCCTGGCTCCTGTATGCC	5	A01 : A03* : A05 : A06** : A09	CGTACAGAGTAGTCAAGCATA GGCGTACAAGGAGTCAAGCAT CGTCAGAGTAGTCAAGCATA GCGTATGAGGAGCCAAGCATA GCGTACAAGGAGTCAAGCATG CGTACAAGGAGTCAAGCATGA GCGTACGAGGAGCCAAGCATG CGTACGAGGAGCCAAGCATGA
ATGGAGTAGAAAATGCAGTCG	1	A02	NO
ATGGGGGTTTTAGGGTGGAG	1	A09	NO
ATGGGTTGATCGGATTGGACA	1	A06	NO
ATGGTGGATCGTAACAGATGA	2	A03 : A06	NO
ATGGTGGTTGCGCATTACGG	2	Scaffold000170 : Scaffold000212	NO
ATGTAATTGCGGGTTCTAAG	4	A01 : A02 : A03 : A07	TTAGAACCCTGCAATTACATA TAGAACCTGCAATTACATAC AGAACCCTGCAATTACATACG
ATGTTCTGGTGGTGCTGGAAA	1	A03	TTTTCCAGCACCACCAGAAC TTCCAGCACCACCAGAACACG TCCAGCACCACCAGAACACGG
ATTATGGAGTGGGGATGGAGA	7	A01 : A01 : A02 : A04 : A06 : A09 : A10	NO
ATTCATGTGGGGGATTGTTGG	3	A07 : A09 : A09	NO
ATTGAATGTCATGCTCTGGG	1	Scaffold000104	NO
ATTGCAGTCTCTTAGGTTGG	6	A01 : A05 : A07 : A09 : Scaffold000211 : Scaffold000616	NO
ATTTAAGTCGTGGAAGTC	1	Scaffold000155	NO
ATTTATCTAAGGTTTAGGGT	7	A01 : A02 : A04 : A04 : A06 : A09 : A09	NO
ATTTACATTTGGGAGTCGTC	1	A08	NO
ATTTGAGATTAGGGATATTG	1	Scaffold000237	NO
CAAGAGTGGTTGAAAAGATTG	1	Scaffold000441	NO
CACGTGTCTACTACTCCAAC	3	A02** : A06** : A09**	TGCTGGAGAAGCAGGGCACGT CTGGAGAAGCAGGGCACGTGC
CAGGATTTAAAGTTGAGGGTT	2	A02 : A09	NO
CAGGTTTCGGCGTCAGCGGCA	4	A01 : A02 : A08 : A09	NO
CATGTGCCATCTTCCCATC	1	A02	AGATGGAGAAGCAGGGCACGT
CCCGCCTGTATCAAGTGAAT	1	A06	CTCGCTTGGTGCAGGTCGGGA CGCTTGGTGCAGGTCGGGAAC
CCCTGTTCTGTTGGAGGTCGC	3	A01 : A03 : A09	NO
CGACCACGAAACGAACAGGGC	1	A03	NO
CGACCCGGTGAGTCAACTCGG	1	A07	NO
CGAGAGTGGTTGAAAAGATTG	1	A05	NO

Table S3A

CGCTGTCCATCCTGAGTTTCA	3	A03* : A04* : A05*	AAAGCTCAGGAGGGATAGCGC
CGCTTGGTGCAGGTCGGAAC	1	A06	CCCGCTTGTATCAAGTGAAT
CGGATCCCGCTTGTATCAAG	1	A06	CGCTTGGTGCAGGTCGGAAC GCTTGGTGCAGGTCGGAAC CTTGGTGCAGGTCGGAAC TGGTGCAGGTCGGAAC
CGGGCAATGAACATTCGGCG	1	Scaffold025992	NO
CGTTCAATAAAGCTGTGGGA	1	A09	TCCACAGCTTTCTGAACTGC CCACAGCTTTCTGAACTGCA CTCCACAGCTTTCTGAACT
CGTACAGAGTAGTCAAGCATG	1	A03	GTGCTGGCTCCCTGTATGCC
CGTTGGTACTGTGAGAGGGTG	2	A03 : A03	NO
CTCAAGAAAGCTGTGGAAAA	1	A08	NO
CTCGGATTCGCTTGGTGCAGG	2	A08 : A09	NO
CTGACAGAAGAGAGTGAGCAC	6	A01 : A01** : A03 : A03 : A04 : A09	TGCTCACGGCTTTCTGTCA GCTCACGGCTTTCTGTCA
CTGGAGAAGCAGGGCACGTGC	1	A09	CACGTGTTCTACTACTCCAAC
CTGGATGACTTAGTAGAAGAC	1	Scaffold000196	NO
CTTGCATATCTTAGGAGCTTT	1	A07	AGTGCTAAGCTATGGATCCC
CTTTGGATTGAAGGGAGCTCT	3	A02* : A07** : A07	GCTCCTTAAAGTTCAAACGAG GGGCTCCTTATAGTTCAAACG GCTCCTTATAGTTCAAACGAG GAGTCCTTAAAGTTCAAACG
GAAGGGCAAGAGGGAGATCGT	1	Scaffold000155	NO
GAAGTGATTTTGTGAAGAGGT	1	A02	NO
GAAGTTAGGGGCAAAACGGTC	1	Scaffold002680	NO
GAATATAAGAACCCTGACTCTT	1	A09	NO
GAATATAAGAAGCTGCTCTTT	1	A01	NO
GAATATAAGAATCCGTCTCTT	1	A04	NO
GAATCTTGATGATGCTGCATC	8	A02 : A02* : A03 : A03* : A03 : A07 : A10* : A10	GCAGCATCATCAAGATTCACA GCAGCATCATTAAAGATTCACA GCAGCATCATTAAAGATTCACA
GAATGGTGACCAGGGAACGGC	1	A04	NO
GAATGTTGTCTGGCTCGAGGA	1	A05	TCGTCGGACCAGGCTTCATT GTCGGACCAGGCTTCATTCCC TGTCGTCGGACCAGGCTTCAT
GAATGTTGTTGGCTCGAAGA	1	A04	GTCGGACCAGGCTTCATTCCC
GACGAATTAGCGAGGAAACGC	10	A01 : A01 : A01 : A03 : A05 : A07 : A08 : A09 : A09 : A09	NO
GAGGAACTCTATTATAAAGGC	1	A10	NO
GAGTTGGAATTGTAGAGGTGC	1	A06	NO
GATATTAGGGAGTTGGGACAC	1	A01	NO
GATCCTCTGAACACTTCATTG	1	A07	NO
GATGACTTAGTAGAAGACTTA	1	Scaffold000196	NO
GATGATGCTTGGGATTTATTT	1	A06	GATAAATCCCAAGCATCATCC ATAAATCCCAAGCATCATCCA
GATTGGACTGAAGGGAGCTCC	1	A03	GAGCTTCCTTTAGTCCATTCA
GCAAGTAGACTTTGGCTCTGT	1	A08	ATTGAGCCAAAGATGACTTGC TTGAGCCAAAGATGACTTGC
GCAAGTAGACTTTGGCTCTGT	1	A08	GAGCCAAAGATGACTTGCCGA
GCAAGTCGACTTTGGCTCTGT	1	A08	NO
GCAAGTTGACTTTGGCTCTGT	5	A06 : A06* : A06* : A08* : A08	ATTGAGCCAAAGATGACTTGC TGAGCCAAAGATGACTTGTGC AGTTGAGCCAAAGATGACTTGC TTGAGCCAAAGATGACTTGC GAGCCAAAGATGACTTGCCGA

Table S3A

GCAAGTTGATTTGGCTCTGT	1	A08	ATTGAGCCAAAGATGACTTGC TTGAGCCAAAGATGACTTGCC GAGCCAAAGATGACTTGCCGA
GCAGCATCATTAAGATTCACA	2	A03** : A10**	TGGGAATCTTGATGATGCTGC GAATCTTGATGATGCTGCATC GGGAATCTTGATGATGCTGCA
GCATCATCATCAAGATTCAGA	1	A09	TGAGAATCTTGATGATGCTGC GAGAATCTTGATGATGCTGCA GAATCTTGATGATGCTGCAGC
GCGTACAAGGAGTCAAGCATG	2	A06 : A09	ATGCCTGGCTCCCTGTATGCC
GCGTACGAGGAGCCAAGCATG	1	A09	ATGCCTGGCTCCCTGTATGCC
GCGTATGAGGAGCCAAGCATA	1	A03	TATGCCTGGCTCCCTGTATGC ATGCCTGGCTCCCTGTATGCC
GCTAATGGACATGGGCTCGCC	1	A05	NO
GCTATGAGATCAAAGGGTTGT	1	A07	NO
GCTCACGGCTCTTCTGTCAG	1	A01	ACTGACAGAAGAGAGTGAGCA CTGACAGAAGAGAGTGAGCAC
GCTCACTGCTCTATCTGTCAG	2	A01* : A03*	CTGACAGAAGAGAGTGAGCAC
GCTCACTGCTCTTCTGTCAG	2	A04** : A09**	ACTGACAGAAGAGAGTGAGCA CTGACAGAAGAGAGTGAGCAC
GCTCTCTAGCCTTCTGTCATC	1	A02	NO
GCTCTCTAGTCTTCTGTCATC	1	A07	TGACAGAAGATAGAGAGCACA
GCTCTCTATACTTCTGTCAAT	1	A01	NO
GCTGGACGACTTAGTAGACGA	1	A01	NO
GCTTGGTGCAGGTCGGGAAC	1	A06	CGGATCCCGCCTTGATCAAG
GGAATGTTGTCTGGATCGAGG	1	A10	NO
GGAATGTTGTTGGCTCGAAG	1	A04	TCGTCGGACCAGGCTTCATTC GTCGGACCAGGCTTCATTCCC
GGAGGCAGCGTTGATCGATC	1	A10	NO
GGCAAGTTGTCCTTCGGCTAC	1	A06	AGCCAAGGATGACTTGCCGGA
GGGCACAGCGGACTGAAACAC	1	A03	NO
GGGCATCTTCTATTGGCACG	1	A09	NO
GGGCTAGTGGTACTTGAGGGG	1	A09	NO
GGGTCGACATGAGAACACATG	1	A03	TCATGTGTTCTCAGGTCACCC ATGTGTTCTCAGGTCACCCCT
GGGTCGATATGAGAACACATG	1	A10	ATGTGTTCTCAGGTCACCCCT TCATGTGTTCTCAGGTCACCC
GGTTGGACTGAAGGGAGCTCC	1	A01	AGCTTCCTTGAGTCCATTGAG
GTAGAGAAGGATCGAGGGACA	1	A03	NO
GTCAACTCGGTGACTCGGTCA	1	A07	NO
GTCACCTTCAATGCGTTGATC	1	A05	NO
GTCGAGCTCGGTAGCGTGGGC	1	A05	NO
GTCGGACCAGGCTTCATTCCC	3	A02** : A04 : A06	AGGGGAATGTTGTCTGGCACG GAATGTTGTCTGGCACGAGGC GGAATGTTGTTGGCTCGAAG GAATGTTGTTGGCTCGAAGA TGGATGATGCTTGGCTCGAGA
GTGAATCAAGCATGGAGCTTC	1	A07	NO
GTGACAGAAGAGAGTGAGCAC	5	A02 : A03 : A06 : A06 : A06	NO
GTGCAAGGCGAGATGTTGGAG	1	Scaffold001940	NO
GTGGGAGGGGTTATGTCTAGG	4	A01 : A02 : A02 : A08	NO
GTGGTTGGTAGCGGTTGTTGG	1	A07	NO
GTGTGAAGTAAGTCTGGGGGA	5	A01 : A10 : Scaffold000392 : Scaffold001236 : Scaffold003415	NO
GTTATTCTATTCCACCTCTTA	1	A08	NO
GTTCCCTGAAACGCTTCATTG	1	A07	NO
TAAGAGATCAAAGCCAGTAGT	1	A04	NO
TAAGTCGTCTGGGAAAAGTTA	1	A01	NO

Table S3A

TAATTGTAGGGACCAAATGC	1	Scaffold000283	NO
TAGTTGGATGCATGCAAAGGT	2	A01 : A10	NO
TATAGTGGATTTTTGAGTGGA	1	A06	NO
TATCTACTGCTTATGCCACCA	1	A09	TAATGGATGCATAACCAGTAG GATGCATAACCAGTAGATATG
TATGATTGTATTCCAGGCTGA	1	Scaffold001161	TGTTATATGGGTCGTGAGTGG
TATGCTTGGACTGAAGGGAGC	1	A02	AGAGCTTTCTTTGGTCCACTC
TCAACTCGGTGACTCGGTCAA	1	A10	NO
TCAAGGCTGGTTCCTACTACATG	1	Scaffold000340	NO
TCATGATCAAGGAAGGGGGCA	1	A01	NO
TCGAGTAACGCCAAGCGCACC	2	A01 : A09	NO
TCGCACGGTTGTCTGACAGGC	2	A06 : A09	NO
TCGCACGGTTGTCTGACAGGT	1	A09	CTCTGCACTGTAGAGGATCGC
TCGCAGGAGAGATAGCGCCAT	3	A03** : A09** : A10**	GATGACGGTATCTCTCCTACG GACGGTATCTCTCCTACGTA TGTTGACGGTATCTCTCCTAC
TCGCTTGGTGCAGGTCGGGAC	2	A01* : A03*	CCGCCTTGCATCAACTGAATC TCCCGCCTTGCATCAACTGAA
TCGTCGTAGGTTAGTTTTGCA	1	A08	NO
TCTGATTGGACTGAAGGGAGC	1	A03	GAGCTTCCTTTAGTCCATTCA
TCTTGATGATGCTGCAGCTGC	1	A10	NO
TCTTTGGATTGAAGGGAGCTC	3	A02 : A07 : A07	NO
TGACAGAAGAGAGTGAGCACT	1	A08	GCTCACCTCTCTTTCTGTGAG
TGACAGAAGATAGAGAGCACA	1	A07	GCTCTCTAGTCTTCTGTGATC
TGACAGAAGATAGAGAGCACT	2	A01* : A03	GCTCTCTATACTTCTGTCAAT
TGACAGCTTGTCTTGTGGGAA	1	A02	NO
TGACCGAGTAGACCGATAGTC	1	A08	CTATCGGTCTACTTGGTCAGC
TGACTCACCGGGTCTGACTCGG	1	A01	NO
TGAGAATCTTGATGATGCTGC	7	A02 : A03 : A04* : A07 : A09* : A10 : A10	GCATCATCTTCAAGATTGAGA GCATCATCATCAAGATTGAGA CAGCACCATCAAGATTGACAT TGCCATCATCAAGATTGACAT
TGAGTTATCATTGGTCTTGTG	1	A03	NO
TGATCGGCAAGTTGCTTGG	1	A05	NO
TGATCTTGTGTTGTGATGATA	1	A03	TTATCATCACAACACTAGATC
TGCCTGGCTCCCTGTATACCA	1	A08	NO
TGGGAATCTTGATGATGCTGC	3	A02 : A03** : A03	TGCAGCATCATTAAAGATTGAC GCAGCATCATTAAAGATTGACA GCAGCATCATCAAGATTGACA TGCCATCATCAAGATTGACAT
TGGGTTGTGATTCAGGCTT	1	A05	NO
TGGTAAGGAGACGGGACTGAG	1	A09	NO
TGGTGGGTTGCGCATTACGGT	1	Scaffold000212	NO
TGGTGTGTTGGATGAGTTGGTG	1	A10	NO
TGGTTGGTTTTGCAGATGCAG	2	A07 : Scaffold000162	NO
TGTACTCGCAGTTTAGACCAG	1	A01	NO
TGTGAATGATGCGGGAGATGT	3	A03 : A04 : A05*	AAATATCTGCCGACTCATCCA ATCTGCCGACTCATCCATCCA AAACATCTGCCGACTCATCCA TCTGCCGACTCATCCATCCA
TGTGTTGTGATGATAATCCGA	1	A03	NO
TGTTTTGTGGTTTTCTACCGA	1	A04	NO
TTAAAAGCTAAGGGACCCTAC	3	A03 : A08 : A10	NO
TTAAAGTGGGGTTCTTAGCGG	1	A08	NO
TTAACTTTAACTAAAAACGC	4	A01 : A04 : A07* : A08	TTTTTTTAGTTAAAAGTTAAG TAGTTTTTTTAGTTAAAAGTT
TTAAGCGTAGGGTCCCTAAC	3	A03 : A03 : A04*	GTTAAGGGACCCTACGCTTAA

Table S3A

TTAGAGGCAACGGTTGATATG	1	A03	NO
TTAGATGACCATCAACAAATA	1	A09	NO
TTAGTTGAGGAATAAATATGC	3	A04 : A04 : A07	NO
TTATCATCACAACTAGATC	1	A03	TTTGATCTTGTGTTGTGATGA TTGATCTTGTGTTGTGATGAT TGATCTTGTGTTGTGATGATA GATCTTGTGTTGTGATGATAA TCTTGTGTTGTGATGATAATC
TTATCCAGGATTTCTTAGGGC	1	A03	NO
TTCAAGTTCTAGATTGGCTCG	1	A05	NO
TTGCTTGGTGCAGGTCGGGA	2	A08 : A09	CCTGCCTTCCATCAACTGAAT
TTGCGACCAGGCTTCATTCCC	1	A09	GACTGTTGTCTGGCTCGAGGA
TTCGTTCTCTATTCTCCATG	1	A01	ATGAAGGAATAGAGAGTGGAA TGAAGGAATAGAGAGTGGAA
TTCTTAGAGTGAGGTTCTTAG	1	A09	NO
TTGACAGAAGAGAGCGAGCAC	1	A10	TGCTATTCTCGTTCTGTCAT GCTATTCTCGTTCTGTCATA
TTGATTCTTGGTGACGTGCTA	1	A02	NO
TTGCCGGAGTGTTGTTTCATCA	1	A06	NO
TTGCTATAGATGGTTTCTGCT	1	A03	ACAGAAACCATCAATAGCAAC GAAACCATCAATAGCAACTAG
TTGGACTGAAGGGAACCTCCCT	1	A08	AGAGTTTCTTAAGTCCATTC
TTGGATCCAAGAGAAGCAAGC	2	A05 : A09	NO
TTGTAAGTCTTCTCAGGTTAG	1	Scaffold000162	NO
TTGTGATCCGAGTGGAGACAA	1	A05	NO
TTGTGCAAGACTAAGAAGCAA	1	A10	TCTTCTTCTTGCACAACC
TTGTGTTGTGATGATAATCCG	1	A03	ATTATCATCACAACTAGAT
TTTAACTAAAACGCTAAGGG	1	A01	NO
TTTAGATCAATAAAAACCCGG	2	A02 : A09	NO
TTTCAAATGTAGACAAAGCA	1	A08	CTTTGTCTATCGTTTGGAAAA TTTGTCTATCGTTTGGAAAAG
TTTCCAAGCCCTCTGTAGATC	1	A08	NO
TTTGAGAATTATCCATGGGGG	1	A09	NO
TTTGATTTGTCCGATGTGGGA	1	A03	NO
TTTGGCATTCTGTCCACCTCC	4	A02 : A06 : A07 : A07	GCTGAGTTGGAACGCAAAATC
TTTTGCGTTTCAACTCGGTCC	1	A07	TGAGGGCTGAGTTGGAACGCA
TTTTGTGGGTTTCTACCGAGT	1	A04	NO

Table S3B

Table S3B. List of the microRNAs predicted from analysis of the *A. thaliana* genome.

The prediction was performed once the prediction using the *B. rapa* genome was done. Location on *A. thaliana* chromosomes of the predicted MIR genes are indicated, as well as identification of putative miRNA*

Sequence	MIR genes	Chromosome	miRNA*
AAGTGGATCCATCTGATGTGA	1	Chr5	NO
AATTGAGCCAAGGATGACTTG	1	Chr3	NO
ACAGAAGATAGAGAGCACAGA	2	Chr1 : Chr1	NO
ACGGTGTGAAGTAAGTCTGGG	1	Chr5	NO
ACTGACAGAAGAGAGTGAGCA	3	Chr2 : Chr4 :Chr4	NO
ATCCGGATCCGTGGCGCAATG	1	Chr3	NO
ATCTGCCGACTCATCCATCCA	1	Chr2	NO
ATGGAGAAGCAGGGCACGTGC	1	Chr5	NO
ATTGGACTGAAGGGAGCTCCC	1	Chr4	NO
CCGCCTTGCATCAACTGAATC	1	Chr4	TTCGCTTGGTGCAGGTCGGGA
CGATCCCCGGCAACGGCGCCA	1	Chr2	NO
CGCTTGGTGCAGGTCGGGAAC	2	Chr4* : Chr5	TCCCGCCTTGCATCAACTGAA
CGGTGTGAAGTAAGTCTGGGG	1	Chr5	NO
CGTACAAGGAGTCAAGCATGA	1	Chr5	ATGCCTGGCTCCCTGTATGCC TATGCCTGGCTCCCTGTATGC
CTCCCGACTTGGAAATGGGCAG	1	Chr1	NO
CTCGGATTGCTTGGTGCAGG	2	Chr4 : Chr5	NO
CTTTGGATTGAAGGGAGCTCT	2	Chr1 : Chr1	NO
GAATGTAGCAGACAACAGCGG	1	chloroplast	NO
GACTCTCGACTATTATCCA	1	Chr5	NO
GAGACAGGGAAACAAGCAGAGC	1	Chr2	NO
GAGCTTTCTTCGGTCCACTCA	1	Chr5	NO
GCGTACAAGGAGTCAAGCATG	1	Chr5	TATGCCTGGCTCCCTGTATGC
GCTCACTCTATCCGTCACC	1	Chr5	GTGACAGAAGAGAGTGAGCAC
GCTTGGACTGAAGGGAGCTCC	1	Chr5	GAGCTTTCTTCGGTCCACTCA
GCTTGGTGCAGGTCGGGAACT	1	Chr5	NO
GGAAGTGAGGATGTTGGGAT	1	Chr5	NO
GGCAAGTTGACTTTGGCTCTG	1	Chr1	ATTGAGCCAAGGATGACTTGC
GGTCGGCCAGTGAGACGGTGG	2	chloroplast : chloroplast	NO
TAACTTATTAGACACCATGAT	1	chloroplast	NO
TATGCTTGGACTGAAGGGAGC	1	Chr5	GAGCTTTCTTCGGTCCACTCA
TCCCGCCTTGCATCAACTGAA	1	Chr4	NO
TCGTCCGACCAGGCTTCATTC	1	Chr3	GACTGTTGTCTGGCTCGAGGA
TCTTTGGATTGAAGGGAGCTC	2	Chr1 : Chr1	NO
TGACAGAAGATAGAGAGCACACA	2	Chr1 : Chr1	NO
TGGACTGAAGGGAGCTCCTTC	1	Chr2	NO
TGGAGTAGTAGAACACGTGCG	1	Chr5	NO
TGGTTTGGCTCTGGAAGTAGG	1	Chr2	NO
TTATGGATTGAAGCAAGCACC	1	Chr3	NO
TTCGAGGGGGTCCCGTTGAAC	1	chloroplast	NO

Table S3B

TTCGCTTGGTGCAGGTCGGGA	2	Chr4* : Chr5	CCGCCTTGCATCAACTGAATC
TTGAGGGGAATGTTGTCTGG	1	Chr5	NO

Table S4

TableS4. identification of 68 siRNAs with very high levels of expression in the S1 generation only, on the basis of raw read sequencing data.

Validation by qRT-PCR was performed on 10 different siRNAs (yellow), and 4 siRNAs were validated by Northern blot analysis (green); 29 of the 68 21-nt siRNAs were identified as corresponding to transcripts retrieved from meiocyte mRNAseq data from the haploid form of an oilseed rape variety (purple).

#	sequence	Raw read counts/generation			Raw read counts/allotetraploid			RPM/allotetraploid		
		Parents	S1	S5	S1_EMZ1	S1_EMZ2	S1_EMZ5	S1_EMZ1	S1_EMZ2	S1_EMZ5
1	GGAGCTACTGGGGCAAGTGCC	4	7112	2	329	70	6713	405	35	3056
2	GGAAGAACACCGAGGAGGGGC	6	7192	1	103	110	6979	127	55	3177
3	TTAGGGACGGACTIONTTTCGGGC	4	4998	0	0	4993	5	0	2475	2
4	AGAACACCGAGGAGGGGCGGA	3	3852	0	213	1063	2576	262	527	1173
5	GCAAGCAACACCTACTGGTCG	0	3936	0	10	4	3922	12	2	1786
6	GAGCTTCGTTACATACACCG	2	2113	0	1	22	2090	1	11	952
7	GGCAGCGGTCTGACTCTT	1	2948	1	1	546	2401	1	271	1093
8	GAGCTACTGGGGCAAGTGCCC	0	2708	0	4	0	2704	5	0	1231
9	GAAAATCGTAAGGGAGAGGTC	0	2815	0	0	0	2815	0	0	1282
10	GTCCCGCGAACGAGAAACGC	1	1136	0	30	31	1075	37	15	489
11	CAACTCGTAGGTTCTTCA	1	1140	1	1107	8	25	1364	4	11
12	TGAAGAGACGTAAACTAGGG	1	1237	0	482	205	550	594	102	250
13	GTTCTATGGAGGACGGCCAT	0	1210	1	7	8	1195	9	4	544
14	GTAGGTGTTGCTTGCCACGGC	0	1417	1	0	40	1377	0	20	627
15	TTTTCGGACCAAGATAGCAG	0	1379	1	0	1379	0	0	684	0
16	GATTTCTGCAGATGGAGATG	0	1130	0	1	5	1124	1	2	512
17	TTTCGGACCAAGATAGCAGC	5	11745	20	0	11733	12	0	5817	5
18	GCGACCCGGACGGTACCAGGC	5	3064	5	92	2321	651	113	1151	296
19	CGTCGAATGAGAGAAGGGCCC	5	3669	5	2691	972	6	3315	482	3
20	CGGACGGTACCAGGCTATGAA	3	2126	1	376	504	1246	463	250	567
21	AGCGGTCTGACTCTTGGC	3	2803	4	25	1931	847	31	957	386
22	ACTTGGATTGAGAAAGACGTA	3	2287	0	0	2280	7	0	1130	3
23	CCGTAGGGGTCTGTAGGGA	4	2380	1	2234	70	76	2752	35	35

Table S4

24	GAAGAACACCGAGGAGGGGCG	0	2075	4	197	118	1760	243	58	801
25	CTCGGGTTGAAAAGGAGTCAG	2	1016	2	672	194	150	828	96	68
26	TTTGGGATTGTGGAATTGGCA	2	1296	0	147	810	339	181	402	154
27	TCTACAGTCCGACGATCAGTA	1	1143	2	0	0	1143	0	0	520
28	TTTGGATTGCTTCTGGACAGC	2	1076	0	22	668	386	27	331	176
29	ACGCGTGACCTTTTAGGGGGC	3	1760	1	4	437	1319	5	217	601
30	AACGAAGGGGGAAGCGTCGGG	2	1155	0	98	916	141	121	454	64
31	AGTCTGGTTTCTGTTAGCAG	0	1444	2	5	695	744	6	345	339
32	TCTTTTCGGACCAAAGATAGC	0	962	0	0	960	2	0	476	1
33	CTTGGATTGAGAAAGACGTAC	1	831	0	0	831	0	0	412	0
34	ACCAGGACGGCCCTCTGAGCC	1	822	0	0	822	0	0	408	0
35	TGCGAGCCTCGGAATTTCTAG	0	840	0	5	264	571	6	131	260
36	GAATGATCGAGGAGCGGTGT	0	857	0	87	33	737	107	16	336
37	GGCAAGCAACACTACTGGTC	0	894	1	0	0	894	0	0	407
38	AGCTTCATGTCAGTTTTGGG	1	796	0	4	341	451	5	169	205
39	TGGGAAGGACTGTTAGGCTGC	1	774	1	15	455	304	18	226	138
40	GACGTTAACTAGGGATACCT	1	761	1	3	123	635	4	61	289
41	CCTGACACACAACGACGACGA	1	703	0	703	0	0	866	0	0
42	AGGCGTATGGACAGGAAAAC	0	739	1	15	628	96	18	311	44
43	AAAAGGTCACGCGTAGAGCGG	1	661	1	58	114	489	71	57	223
44	GGTTAGGGACGGACTTTTCGG	1	661	1	0	654	7	0	324	3
45	CGAAAGGGATAGGGCTGGCGC	1	611	0	0	610	1	0	302	0
46	GCCCTTCTCATAACGACGGT	1	637	1	5	0	632	6	0	288
47	TTCAGACGGGCGACCCGGACG	1	668	1	629	13	26	775	6	12
48	GCTACAAAAGAACCTACTGG	0	610	0	0	30	580	0	15	264
49	TGTAGTCGGATGAAGGTCAAC	0	640	0	40	370	230	49	183	105
50	TAGTCTGGCTGAGCTACCGGT	0	640	0	0	640	0	0	317	0
51	AACTCGGTAGGTTCTTGAC	0	665	0	658	3	4	811	1	2
52	CCGAGTTGGATAAGCGAATCG	0	699	0	15	680	4	18	337	2

Table S4

	53	GGAAACTCGGGGGGCGACAT	0	689	0	1	4	684	1	2	311
	54	AGGTCACGCCTAGAGCGGAAG	1	520	0	61	93	366	75	46	167
	55	GACTCCTTTTCAACCCGAGGG	1	501	0	0	6	495	0	3	225
	56	GGTCGTACAGAGGAGAACCTC	1	556	0	1	387	168	1	192	76
	57	CGGACCAAAGATAACAGCCTG	1	552	0	1	1	550	1	0	250
	58	CAGACATGGAACCTAACGGCT	1	502	0	489	2	11	602	1	5
	59	CTTCTCTCATACGACGGGGTA	1	553	0	0	0	553	0	0	252
	60	GAAGCTCTTGAGGAAACCAG	1	557	1	3	31	523	4	15	238
	61	ATGACTCTTGGCCATAGGAC	0	562	0	466	41	55	574	20	25
	62	TTCGGACCAAAGATAGCAGCC	0	518	1	0	518	0	0	257	0
	63	CTAACCAGGACGGCCCTCTGA	0	597	1	0	597	0	0	296	0
	64	CCTACTGGTCGAGCGTTTCTC	0	564	0	401	18	145	494	9	66
	65	TCCGAAAAAGACAGCCTGAAA	0	506	0	506	0	0	623	0	0
	66	GAAACTCGGGGGGCGACATC	0	509	0	4	4	501	5	2	228
	67	ATACACCGGAGAGACAGGGCC	0	518	0	463	24	31	570	12	14
	68	GATGGGTTTTGCCGTGGCAAG	0	511	0	1	2	508	1	1	231
	TOTAL		81	109548	66	13485	41432	54631	16614	20540	24873

Table S5

Table S5. List of the primers used for qRT-PCR validation of sRNA expression levels and mRNA/TE transcript amounts.

Only primers leading to exploitable results are provided.

<i>Forward primer</i>	<i>5'→3' Sequence</i>	<i>Reverse primer</i>	<i>5'→3' Sequence</i>
References			
Aq-F	TTCAATGTCCCTGCCATGTA	Aq-R	GAGACGGAGGATAGCGTGAG
EF2-F	AAGCTGGTTCTGAATATGC	EF2-R	GAGTACATTGAGAGGGATGA
miRNA_159a	TTTGGATTGAAGGGAGCTCTA	miScript universal primer	
miRNA_160a	TGCCTGGCTCCCTGTATGCCA	miScript universal primer	
miRNA_166a	TCGGACCAGGCTTCATCCCC	miScript universal primer	
21-nt siRNAs and putative transcript targets			
siRNA21_01fow	GGAGCTACTGGGGCAAGTGCC	miScript universal primer	
siRNA21_12fow	TGAAGAGACGTTAAACTAGGG	miScript universal primer	
siRNA21_18fow	GCGACCCGGACGGTACCAGGC	miScript universal primer	
siRNA21_20fow	CGGACGGTACCAGGCTATGAA	miScript universal primer	
siRNA21_21fow	AGCGGTCTGATGACTCTTGGC	miScript universal primer	
transcript_siRNA21-21U	TACGGGAGTTGGTCAAT	transcript_siRNA21-21L	GGTGATGCTTGCCACGG
siRNA21_26fow	TTTGGGATTGTGGAATTGGCA	miScript universal primer	
virus-like transcript_siRNA21_26aU	CGGGTTCCTCGTCTAC	transcript_siRNA21_26aL	CTCTACAAAGGCAATGGTT
transcript_siRNA21_26bU	GGCATCATCCGTGGATT	transcript_siRNA21_26bL	TCCCAAATAGGTAGACGAG
siRNA21_28fow	TTTGGATTGCTTCTGGACAGC	miScript universal primer	
siRNA21_31fow	AGTCTGGTTTCTGTTAGCAG	miScript universal primer	
siRNA21_39fow	TGGGAAGGACTGTTAGGCTGC	miScript universal primer	
virus-like transcript_siRNA21_39aU	ACGTGACTGCCCAATGCT	transcript_siRNA21_39aL	AGCCTAACAGTCCTCCCA
transcript_siRNA21_39bU	TCGGATTCCACTCGTGAA	transcript_siRNA21_39bL	AACAGTCCTCCAGTT
siRNA21_56fow	GGTCGTACAGAGGAGAACCTC	miScript universal primer	
transcript_siRNA21_56aU	ATTCGCTTAGGGTCGTA	transcript_siRNA21_56bL	TAAACTCATGGGGAGAT
24-nt siRNAs identical to 21-nt siRNA sequences			

Table S5

siRNA24_c1-2fwd (21-nt siRNA21_01)	AGGAGCTACTGGGGCAAGTGCCCA	miScript universal primer	
siRNA24_c20-3fwd	CGGACGGTACCAGGCTATGAAGGT	miScript universal primer	
siRNA24_c20-9fwd	CGGACGGTACCAGGCTATGAAGGG	miScript universal primer	
24-nt TE-derived siRNAs and putative TE targets			
LTR-siRNA24_16fwd	GAGCTGGAGTAGTGGAAGGATGGG	miScript universal primer	
LTR target; TE_siRNA24_16fwdU	GATTTGGCTACCTATGTCAC	TE_siRNA24_16fwdL	GCAATCACCACCCGACCTTT
TIR-siRNA24_22fwd	GAACTTGTGGGGTGAATAGGGCAA	miScript universal primer	
TIR target; TE17_siRNA24_22fwdU	GCCAAACCAGCAGACACCTAT	TE17_siRNA24_22fwdL	CGCCACAGAGGACGAACTA
TIR-siRNA24_36fwd	AGGGTTTAGGGTTTAGGGTTTAGA	miScript universal primer	
TIR target; TE_siRNA24_36fwdU	ACCTTGAGCCCATAGTTGTT	TE_siRNA24_36fwdL	TGTGCTTGCACGAGAGTTT
TIR-siRNA24_37fwd	AGAGTTGAGAAGTGAGGTTTTGGG	miScript universal primer	

Supplemental Figure S1. Pedigree of the three resynthesized *B. napus* allotetraploids analysed in the present study, and identification of the genotypes chosen for RNA extraction and subsequent construction of small RNA libraries and Illumina sequencing.

A total of eleven small RNA libraries were constructed and sequenced, corresponding to RNA extracts from the diploid progenitors (with two biological replicates for each progenitor; with *B. oleracea* in blue and *B. rapa* in yellow) as well as a 1:1 parental RNA mix used as a technical replicate (in green), and RNA extracts from the three resynthesized *B. napus* allotetraploids surveyed at two different generations of selfing S1 (in red) and S5 (in purple).

Supplemental Figure S2. Choice of appropriate reference genes for normalizing expression of small RNAs by qRT-PCR in our experimental design.

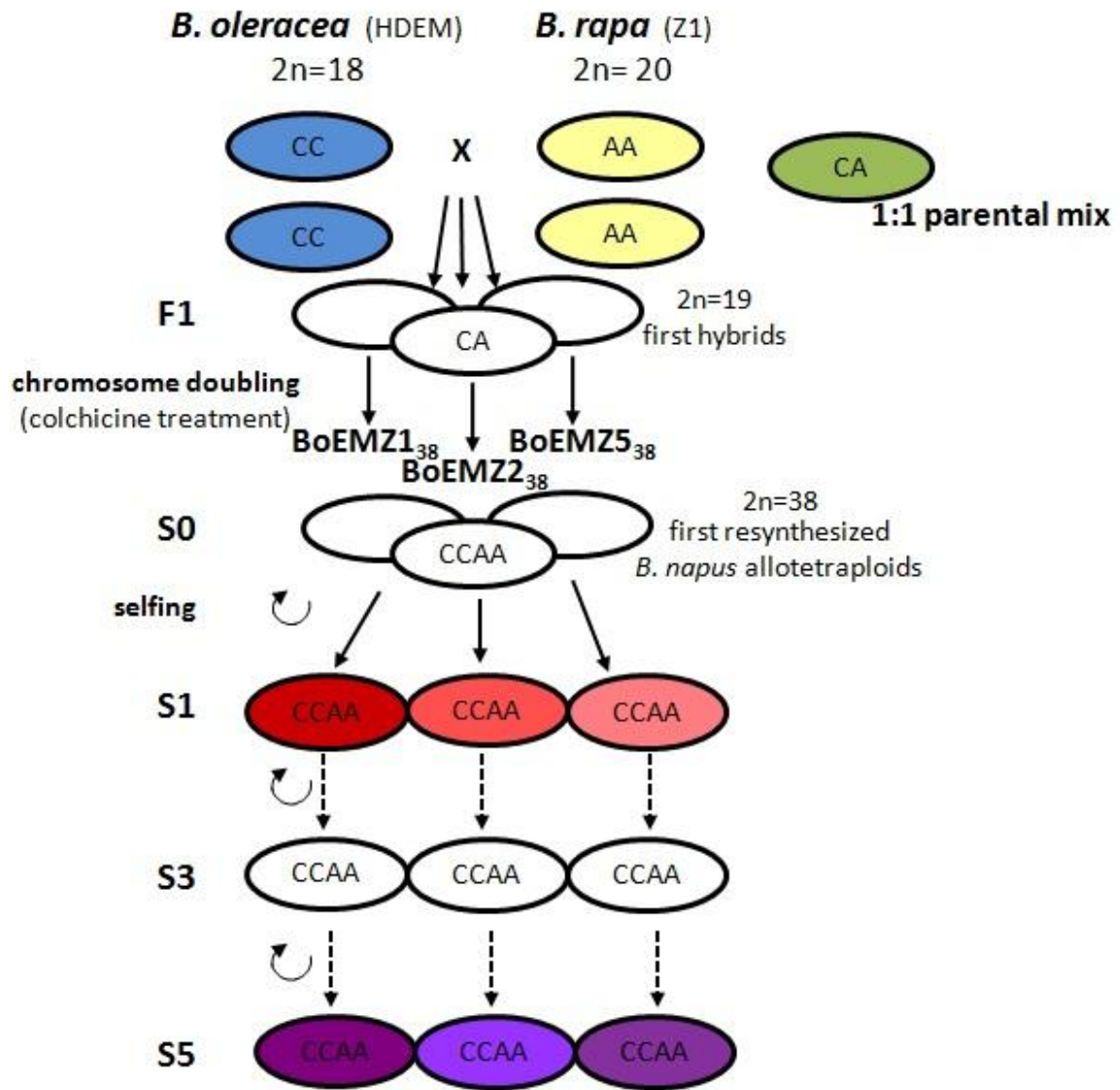
Page 1/6: Average expression stability values (M) of reference targets using geNorm. It is recommended to choose reference genes with M values below the limit of 1.5, with lower M values corresponding to more stable gene expression and thus more appropriate reference genes. Genes coding for Actin *Act2/7* (Marmagne et al. 2010) and Elongation Factor 2-alpha *EF2* were selected as reference targets, in addition to the three microRNAs miRNA159a, miRNA160a and miRNA166a, which were predicted as additive according to our bioinformatic data and previously published studies (e.g. Ha et al. 2009). The M values for these 5 candidate reference genes were all <1.3 suggesting the stability of their expression across the 31 samples representative of the six generations (diploid parents, F1, S0, S1, S3 and S5) which were analysed. For more stable expression, the three microRNAs with M values <0.56 were retained as reference genes.

From page 2/6 to page 6/6: Melting curves showing single peaks generated for the amplicons of each reference target, from RNAs isolated from the six generations under study:

the two diploid parents, the F1 hybrids, the first resynthesized *B. napus* allotetraploids, and the generations of selfing S1, S3 and S5.

Supplemental Figure S3. Validation of 21-nt siRNAs predicted as specifically and highly expressed in the S1 generation of the resynthesized *B. napus* allotetraploids in comparison with the parental and S5 generations, by qRT-PCR.

The hybrids F1 and the generations S0 and S3 were included in the analysis. All the melting curves are provided for the different genotypes surveyed within one generation. We mainly observed similar qRT-PCR results for the F1 hybrids and the S0 and S1 generations, with significant expression levels for the tested siRNAs and corresponding transcripts, while nearly no amplicons (with aspecific amplifications generating melting curves with multi-peaks or displaying very high C_q values) were obtained in the diploid parents as well as the S3 and S5 generations.



Supplemental Figure S1

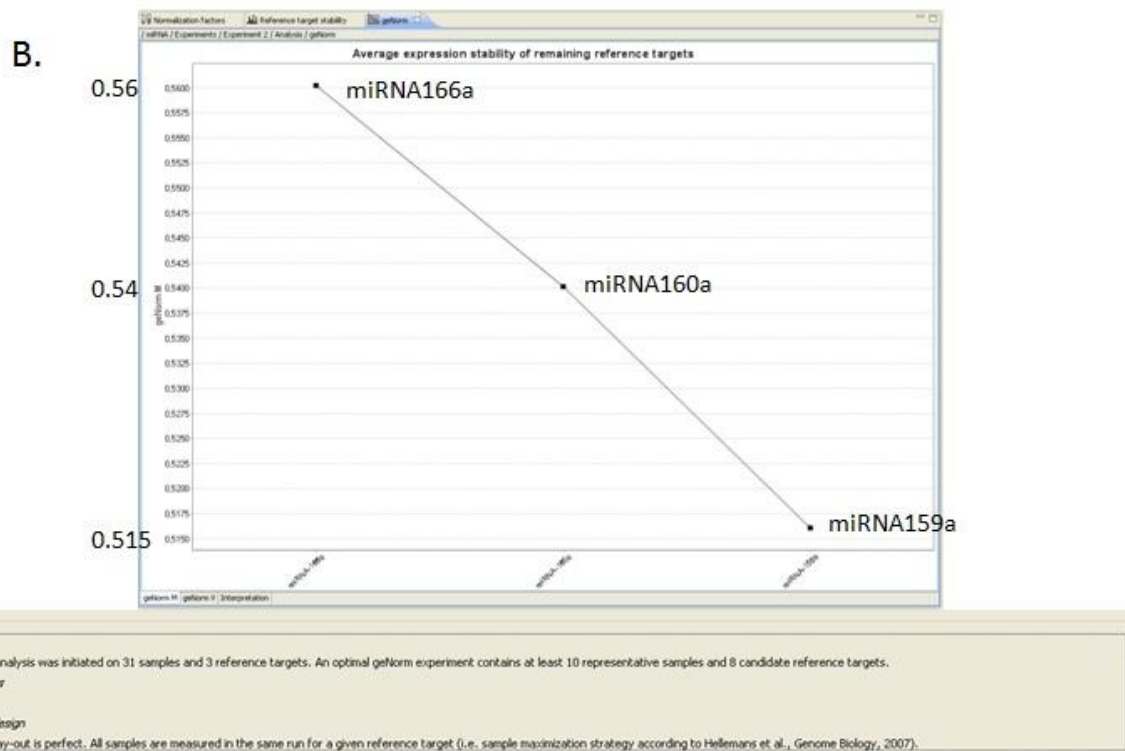
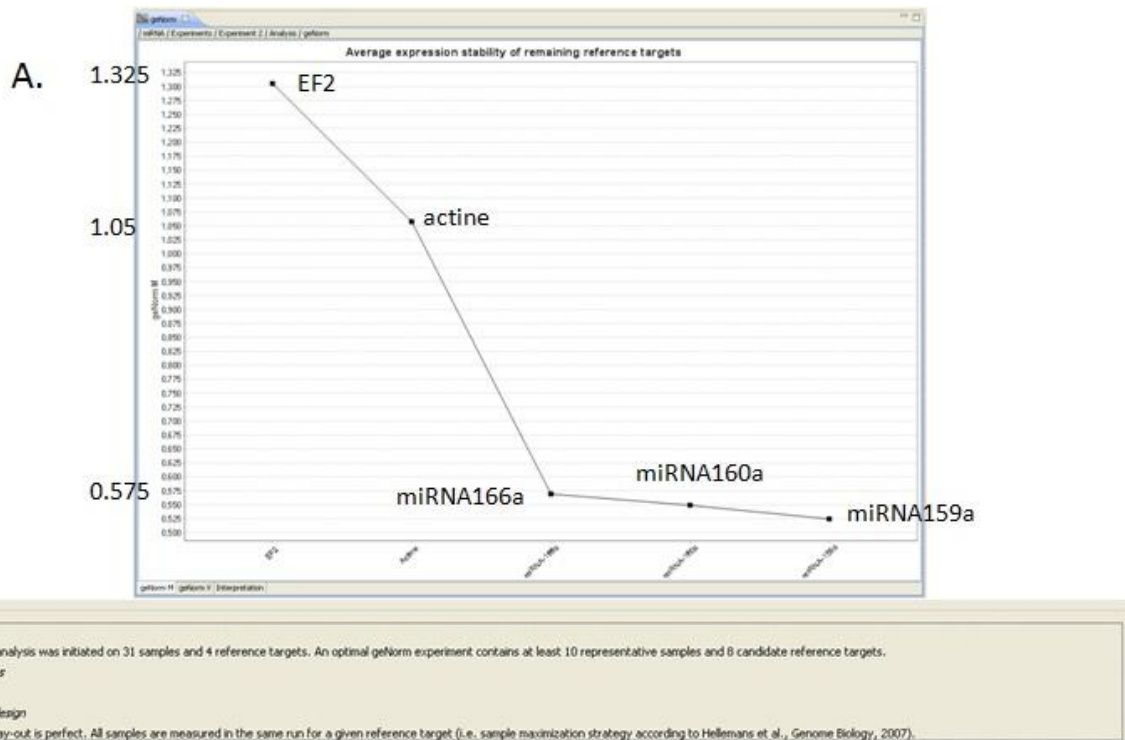


Fig. S2; page 1/6

Actin gene *ACT2/7*

F: TTCAATGTCCTGCCATGTA
R: GAGACGGAGGATAGCGTGAG

55°C

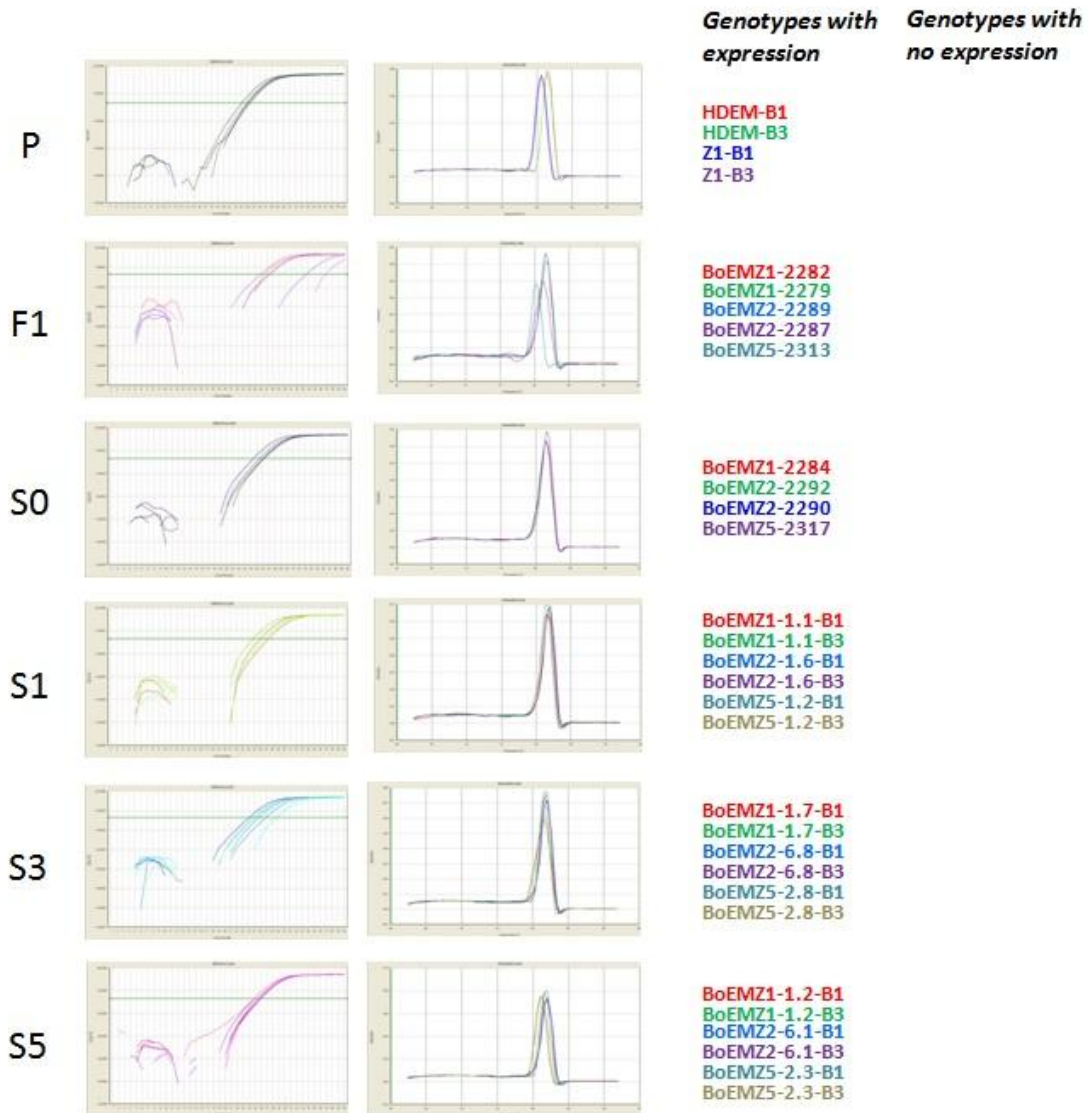


Fig. S2; page 2/6

Supplemental Figure S2

Elongation factor *EF2*

F: AAGCTGGTTCTGAATATGC
R: GAGTACATTGAGAGGGATGA

55°C

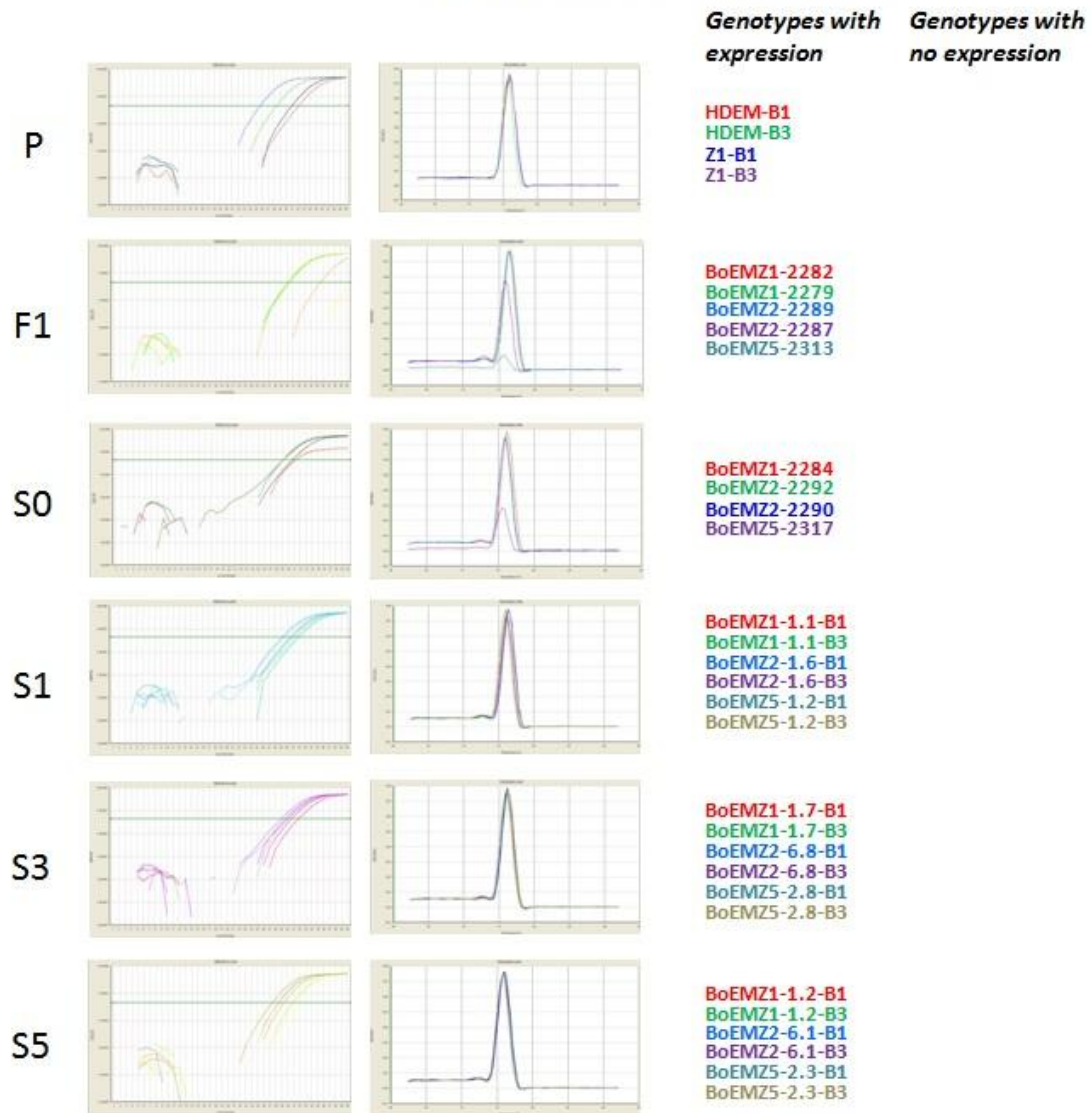


Fig. S2; page 3/6

Supplemental Figure S2

miRNA159a

TTTGGATTGAAGGGAGCTCTA

55°C

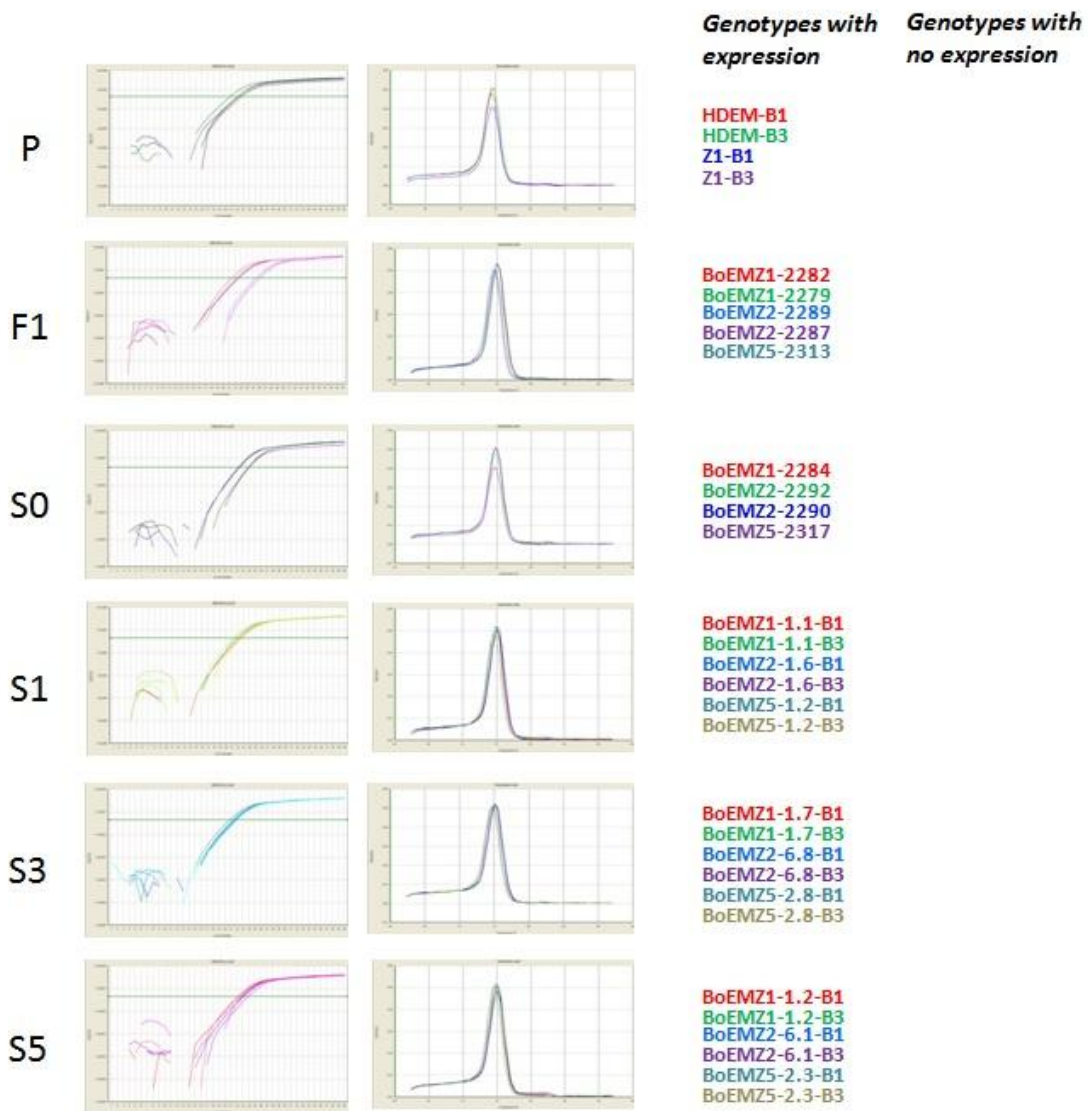


Fig. S2; page 4/6

Supplemental Figure S2

miRNA160a

TGCCTGGCTCCCTGTATGCCA

55°C

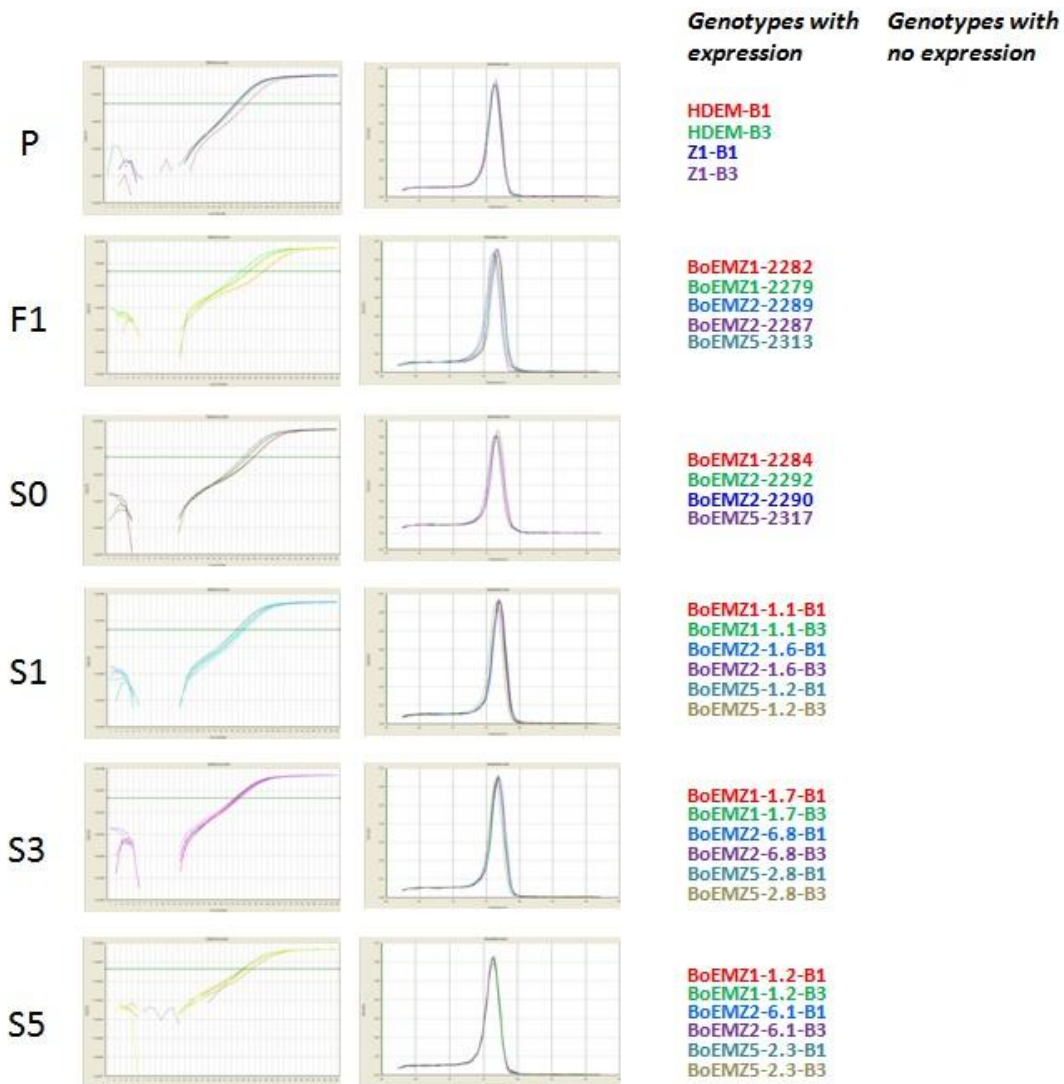


Fig. S2; page 5/6

Supplemental Figure S2

miRNA166a

TCGGACCAGGCTTCATTCCCC

55°C

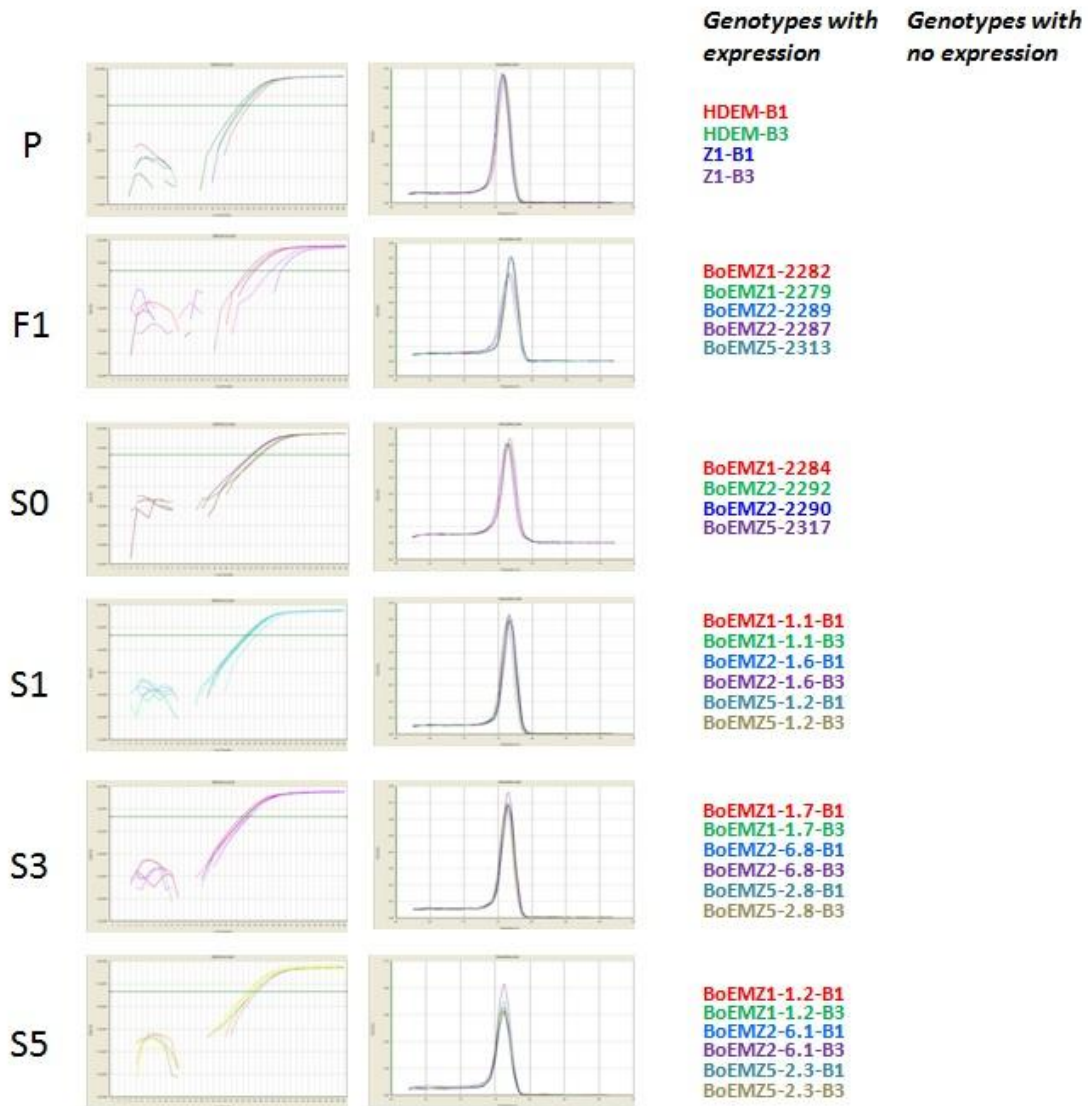


Fig. S2; page 6/6

Supplemental Figure S2

siRNA21.01fow

GGAGCTACTGGGGCAAGTGCC

65°C

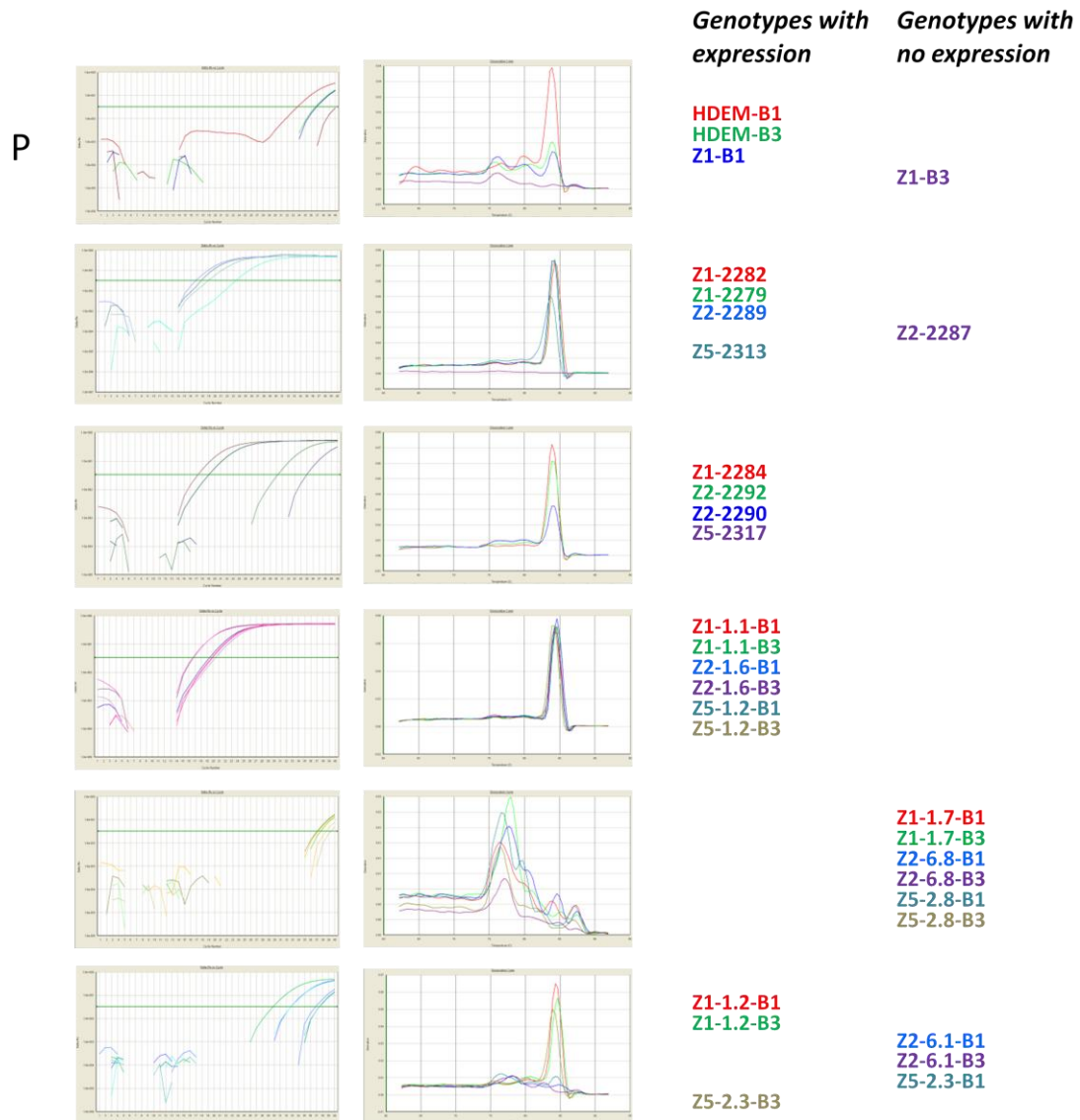


Fig. S3; page 1/16

Supplemental Figure S3

siRNA21.12fow

TGAAGAGACGTAAACTAGGG

60°C

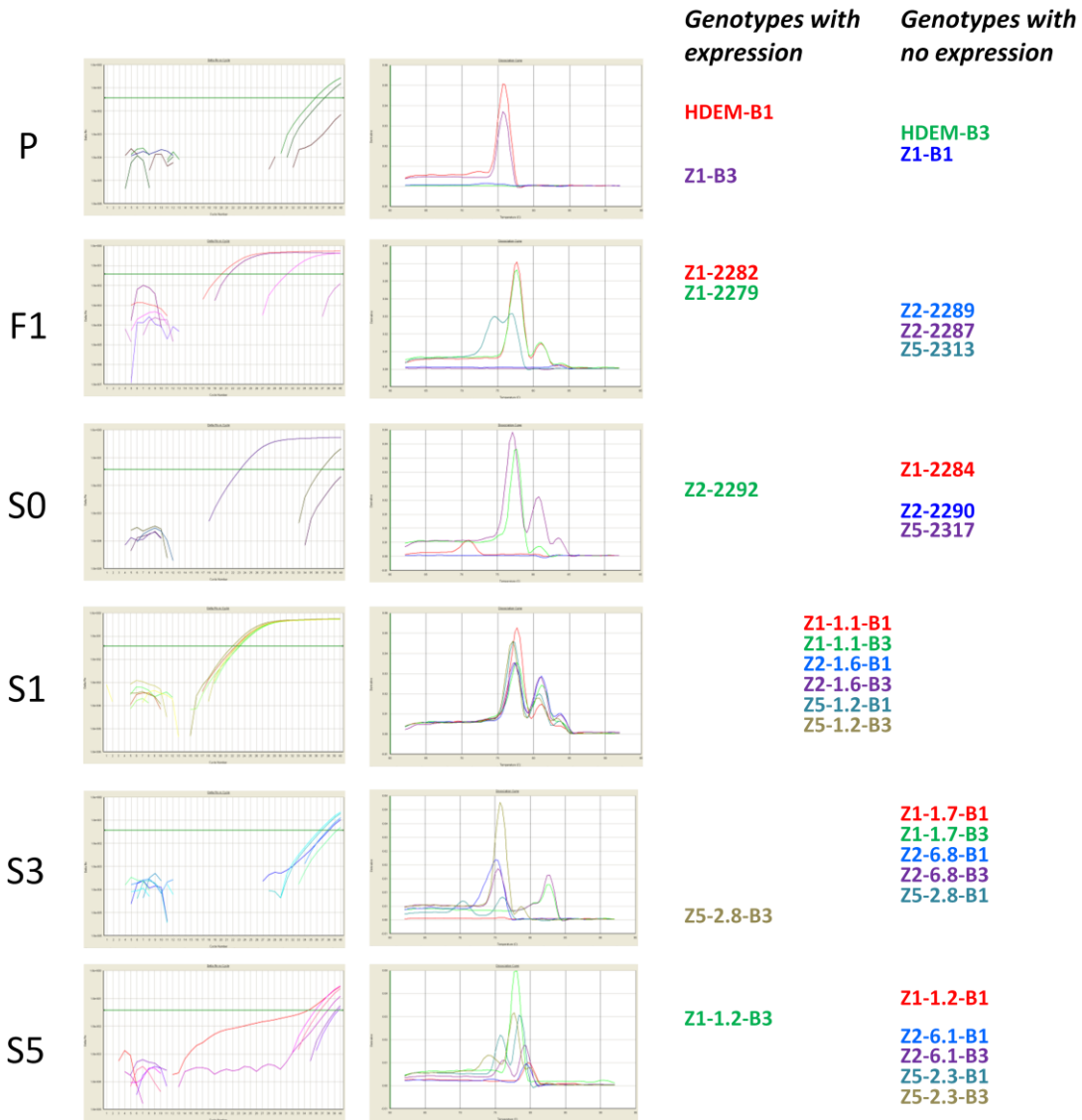


Fig. S3; page 2/16

Supplemental Figure S3

siRNA21.18fow

GCGACCCGGACGGTACCAGGC

65°C

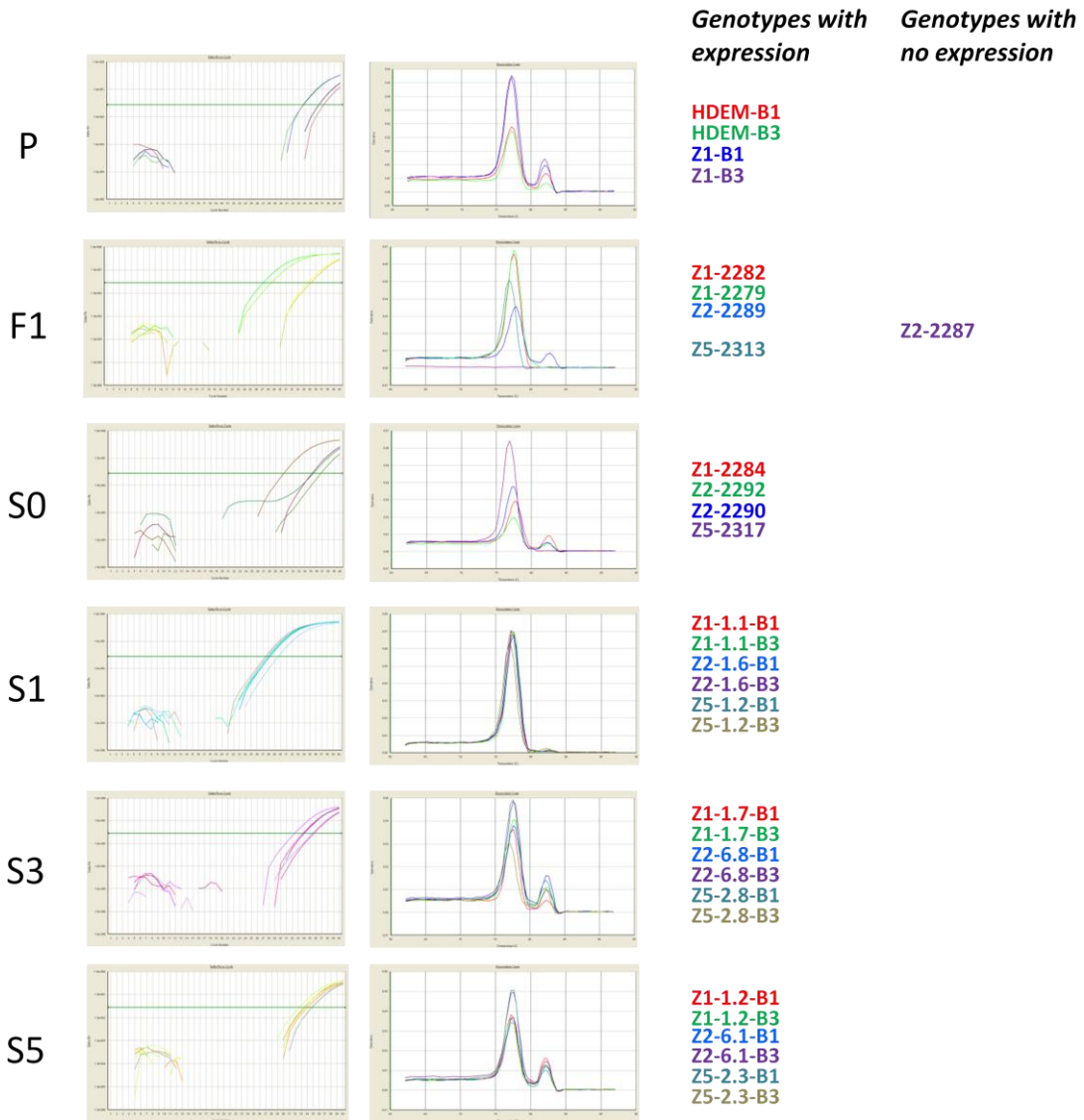


Fig. S3; page 3/16

Supplemental Figure S3

siRNA21.20fow

AGCGGTCTGATGACTCTTGGC

65°C

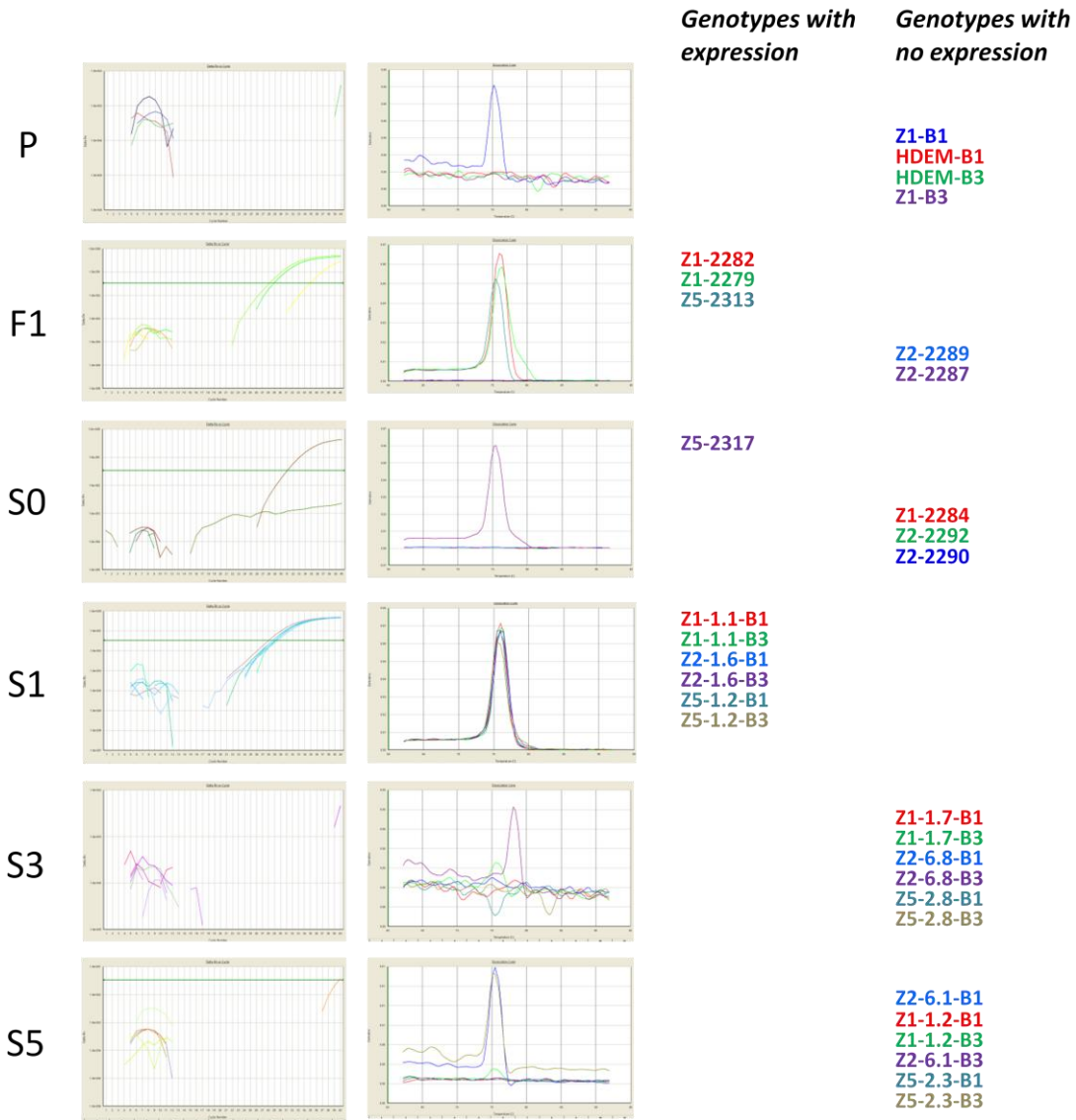


Fig. S3; page 4/16

Supplemental Figure S3

siRNA21.21fow

AGCGGTCTGATGACTCTTGGC

65°C

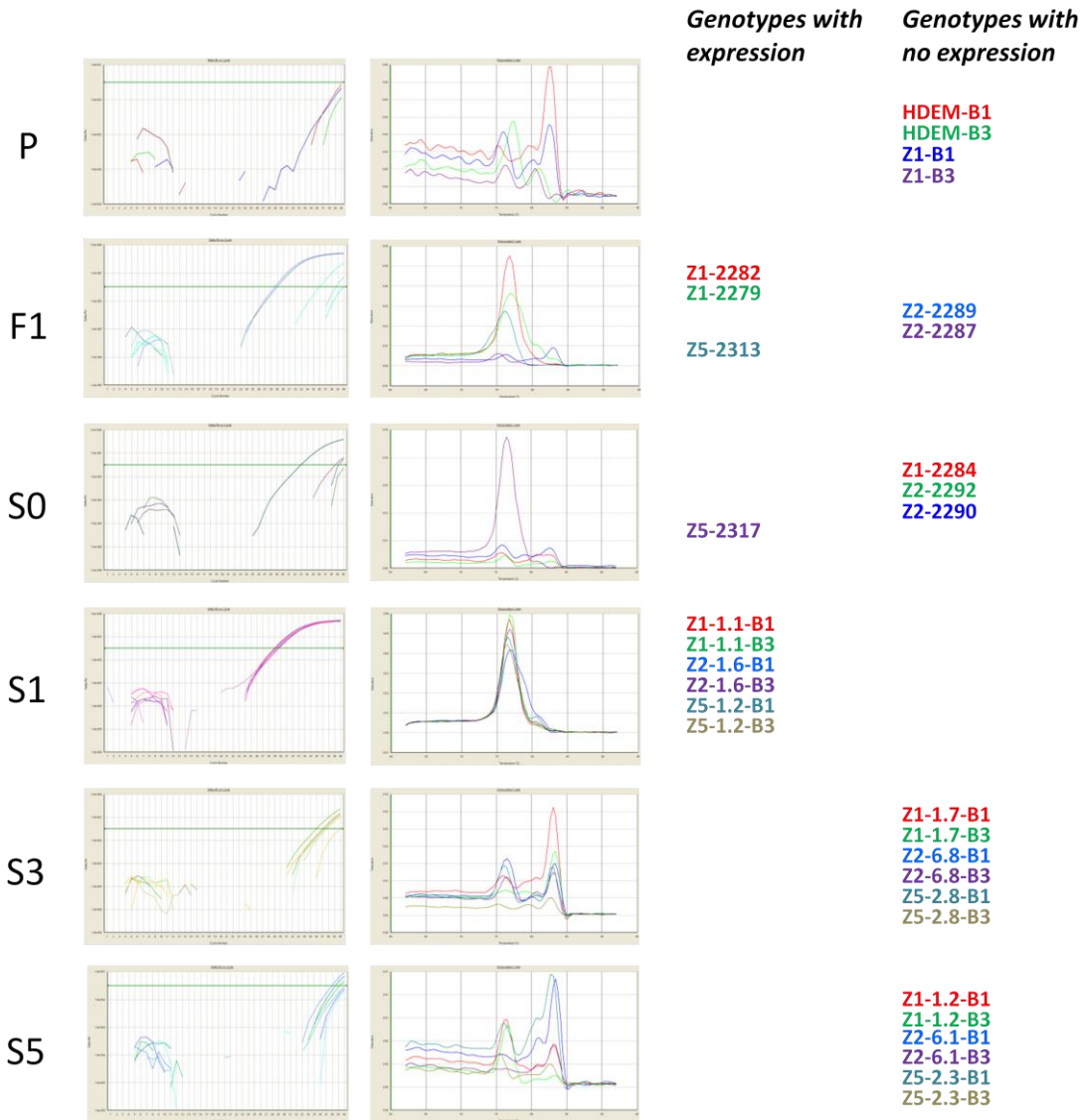


Fig. S3; page 5/16

Supplemental Figure S3

siRNA21.26fow

TTTGGGATTGTGGAATTGGCA

60°C

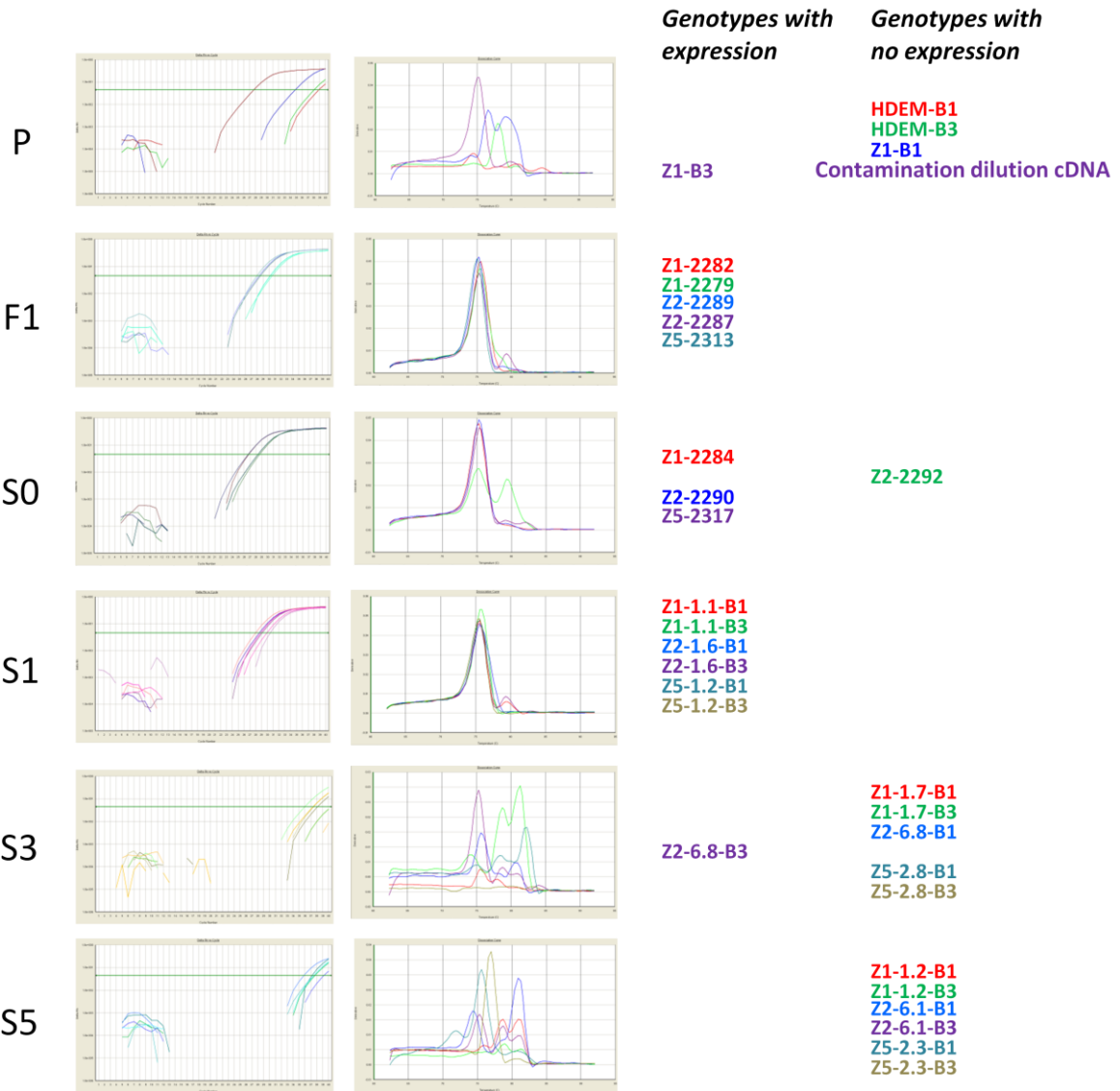


Fig. S3; page 6/16

Supplemental Figure S3

siRNA21.28fow

TTTGGATTGCTTCTGGACAGC

65°C

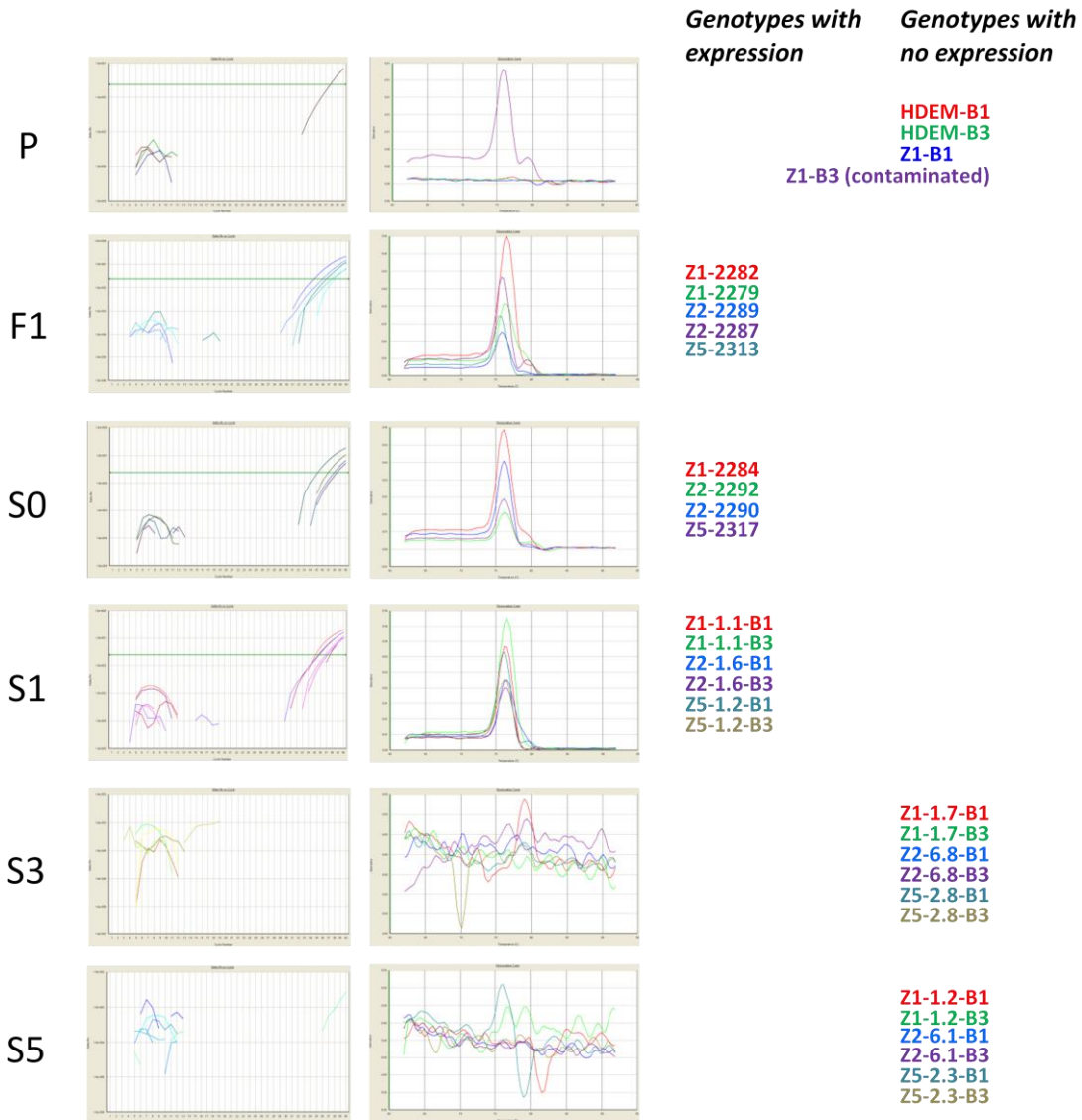


Fig. S3; page 7/16

Supplemental Figure S3

siRNA21.31fow

AGTCTGGTTTCTGTTTAGCAG

60°C

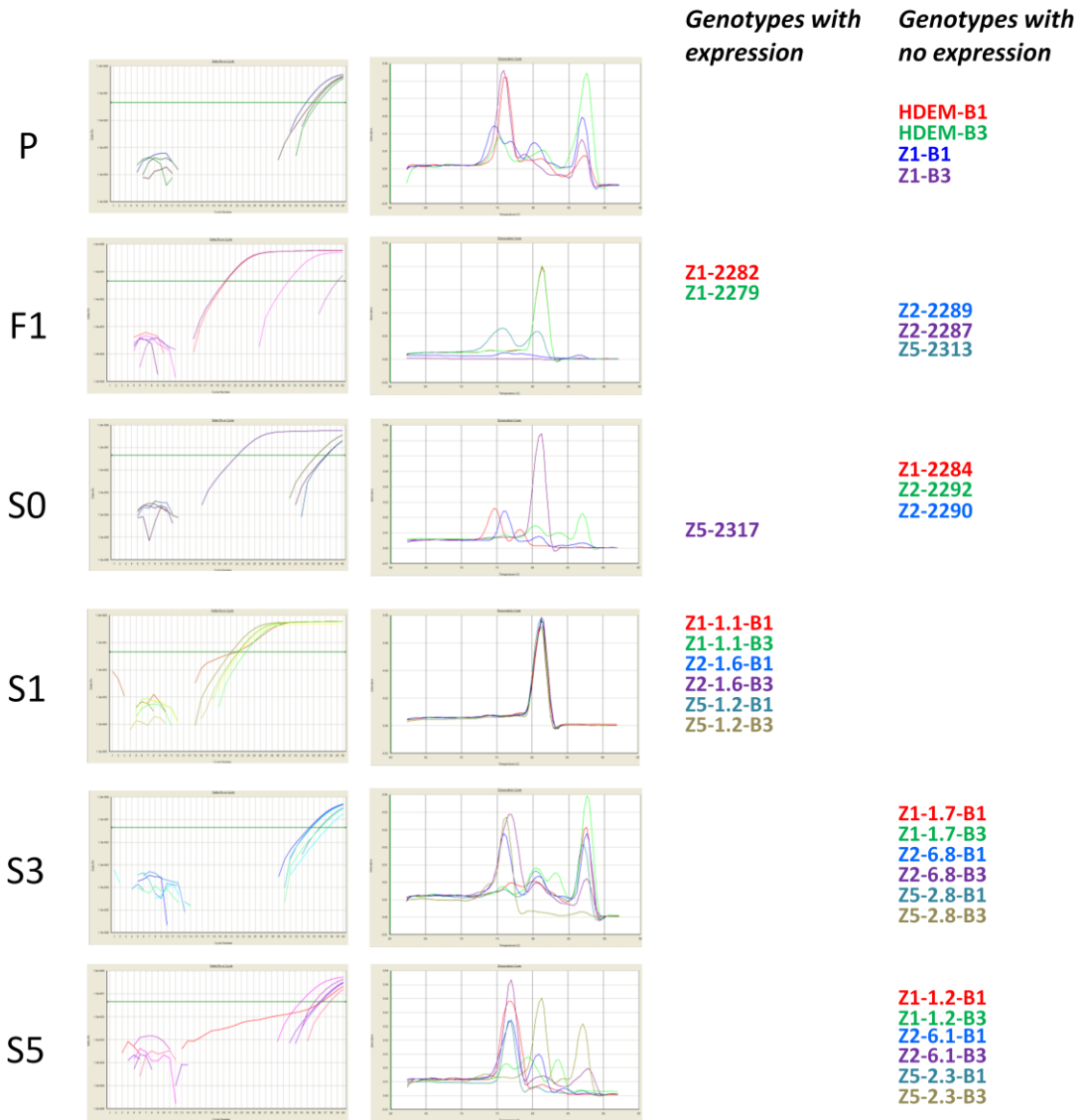


Fig. S3; page 8/16

Supplemental Figure S3

siRNA21.39fow

TGGGAAGGACTGTTAGGCTGC

65°C

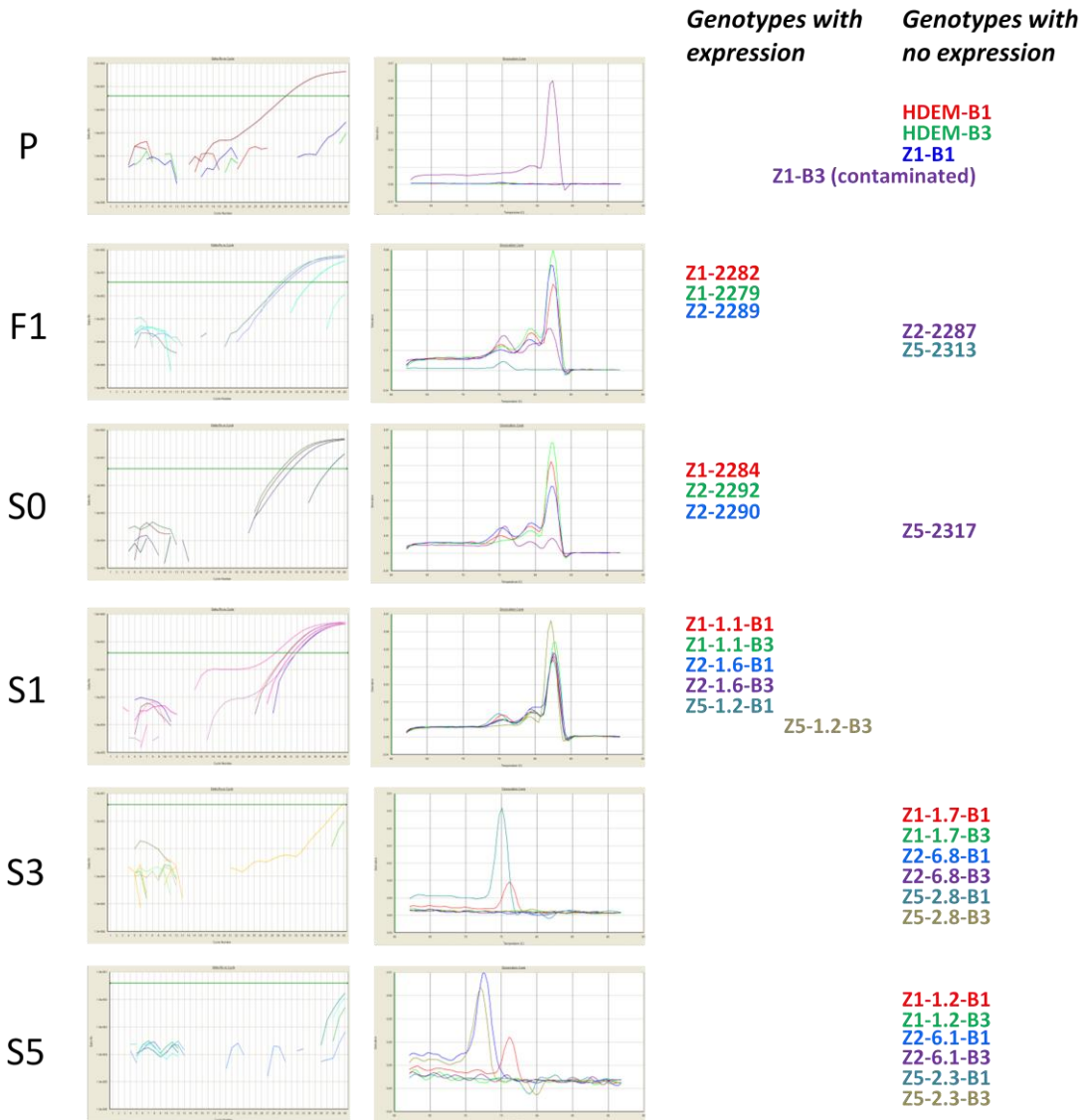


Fig. S3; page 9/16

Supplemental Figure S3

siRNA21.56fow

GGTCGTACAGAGGAGAACCTC

65°C

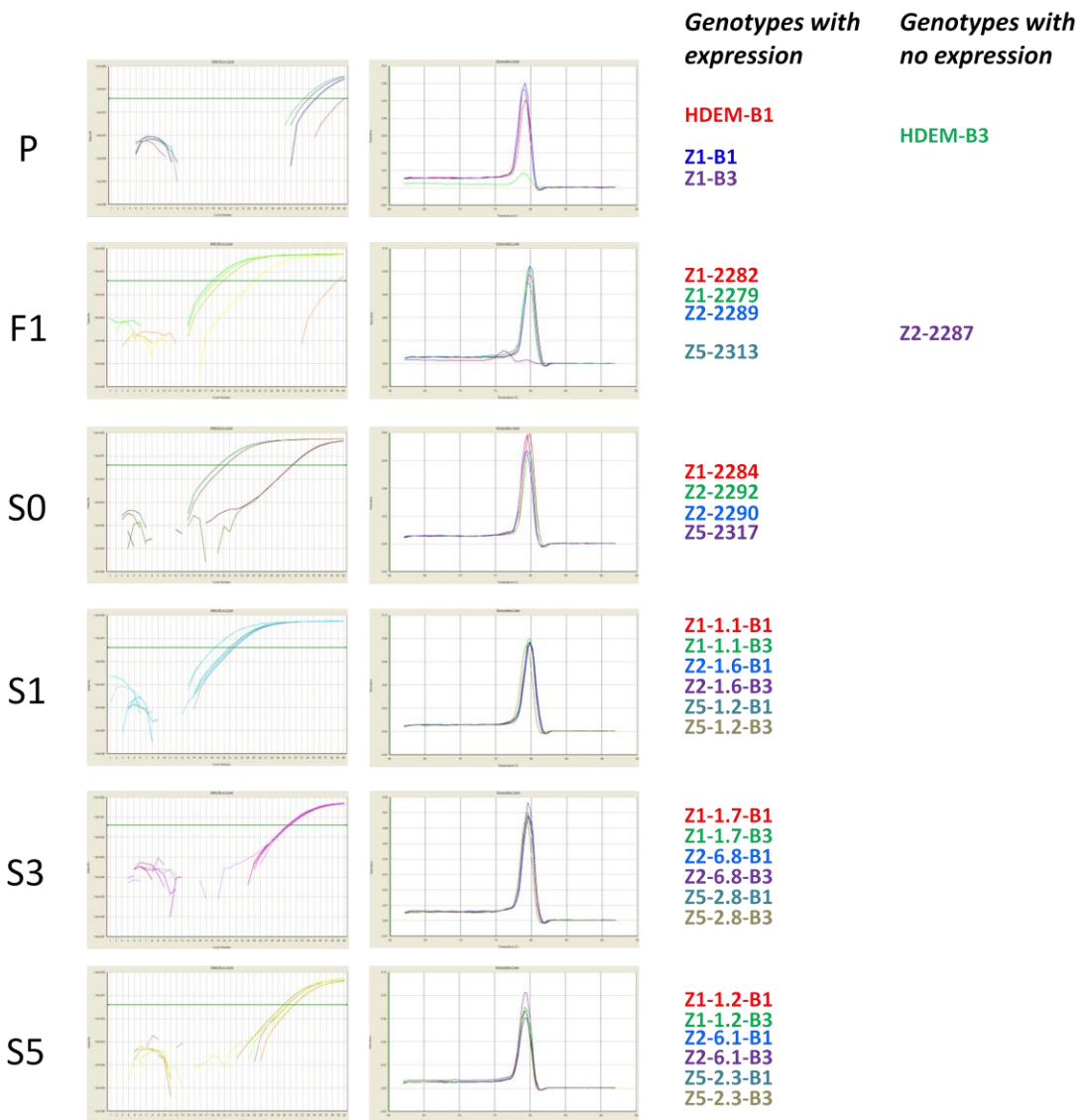


Fig. S3; page 10/16

Supplemental Figure S3

transcript_siRNA21-21
transcript_siRNA21-31

F: TACGGGAGTTGGTCAAT
R: GGTGATGCTTGCCACGG

55°C

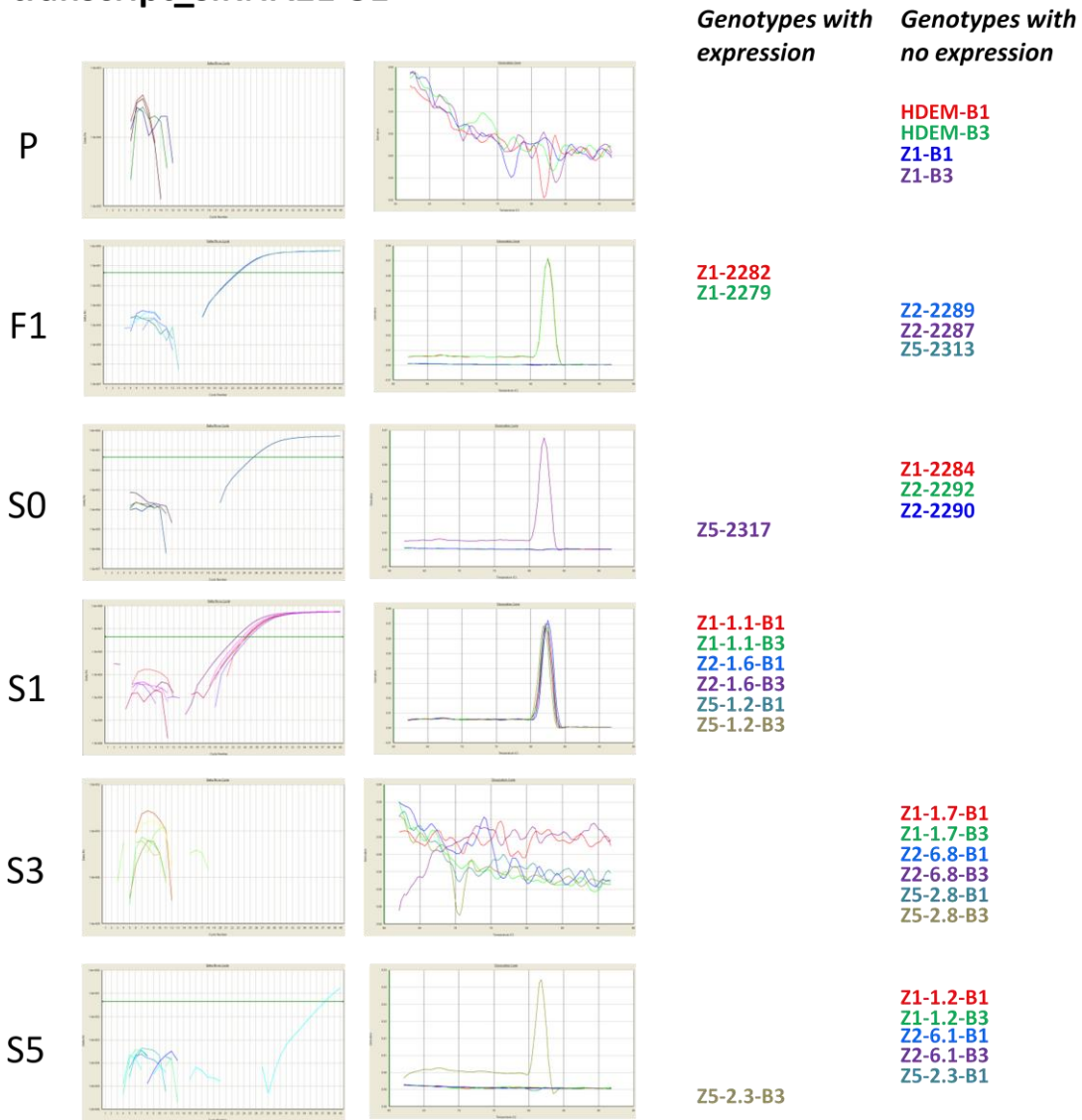


Fig. S3; page 11/16

Supplemental Figure S3

transcript_siRNA21-26a

F: CGGGTTCCTCGTCTAC
R: CTCTACAAAGGCAATGGTT

55°C

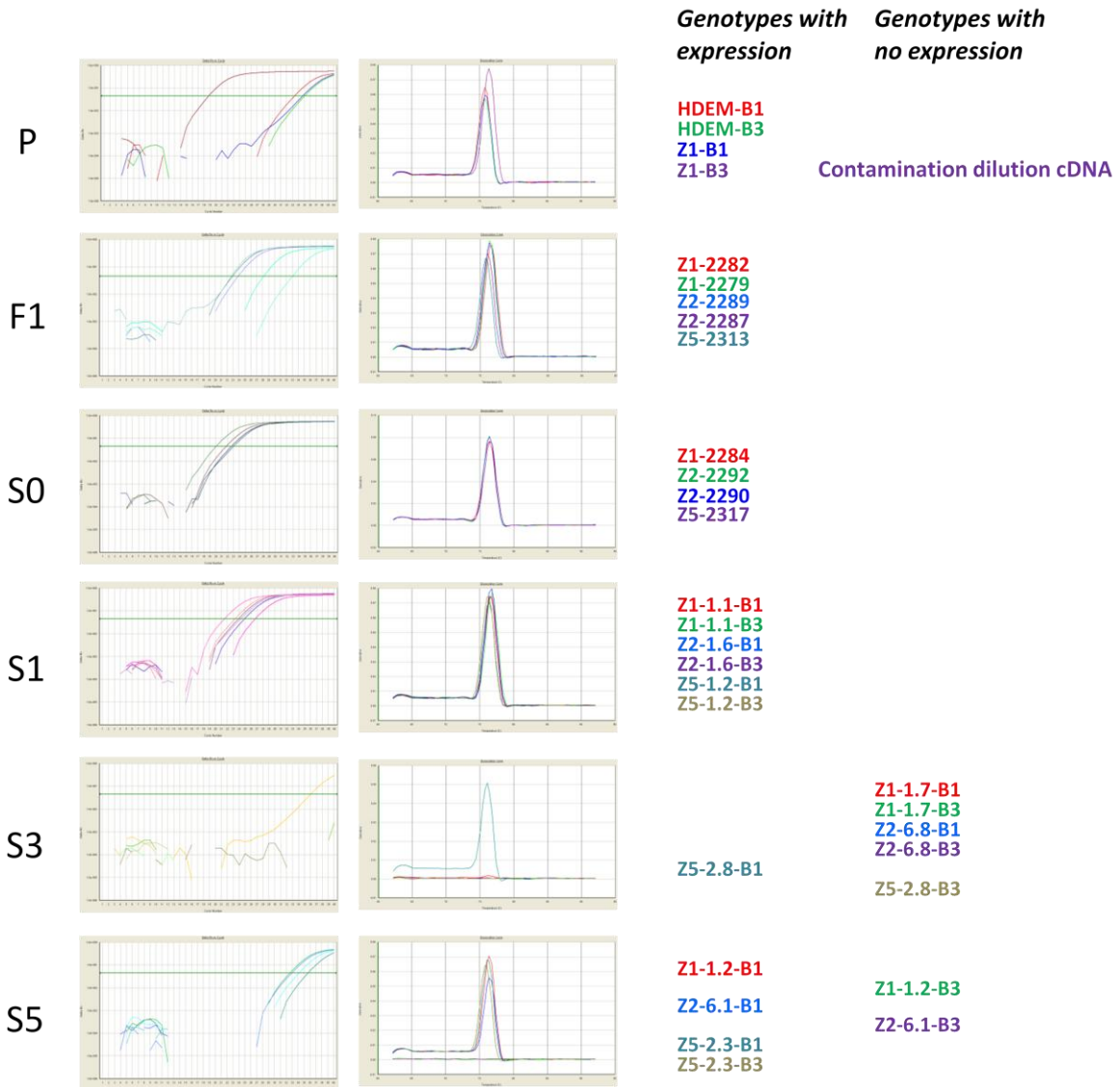


Fig. S3; page 12/16

Supplemental Figure S3

transcript_siRNA21-26b

F: GGCATCATCCGTGGATT
R: TCCCAAATAGGTAGACGAG

55°C

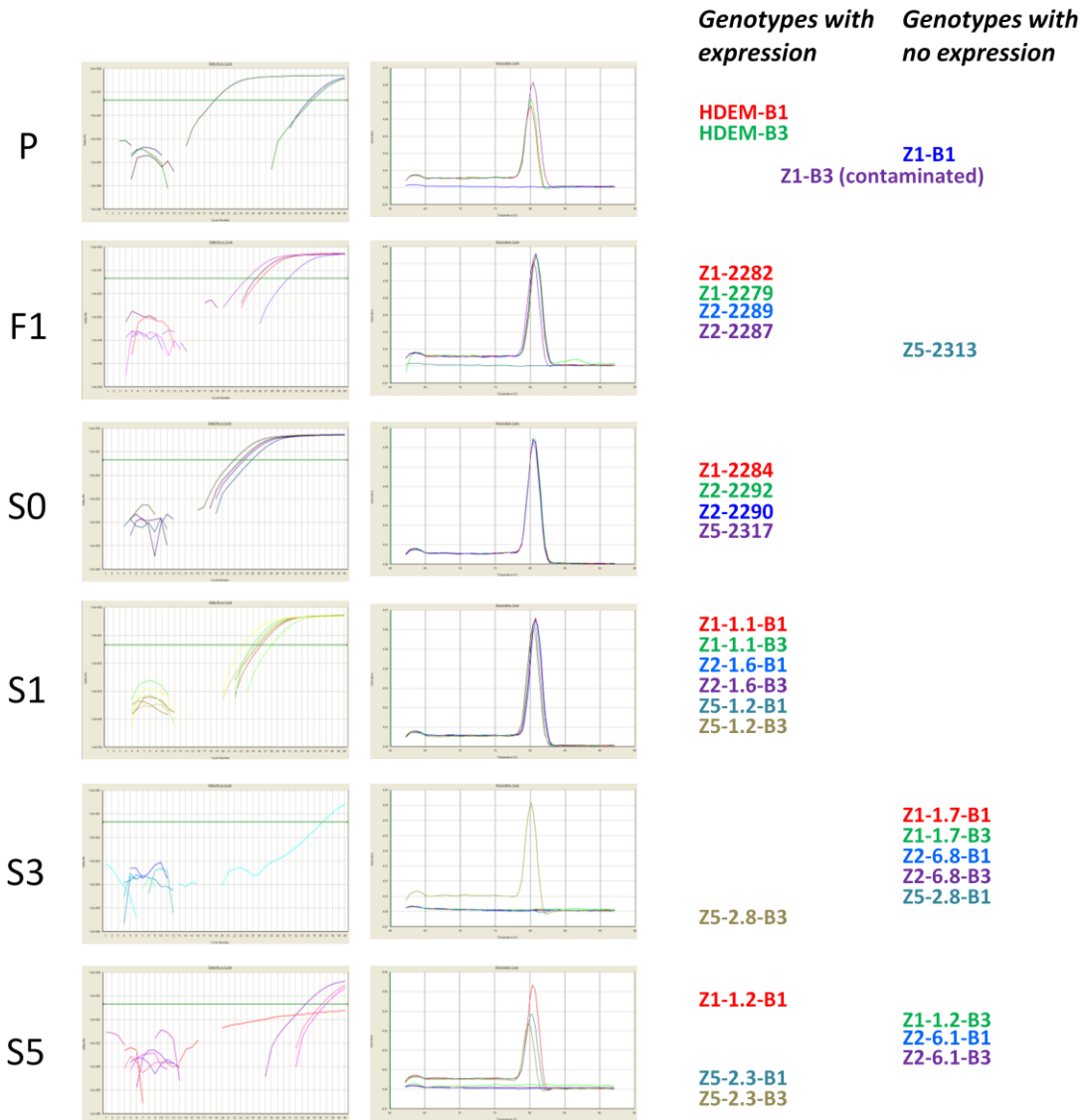


Fig. S3; page 13/16

Supplemental Figure S3

transcript_siRNA21-39a

F: ACGTGACTGCCCCAATGCT
R: AGCCTAACAGTCCTTCCCA

55°C

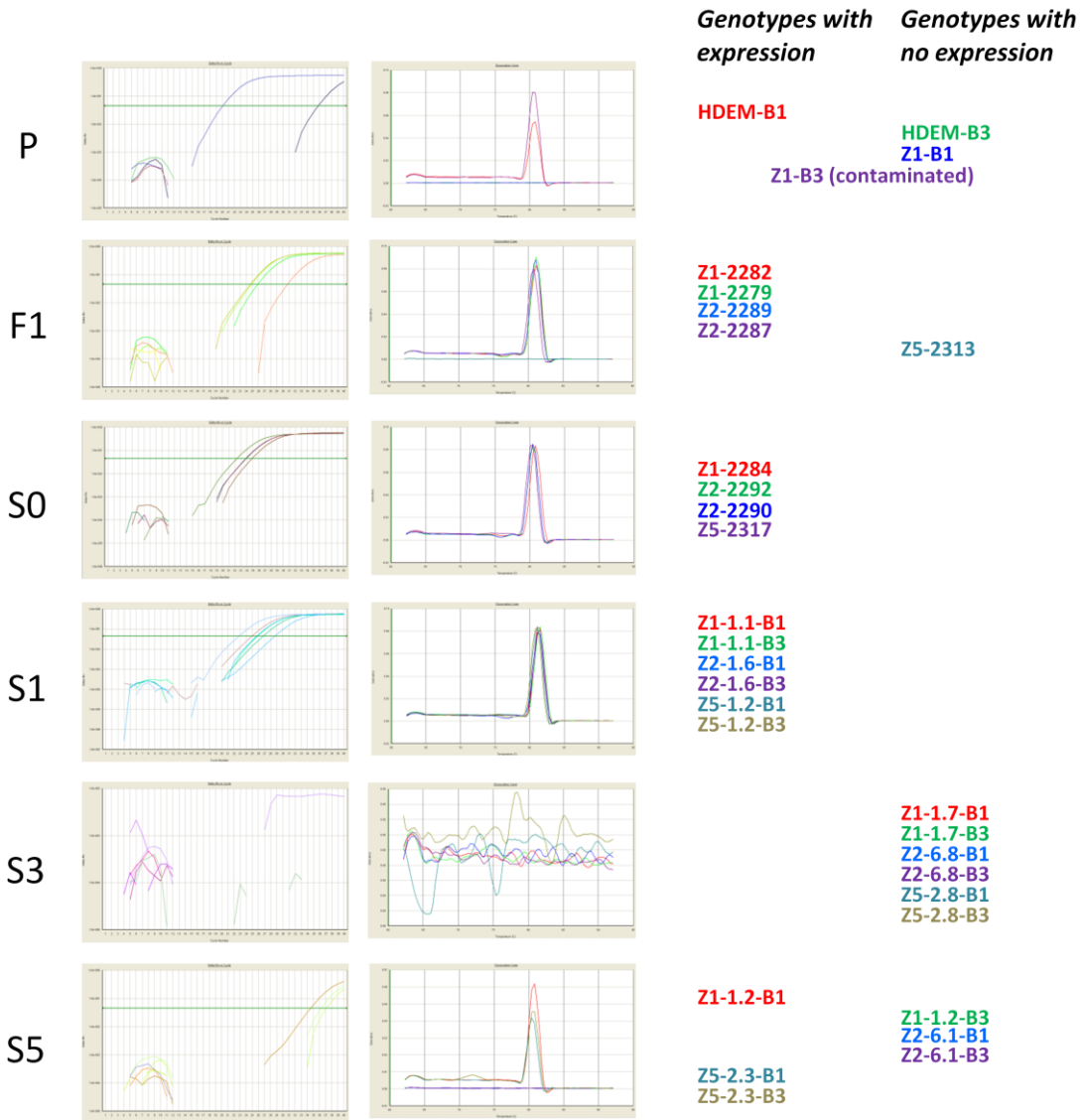


Fig. S3; page 14/16

Supplemental Figure S3

transcript_siRNA21-39b

F: TCGGATTCCACTCGCTGAA
R: AACAGTCCTTCCCAGTT

55°C

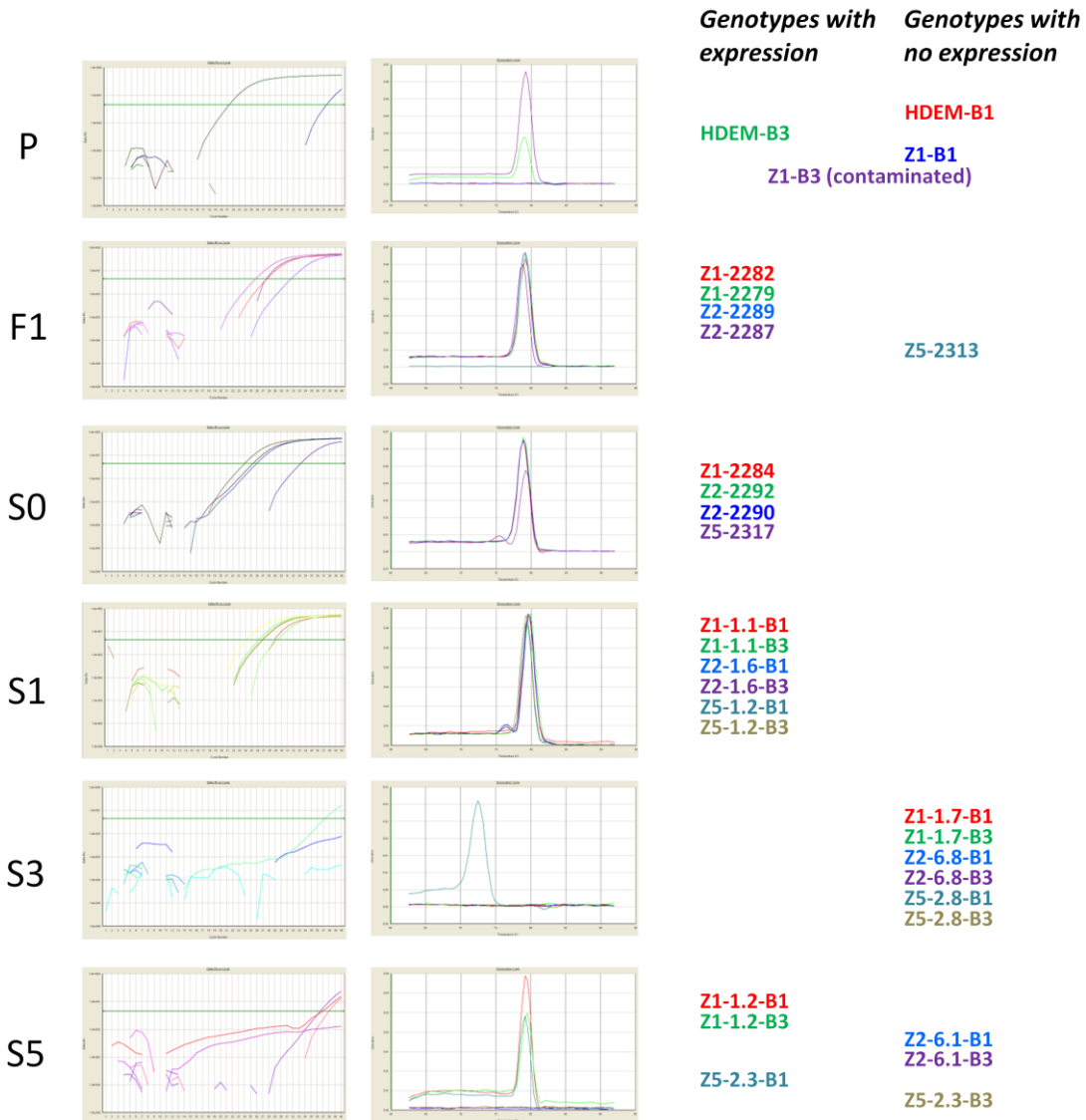


Fig. S3; page 15/16

Supplemental Figure S3

transcript_siRNA21-56b

F: ATTCGCTTAGGGTCGTA
R: TAAACTCATGGGGAGAT

55°C

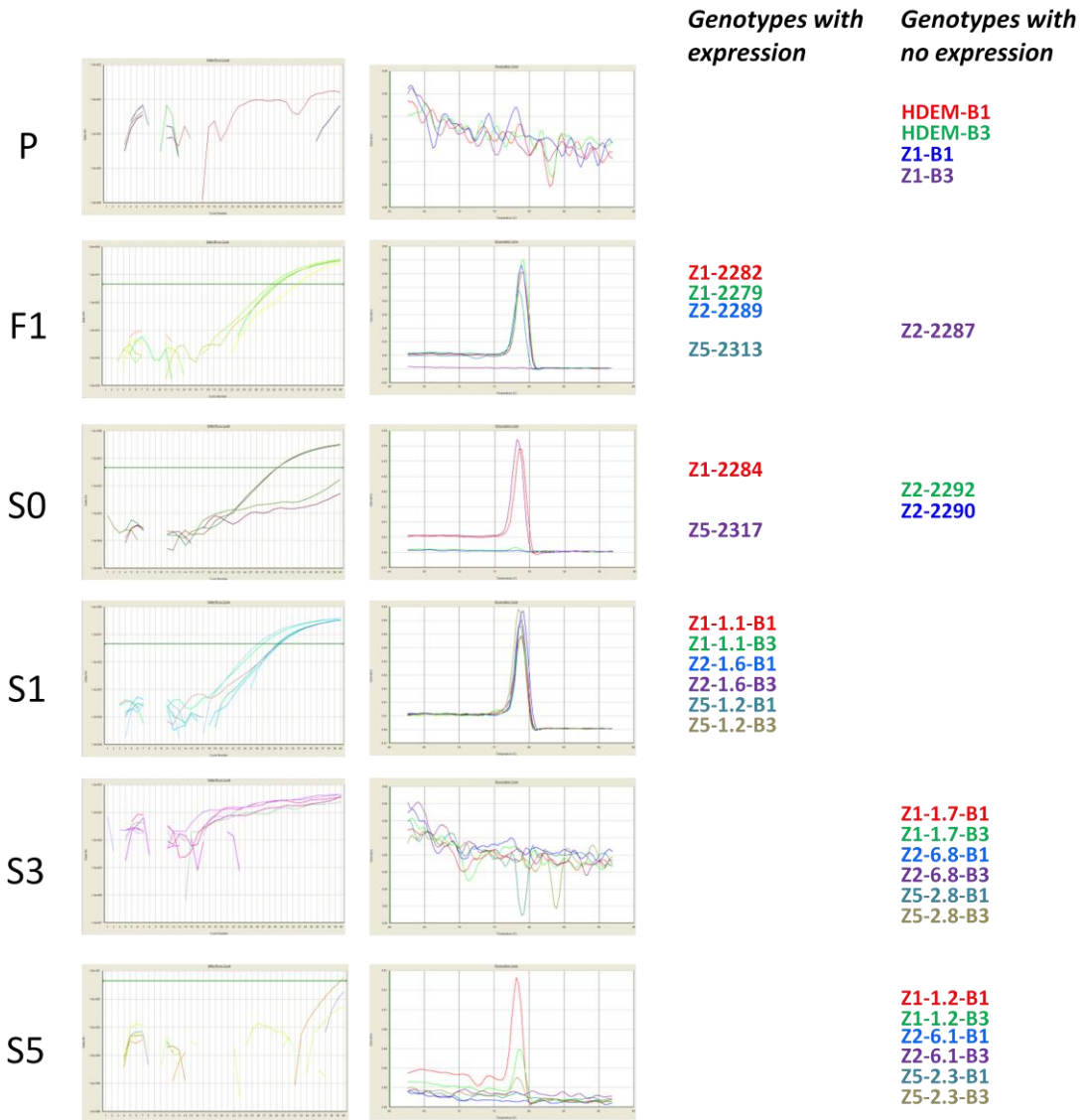


Fig. S3; page 16/16

Supplemental Figure S3

CONCLUSIONS ET PERSPECTIVES

C ONCLUSIONS ET PERSPECTIVES

L'ensemble de mes travaux de thèse a concerné l'évolution du génome allopolyploïde du colza, en m'intéressant tout particulièrement aux premières phases de formation de ce génome allotétraploïde à partir des deux génomes diploïdes du chou et de la navette. L'intérêt de disposer de colzas néo-synthétisés est de mimer l'événement de spéciation de *B. napus*, et d'essayer de décrire puis d'intégrer les mécanismes moléculaires impliqués dans la coexistence de ces deux génomes différenciés. Deux étapes peuvent ainsi être distinguées : **l'hybridation**, correspondant à l'étape de mise en contact des deux génomes et caractérisée par son « choc génomique » (McClintock 1984) ; et **l'allopolyploïdisation**, où l'on assiste à la duplication complète des deux génomes concernés, permettant de passer du génome homoploïde (AC) au génome allopolyploïde (AACC). Alors que bon nombre d'études menées sur les systèmes polyploïdes végétaux et la compréhension de leur formation et évolution ont concerné la régulation de l'expression des gènes (essentiellement au niveau des transcrits, et parfois à celui des protéines voire des phénotypes), je me suis intéressée aux éléments considérés de façon générale comme non codants : **les éléments transposables**, et de manière plus spécifique à mon projet de thèse, **les petits ARN non codants**.

Un fort impact de l'allopolyploïdie sur les petits ARN non codants n'autorisant qu'une réaction modérée des éléments transposables à cet événement

J'ai démontré pour *Bot1*, un élément CACTA de classe II de l'ordre des transposons à TIR, que peu d'événements de transposition apparaissaient au cours des premières générations des néo-allopolyploïdes (toutes identifiées en S2), suggérant un relâchement du contrôle habituel de ce transposon, par ailleurs spécifique du génome C du chou. Par ailleurs, les profils SSAP développés sur *Bot1* et les deux autres éléments transposables étudiés, un rétrotransposon à LTR (*Athila*) et un MITE (*BraSto* – également de l'ordre des transposons à TIR), ont suggéré des modifications des niveaux de méthylation de l'ADN aux sites d'insertion de ces différents ET. Néanmoins, aucune transposition n'avait pu être mise en évidence pour *Athila* et *BraSto*. Ces résultats nous ont donc amenés à considérer que l'impact de l'allopolyploïdie sur ces trois ET était relativement modéré (Sarilar et al. 2013). L'analyse que j'ai ensuite menée sur les petits ARN non codants en réponse à ce même événement d'allopolyploïdie, nous permet de comprendre un peu mieux la réaction de ces éléments transposables.

Pour les transposons à TIR, il est apparu que les populations de 21-nt originaires de ces ET et impliqués dans leur régulation au niveau post-transcriptionnel, n'étaient pas significativement affectées par l'événement d'allopolyploïdie (du moins d'un point de vue global). Par contre les populations de 24-nt, ciblant les mêmes ET mais au niveau transcriptionnel, par méthylation des sites où sont insérés ces transposons à TIR, présentaient une réduction de leurs effectifs préférentiellement en S1, et éventuellement en S0 comme

clairement illustré avec TIR-siRNA24_37 de 24-nt, donné en Figure 4 de l'article Martinez Palacios et al. (en préparation), avec des niveaux similaires à l'attendu des moyennes parentales en S5. Nous pouvons donc établir un scénario où l'allopolyploïdisation, et de façon moindre l'hybridation, aboutit à une réduction du contrôle transcriptionnel des transposons à TIR en générations très précoces – moins de siRNAs de 24-nt impliqués dans l'hétérochromatinisation des régions riches en ET. Par contre, le contrôle post-transcriptionnel n'est que peu modifié mais ne peut répondre de manière complètement efficace au nombre des transposons nouvellement transcrits à réprimer, autorisant ainsi quelques événements de transposition. C'est ce que j'ai observé avec l'élément *Bot1*. Cependant, le contrôle transcriptionnel se remet très rapidement en place et commence à revenir à des niveaux habituels dès la génération S3, bloquant ainsi très vite l'activité de transposition des transposons à TIR potentiellement actifs (Figure 4.1.). Ce scénario a l'avantage de permettre de comprendre que l'élément *Bot1*, spécifique du génome C, n'ait pas envahi le génome A qui en était dépourvu, lors de la formation du génome du colza (Alix et al. 2008) : les mécanismes de contrôle par petits ARN, réagissant très vite après allopolyploïdisation, n'ont pas permis d'explosion de transposition, mais éventuellement quelques transpositions ponctuelles.

En ce qui concerne **les rétrotransposons à LTR** (Figure 4.1.), le mécanisme de répression mis en place de suite après l'événement d'allopolyploïdie apparaît encore plus efficace. Nous avons démontré qu'il y avait production immédiate de petits ARN de 21-nt, à rôle de dégradation des transcrits (PTGS) des rétrotransposons à LTR ayant eu la possibilité de réaliser leur transcription, empêchant par là-même toute possibilité de nouvelle transposition. Par la suite, en génération S5 se surimpose un mécanisme de répression transcriptionnelle (TGS) avec une production fortement accrue des populations de 24-nt qui vont réprimer les ET par hétérochromatinisation des régions où ils sont localisés. La transcription des rétrotransposons à LTR se trouve à nouveau sous contrôle.

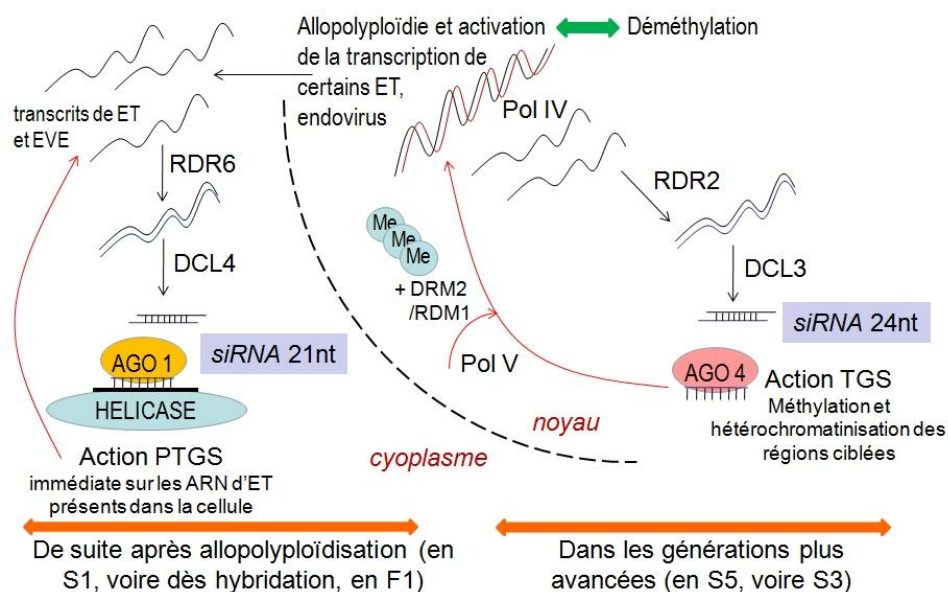


Figure 4.1. Scénario proposé pour la mise en place du contrôle des éléments transposables et de certains endovirus, suite à l'événement d'allopolyploïdie du colza.

Cette répression globale des ET par les populations de petits ARN qui se mobilisent de suite après l'événement d'allopolyploïdisation, voire même au cours du processus avec une mobilisation de certaines des populations de petits ARN dès l'hybridation, apparaît comme une des étapes clés de la réussite de la formation d'un génome allopolyploïde – correspondant à 'l'effet tampon' décrit par Ha et al. (2009) dans leur étude menée sur les allopolyploïdes d'*Arabidopsis*. Cependant, comme nous l'avons clairement mis en évidence, les niveaux de mobilisation des petits ARN sont différents selon leur ET d'origine et par conséquent leur ET cible : cette spécificité peut expliquer les niveaux de transcrits des ET complètement imprévisibles suite à l'événement d'allopolyploïdie, où seule une règle d'identité de séquence semble être à l'origine du contrôle subit par les ET (Sarilar 2011). Cette caractéristique est peut-être aussi à l'origine de la diversité des remaniements structuraux auxquels est soumis le génome allopolyploïde au cours de ses toutes premières générations (Szadkowski et al. 2010), où les ET sont connus comme étant de bons médiateurs (e.g. chez le blé, Chantret et al. 2005).

Une réactivation potentielle d'endovirus dont la répression immédiate par petits ARN assurerait la réussite de l'hybridation interspécifique et de l'événement d'allopolyploïdie

Un résultat particulièrement original est la mise en évidence de la transcription de séquences de type viral dès la confrontation des deux génomes différenciés chez l'hybride F1, et maintenue au cours des générations précoces de l'allopolyploïde (S0 et S1). Chez les colzas synthétiques étudiés, affichant un niveau de fertilité satisfaisant et une viabilité sur déjà cinq générations d'autofécondation, nous avons au préalable identifié des populations de petits ARN de 21-nt correspondant à ces endovirus (et qui nous ont d'ailleurs permis d'identifier ces séquences virales) et qui présentaient la même dynamique d'expression que ces séquences virales qu'elles sont supposées réprimer *a priori* dès l'hybridation (Figure 4.1.).

Les différents résultats que nous avons ainsi obtenus nous amènent à l'hypothèse que des éléments viraux endogènes (EVE) du génome de *Brassica* se trouveraient réactivés en réponse au stress engendré par l'hybridation interspécifique à l'origine du nouveau génome allopolyploïde. Les EVE ont été mis en évidence récemment pour l'ensemble des génomes eucaryotes ; ils correspondent à des cas d'exaptation et de domestication de virus, tous les types de virus étant concernés (Feschotte & Gilbert, 2012). Devenus endogènes, ils peuvent s'exprimer en réponse aux stress (comme chez *Musa*, Côte et al. 2010). Une hypothèse originale, que nous avançons ici, serait que les réactions auto-immunes avec nécroses responsables de l'incompatibilité hybride entre espèces (Bomblies & Weigel, 2007) résulteraient de la réactivation de certains EVE : exprimés, sans rétrocontrôle par petits ARN, ces EVE conduiraient la plante hybride à répondre à cette « invasion virale » par une réaction d'apoptose généralisée. Dans le cas des colzas synthétiques que nous avons étudiés, l'expression de petits ARN spécifiques des EVE réactivés permettrait de juguler ces endovirus et d'assurer la viabilité du nouvel allopolyploïde. Il serait donc intéressant de tester cette hypothèse.

A.-M. Chèvre et son équipe (IGEEP, INRA Rennes) ont à disposition des hybrides de colza impliquant les colzas synthétiques S0, pour lesquels nous avons démontré la présence de transcrits de type viral. Il s'avère que ces croisements entre génomes de type colza

conduisent à des plantes qui meurent précocement, développant des symptômes de viroses dont les causes demeurent inconnues (Figure 4.2.). Le séquençage haut débit des populations



Figure 4.2. Développement de nécroses précoces chez des hybrides « colza synthétique S0 BoEMZ₂₃₈ × variété de colza Darmor ». Photos © A.-M. Chèvre.

de petits ARN et conjointement des transcrits de ces hybrides, au tout début de leur développement, permettrait de vérifier notre hypothèse d'une réactivation de certains EVE, dont le contrôle par petits ARN s'avère ici inefficace voire inexistant. Au préalable, il sera cependant indispensable de décrire plus précisément les EVE que nous suspectons d'être réactivés puis réprimés chez les colzas synthétiques, par étude globale au niveau du génome des *Brassica*. Leur accès semblant difficile par séquençage massif du génome, sans doute en lien avec la présence de ces éléments dans des régions fortement hétérochromatiques, une approche par analyse de BAC pourrait être envisagée – une banque BAC de la variété de colza Darmor est disponible au CNRGV de Toulouse et a déjà été utilisée dans le cadre d'un autre projet au laboratoire.

Comprendre la mobilisation des petits ARN non codants dès l'événement d'hybridation interspécifique

Dans le cadre de mon projet de thèse, seules des générations d'allopolypléïdes néo-synthétisés ont été étudiées par rapport aux parents diploïdes, pour leur contenu global en populations de petits ARN. Quelques expérimentations par RT-PCR quantitative ont été réalisées sur les hybrides F1, démontrant que l'hybridation interspécifique avait un impact important sur certains de ces petits ARN (tels que sur les petits ARN de 21-nt correspondant aux séquences de type endovirus). Des études de protéomique comparative menées sur ces mêmes colzas synthétiques au laboratoire avaient d'ailleurs mis en évidence que près de 90% des modifications observées au niveau des protéomes des allotétraploïdes, en comparaison de ceux des progéniteurs diploïdes, se mettaient en place dès l'hybride F1 (Albertin et al. 2006). Une étude complémentaire sur les transcrits devait aussi préciser que ces modifications étaient dans 60% des cas liées à des modifications post-transcriptionnelles (Marmagne et al. 2010). C'est ainsi que j'aurais particulièrement voulu intégrer ces différents hybrides F1 à mon étude de l'impact global de l'allopolypléïdie sur les petits ARN, en dissociant l'hybridation de la duplication de génome. Nous avons donc au cours de ma thèse initié cette analyse, par la mise en place d'une collaboration avec la plateforme de séquençage à haut débit d'ImaGif (équipe de C. Thermes) au CNRS de Gif-sur-Yvette. La mise au point de la technique de séquençage des petits ARN, visant à optimiser la détection et l'évaluation quantitative des séquences de micro ARN (souvent « diluées » dans les séquences de siRNAs) est actuellement en cours.

L'accès aux données de la distribution des petits ARN dans les hybrides F1 nous renseignera quant à l'impact de l'hybridation par rapport à celui de la duplication de génome, que nous avons caractérisé à la génération S1 par rapport à la génération S5. Ces rôles différenciés de l'hybridation et de l'allopolyploïdisation sur l'expression des petits ARN ont déjà été observés avec l'étude de néo-allopolyploïdes de blé au niveau de l'expression des siRNAs où seule l'allopolyploïdisation semblait présenter un impact majeur (Kenan-Eichler et al. 2011). Il sera intéressant d'étudier la généralité de ce processus au niveau du système polyploïde colza, pour lequel nous avons cependant déjà observé une mobilisation de certaines populations de petits ARN de 21-nt dès l'hybridation.

Passer d'une étude globale à une étude ciblée de la réponse des populations de petits ARN non codants à l'allopolyploïdie

Mon projet de thèse représente la première étude de la dynamique d'expression des petits ARN non codants en réponse à l'allopolyploïdie du colza, menée **à une échelle globale**. Elle a permis de démontrer la mise en place d'une réponse immédiate et transitoire de ces petits ARN à l'allopolyploïdie, les petits ARN ciblant les éléments transposables et autres endovirus du génome des *Brassica* présentant les plus gros changements d'expression. Mais un des atouts de cette étude est d'avoir fourni une importante base de données regroupant une collection de petits ARN pouvant s'exprimer chez les *Brassica*. Par exemple, mon travail a permis de prédire pas moins de 274 nouveaux miRNAs, représentant autant de séquences d'intérêt à étudier quand on s'intéresse à la régulation de l'expression des gènes chez *Brassica*. Ce travail est également une base de départ essentielle pour affiner la compréhension de certaines régulations observées au cours de l'événement d'allopolyploïdie du colza, **à une échelle plus ciblée** :

➤ **La régulation des ADN ribosomiques et la dominance nucléolaire** - Les ADNr composent les régions NOR, *nucleolar organizer regions*, sites majeurs de l'expression des gènes puisqu'ils correspondent à l'endroit où sont synthétisés les ribosomes impliqués dans la traduction des protéines. Or chez des hybrides et allopolyploïdes (notamment chez *A. suecica*), il a été démontré que les gènes codant les ARNr issus d'un des parents sont souvent réprimés : on parle de **dominance nucléolaire**. Il a ainsi été démontré que chez les colzas néo-synthétisés, ce sont les ADNr du génome C du chou qui apparaissent réprimés, et ce dès la formation des hybrides F1 (Ksiazczyk et al. 2011). Il est apparu que cette répression était le résultat d'une augmentation du niveau de méthylation aux sites NOR du génome C, alors que ceux du génome A, toujours actifs, étaient plutôt déméthylés. Il apparaît aujourd'hui pertinent d'envisager un rôle des petits ARN dans la mise en place de l'hyperméthylation des sites NOR du génome C, par un processus de type RdDM (cf. 3.2.3.). Les données de séquençage de banques de petits ARN que j'ai analysées ont notamment été filtrées par rapport aux séquences correspondant aux ADNr ; les étudier de manière ciblée permettra d'identifier des populations candidates pour leur implication dans la répression différentielle des sites NOR du génome C. A noter que j'ai déjà pu observer une surabondance des petits ARN de 21-nt ciblant spécifiquement les ARNr, par rapport à ce que j'ai pu identifier dans la fraction de 24-nt, suggérant la mise en place d'une régulation post-transcriptionnelle de l'expression des gènes codant les ARN ribosomiques qui mériterait d'être étudiée plus en détails. Cette analyse

contribuerait à une meilleure connaissance des mécanismes de la dominance nucléolaire, étape nécessaire à la régulation de la machinerie de la traduction protéique du néo-allopolyploïde. Un projet en collaboration avec A. Kovarik en République Tchèque a d'ailleurs été déposé sur le sujet (projets bilatéraux du programme BARRANDE 2014).

➤ **Les modifications de la régulation de l'expression des gènes et l'hétérosis** – L'un des avantages majeurs des hybrides et allopolyploïdes est l'expression de l'hétérosis (défini à l'origine chez des hybrides intra-spécifiques, comme le jeu de la confrontation d'allèles différents dans un seul et même génome). Ce phénomène (être meilleur que le meilleur des deux parents) résulte de modifications de l'expression des gènes. Comprendre les mécanismes moléculaires à l'origine de l'expression de la **vigueur hybride** demeure un enjeu majeur en génétique végétale, surtout au niveau des espèces cultivées. Il serait intéressant de s'intéresser préférentiellement aux petits ARN de type micro ARN, connus pour être impliqués dans la régulation de l'expression des gènes. L'analyse globale n'a pas mis en évidence de réponse majeure de la fraction des micro ARN à l'allopolyploïdie, mais désormais un travail très ciblé peut être envisagé pour aller rechercher parmi les séquences annotées comme miRNAs (connus et prédits), celles qui présentent des profils différentiels. Identifier les miRNAs présentant des profils d'expression spécifiques en réponse à l'allopolyploïdie puis caractériser leurs cibles géniques potentielles (par analyse de banques de dégradome, par exemple) contribuerait à repérer les cibles candidates de l'hétérosis.

Enfin, je tiens à faire remarquer que mon projet de thèse a été rendu possible par l'accès aux nouvelles techniques de séquençage haut débit (NGS, *next generation sequencing*) permettant d'accéder à l'ensemble des petits ARN non codants exprimés à un moment donné dans un génotype. J'ai également profité de cette technologie au moment de l'exploitation de données *RNAseq* du transcriptome de méiocytes de variétés de colza (coll. A. Lloyd, E. Jenczewski – IJPB, INRA Versailles), conduisant à l'identification de transcrits de séquences virales exprimées chez des haploïdes et des néo-allopolyploïdes de colza. Ces approches haut débit sont un moyen efficace pour générer de l'information à partir de laquelle des questions biologiques peuvent être posées. Les approches NGS représentent ainsi un énorme atout pour répondre à des phénomènes biologiques qui ne pouvaient pas être abordés (ni envisagés) avant le développement de ces techniques.

REFERENCES BIBLIOGRAPHIQUES

Bibliographie

- Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4649–4654.
- Adams, K.L., Percifield, R., Wendel, J.F., 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168, 2217–2226.
- Adams, K.L., Wendel, J.F., 2005. Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8, 135–141.
- Addo-Quaye, C., Snyder, J.A., Park, Y.B., Li, Y.-F., Sunkar, R., Axtell, M.J., 2009. Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA* 15, 2112–2121.
- Ainouche, M.L., Baumel, A., Salmon, A., Yannic, G., 2003. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytol.* 161, 165–172.
- Albertin, W., Alix, K., Balliau, T., Brabant, P., Davanture, M., Malosse, C., Valot, B., Thiellement, H., 2007. Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization. *BMC Genomics* 8, 56.
- Albertin, W., Balliau, T., Brabant, P., Chèvre, A.-M., Eber, F., Malosse, C., Thiellement, H., 2006. Numerous and Rapid Nonstochastic Modifications of Gene Products in Newly Synthesized *Brassica napus* Allotetraploids. *Genetics* 173, 1101–1113.
- Albertin, W., Brabant, P., Catrice, O., Eber, F., Jenczewski, E., Chèvre, A.-M., Thiellement, H., 2005. Autopolyploidy in cabbage (*Brassica oleracea* L.) does not alter significantly the proteomes of green tissues. *Proteomics* 5, 2131–2139.
- Alix, K., Joets, J., Ryder, C.D., Moore, J., Barker, G.C., Bailey, J.P., King, G.J., (Pat) Heslop-Harrison, J.S., 2008. The CACTA transposon *Bot1* played a major role in *Brassica* genome divergence and gene proliferation. *Plant J.* 56, 1030–1044.
- Alix, K., Ryder, C.D., Moore, J., King, G.J., Pat Heslop-Harrison, J.S., 2005. The genomic organization of retrotransposons in *Brassica oleracea*. *Plant Mol. Biol.* 59, 839–851.
- Allen, E., Xie, Z., Gustafson, A.M., Carrington, J.C., 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121, 207–221.
- Allender, C.J., King, G.J., 2010. Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biol.* 10, 54.
- Ambros, V., 2001. microRNAs: tiny regulators with great potential. *Cell* 107, 823–826.
- Andorf, S., Selbig, J., Altmann, T., Poos, K., Witucka-Wall, H., Repsilber, D., 2009. Enriched partial correlations in genome-wide gene expression profiles of hybrids (*A. thaliana*): a systems biological approach towards the molecular basis of heterosis. *Theor. Appl. Genet.* 120, 249–259.
- Axtell, M.J., 2013. Classification and Comparison of Small RNAs from Plants. *Annu. Rev. Plant Biol.* 64, 137–159.
- Barber, W.T., Zhang, W., Win, H., Varala, K.K., Dorweiler, J.E., Hudson, M.E., Moose, S.P., 2012. Repeat associated small RNAs vary among parents and following hybridization in maize. *Proc. Natl. Acad. Sci.* 109, 10444–10449.
- Baroux, C., Pien, S., Grossniklaus, U., 2007. Chromatin modification and remodeling during early seed development. *Curr. Opin. Genet. Dev.* 17, 473–479.
- Baroux, C., Raissig, M.T., Grossniklaus, U., 2011. Epigenetic regulation and reprogramming during gamete formation in plants. *Curr. Opin. Genet. Dev.* 21, 124–133.

- Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- Bernstein, E., Caudy, A.A., Hammond, S.M., Hannon, G.J., 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366.
- Birchler, J.A., Auger, D.L., Riddle, N.C., 2003. In Search of the Molecular Basis of Heterosis. *Plant Cell* 15, 2236–2240.
- Bird, C.P., Stranger, B.E., Liu, M., Thomas, D.J., Ingle, C.E., Beazley, C., Miller, W., Hurles, M.E., Dermitzakis, E.T., 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biol.* 8, R118.
- Bomblies, K., Weigel, D., 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet.* 8, 382–393.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., Zhu, J.-K., 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 123, 1279–1291.
- Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., Voinnet, O., 2008. Widespread Translational Inhibition by Plant miRNAs and siRNAs. *Science* 320, 1185–1190.
- Buggs, R.J.A., Zhang, L., Miles, N., Tate, J.A., Gao, L., Wei, W., Schnable, P.S., Barbazuk, W.B., Soltis, P.S., Soltis, D.E., 2011. Transcriptomic Shock Generates Evolutionary Novelty in a Newly Formed, Natural Allopolyploid Plant. *Curr. Biol.* 21, 551–556.
- Bushell, C., Spielman, M., Scott, R.J., 2003. The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species. *Plant Cell* 15, 1430–1442.
- C.L. Mc Grath et Lynch, 2012. Evolutionary significance of whole-genome duplication, in: *Ploidy and Genome Evolution*. Pamela S. Soltis et Douglas E. Soltis.
- Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., Martienssen, R.A., 2012. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell* 151, 194–205.
- Calarco, J.P., Martienssen, R.A., 2011. Genome reprogramming and small interfering RNA in the Arabidopsis germline. *Curr. Opin. Genet. Dev.* 21, 134–139.
- Castel, S.E., Martienssen, R.A., 2013. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* 14, 100–112.
- Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.-F., Cattivico, L., Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P., Chalhou, B., 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 17, 1033–1045.
- Chen, Z.J., 2007. Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. *Annu. Rev. Plant Biol.* 58, 377–406.
- Chen, Z.J., 2010. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15, 57.
- Chen, Z.J., 2013. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14, 471–482.
- Chen, Z.J., Comai, L., Pikaard, C.S., 1998. Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in Arabidopsis allopolyploids. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14891–14896.
- Chen, Z.J., Pikaard, C.S., 1997. Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3442–3447.

- Cifuentes, M., Eber, F., Lucas, M.-O., Lode, M., Chevre, A.-M., Jenczewski, E., 2010. Repeated Polyploidy Drove Different Levels of Crossover Suppression between Homoeologous Chromosomes in *Brassica napus* Allohaploids. *PLANT CELL ONLINE* 22, 2265–2276.
- Comai, L., 2000. Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* 43, 387–399.
- Comai, L., 2005. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846.
- Crow, K.D., Wagner, G.P., 2006. What Is the Role of Genome Duplication in the Evolution of Complexity and Diversity? *Mol. Biol. Evol.* 23, 887–892.
- Cui, L., 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749.
- Cuperus, J.T., Fahlgren, N., Carrington, J.C., 2011. Evolution and Functional Diversification of MIRNA Genes. *Plant Cell* 23, 431–442.
- Dahal, D., Mooney, B.P., Newton, K.J., 2012. Specific changes in total and mitochondrial proteomes are associated with higher levels of heterosis in maize hybrids: *2D DIGE of heterosis in maize*. *Plant J.* 72, 70–83.
- Dobzhansky, T., 1936. Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila Pseudoobscura* Hybrids. *Genetics* 21, 113–135.
- Doyle, J.J., Flagel, L.E., Paterson, A.H., Rapp, R.A., Soltis, D.E., Soltis, P.S., Wendel, J.F., 2008. Evolutionary Genetics of Genome Merger and Doubling in Plants. *Annu. Rev. Genet.* 42, 443–461.
- Dunoyer, P., Brosnan, C.A., Schott, G., Wang, Y., Jay, F., Alioua, A., Himber, C., Voinnet, O., 2010. An endogenous, systemic RNAi pathway in plants. *EMBO J.* 29, 1699–1712.
- Dunoyer, P., Schott, G., Himber, C., Meyer, D., Takeda, A., Carrington, J.C., Voinnet, O., 2010. Small RNA Duplexes Function as Mobile Silencing Signals Between Plant Cells. *Science* 328, 912–916.
- Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498.
- Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., Carrington, J.C., 2010. MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22, 1074–1089.
- Fawcett, J.A., Maere, S., Van de Peer, Y., 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc. Natl. Acad. Sci.* 106, 5737–5742.
- Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.
- Flagel, L., Udall, J., Nettleton, D., Wendel, J., 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 6, 16.
- Flagel, L.E., Wendel, J.F., 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 186, 184–193.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.

- Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., Mummenhoff, K., 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* 16, 108–116.
- Fujimoto, R., Taylor, J.M., Shirasawa, S., Peacock, W.J., Dennis, E.S., 2012. Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci.* 109, 7109–7114.
- Gaeta, R.T., Yoo, S.-Y., Pires, J.C., Doerge, R.W., Chen, Z.J., Osborn, T.C., 2009. Analysis of Gene Expression in Resynthesized *Brassica napus* Allopolyploids Using *Arabidopsis* 70mer Oligo Microarrays. *PLoS ONE* 4.
- Groszmann, M., Greaves, I.K., Albertyn, Z.I., Scofield, G.N., Peacock, W.J., Dennis, E.S., 2011. Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci.* 108, 2617–2622.
- Guo, M., 2004. Allelic Variation of Gene Expression in Maize Hybrids. *PLANT CELL ONLINE* 16, 1707–1716.
- Guo, M., Davis, D., Birchler, J.A., 1996. Dosage Effects on Gene Expression in a Maize Ploidy Series. *Genetics* 142, 1349–1355.
- Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K.D., Chapman, E.J., Carrington, J.C., Chen, X., Wang, X.-J., Chen, Z.J., 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17835–17840.
- Haag, J.R., Pikaard, C.S., 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* 12, 483–492.
- Haig, D., Westoby, M., 2006. An earlier formulation of the genetic conflict hypothesis of genomic imprinting. *Nat. Genet.* 38, 271.
- Hamilton, A.J., Baulcombe, D.C., 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952.
- Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J., 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293–296.
- Hammond, S.M., Caudy, A.A., Hannon, G.J., 2001. Post-transcriptional gene silencing by double-stranded RNA. *Nat. Rev. Genet.* 2, 110–119.
- He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., Qi, Y., Chen, R., Deng, X.-W., 2010. Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *Plant Cell* 22, 17–33.
- Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J., Hiscock, S.J., 2006. Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. *Curr. Biol. CB* 16, 1652–1659.
- Hegarty, M.J., Hiscock, S.J., 2008. Genomic Clues to the Evolutionary Success of Polyploid Plants. *Curr. Biol.* 18, R435–R444.
- Hoecker, N., Lamkemeyer, T., Sarholz, B., Paschold, A., Fladerer, C., Madlung, J., Wurster, K., Stahl, M., Piepho, H.-P., Nordheim, A., Hochholdinger, F., 2008. Analysis of nonadditive protein accumulation in young primary roots of a maize (*Zea mays* L.) F1-hybrid compared to its parental inbred lines. *PROTEOMICS* 8, 3882–3894.
- Hovav, R., Udall, J.A., Chaudhary, B., Rapp, R., Flagel, L., Wendel, J.F., 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc. Natl. Acad. Sci.* 105, 6191–6195.
- Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., Carrington, J.C., 2007. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* 19, 926–942.

- Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., Zilberman, D., 2009. Genome-Wide Demethylation of Arabidopsis Endosperm. *Science* 324, 1451–1454.
- Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., Zilberman, D., 2012. Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. *Science* 337, 1360–1364.
- Iida, S., Morita, Y., Choi, J.-D., Park, K.-I., Hoshino, A., 2004. Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories. *Adv. Biophys.* 38, 141–159.
- Incarbone, M., Dunoyer, P., 2013. RNA silencing and its suppression: novel insights from in planta analyses. *Trends Plant Sci.* 18, 382–392.
- Jackson, S., Chen, Z.J., 2010. Genomic and Expression Plasticity of Polyploidy. *Curr. Opin. Plant Biol.* 13, 153–159.
- Jeffrey Chen, Z., Ni, Z., 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 28, 240–252.
- Jenczewski, E., 2013. Evolution: He Who Grabs Too Much Loses All. *Curr. Biol.* 23, R961–R963.
- Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100.
- Jones-Rhoades, M.W., 2011. Conservation and divergence in plant microRNAs. *Plant Mol. Biol.* 80, 3–16.
- Jones-Rhoades, M.W., Bartel, D.P., Bartel, B., 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53.
- Jullien, P.E., Berger, F., 2008. L’empreinte parentale, un phénomène épigénétique essentiel pour la reproduction des plantes. *médecine/sciences* 24, 753–757.
- Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., Berger, F., 2012. DNA Methylation Dynamics during Sexual Reproduction in *Arabidopsis thaliana*. *Curr. Biol.* 22, 1825–1830.
- Kashkush, K., Khasdan, V., 2007. Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics* 177, 1975–1985.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Carrington, J.C., 2007. Genome-Wide Profiling and Analysis of Arabidopsis siRNAs. *PLoS Biol.* 5.
- Katiyar-Agarwal, S., Jin, H., 2007. Discovery of pathogen-regulated small RNAs in plants. *Methods Enzymol.* 427, 215–227.
- Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr, Zhu, J.-K., Staskawicz, B.J., Jin, H., 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18002–18007.
- Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M., Levy, A.A., 2011. Wheat Hybridization and Polyploidization Results in Deregulation of Small RNAs. *Genetics* 188, 263–272.
- Książczyk, T., Kovarik, A., Eber, F., Huteau, V., Khaitova, L., Tesarikova, Z., Coriton, O., Chèvre, A.-M., 2011. Immediate unidirectional epigenetic reprogramming of NORs occurs independently of rDNA rearrangements in synthetic and natural forms of a polyploid species *Brassica napus*. *Chromosoma* 120, 557–571.

- Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220.
- Le Trionnaire, G., Twell, D., 2010. Small RNAs in angiosperm gametophytes: from epigenetics to gamete development. *Genes Dev.* 24, 1081–1085.
- Lee, H.-S., Chen, Z.J., 2001. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6753–6758.
- Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854.
- Leitch, A.R., Leitch, I.J., 2008. Genomic Plasticity and the Diversity of Polyploid Plants. *Science* 320, 481–483.
- Lelandais-Brière, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J., Crespi, M., 2009. Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* 21, 2780–2796.
- Levy, A.A., Feldman, M., 2002. The impact of polyploidy on grass genome evolution. *Plant Physiol.* 130, 1587–1593.
- Lippman, Z., Martienssen, R., 2004. The role of RNA interference in heterochromatic silencing. *Nature* 431, 364–370.
- Lisch, D., 2009. Epigenetic Regulation of Transposable Elements in Plants. *Annu. Rev. Plant Biol.* 60, 43–66.
- Lukens, L.N., Pires, J.C., Leon, E., Vogelzang, R., Oslach, L., Osborn, T., 2006. Patterns of Sequence Loss and Cytosine Methylation within a Population of Newly Resynthesized *Brassica napus* Allopolyploids. *Plant Physiol.* 140, 336–348.
- Lynch, M., 2007. The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* 8, 803–813.
- Lynch, M., Conery, J., 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* 290, 1151–1155.
- Lysak, M.A., 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Res.* 15, 516–525.
- Madlung, A., Comai, L., 2004. The Effect of Stress on Genome Regulation and Structure. *Ann. Bot.* 94, 481–495.
- Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., Bartel, D.P., 2004. MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. *EMBO J.* 23, 3356–3364.
- Marcon, C., Schützenmeister, A., Schütz, W., Madlung, J., Piepho, H.-P., Hochholdinger, F., 2010. Nonadditive Protein Accumulation Patterns in Maize (*Zea mays* L.) Hybrids during Embryo Development. *J. Proteome Res.* 9, 6511–6522.
- Marmagne, A., Brabant, P., Thiellement, H., Alix, K., 2010. Analysis of gene expression in resynthesized *Brassica napus* allotetraploids: transcriptional changes do not explain differential protein regulation. *New Phytol.* 186, 216–227.
- Martienssen, R.A., 2010. Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interplod hybrids of *Arabidopsis*. *New Phytol.* 186, 46–53.
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., Matzke, A.J., 2009. RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* 21, 367–376.
- McClintock, B., 1984. The significance of responses of the genome to challenge. *Science* 226, 792–801.
- Mello, C.C., Conte, D., Jr, 2004. Revealing the world of RNA interference. *Nature* 431, 338–342.

- Melnyk, C.W., Molnar, A., Bassett, A., Baulcombe, D.C., 2011. Mobile 24 nt Small RNAs Direct Transcriptional Gene Silencing in the Root Meristems of *Arabidopsis thaliana*. *Curr. Biol.* 21, 1678–1683.
- Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., Poethig, R.S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., Zhu, J.-K., 2008. Criteria for Annotation of Plant MicroRNAs. *PLANT CELL ONLINE* 20, 3186–3190.
- Mirouze, M., 2012. The Small RNA-Based Odyssey of Epigenetic Information in Plants: From Cells to Species. *DNA Cell Biol.* 31, 1650–1656.
- Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the *agouti* locus in the mouse. *Nature Genet.* 23, 314–318 (1999).
- Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., Baulcombe, D.C., 2009. Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* 460, 283–286.
- Mun, J.-H., Kwon, S.-J., Seol, Y.-J., Kim, J.A., Jin, M., Kim, J.S., Lim, M.-H., Lee, S.-I., Hong, J.K., Park, T.-H., Lee, S.-C., Kim, B.-J., Seo, M.-S., Baek, S., Lee, M.-J., Shin, J.Y., Hahn, J.-H., Hwang, Y.-J., Lim, K.-B., Park, J.Y., Lee, J., Yang, T.-J., Yu, H.-J., Choi, I.-Y., Choi, B.-S., Choi, S.R., Ramchiary, N., Lim, Y.P., Fraser, F., Drou, N., Soumpourou, E., Trick, M., Bancroft, I., Sharpe, A.G., Parkin, I.A., Batley, J., Edwards, D., Park, B.-S., 2010. Sequence and structure of *Brassica rapa* chromosome A3. *Genome Biol.* 11, R94.
- Napoli, C., Lemieux, C., Jorgensen, R., 1990. Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* 2, 279–289.
- Ng, D.W.-K., Lu, J., Chen, Z.J., 2012. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr. Opin. Plant Biol.* 15, 154–161.
- Ni, Z., Kim, E.-D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., Chen, Z.J., 2008. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457, 327–331.
- Ossowski, S., Schwab, R., Weigel, D., 2008. Gene silencing in plants using artificial microRNAs and other small RNAs: Engineering small RNA-mediated gene silencing. *Plant J.* 53, 674–690.
- Otto, S.P., Whitton, J., 2000. POLYPLOID INCIDENCE AND EVOLUTION. *Annu. Rev. Genet.* 34, 401–437.
- Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B., Grandbastien, M.-A., 2010a. Impact of transposable elements on the organization and function of allopolyploid genomes: Research review. *New Phytol.* 186, 37–45.
- Parisod, C., Holderegger, R., Brochmann, C., 2010b. Evolutionary consequences of autopolyploidy: Research review. *New Phytol.* 186, 5–17.
- Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.-A., Ainouche, M., 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol.* 184, 1003–1015.
- Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., Poethig, R.S., 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3691–3696.
- Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q.T., Lim, K.Y., Kovarik, A., Leitch, A.R., Grandbastien, M.-A., Mhiri, C., 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytol.* 186, 135–147.

- Pontes, O., Neves, N., Silva, M., Lewis, M.S., Madlung, A., Comai, L., Viegas, W., Pikaard, C.S., 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proc. Natl. Acad. Sci. U. S. A.* 101, 18240–18245.
- Pumphrey, M., Bai, J., Laudencia-Chingcuanco, D., Anderson, O., Gill, B.S., 2009. Nonadditive Expression of Homoeologous Genes Is Established Upon Polyploidization in Hexaploid Wheat. *Genetics* 181, 1147–1157.
- Qi, B., Huang, W., Zhu, B., Zhong, X., Guo, J., Zhao, N., Xu, C., Zhang, H., Pang, J., Han, F., Liu, B., 2012. Global transgenerational gene expression dynamics in two newly synthesized allohexaploid wheat (*Triticum aestivum*) lines. *BMC Biol.* 10, 3.
- Ramachandran, V., Chen, X., 2008. Small RNA metabolism in *Arabidopsis*. *Trends Plant Sci.* 13, 368–374.
- Ramsey J., Schemske DW, 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29: 467–501.
- Randolph, L.F., 1941. Genetic Characteristics of the B Chromosomes in Maize. *Genetics* 26, 608–631.
- Rong, J., Feltus, F.A., Liu, L., Lin, L., Paterson, A.H., 2010. Gene copy number evolution during tetraploid cotton radiation. *Heredity* 105, 463–472.
- Ruiz-Ferrer, V., Voinnet, O., 2009. Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.* 60, 485–510.
- Salmon, A., Ainouche, M.L., Wendel, J.F., 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol. Ecol.* 14, 1163–1175.
- Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., Feuillet, C., 2008. Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *PLANT CELL ONLINE* 20, 11–24.
- Sarilar, V., Marmagne, A., Brabant, P., Joets, J., Alix, K., 2011. BraSto, a Stowaway MITE from Brassica: recently active copies preferentially accumulate in the gene space. *Plant Mol. Biol.* 77, 59–75.
- Sarilar, V., Palacios, P.M., Rousselet, A., Ridet, C., Falque, M., Eber, F., Chèvre, A.-M., Joets, J., Brabant, P., Alix, K., 2013. Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized Brassica napus allotetraploids. *New Phytol.* 198, 593–604.
- Saze, H., Tsugane, K., Kanno, T., Nishimura, T., 2012. DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. *Plant Cell Physiol.* 53, 766–784.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L.,

- Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu, Y., Jeddelloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., Wilson, R.K., 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326, 1112–1115.
- Schranz, M., Lysak, M., Mitchellolds, T., 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* 11, 535–542.
- Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., Peng, Z., He, G., Zhong, S., Qi, Y., Terzaghi, W., Deng, X.W., 2012. Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two Arabidopsis Ecotypes and Their Reciprocal Hybrids. *Plant Cell* 24, 875–892.
- Shivaprasad, P.V., Dunn, R.M., Santos, B.A., Bassett, A., Baulcombe, D.C., 2011. Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J.*
- Si-Ammour, A., Windels, D., Arn-Boulidoires, E., Kutter, C., Ailhas, J., Meins, F., Vazquez, F., 2011. miR393 and Secondary siRNAs Regulate Expression of the TIR1/AFB2 Auxin Receptor Clade and Auxin-Related Development of Arabidopsis Leaves. *PLANT Physiol.* 157, 683–691.
- Simillion, C., Vandepoele, K., Van Montagu, M.C.E., Zabeau, M., Van de Peer, Y., 2002. The hidden duplication past of Arabidopsis thaliana. *Proc. Natl. Acad. Sci.* 99, 13627–13632.
- Siomi, H., Siomi, M.C., 2010. Posttranscriptional Regulation of MicroRNA Biogenesis in Animals. *Mol. Cell* 38, 323–332.
- Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285.
- Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A., Martienssen, R.A., 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell* 136, 461–472.
- Soltis, D.E., Soltis, P.S., Bennett, M.D., Leitch, I.J., 2003. Evolution of genome size in the angiosperms. *Am. J. Bot.* 90, 1596–1603.
- Soltis, P.S., Soltis, D.E., 2004. The origin and diversification of angiosperms. *Am. J. Bot.* 91, 1614–1626.
- Soltis, P.S., Soltis, D.E., 2009. The Role of Hybridization in Plant Speciation. *Annu. Rev. Plant Biol.* 60, 561–588.
- Song, K., Lu, P., Tang, K., Osborn, T.C., 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7719–7723.
- Song, L., Han, M.-H., Lesicka, J., Fedoroff, N., 2007. Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5437–5442.
- Stebbins, G.L., Price, H.J., 1971. The Developmental Genetics of the CALCAROIDES Gene in Barley. I. Divergent Expression at the Morphological and Histological Level. *Genetics* 68, 527–538.

- Stupar, R.M., Springer, N.M., 2006. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. *Genetics* 173, 2199–2210.
- Sun, G., 2011. MicroRNAs and their diverse functions in plants. *Plant Mol. Biol.* 80, 17–36.
- Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., Schnable, P.S., 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6805–6810.
- Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Coriton, O., Jenczewski, E., Chèvre, A.M., 2011. Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytol.* 191, 884–894.
- Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Huneau, C., Belcram, H., Coriton, O., Manzaneres-Dauleux, M.J., Delourme, R., King, G.J., Chalhou, B., Jenczewski, E., Chèvre, A.-M., 2010. The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol.* 186, 102–112.
- Tate, J.A., Joshi, P., Soltis, K.A., Soltis, P.S., Soltis, D.E., 2009. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biol.* 9, 80.
- The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Thiel, T., Graner, A., Waugh, R., Grosse, I., Close, T.J., Stein, N., 2009. Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. *BMC Evol. Biol.* 9, 209.
- Tirosh, I., Reikhav, S., Levy, A.A., Barkai, N., 2009. A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science* 324, 659–662.
- Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L., Vandepoele, K., 2009a. The flowering world: a tale of duplications. *Trends Plant Sci.* 14, 680–688.
- Van de Peer, Y., Maere, S., Meyer, A., 2009b. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732.
- Vaucheret, H., 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* 20, 759–771.
- Vaucheret, H., 2008. Plant ARGONAUTES. *Trends Plant Sci.* 13, 350–358.
- Vaucheret, H., Fagard, M., 2001. Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends Genet. TIG* 17, 29–35.
- Vazquez, F., Blevins, T., Ailhas, J., Boller, T., Meins, F., Jr, 2008. Evolution of Arabidopsis MIR genes generates novel microRNA classes. *Nucleic Acids Res.* 36, 6429–6438.
- Vazquez, F., Hohn, T., 2013. Biogenesis and Biological Activity of Secondary siRNAs in Plants. *Scientifica* 2013.
- Vazquez, F., Legrand, S., Windels, D., 2010. The biosynthetic pathways and biological scopes of plant small RNAs. *Trends Plant Sci.* 15, 337–345.
- Veitia, R.A., Bottani, S., Birchler, J.A., 2008. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 24, 390–397.
- Voinnet, O., 2009. Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell* 136, 669–687.
- Wang, J., Tian, L., Lee, H.-S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn, T.C., Doerge, R.W., Comai, L., Chen, Z.J., 2006. Genomewide Nonadditive Gene Regulation in Arabidopsis Allotetraploids. *Genetics* 172, 507–517.
- Wang, J., Tian, L., Madlung, A., Lee, H.-S., Chen, M., Lee, J.J., Watson, B., Kagochi, T., Comai, L., Chen, Z.J., 2004. Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. *Genetics* 167, 1961–1973.
- Wang, X., Torres, M.J., Pierce, G., Lemke, C., Nelson, L.K., Yuksel, B., Bowers, J.E., Marler, B., Xiao, Y., Lin, L., Epps, E., Sarazen, H., Rogers, C., Karunakaran, S., Ingles, J.,

- Giattina, E., Mun, J.-H., Seol, Y.-J., Park, B.-S., Amasino, R.M., Quiros, C.F., Osborn, T.C., Pires, J., Town, C., Paterson, A.H., 2011. A physical map of Brassica oleracea shows complexity of chromosomal changes following recursive paleopolyploidizations. *BMC Genomics* 12, 470.
- Wassenegger, M., Krczal, G., 2006. Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci.* 11, 142–151.
- Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B., Powell, W., 1997. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet. MGG* 253, 687–694.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982.
- Wierzbicki, A.T., Cocklin, R., Mayampurath, A., Lister, R., Rowley, M.J., Gregory, B.D., Ecker, J.R., Tang, H., Pikaard, C.S., 2012. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev.* 26, 1825–1836.
- Wittkopp, P.J., Haerum, B.K., Clark, A.G., 2004. Evolutionary changes in cis and trans gene regulation. *Nature* 430, 85–88.
- Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L.H., 2009. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci.* 106, 13875–13879.
- Yaakov, B., Kashkush, K., 2011. Methylation, Transcription, and Rearrangements of Transposable Elements in Synthetic Allopolyploids. *Int. J. Plant Genomics* 2011.
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., Chen, X., 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* 307, 932–935.
- Zanca, A.S., Vicentini, R., Ortiz-Morea, F.A., Del Bem, L.E., da Silva, M.J., Vincentz, M., Nogueira, F.T., 2010. Identification and expression analysis of microRNAs and targets in the biofuel crop sugarcane. *BMC Plant Biol.* 10, 260.
- Zaratiegui, M., Martienssen, R.A., 2012. SnapShot: Small RNA-Mediated Epigenetic Modifications. *Cell* 151, 456–456.e1.
- Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., Gonzalez, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., Meyers, B.C., 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* 25, 2540–2553.
- Zhang, X., Borevitz, J.O., 2009. Global Analysis of Allele-Specific Expression in Arabidopsis thaliana. *Genetics* 182, 943–954.
- Zhang, X., Xia, J., Lii, Y., Barrera-Figueroa, B., Zhou, X., Gao, S., Lu, L., Niu, D., Liang, W., Chen, Z., Leung, C., Wong, T., Guo, J., Li, Y., Liu, R., Zhu, J.-K., Zhang, W., Jin, H., Zhang, H., 2012. Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biol.* 13, R20.
- Zhao, X.P., Si, Y., Hanson, R.E., Crane, C.F., Price, H.J., Stelly, D.M., Wendel, J.F., Paterson, A.H., 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res.* 8, 479–492.
- Zheng, H., Qiyang, J., Zhiyong, N., Hui, Z., 2013. Prediction and identification of natural antisense transcripts and their small RNAs in soybean (*Glycine max*). *BMC Genomics* 14, 280.

Response of non-coding components of the genome – transposable elements and small non-coding RNAs – to a new allopolyploidisation event: the genome of oilseed rape (*Brassica napus*) as a model of study.

Abstract. The evolutionary success of polyploid species is partly due to the dynamic changes in genome organization and gene expression patterns that occur at the onset of the polyploid formation. These changes are promoted by the merging of divergent genomes into a single nucleus (i.e. allopolyploidy) that causes a “genomic shock”; they are thought to provide a rich source of new genetic material upon which selection can act to promote adaptation and evolution. Many studies have thus aimed to uncover molecular mechanisms that are responsible for the evolutionary success of allopolyploid species, most of them focusing on gene expression changes. In the present PhD thesis, my interest has been concentrated on the non-coding components of the genome: transposable elements and small non-coding RNAs. My study involves oilseed rape (*Brassica napus*, AACC), a relatively young allopolyploid species that originated from hybridizations between *B. rapa* (AA) and *B. oleracea* (CC). Specifically, I have used resynthesized *B. napus* polyploids advanced by self-pollination of single plants for several generations; I have analyzed these plants at different generations for genomic changes accompanying polyploid formation and subsequent evolution.

In a first part, sequence-specific amplification polymorphism (SSAP) targeting the C genome-specific transposable element *Bot1*, was used to evaluate transposition rate of *Bot1* in resynthesized *B. napus* in comparison with the diploid parents. Only a few transposition events were identified. When combined with the results obtained for two other TEs, this work suggests that allopolyploidy has only a moderate impact on TE transposition and restructuring. The changes observed in SSAP profiles led us to hypothesize that some of them resulted from changes in DNA methylation, resulting in rare but highly specific TE activation and transposition.

In a second part, I have concentrated on small non-coding RNAs (sRNAs), which are thought to mediate different aspects of the response to the “genomic shock” induced by allopolyploid formation. Comprehensive analyses of sRNA expression in resynthesized *B. napus* allopolyploids have been carried out by deep sequencing sRNAs from 11 libraries prepared from stems of three allotetraploids (surveyed at the two generations S1 and S5) and the two diploid parents. Characterization of sRNA distributions in these plants indicates that sRNAs show an immediate but transient response to allopolyploidy. The sRNAs derived from transposable elements (down-regulated in the S1) or targeting unknown sequences (no Blast hit against any available public database) were particularly affected. The use of *B. napus* mRNAseq data revealed that these latest unknown candidates, which are 21-nt long and over-expressed in the earliest generations (F1, S0, S1) were derived from endogenous viral elements (EVE). We confirmed that these EVEs showed the same expression patterns as the 21-nt long sRNAs that specifically target them (over-expression in the F1, S0 and S1). These results suggest that (at least) some EVEs might be reactivated as a response to the merging of divergent genomes (in interspecific hybrids and newly formed allopolyploids).

Altogether, our results have demonstrated a succession of sRNA pathways that counteract the reactivation of some specific TEs and/or EVEs at the onset of polyploid formation; reactivated TEs and/or EVEs being immediately repressed at the post-transcriptional level (PTGS), and then fully repressed by transcriptional gene silencing (TGS) in the subsequent generations. Such data lead to hypothesize that sRNAs are essential to overcome interspecific hybrid incompatibilities due to the uncontrolled and deleterious reactivation of TEs / EVEs. Therefore, sRNAs should be considered as the guardians of genome integrity even in newly-formed allopolyploids.

Key words. Allopolyploidy, *Brassica*, transposable elements, endogenous viral elements (EVEs), micro RNAs (miRNAs), small interfering RNAs (siRNAs), small non-coding RNAs, next generation sequencing (NGS), sequence-specific amplification polymorphism (SSAP).

Réponse des agents non codants du génome – éléments transposables et petits ARN – à un événement d'allopolyploïdie : le génome du colza (*Brassica napus*) comme modèle d'étude.

Résumé. Le succès évolutif de la polyploïdie, notamment de l'allopolyploïdie (où la duplication de génome complet est associée à une hybridation entre génomes différenciés) est en partie lié au fait que cet événement s'accompagne de nombreux changements dans l'organisation du génome et la régulation de l'expression des gènes. On parle du « choc génomique » de l'hybridation interspécifique et de l'allopolyploïdie. Ces sources de diversité génétique, à la fois structurale et fonctionnelle, apparaissent utiles et nécessaires à l'adaptation et l'évolution des espèces. Alors que de nombreuses études portant sur la compréhension des mécanismes moléculaires à l'origine du succès des allopolyploïdes ont concerné les modifications de l'expression des gènes, mes travaux de thèse ont porté sur les agents non codants du génome que sont les éléments transposables et les petits ARN non codants. Le modèle d'étude est le colza (*Brassica napus*, AACC), espèce allotétraploïde issue de l'hybridation entre les espèces diploïdes navette (*B. rapa*, AA) et chou (*B. oleracea*, CC). Nous disposons de colzas néo-synthétisés, étudiés à différentes générations d'autofécondation, permettant de caractériser les changements génomiques accompagnant la formation puis l'évolution du génome néo-allopolyploïde.

Une étude a tout d'abord été menée sur un élément transposable (ET) spécifique du génome C, *Bot1*, en vue d'identifier de nouvelles transpositions survenant chez les colzas néo-synthétisés par rapport aux parents diploïdes, par une approche SSAP. Quelques rares événements de transposition ont été identifiés. Ces résultats, confrontés à ceux obtenus sur deux autres ET, ont permis de mettre en évidence un impact modéré de l'allopolyploïdie sur la transposition de ces différents ET. Par contre, il est apparu que des changements de méthylation auraient accompagné cette allopolyploïdisation, sans doute à l'origine de la réactivation et la transposition de quelques copies de *Bot1*.

Les petits ARN non codants ont été suggérés comme impliqués dans les différents événements génomiques accompagnant la formation d'un génome allopolyploïde. Pour étudier la dynamique d'expression des petits ARN chez des colzas néo-synthétisés pris à deux générations d'autofécondation (S1, S5) en comparaison de leurs parents diploïdes, j'ai exploité des données de séquençage haut débit obtenues pour 11 banques construites à partir des tiges de ces différents génotypes. J'ai ainsi démontré, qu'à une échelle globale, les petits ARN présentaient une réponse immédiate mais transitoire à l'événement d'allopolyploïdie. Les fractions particulièrement affectées par l'allopolyploïdie se sont révélées correspondre (1) à des petits ARN interférents dérivés d'éléments transposables avec une baisse de leur abondance en génération précoce S1, et (2) à des populations de petits ARN de 21 nucléotides exprimées uniquement de manière très précoce, de l'hybride F1 à la génération S1. Nous avons notamment identifié des transcrits de type viral correspondant à ces petits ARN de 21-nt, et présentant les mêmes profils d'expression (de l'hybride F1 à la génération S1), suggérant une réactivation d'éléments viraux endogènes (EVE) en réponse à l'hybridation et l'allopolyploïdie.

L'ensemble de mon étude a démontré la mise en place d'une succession des voies de régulation par petits ARN où ET et EVE, réactivés au niveau transcriptionnel, sont immédiatement soumis à une répression post-transcriptionnelle (PTGS), renforcée ensuite par une répression de leur transcription (TGS). L'hypothèse d'une absence de cette régulation par petits ARN lors des phénomènes de nécrose et létalité hybride, amène à envisager ces populations de petits ARN comme les clés de la réussite de la formation d'un génome hybride, où la répression immédiate et efficace des ET et autres endovirus, réactivés suite au choc génomique, se révèle être une nécessité.

Mots clés. Allopolyploïdie, *Brassica*, éléments transposables, éléments viraux endogènes (EVE), micro ARN (miRNAs), petits ARN interférents (siRNAs), petits ARN non codants, séquençage haut débit (NGS), technique SSAP (*sequence-specific amplification polymorphism*).