

Etude du contrôle des éléments transposables par la méthylation de l'ADN chez Arabidopsis thaliana

Mathilde Etcheverry

▶ **To cite this version:**

Mathilde Etcheverry. Etude du contrôle des éléments transposables par la méthylation de l'ADN chez Arabidopsis thaliana. Sciences agricoles. Université Paris Sud - Paris XI, 2013. Français. NNT : 2013PA112208 . tel-01124138

HAL Id: tel-01124138

https://theses.hal.science/tel-01124138

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.

UNIVERSITE DE PARIS-SUD
FACULTE DES SCIENCES D'ORSAY
Ecole doctorale : 426- Gènes, Génomes, Cellules
Laboratoire d'épigénétique et d'épigénomique chez Arabidopsis
Discipline : Biologie

THESE DE DOCTORAT

Présentée pour obtenir le grade de
DOCTEUR EN SCIENCES DE L'UNIVERSITE PARIS XI

Soutenue le 30 Septembre 2013

par
Mathilde Etcheverry

Etude du contrôle des éléments transposables par la méthylation de l'ADN chez *Arabidopsis thaliana*

Directeur de thèse **: Vincent Colot**

JURY

Pascale LESAGE	Rapporteur
Daniel GRIMANELLI	Rapporteur
Stéphane RONSSERAY	Examinateur
Pierre CAPY	Examinateur
Vincent COLOT	Directeur de thèse

Remerciements

Avant toute chose, merci à Vincent, mon directeur de thèse sans qui rien n'aurait été possible. Merci de la confiance que tu m'as accordée, des libertés que tu m'as laissées tout en restant disponible. Merci surtout d'avoir entretenu ma curiosité scientifique, encouragé ma détermination et mon exigence m'ayant ainsi permis de progresser et de me dépasser.

Je tiens à remercier tout spécialement et très sincèrement Felipe Teixeira qui m'a encadrée au jour le jour pendant mon stage de M2 et le début de ma thèse, je lui dois énormément.

Je veux également remercier toutes les personnes, et elles sont nombreuses, que j'ai eu le plaisir de côtoyer, avec qui j'ai pu interagir et travailler au cours de ces quatre années. Tout d'abord, l'ensemble des membres de l'équipe A2e présents et passés. J'ai en effet bénéficié de leur aide et leur soutien tout au long de ma thèse et j'ai eu la chance d'échanger avec chacun d'entre eux, que ce soit sur le plan scientifique et/ou personnel. Mes premières pensées vont à Sandra et Alexis, thésards en même temps que moi, avec qui j'ai partagé au quotidien tous les aléas de la thèse, aussi bien les coups de cafard que les joies, les déceptions que les réussites. Merci à Erwann que j'ai sollicité bien plus que de raison ; à François qui, malgré sa prétendue mauvaise humeur, a toujours pris le temps de répondre mes questions et de discuter ; à Antoine pour son aide et ses conseils ; à Barbara qui a toujours été là dans les moments difficiles ; à Martine pour les discussions variées et animées ; à Benoît pour m'avoir laissée envahir son bureau avec mon fouillis et enfin, merci à Amanda pour l'aide dans les ultimes expériences précipitées de fin de thèse et qui, je l'espère, poursuivra les projets que je laisse inachevés. Merci infiniment à tous les bio-informaticiens, en particulier à Evelyne et Tony, pour leur patience et leur aide précieuse. J'ai également eu l'opportunité d'encadrer plusieurs étudiants et stagiaires : Jérémie, Luc, Adèle et Milos qui m'ont apporté au moins autant que ce que je leur ai apporté.

Merci à tous les membres de la section Génomique Environnementale et Evolutive et plus particulièrement aux équipes de Frédy Barneche et Lionel Navarro. Un grand merci à Clara, Stéphanie et Anne-Sophie pour tous les bons moments passés ensemble, en espérant qu'il y en ait d'autres. Merci à Angélique pour ses remarques et conseils avisés. Merci à Sophie de Peindray et Camille François pour leur disponibilité et leur aide dans toutes les démarches administratives.

Je tiens également à remercier toutes les personnes avec qui j'ai eu la chance de collaborer sur différents projets. Arthur Gilly, Amine Madoui, Jean-Marc Aury, Karine Labadie et Patrick Wincker du Genoscope pour le séquençage des epiRIL et l'analyse des données sur lesquelles sont basés la majorité de mes analyses et résultats. J'ai également eut le plaisir de contribuer aux travaux d'Olivier Voinnet et Arturo Mari-Ordonnez (ETH, Zurich) sur *EVADE*, ainsi qu'à ceux de l'équipe de Tetsuji Kakutani (NIG, Shizuoka) sur *VANDAL21.* Enfin, merci à toutes les personnes qui ont participé plus ou moins directement au crible génétique *FWA* : Christophe Himber (IBP, Strasbourg) pour la mutagénèse EMS ainsi que Christine Camilleri et Lilian Dahuron (INRA de Versailles).

Je voudrais encore remercier les membres de mon comité de thèse, Stéphane Ronsseray, Allison Mallory, Pascale Lesage et ma tutrice Déborah Bourc'his qui se sont réunis à deux reprises pour évaluer mon travail et m'apporter de précieux conseils.

Je remercie vivement mon jury de thèse, mes rapporteurs Pascale Lesage et Daniel Grimanelli, ainsi que mon examinateur Stéphane Ronsseray et le président du Jury, Pierre Capy. Merci de l'attention et de l'intérêt que vous avez porté à mon travail. La discussion de ma soutenance de thèse a été, grâce à vous, très stimulante et enrichissante.

Pour finir, je remercie chaleureusement mes proches, ma famille, mes amis et surtout Jean-Philippe pour leur soutien constant. Ils ont été les premiers à subir mon stress et, bien que je les aie souvent négligés durant ces quatre années, ont toujours été les premiers à me réconforter, me conseiller et m'encourager. Je n'y serais pas arrivée sans vous, merci pour tout.

Sommaire

Abréviations **3**

1 INTRODUCTION **5**

1.1 Présentation de l'organisme model *Arabidopsis thaliana* **6**

1.2 Les éléments transposables (ET) **7**
1.2.1 Définition, découverte et généralités 7
1.2.2 Classification des ET et mécanismes de transposition 7
1.2.3 Les ET d'*Arabidopsis thaliana* 12

1.3 Contrôle des ET via des mécanismes épigénétiques chez Arabidopsis **15**
1.3.1 Etats chromatiniens chez Arabidopsis : généralités 15
1.3.2 La méthylation de l'ADN 17
1.3.3 Autres marques et acteurs associés aux ET 21
1.3.4 Conséquences de la perte du contrôle épigénétique sur les ET 23

1.4 Dynamique du contrôle des ET chez les plantes : au cours du développement et en réponse à l'environnement **25**
1.4.1 Contrôle des ET au moment de la reproduction 25
1.4.2 Comportement des ET en réponses à des variations environnementales 26

1.5 Rôle des ET dans la régulation génique **28**
1.5.1 Implication des ET dans la régulation génique 28
1.5.2 Conséquences de la régulation épigénétique des ET sur les gènes situés à proximité 28
1.5.3 Les éléments transposables à l'origine de la création d'épimutations 30
1.5.4 Impact de la perturbation du contrôle épigénétique des ET sur la régulation des gènes chez *Arabidopsis thaliana* 31

1.6 Objectifs de la thèse **33**

2 ETUDE DES CONSEQUENCES D'UNE PERTE DE METHYLATION DE L'ADN SUR LA MOBILISATION DES ELEMENTS TRANSPOSABLES **34**

2.1 Identification des évènements de transposition ayant eu lieu lors de la génération des epiRIL **35**

2.2 Résultats du séquençage du génome des epiRIL : identité des ET mobilisés, fréquence et dynamique de mobilisation **55**
2.2.1 Analyses globales de la mobilisation des ET dans les epiRIL et le mutant *ddm1* 55
2.2.2 Identification, caractérisation et propriétés des locus donneurs 58
2.2.3 Dynamique de transposition des ET dans les epiRIL 60
2.2.4 Etude comparative des différentes familles comptant des ET mobiles dans les epiRIL 61
2.2.5 Mobilisation des ET en fonction de l'origine parentale du locus donneur et de la présence d'insertions partagées 64
2.2.6 Etude de l'activation en trans des transposons à ADN *ATENSPM3* et *VANDAL21* 66

2.3 Distribution des nouvelles insertions d'ET **68**
2.3.1 Distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis 69
2.3.2 Distribution des nouvelles insertions d'ET en fonction des états chromatiniens 70
2.3.3 Distribution des nouvelles insertions d'ET par rapport aux gènes 72

2.4 Discussion **75**
2.4.1 Remarques générales sur l'approche expérimentale 75
2.4.2 Eléments mobilisés suite à la perte du contrôle épigénétique 76
2.4.3 Distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis 77

3 ETUDE DE L'ETABLISSEMENT DU CONTROLE EPIGENETIQUE SUR LES ELEMENTS TRANSPOSABLES **79**

3.1 Etude de l'établissement de la méthylation de l'ADN sur les copies d'ET nouvellement insérées **80**
3.1.1 Etude comparative de deux rétroéléments à LTR de type COPIA : *ATCOPIA78* et *ATCOPIA93* 80
3.1.2 Etude de la mise en place du contrôle épigénétique sur les transposons à ADN 101
3.1.3 Conclusions et perspectives 105

3.2 Identification de régulateurs du processus d'établissement/rétablissement de la méthylation de l'ADN par crible génétique **107**
3.2.1 Description du crible 107
3.2.2 Caractérisation du locus *FWA* dans les epiRIL et analyses préliminaires 108
3.2.3 Mutagénèse et phénotypage 109
3.2.4 Analyses complémentaires et résultats 111
3.2.5 Discussion et perspectives 112

4 INFLUENCE DE LA METHYLATION DES ELEMENTS TRANSPOSABLES SUR LES GENES SITUES A PROXIMITE **113**

4.1 Conséquences de la perte de la méthylation de l'ADN des ET sur l'expression des gènes situés à proximité **114**

4.2 Influence des nouvelles insertions d'ET sur l'expression des gènes situés à proximité avant et après établissement du contrôle épigénétique **129**

5 CONCLUSIONS GENERALES ET PERSPECTIVES **130**

5.1 Conclusions générales **131**

5.2 Perspectives **133**
5.2.1 Contrôle de la mobilisation des ET chez *Arabidopsis thaliana* 133
5.2.2 Prolifération et élimination des ET chez *Arabidopsis thaliana* 134

6 MATERIEL ET METHODES **136**

RÉFÉRENCES **145**

ANNEXES **159**

Abréviations

5-meC	5méthylcytosine
Ac-Ds	Activator-Dissociation
ADN	Acide désoxyribonucléique
ADNc	Acide désoxyribonucléique complémentaire
AGO	Argonaute
ARN	Acide ribonucléique
ARNi	ARN interférence
ARNm	ARN messager
ARNt	ARN de transfert
ATP	Adénosine triphosphate
ATXR5/6	ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5/6
Avy	agoutiviableyellow
BLAST	Basic Local Alignment Search Tool
BLASTp	protein-protein BLAST
BNS	BONSAI
BS	Bisulfite
CHIP	CHromatin Immuno Precipitation
CMT3	CHROMOMETHYLASE 3
CNR	COLORLESSNON-RIPENING
CS1/2/3/4	ChromatinState1/2/3/4
DCL1-4	Dicer-like1-4
DDM1	DECREASEINDNAMETHYLATION1
DIRS	Dictyostelium intermediate repeat sequences
DME	DEMETER
DML	DEMETER-LIKE
Dnmt1	DNA methyltransferase 1
Dnmt3	DNA methyltransferase 1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
EN/SPM	enhancer/supressor of mutator
epiRIL	Epigenetic Recombinant Inbreed Lines
ET	Elément transposable
FWA	FLOWERING WAGENINGEN
GFF	Generic Format File
H1/2A/2B/3/4	Histone 1/2A/2B/3/4
HDA6	HISTONE DEACETYLASE6
HEN1	HUA ENHANCER 1
het-siARN	heterochromatic-siRNA
HMM	Hidden Markov Model
IBM1	INCREASE INBONSAI METHYLATION 1
JmJ14	JumonJi14
KYP	KRYPTONITE
LINE	Long Interspersed Nuclear Elements
LTR	longterminalrepeat
MEA	MEDEA
MeDIP	Methylated DNA immunoprecipitation
MET1	METHYLTRANSFERASE1
miARN	microARN
MITE	MiniatureInverted-repeatTransposableElements
MOM1	MORPHEUS MOLECULE 1
MORC	Microrchidia

MS	milieu de Murashige et Skoog
Mu	Mutator
MuK	Mu killer
nt	nucléotide
ORF	Open Reading Frame
PBS	primer binding site
PCR	Polymerase Chain Reaction
PLE	Penelope-Like Element
PolII/III/IV/V	RNA polymeraseII/III/IV/V DNA dependant
polyA	polyAdénylé
PPT	polypurine tract
PRC2	Polycomb Repressive Complex 2
PSI-BLAST	Position-Specific Iterated BLAST
PTGS	Post-Transcriptionnal Gene Silencing
qPCR	quantitative Polymerase Chain Reaction
QTL	Quatitative trait locus
RBR1	RETINOBLASTOMA-RELATED 1
RdDM	RNA directed DNA methylation
RDR2/6	RNA DEPENDENT RNA POLYMERASE 2/6
RdRP	ARN polymérase ARN dépendante
RING	Really Interesting New Gene
RISC	RNA induced silencing complex
ROS1	REPRESSOR OF SILENCING 1
RT	Reverse Transcriptase
SDC	SUPRESSOR OF DRM1 DRM2 CMT3
SHH1	SAWADEE HOMEODOMAIN HOMOLOG 1
siARN	Short-interfering RNA
SINE	Short Interspersed Nuclear Elements
SRA	SET- and RING-associated
sRNA-seq	Small RNA Sequencing
SUP	SUPERMAN
SUVH4/5/6	SUPRESSOR OF VARIAGATION HOMOLOG 4/6/9
SWI/SNF2	SWItch/Sucrose NonFermentable
ta-siARN	Trans-acting siRNA
T-DNA	ADN de transfert
TGS	TranscriptionnalGeneSilencing
TIR	TerminalInvertedRepeat
TNPA/B/C/D	Transposition regulatory Protein A/B/C/D
TPRT	Target-site-Primed Reverse Transcription
TSD	Target Site Duplication
Uhrf1	Ubiquitin-like containing PHD and RING finger domains 1
UTR	UnTranslated Region
VIM1	VARIANTINMETHYLATION1
WT	WildType

1
Introduction

1.1 Présentation de l'organisme model *Arabidopsis thaliana*

Arabidopsis thaliana est une plante annuelle de la famille des *Brassicaceae* (*Angiospermes*, *Dicotylédones vraies*) à laquelle appartiennent également le chou et le colza. Certaines de ses caractéristiques comme sa petite taille, sa grande prolificité (plusieurs milliers de graines produites par plante), son cycle de reproduction court (2 mois) ainsi qu'un mode de reproduction essentiellement autogame (98% d'autogamie) permettant l'obtention de lignées pures en font un organisme model de laboratoire idéal.

De plus, son génome diploïde d'une taille de 125MB environs répartit en 5 chromosomes est le plus petit génome végétal connu ce qui facilite les approches génomiques. Il est entièrement séquencé à l'exception des répétitions centromériques (114,5Mb/125Mb) qui sont composées d'un très grand nombre de répétitions en tandem d'une séquence de 180bp (AGI 2000). Les analyses in silico ont permis l'annotation de quelques 28000 gènes codant des protéines.

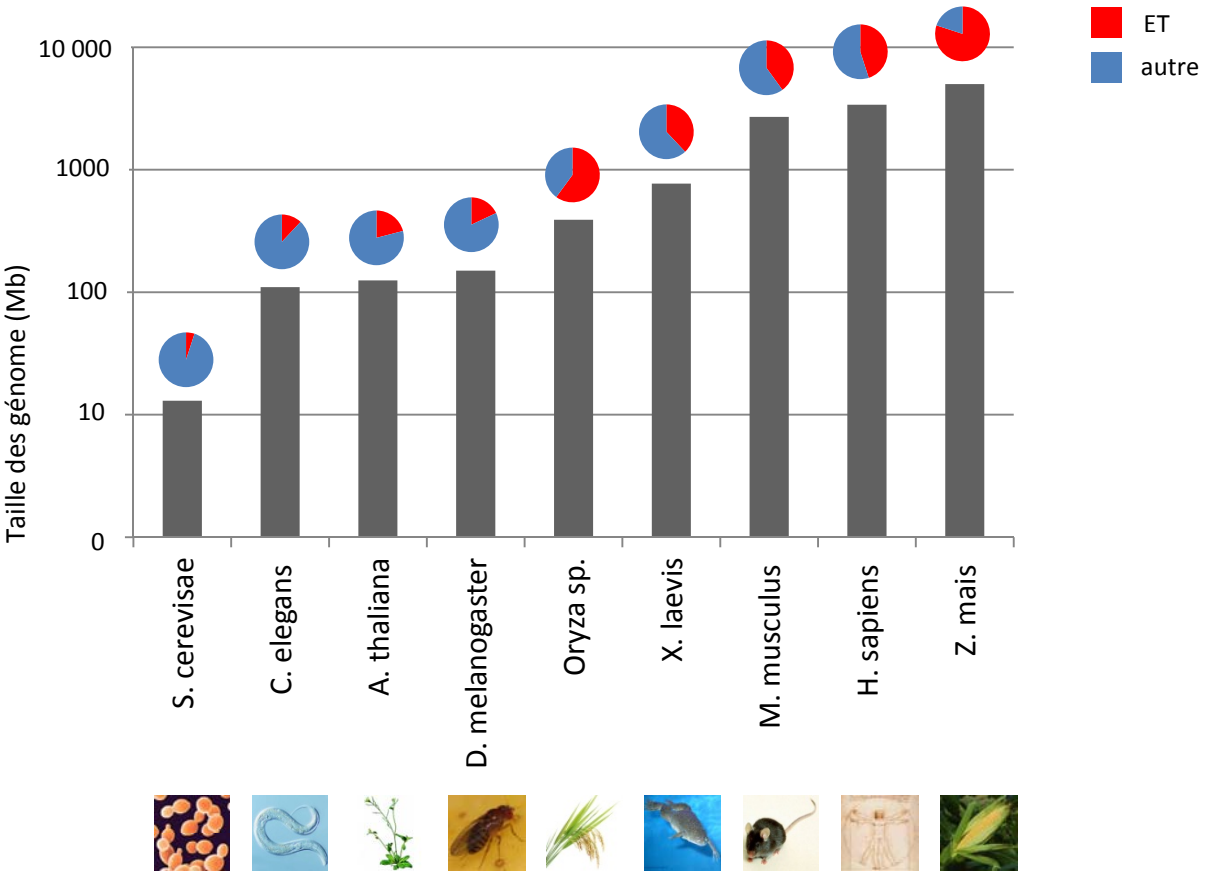


Figure 1.1 : Variation de la taille des génomes et de la fraction composée d'ET entre différentes espèces.
D'après Vierra et al. 2002, Tenaillon et al. 2010 et Buisine et al. 2008.

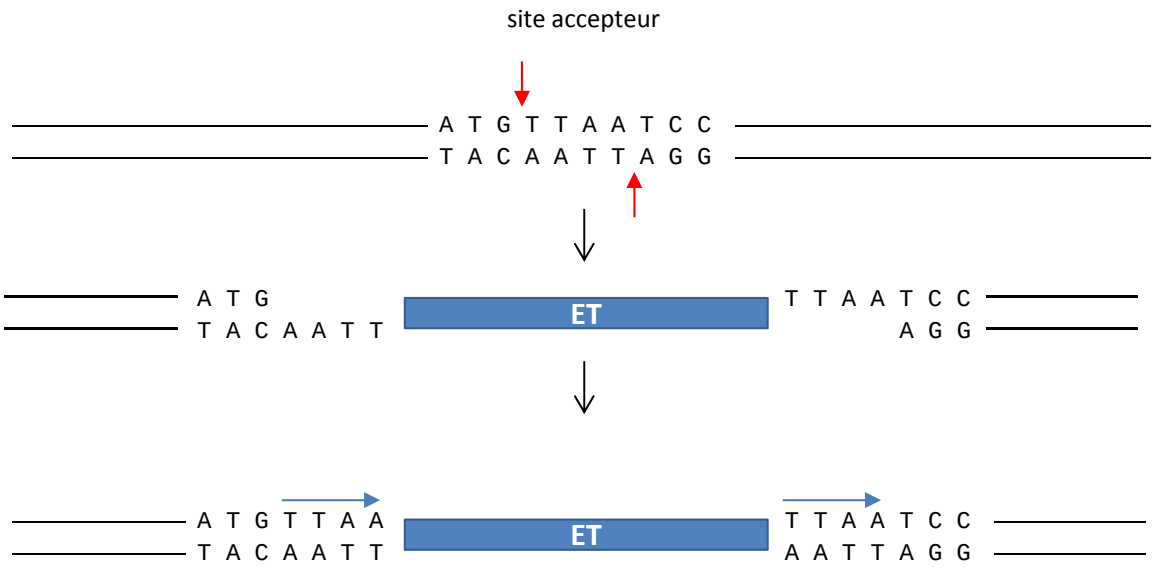


Figure 1.2 : Création de la Target Site Duplication lors de l'insertion d'un ET.
Les flèches rouges représentent la coupure par l'intégrase ou la transposase, les bleus représentent la TSD.

1.2 Les éléments transposables (ET)

1.2.1 Définition, découverte et généralités

Les ET sont des séquences répétées dispersées qui ont la capacité de se déplacer au sein d'un génome par un mécanisme appelé transposition. Ils ont été mis en évidence à la fin des années 1940 par Barbara McClintock suite à l'observation de mutations instables chez le maïs. La première mise en évidence de la transposition a été faite sur le système *Ac-Ds* (*activator-dissociation*) où la présence d'un élément *Ac* induit l'excision de l'élément *Ds* qui génère une cassure chromosomique. Cette cassure a pour conséquence la délétion d'un gène impliqué dans la régulation de la voie des anthocyanes (pigment) de l'albumen du grain de mais. De par la conséquence phénotypique associée à leur mobilisation, ces éléments ont alors été appelés éléments de contrôle (McClintock, non publié, repris dans Fedoroff 2013). Par la suite, de nombreux autres éléments transposables ont étés identifiés chez de nombreux organismes.

Depuis quelques années, le séquençage des génomes de nombreuses espèces a pu mettre en évidence que les ET et leurs reliques sont des composants majeurs des génomes eucaryotes allant de ~3% chez la levure *S. cerevisae* et pouvant atteindre jusqu'à 45% chez l'Homme et >80% chez certain végétaux (fig. 1.1) (Vieira et al. 2012). De fait, la proportion de séquences répétées d'un génome, dont une grande partie sont des vestiges d'ET, est positivement corrélée avec la taille de celui-ci. Ce phénomène, est à l'origine du paradoxe de la valeur C qui illustre l'absence de corrélation entre la « complexité » d'un organisme et la taille de son génome.

1.2.2 Classification des ET et mécanismes de transposition

Les ET peuvent être classés selon leur mode de transposition dans deux catégories majeures : les ET de type I, ou rétroéléments, dont la mobilisation fait intervenir un intermédiaire ARN et les ET de type II, ou transposons à ADN, qui utilisent un intermédiaire ADN pour leur transposition.

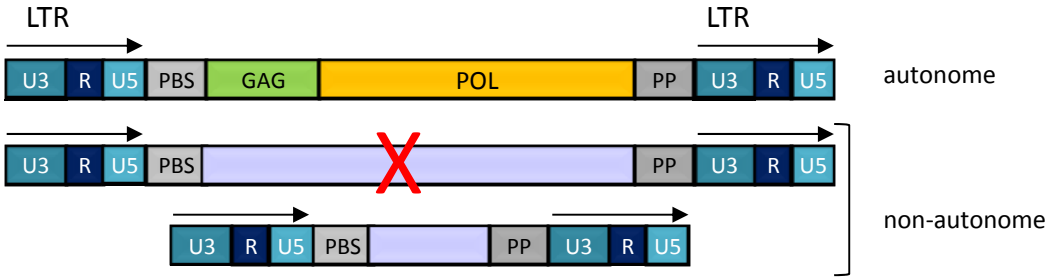


Figure 1.3: Structure générale des rétrotransposons à LTR.
Les domaines PBS et PP sont nécessaires à l'étape de transcription inverse (voir fig. 1.6). Adapté de Feschotte et al. 2002 et Slotkin and Martienssen 2007.

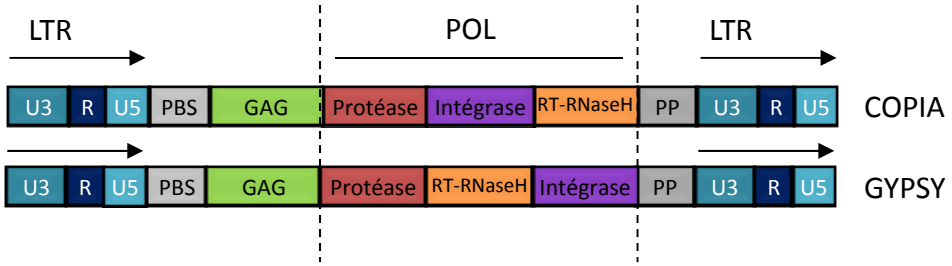


Figure 1.4 : Structure détaillée des rétrotransposons à LTR GYPSY et COPIA.
Adapté de Havecker, Gao and Voytas 2004 et Sabot and Schulman 2006.

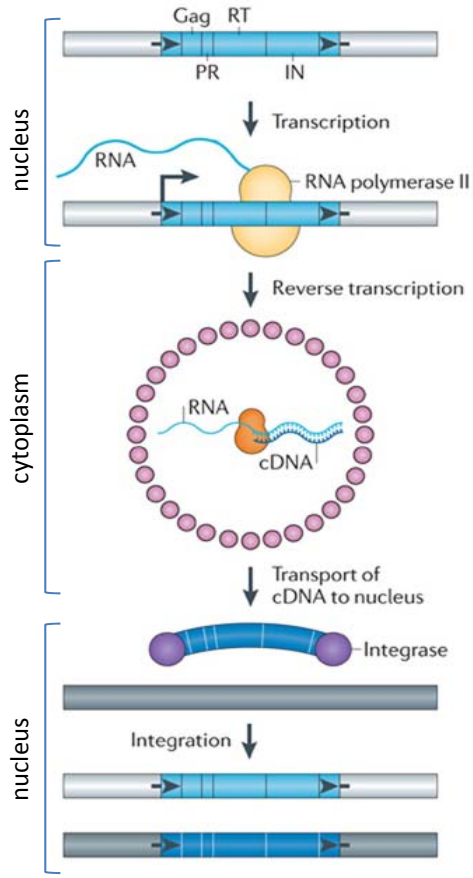


Figure 1.5 : Mécanisme de transposition des rétroéléments à LTR.
D'après Levin et al. 2011.

Par ailleurs, les ET peuvent être regroupés en superfamilles puis en familles en fonction de leur structure, de la similarité de leur séquence nucléotidique ou protéique ou encore de la longueur de la séquence cible dupliquée (TSD, Target Site Duplication) lors de la transposition (fig. 1.2) (Wicker et al. 2007).

1.2.2.1 Les ET de type I ou rétroéléments

Les ET de type I sont caractérisés par un mode de transposition dit de « copier/coller » faisant intervenir un intermédiaire ARN et une étape de transcription inverse. Ce mode de transposition engendre systématiquement l'augmentation du nombre de copies ce qui peut expliquer en partie la prédominance de ce type d'ET dans les génomes eucaryotes. Il existe deux grandes classes de rétroéléments : ceux à grandes répétitions terminales (LTR, Long Terminal Repeat) et ceux sans LTR qui diffèrent également par leur mode de rétrotransposition.

Les rétroéléments à LTR sont très étroitement apparentés aux rétrovirus. Les LTR sont orientées dans le même sens et leur taille peut varier de quelques centaines à plus d'un millier de paires de bases. Elles constituent les régions promotrices de ces ET. Comme chez les rétrovirus, elles contiennent trois domaines consécutifs U3, R et U5. Les éléments autonomes contiennent au moins deux ORF, l'une correspondant au gène *GAG* qui code une polyprotéine de capside, l'autre au gène *POL* qui code une polyprotéine impliquée dans les diverses étapes de la rétrotransposition (fig. 1.3). Cette polyprotéine présente les activités enzymatiques suivantes :

- Une protéase qui permet le clivage de la polyprotéine *POL* et de la polyprotéine *GAG*.
- Une transcriptase inverse (RT) qui catalyse la transcription inverse.
- Une RNase H qui dégrade l'ARN quand il est sous forme d'hétéroduplexe avec l'ADNc produit par la transcription inverse.
- Une intégrase qui a notamment une activité d'endonucléase et qui permet l'intégration de l'ADNc double brin dans le génome.

Deux superfamilles de rétroéléments à LTR sont présentes chez les plantes et sont déterminées sur la base de l'organisation des différents domaines au sein du gène *POL* : les

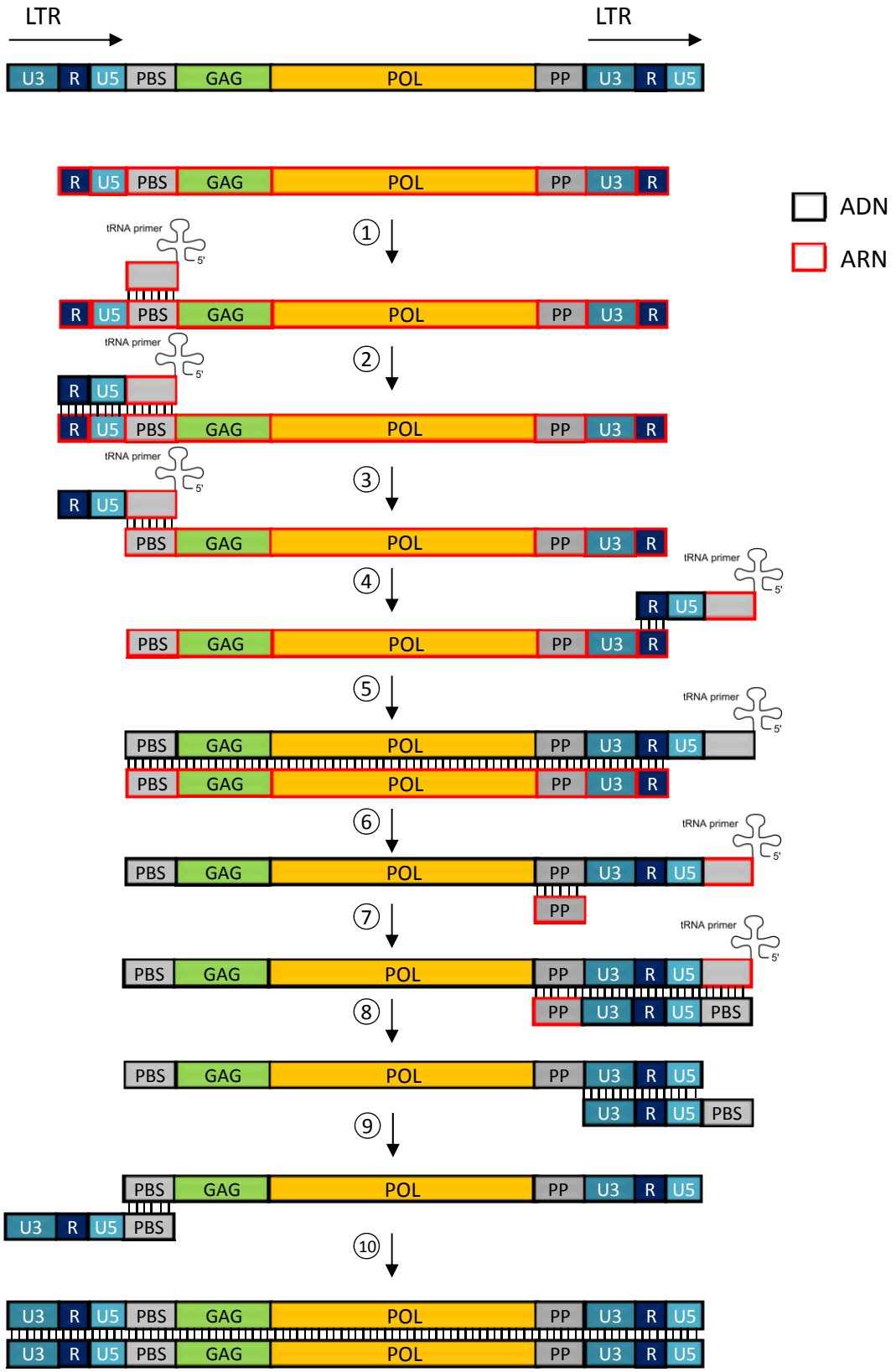


Figure 1.6 : Etapes de la transcription inverse des rétroéléments à LTR et des rétrovirus.
(1) Hybridation d'un ARNt cellulaire sur la région PBS (Primer Binding Site). (2) Synthèse d'un fragment d'ADN. (3) Digestion de la région RU-5 par la RNAse H. (4) Premier saut : hybridation du fragment d'ADN sur le R (ARN) en 3'. (5) Synthèse d'un brin d'ADN à partir de l'extrémité 3'. (6) Digestion de l'ARN viral par la RNAse H à l'exception de la région PP (Polypurine section). (7) Synthèse d'un second brin d'ADN. (8) Digestion de l'ARN par la RNAse H, ce qui engendre le décrochage de l'ARNt. (9) Deuxième saut : hybridation des deux régions PBS. (10) Extension des deux brins d'ADN.

LTR/COPIA (*Pseudoviridae*) et les LTR/GYPSY (*Metaviridae*) (fig. 1.4) (Havecker et al. 2004, Sabot and Schulman 2006).

La mobilisation des rétroéléments à LTR, similaire à celle des rétrovirus (fig. 1.5), fait intervenir une étape de transcription inverse (dont les étapes sont détaillées fig. 1.6), qui a lieu au sein de la capside, dans le cytoplasme. Deux ARN sont en général présents au sein d'une même capside et la RT est alors capable de « sauter » d'une matrice ARN à une autre. Si les deux ARN proviennent de locus distincts cela peut aboutir à la création d'un ADNc composite de plusieurs éléments donneurs (Goodrich and Duesberg 1990, Jordan and McDonald 1998).

Les rétrotransposons sans LTR regroupent principalement les LINE (Long Interspersed Nuclear Elements) et les SINE (Short Interspersed Nuclear Elements) (fig. 1.7). Les LINE ont une taille d'environ 6-7kb et contiennent deux ORF, l'ORF1 codant une protéine chaperonne d'acide nucléique et l'ORF2 la RT et l'endonuclease. Les LINE présentent également une région 5'UTR qui contient la région promotrice et une courte région 3' suivie d'un signal de polyadénylation et d'une séquence polyA issue de la transcription inverse de l'extrémité polyA du transcrit. Les SINE sont des séquences répétées de taille et de structure variables allant de quelques centaines de paires de bases (pour la plupart) à plusieurs kb. Tout comme les LINE, ils présentent une séquence polyA mais à l'inverse de ces derniers, ils ne codent pas les protéines nécessaires à leur transposition. De fait, ils utilisent la machinerie des LINE pour leur mobilisation et sont, par conséquent, non-autonomes. Du fait de cette mobilisation en trans, la région 3' de certains SINE présente des homologies de séquence avec les régions 3' des LINE au sein d'un même organisme (Ogiwara et al. 1999). Enfin, la majorité des SINE décrits (et notamment tous ceux identifiés dans les génomes végétaux) présentent intrinsèquement des promoteurs Pol III, une grande partie d'entre eux dérivant d'ARN de transfert (Feschotte et al. 2002, Deragon and Zhang 2006).

Le cycle de rétrotransposition des LINE (fig. 1.8) diffère à plusieurs niveaux de celui des rétroélément à LTR. Après l'export dans le cytoplasme, l'ARN messager s'associe avec les protéines qu'il code pour former un complexe ribonucléoprotéique qui est ensuite réimporté dans le noyau. L'étape de transcription inverse a lieu au sein du noyau et est couplée à l'étape d'insertion par un processus appelé « target-site-primed reverse

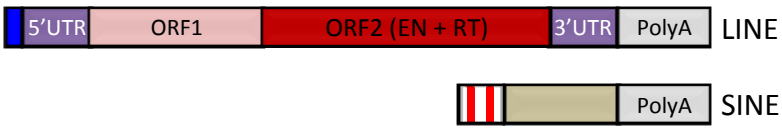


Figure 1.7 : Structure des rétrotransposons sans LTR LINE et SINE.
Les barres rouges représentent les sites de fixation de POL III. Adapté de Feschotte et al. 2002 et Slotkin et al. 2007.

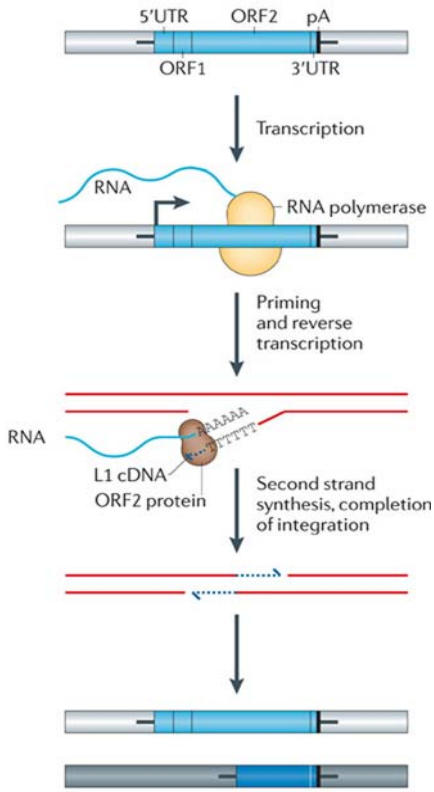


Figure 1.8 : Mécanisme de transposition des rétroéléments sans LTR.
D'après Levin et al. 2011.

transcription » (TPRT) (Cost et al. 2002, Kazazian 2004). L'endonucléase coupe préférentiellement des séquences riches en T et l'extrémité polyA de l'ARNm peut alors s'associer à celle-ci par complémentarité des bases (Levin and Moran 2011). Il est courant que la transcription inverse ne soit pas complète, donnant alors lieu à des nouvelles insertions tronquées en 5', par conséquent incapables de se remobiliser, de façon autonome du moins.

Il existe également d'autres types de rétroéléments sans LTR, présents notamment chez les plantes, dont les mécanismes de transposition sont encore énigmatiques. Les DIRS (*Dictyostelium* intermediate repeat sequences) qui possèdent une tyrosine recombinase à la place de l'intégrase et les PLE (*Penelope*-like element) dont la RT présente plus de similarité avec une télomérase qu'avec les RT des LINE ou des rétroéléments à LTR (Wicker et al. 2007).

1.2.2.2 Les ET de type II ou transposons à ADN

Le type II regroupe deux sous-classes d'ET (fig. 1.9).

Les transposons de la sous-classe 1 (fig. 1.9a) se déplacent dans le génome par un mécanisme de couper/coller qui consiste en l'excision d'un élément à son site donneur et son insertion à autre endroit du génome, le site accepteur (fig. 1.10a). La transposase assure l'excision et l'intégration du transposon car elle possède une activité endonucléase ainsi qu'une activité ligase qui permet la fusion des extrémités du TE avec la séquence acceptrice. Les éléments autonomes présentent au moins une ORF correspondant au gène codant la transposase et la présence de courtes séquences répétées inversées TIR (Tandem Inverted Repeats) à leurs extrémités. Les éléments non-autonomes présentent également des TIR mais ne codent aucune transposase ou alors une forme défectueuse de celle-ci. C'est notamment le cas des MITE (Miniature Inverted Repeats) qui ne sont composés que des TIR (Bureau and Wessler 1994) et sont mobilisés en trans par une transposase codée par un élément autonome.

Contrairement au mode de transposition copier/coller des rétroélélments, le mode couper/coller n'induit pas systématiquement une augmentation du nombre de copies. En

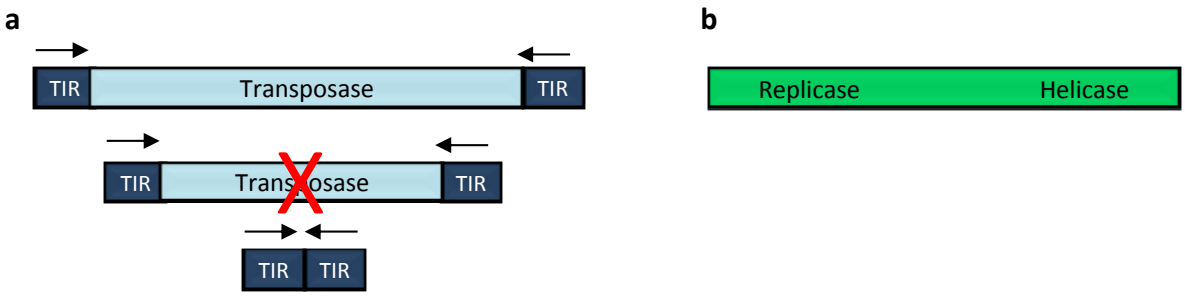


Figure 1.9 : Structure générale des transposons à ADN.
(**a**) Transposons à ADN de la sous-classe 1. (**b**) Helitrons (sous-classe 2). Adapté de Feschotte et al. 2002.

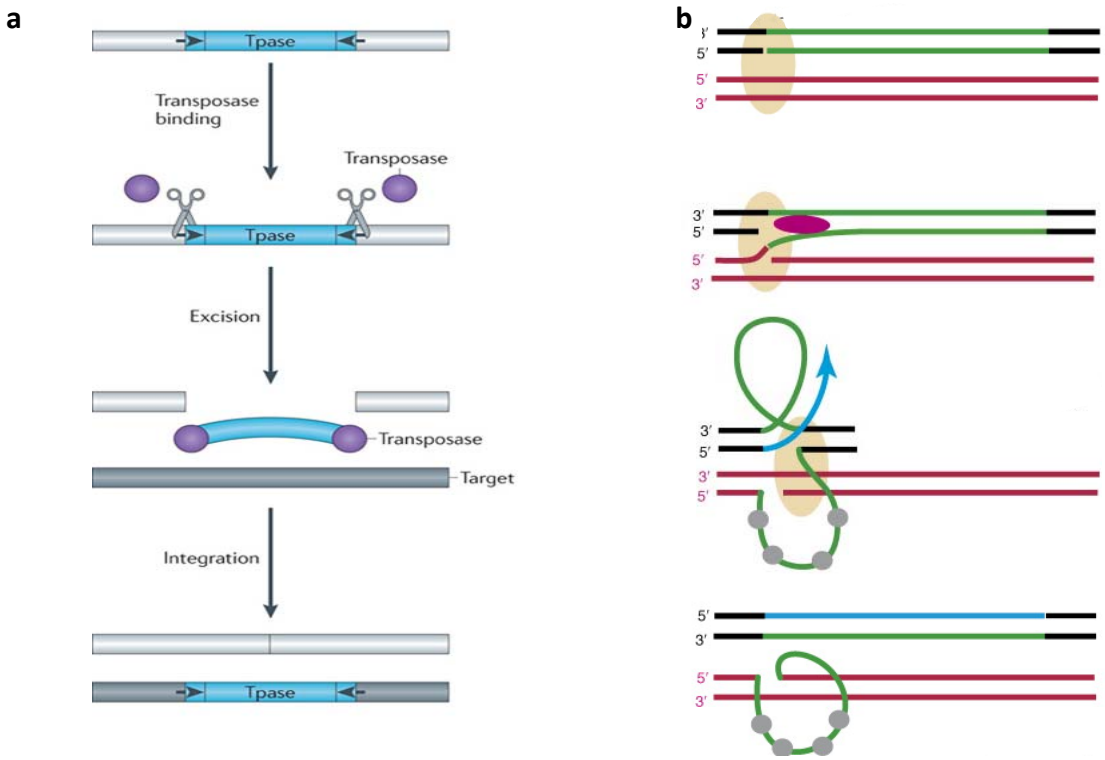


Figure 1.10 : Mécanisme de transposition des transposons à ADN.
(**a**) Transposition par couper/coller des transposons à ADN de la sous-classe 1 (Levin et al. 2011). (**b**) Mode de transposition réplicatif présumé des Helitrons (Kapitonov et al. 2007).

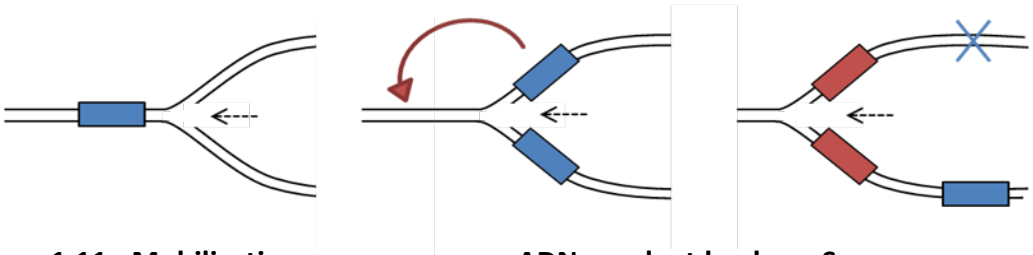


Figure 1.11 : Mobilisation des transposon ADN pendant la phase S.
Le transposon s'excise après passage de la fourche de réplication et s'insère en amont de celle-ci (illustré ici) ou sur une autre région de génome non encore répliquée. Après division, l'une des deux cellules filles contiendra alors une copie supplémentaire du transposon tandis que l'autre ne présentera que la copie transposée.

effet, si la réparation du site donneur se fait par « non-homologous end joining », la copie au site donneur est perdue. En revanche, si la réparation se fait par recombinaison homologue (à partir de la séquence de la chromatide sœur ou du chromosome homologue) alors la copie du site donneur est conservée. Si le choix du mode de réparation qui suit l'excision n'est pas un phénomène encore bien caractérisé, il semble varier d'une famille de transposon à ADN à l'autre. Enfin, une mobilisation pendant la phase de réplication peut engendrer l'augmentation du nombre de copies quelque soit le mode de réparation adopté au locus donneur. Cela a notamment été décrit pour l'élément *Ac* qui s'excise après le passage de la fourche de réplication pour aller s'insérer dans une région encore non-répliquée (fig. 1.11) (Fedoroff 2013).

Il faut noter que certains transposons à ADN présentent plusieurs ORF ou encore une ORF donnant lieu à plusieurs protéines par épissage alternatif du transcrit. Si les fonctions de ces protéines ne sont pas toujours caractérisées, certaines d'entre elles semblent jouer un rôle crucial dans la régulation de l'activité des ET. Ainsi, la protéine TNPA issue de l'épissage alternatif de l'unique ORF des éléments Spm du maïs est non seulement impliquée dans l'activation transcriptionnelle de ces éléments, mais peut également avoir un rôle de régulateur négatif de la transposition en se fixant en aval de la région promotrice et bloquant ainsi la transcription (Masson et al. 1989, Masson et al. 1991, Schlappi et al. 1993). De même, chez la drosophile, l'épissage alternatif du gène codant la transposase de l'élément P conduit à la production d'un répresseur de la transposition (Laski et al. 1986, Misra and Rio 1990).

La sous-classe 2 est représentée par deux groupes d'ET : les Helitrons (fig. 1.9b) et les Mavericks. Les Helitrons ont été identifiés *in silico* et forment un groupe étendu, diversifié et parfois abondant d'ET, notamment chez Arabidopsis (Kapitonov and Jurka 2001). Ils sont cependant difficiles à reconnaitre car ils ne présentent pas de structures terminales particulières et ne génèrent pas de TSD. Ils contiennent deux domaines enzymatiques caractéristiques, un initiateur de réplication Rep et un domaine hélicase Hel présents sur une même protéine RepHel (Kapitonov and Jurka 2001, Kapitonov and Jurka 2007). Leur mode de transposition réplicatif, encore peu connu, se ferait par un mécanisme de cercle roulant (rolling circle) de façon analogue aux transposons bactériens (fig. 1.10b). Il apparait donc que, bien que classés dans le type II, les hélitrons ne présentent aucune similarité que ce soit

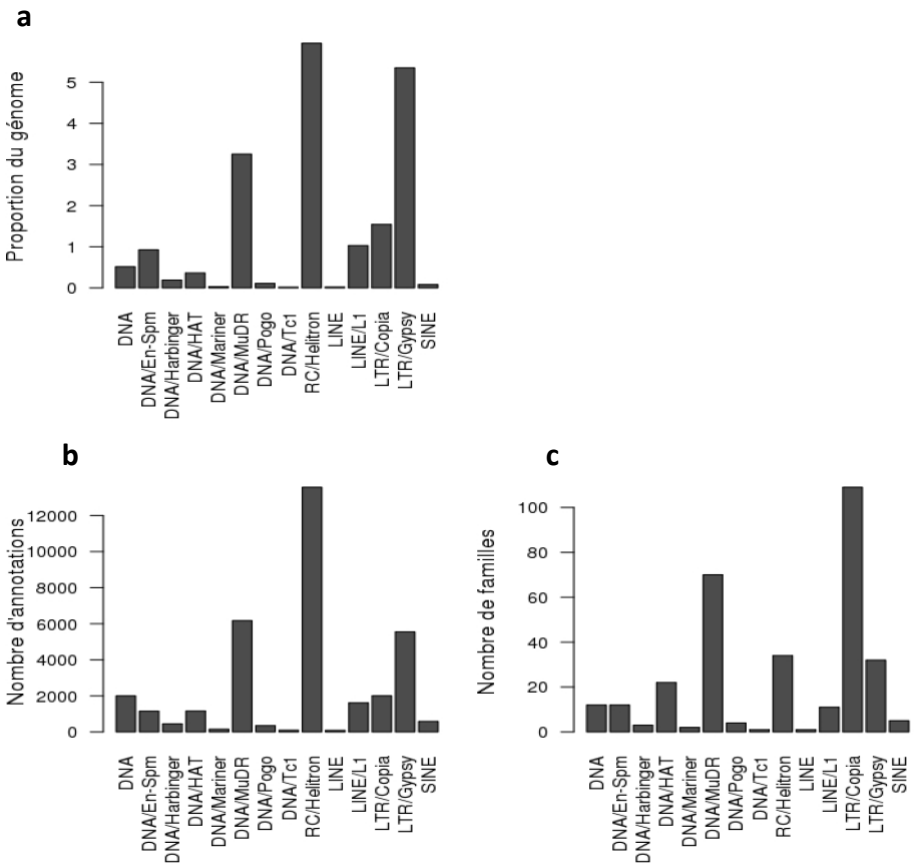


Figure 1.12 : Diversité des superfamilles d'ET d'Arabidopsis.
(**a**) Proportion du génome correspondant à chaque superfamille. (**b**) Nombre d'annotations par superfamille. (**c**) Nombre de familles pour chaque superfamille. Analyse effectuée par Alexis Sarazin à partir des données de Buisine et al. 2008.

au niveau structural ou au niveau du mode de transposition avec les transposon à ADN de la sous classe I décrits plus haut. Enfin, les Mavericks (ou Polintrons), sont des transposons eucaryotes découverts récemment. Ils peuvent contenir jusqu'à 11 gènes dont une ADN polymérase de type viral et une intégrase présentant des similarités avec celles de certains éléments de type I. Leur mode de transposition serait l'excision sous forme d'ADN simple brin, synthèse du double brin par l'ADN polymérase qu'ils codent et intégration (Kapitonov and Jurka 2006, Wicker et al. 2007, Pritham et al. 2007). Aucun Maverick n'a encore été identifié chez les plantes.

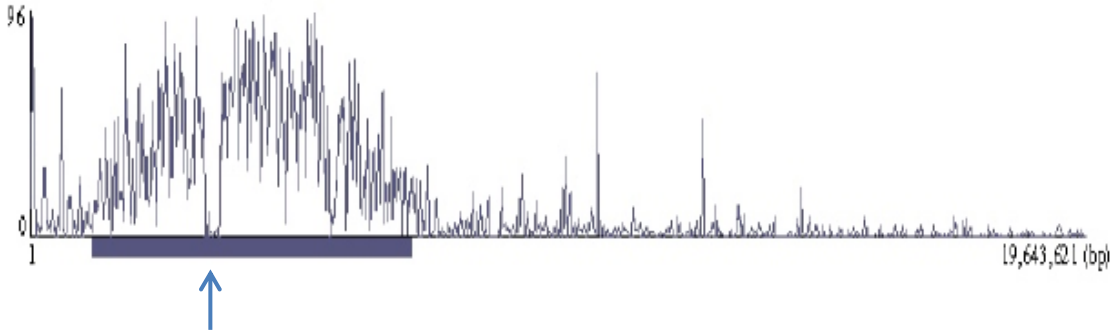
1.2.3 Les ET d'*Arabidopsis thaliana*

La dernière annotation des ET du génome d'*Arabidopsis thaliana* (Ahmed et al. 2011) indique qu'environ 21% (25Mb) de celui-ci est composé de séquences d'ET. Cette relativement faible proportion d'ET par rapport à d'autres génomes de plantes ou de mammifères facilite leur analyse fine et leur caractérisation. Au nombre de 31 245, ils appartiennent à 318 familles et 14 superfamilles.

Parmi ces superfamilles, certaines sont peu représentées chez Arabidopsis (fig. 1.12). C'est notamment le cas pour les éléments a ADN Harbinger, Mariner, Pogo, Tc1, ainsi que pour les SINE. D'autres, en revanche, comportent un nombre important de copies et sont très diversifiées comme les superfamilles Helitron et MuDR. Enfin, en ce qui concerne les rétroéléments a LTR, le génome d'Arabidopsis présente un faible nombre de séquences de la superfamille LTR/COPIA par rapport à la superfamille LTR/GYPSY. Cependant la première est beaucoup plus diversifiée avec 109 familles contre seulement 32 pour la superfamille LTR/ GYPSY (fig. 1.12).

Différentes études sur les populations d'ET d'Arabidopsis ont cherché à estimer leur âge relatif et les périodes auxquelles ils ont envahi le génome. Certaines familles d'Helitron ont envahi le génome récemment, après la divergence entre *A.thaliana* et *A.lyrata* (~10 millions d'années) (Hollister and Gaut 2007) tandis que la superfamille SINE, qui est peu abondante et faiblement diversifiée dans le génome d'*A.thaliana*, l'aurait envahi il y a 12 à 20 millions d'années (Lenoir et al. 2001, Lenoir et al. 2005). Par ailleurs, l'étude de similarité réalisée sur

a



b

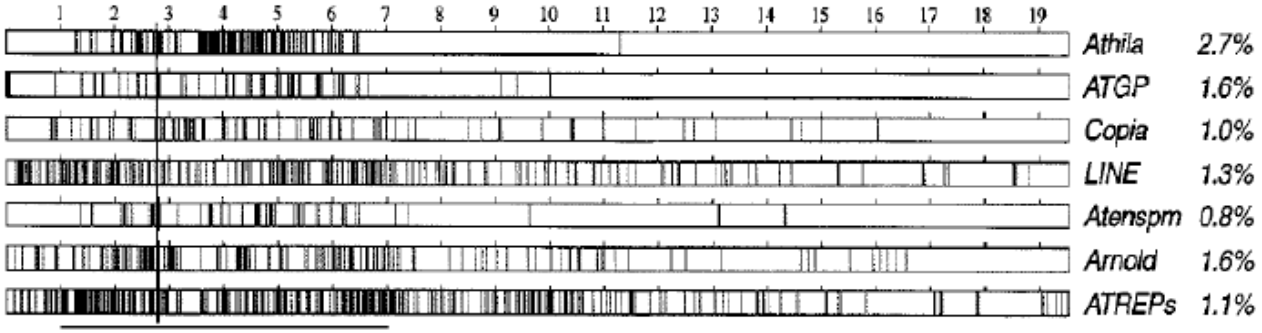


Figure 1.13 : Distribution des ET le long du chromosome 2 d'*Arabidopsis thaliana*.
(**a**) Proportion d'ADN correspondant à des séquences d'ET pour des fenêtres de 25kb le long du chromosome 2 (d'après Kapitonov and Jurka 2002). Les régions péricentromériques sont indiquées par un rectangle bleu. Le centromère (flèche bleue) n'est pas séquencé car il est essentiellement composé de plusieurs dizaines de milliers de copies en tandem d'une même séquence satellite de 180 bp. (**b**) Distribution de différentes classes d'ET le long du chromosome 2. Chaque ligne verticale correspond à une fenêtre de 17kb contenant au moins une séquence d'ET. Les éléments *ATHILA* et *ATGP* correspondent à la superfamille LTR/GYPSY, les éléments *ATENSPM* et *ARNOLD* sont des transposons à ADN et les *ATREP* des Helitrons. D'après Kapitonov et al. 1999.

des séquences LTR des superfamilles LTR/Gypsy et LTR/Copia indique que la seconde a colonisé le génome d'Arabidopsis et s'est diversifiée plus récemment que la première (Peterson-Burch et al. 2004).

Les ET sont majoritairement présents au niveau des régions péricentromériques des chromosomes d'Arabidopsis ainsi qu'au niveau du renflement hétérochromatique localisé sur le bras court du chromosome 4 de l'accession Columbia étudiée au laboratoire. Cependant la distribution des ET le long des chromosomes varie d'une famille à l'autre (fig. 1.13) (AGI 2000, Kapitonov and Jurka 1999). Cette situation contraste avec celle observée chez le maïs où les ET appartenant aux différentes familles présentent des distributions bien plus distinctes. Ainsi, les rétroéléments LTR/Gypsy sont principalement localisés au niveau des régions préricentromériques alors que les LTR/COPIA y sont sous-représentés. Les éléments CACTA (transposon à ADN EN/SPM) sont quant à eux également distribués le long des chromosomes (Schnable et al. 2009). La suraccumulation des ET au niveau des régions péricentromériques chez *Arabidopsis thaliana* pourrait s'expliquer par un biais d'insertion dans ces régions ou par l'élimination par sélection naturelle des évènements d'insertion le long des bras chromosomiques du fait de l'effet délétère sur les gènes situés à proximité. La comparaison de l'organisation des génomes d'*Arabidopsis thaliana* et d'*Arabidopsis lyrata* indique une activité plus récente des rétroéléments à LTR dans la seconde espèce ainsi qu'une distribution différente des ET le long des deux génomes. De fait, la proportion de gènes avec un ET situé à proximité est plus élevée et la distance aux gènes plus courte chez *A. lyrata* que chez *A. thaliana* (Hu et al. 2011). Ces observations sont donc en accord avec un rôle prépondérant de la sélection naturelle dans la distribution inégale des ET le long des chromosomes d'*A. thaliana* et *A*. *lyrata*.

Il est notable que la taille et l'organisation du génome d'Arabidopsis ne sont pas typiques des angiospermes. De fait, les génomes de plantes sont généralement plus grands et plus complexes. La structure des génomes végétaux est souvent représentée de façon imagée comme « des ilots de gènes entourés d'océans d'ADN répété ». Plus les génomes sont grands, plus les îlots sont petits et les océans vastes. Notamment, chez le maïs (2400Mb) et le blé (6000Mb pour chacun des trois génomes composant le blé hexaploïde), les îlots contiennent un ou deux gènes et sont d'une taille inférieure à 10 kb en moyenne alors que les océans interstitiels font de 20 à 100 kb (Bennetzen et al. 1994, Devos et al. 2005,

Schnable et al. 2009., Wicker et al. 2001). Chez Arabidopsis en revanche, on observe plutôt des îlots géniques géants et de rares et petits océans d'ET excepté dans les régions péricentromériques. Dès lors, la densité moyenne en ET le long des génomes du mais et du blé correspond plus à celle des régions péricentromériques qu'à celle des bras d'*Arabidopsis thaliana*. Il existe cependant également des exemples de génomes de taille intermédiaire, comme celui du sorgho (700 Mb) qui présente une plus grande quantité d'ET mais une organisation semblable à celle du génome d'Arabidopsis (Paterson et al. 2009).

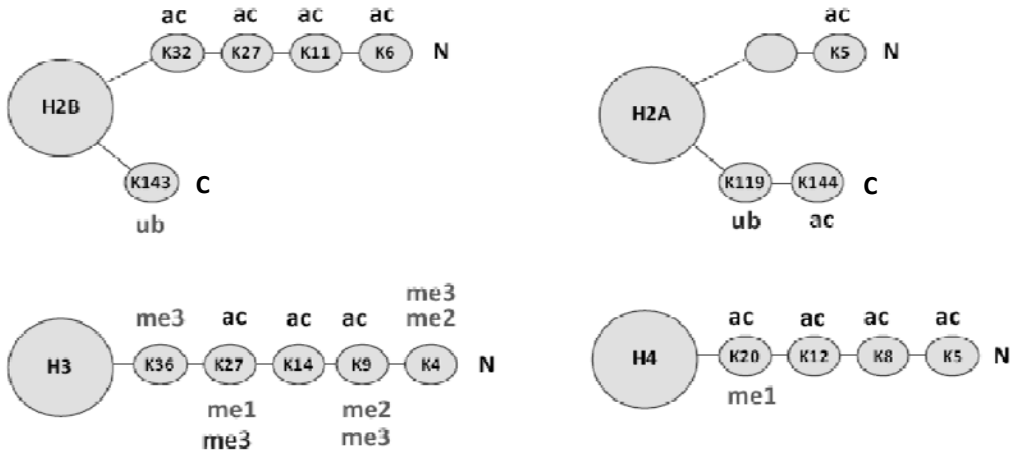


Figure 1.14 : Principales marques d'histones présentes chez Arabidopsis.
Représentation non exhaustive des résidus des histones H2A, H2B, H3 et H4 pouvant être acétylés (ac), méthylés (me), ou ubiquitinés (ub) chez Arabidopsis.

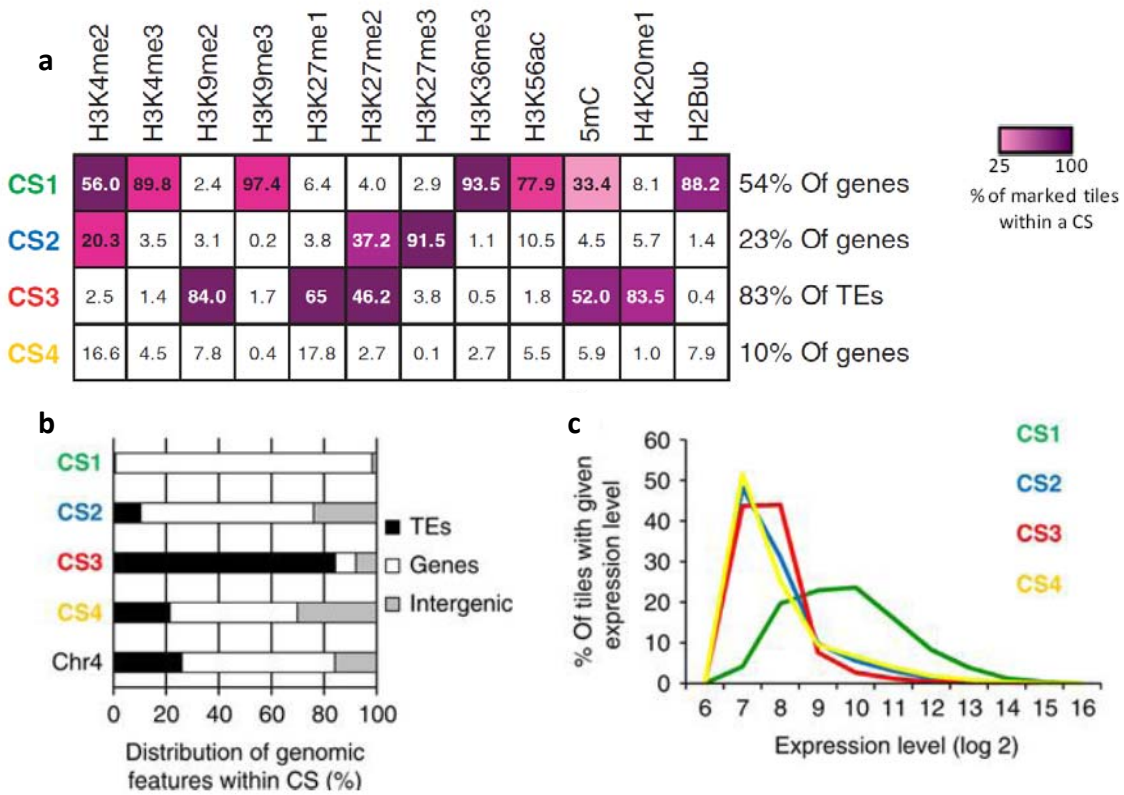


Figure 1.15 : Etats chromatiniens chez Arabidopsis.
(**a**) Composition des 4 principaux états chromatiniens (CS1 à 4) définis partir de l'analyse de 11 marques d'histones ainsi que la méthylation de l'ADN. La distribution des 12 marques chromatiniennes sur les 4 CS est indiquée par la couleur de la case. Le degré d'homogénéité de chaque CS est indiqué par le pourcentage de sondes du CS associées à chacune des 12 marques (nombres dans les cases). (**b**) Proportion relative de gènes et de TE dans chaque CS. (**c**) Relation entre les CS et le niveau d'expression des gènes. Le pourcentage de sondes associées à chaque CS est donné pour chaque niveau d'expression. D'après Roudier et al. 2011.

1.3 Contrôle des ET via des mécanismes épigénétiques chez Arabidopsis

De par leur capacité à se déplacer dans le génome, les ET sont potentiellement hautement mutagènes et par conséquent, constituent une menace pour l'intégrité des génomes au sein desquels ils résident. Cependant, relativement peu de copies d'ET sont mobilisables au sein des génomes, et celles-ci sont le plus souvent maintenues dans un état réprimé par tout un ensemble de mécanismes, parmi lesquels la méthylation de l'ADN et autres marques chromatiniennes ainsi que l'interférence ARN (ARNi) jouent un rôle important.

1.3.1 Etats chromatiniens chez Arabidopsis : généralités

La majorité des régulations de l'activité du génome est due à des modifications de la chromatine. L'unité fonctionnelle de la chromatine est le nucléosome, constitué de 147 bp d'ADN enroulées autour d'un octamère d'histones (2 copies de chaque histone H2A, H2B, H3 et H4). L'histone H1 est quant à elle localisée à l'extérieur du nucléosome, au niveau de l'ADN « linker » qui relie les nucléosomes entre eux. La chromatine peut être modifiée de plusieurs façons. Tout d'abord, par incorporation de formes variantes des histones, notamment H3 et H2A. Ensuite, par modification post traductionnelle de certains résidus, particulièrement ceux situés à l'extrémité N-terminale des Histones H3 et H4 (fig. 1.14). Enfin chez de nombreuses espèces, notamment les plantes et les mammifères, les cytosines peuvent être méthylées en position 5 du cycle aromatique. Les combinaisons de ces différentes modifications confèrent à la chromatine des propriétés diverses se traduisant par des variations de niveaux de compaction et d'activité transcriptionnelle qui définissent ainsi différents états chromatiniens.

Chez Arabidopsis, quatre états chromatiniens distincts ont été caractérisés sur la base de l'étude de 12 marques chromatiniennes. Les données ont été obtenues par des expériences d'immunoprécipitation de la chromatine et de l'ADN méthylé suivies d'hybridation sur puce à ADN (ChIP-chip et MeDIP-chip, (Roudier et al. 2011) (fig. 1.15).

- Le CS1 (Chromatine State 1) est principalement associé aux gènes transcriptionellement actifs. Il est caractérisé par la di- et la triméthylation de la lysine 4 de l'histone H3 (H3K4me2

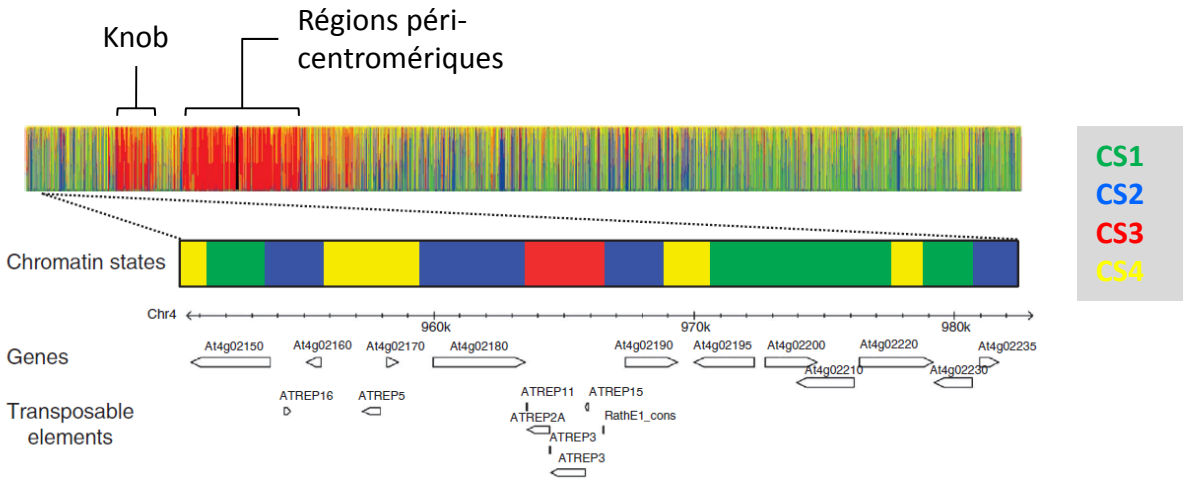


Figure 1.16 : Distribution des quatre états chromatiniens le long du chromosome 4 d'Arabidopsis.

Les états chromatiniens forment généralement des domaines de petite taille à l'exception du CS3 qui forme également de larges blocs au niveau des régions péricentromériques et du renflement hétérochromatique (knob). D'après Roudier et al. 2011.

et H3K4me3), la triméthylation de la lysine 9 de l'histone H3 (H3K9me3), la monoubiquitination de l'histone H2B (H2Bub), la triméthylation de la lysine 36 de l'histone H3 (H3K36me3) et l'acétylation de la lysine 56 de l'histone H3. Il a également été montré que l'acétylation d'autres résidus des histones H3 et H4 est fortement associée aux régions transcriptionnellement actives (Fuchs et al. 2006). Enfin, la méthylation de l'ADN est trouvée sur un sous-ensemble des domaines CS1.

- Le CS2, principalement caractérisé par la triméthylation de la lysine 27 de l'histone H3 (H3K27me3), est typiquement associé aux gènes réprimés. H3K27me3 est déposé par le complexe PRC2 (Polycomb Repressive Complex 2), présent chez tous les organismes multicellulaires.

- Le CS3, regroupe la mono et diméthylation de la lysine K27 de l'histone H3 (H3K27me1 et H3K27me2), la diméthylation de la lysine 9 de l'histone H3 (H3K9me2), la monométhylation de la lysine 20 de l'histone H4 (H4K20me1) et la méthylation de l'ADN. Cet état cible essentiellement les ET et correspond cytologiquement aux régions formant l'hétérochromatine. Cet état sera détaillé plus loin.

- Le CS4, enfin, n'est pas particulièrement enrichi en aucune marque et n'est pas spécifiquement associé aux gènes ni aux ET, mais présente une légère préférence pour les régions intergéniques. Il est associé le plus souvent à l'absence d'activité transcriptionnelle.

Typiquement, les états chromatiniens CS1, CS2 et CS4 forment des domaines de petite taille (de l'ordre du gène pour CS1 et CS2) organisés sans ordre apparent le long du génome. Le CS3, en revanche, forme de larges blocs au niveau des régions péricentromériques et du renflement hétérochromatique où la vaste majorité des ET sont accumulés. Néanmoins, le CS3 est également localisé au niveau des ET isolés dans les bras chromosomiques, formant ainsi de petits ilots d'hétérochromatine (fig. 1.16).

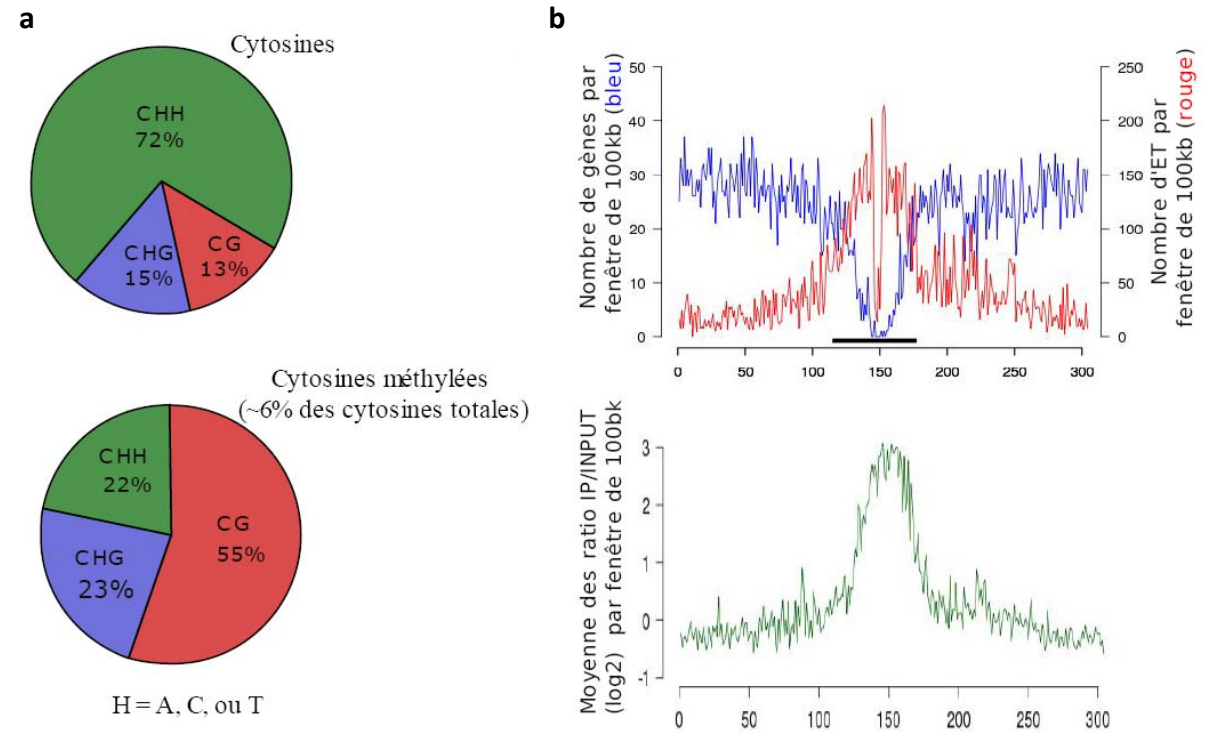


Figure 1.17 : Méthylation de l'ADN chez *Arabidopsis thaliana*.
(**a**) Répartition des cytosines du génome (en haut) et des cytosines méthylées (en bas) dans les trois contextes CG, CHG et CHH. Daprès Cokus et al. 2008 et Lister et al. 2008. (**b**) Distribution des gènes, ET et de la méthylation de l'ADN le long du chromosome 1 d'*Arabidopsis thaliana.*

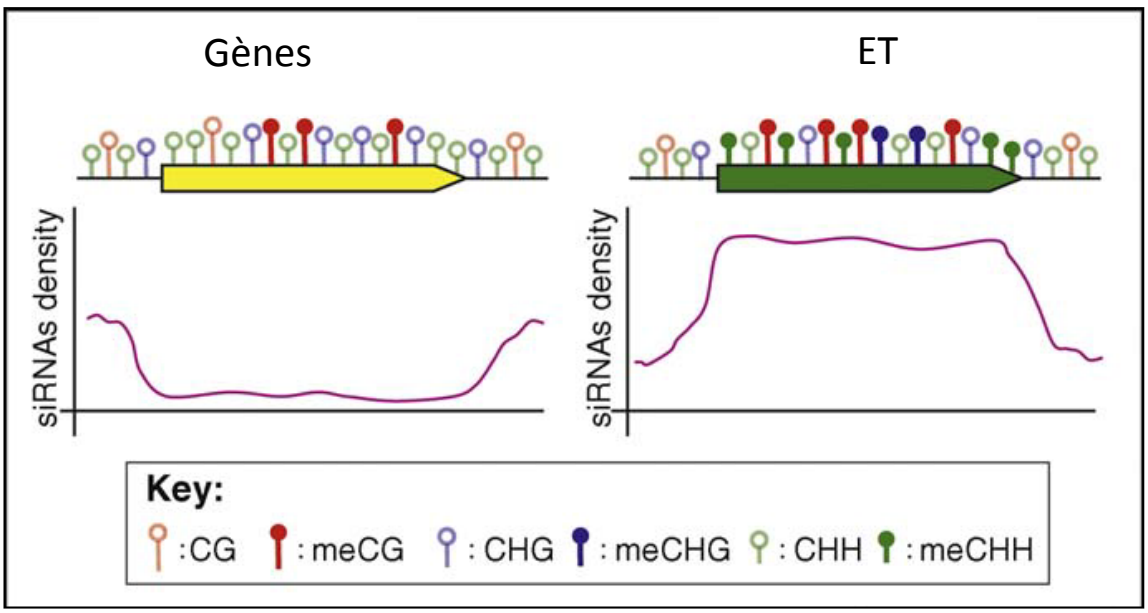


Figure 1.18 : Profils types de méthylation des cytosines le long des gènes et des ET chez *Arabidopsis thaliana*.
D'après Roudier et al. 2009.

1.3.2 La méthylation de l'ADN

Parmi les différentes marques associées aux séquences répétées (marques du CS3), la mieux caractérisée est la méthylation de l'ADN qui joue un rôle majeur dans la répression des ET chez les plantes et les mammifères. Cependant, contrairement à ce qui est observé dans la plupart des cellules chez les mammifères, la méthylation de l'ADN chez les plantes n'est pas restreinte aux dinucléotides CG, mais cible tous les contextes nucléotidiques : CG, CHG et CHH (où H représente n'importe quel nucléotide sauf G) (fig. 1.17a). D'autre part, si la méthylation de l'ADN cible la majorité des cytosines des génomes mammifères de façon constitutive (à l'exception des ilots CpG situés dans les promoteurs des gènes qui ne sont méthylés que de manière spécifique au cours du développement (Suzuki and Bird 2008), elle n'affecte que 6% des cytosines chez Arabidopsis, ce qui témoigne d'un ciblage plus spécifique de la marque chez cet organisme où elle est principalement trouvée au niveau des séquences répétées (fig. 1.17b) (Lister et al. 2008, Cokus et al. 2008, Lippman et al. 2004, Zhang et al. 2006). En effet, si la méthylation génique existe bien aussi chez Arabidopsis, elle est restreinte au corps des gènes, ne touche que 30% d'entre eux et est presque exclusivement de type CG (fig. 1.18 et fig. 1.15). (Zhang et al. 2006, Vaughn et al. 2007, Zilberman et al. 2007, Lister et al. 2008, Cokus et al. 2008, Saze and Kakutani 2011).

1.3.2.1 Maintenance de la méthylation de l'ADN

La maintenance de la méthylation de l'ADN des ET est dépendante du contexte nucléotidique des cytosines. La méthylation des sites CG est principalement catalysée par la METHYLTRANSFERASE1 (MET1) (Henderson and Jacobsen 2007), l'homologue de Dnmt1 des mammifères, qui reconnait l'ADN hémiméthylé après réplication (Goll and Bestor 2005). Plusieurs autres protéines sont requises pour la maintenance de la méthylation en contexte CG (Law and Jacobsen 2010), notamment VARIANT IN METHYLATION 1 (VIM1) qui contient un domaine SRA (SET and RING Associated) et dont l'homologue mammifère est Uhrf1. Son rôle supposé est le recrutement de MET1/Dnmt1 au niveau des CG hemiméthylés (Bostick et al. 2007, Sharif et al. 2007, Woo et al. 2007, Woo et al. 2008) (fig. 1.19 et 1.20).

La maintenance de la méthylation des sites CHG est majoritairement assurée par CHROMOMETHYLASE 3 (CMT3), une ADN-méthyltransférase spécifique des plantes (Henikoff

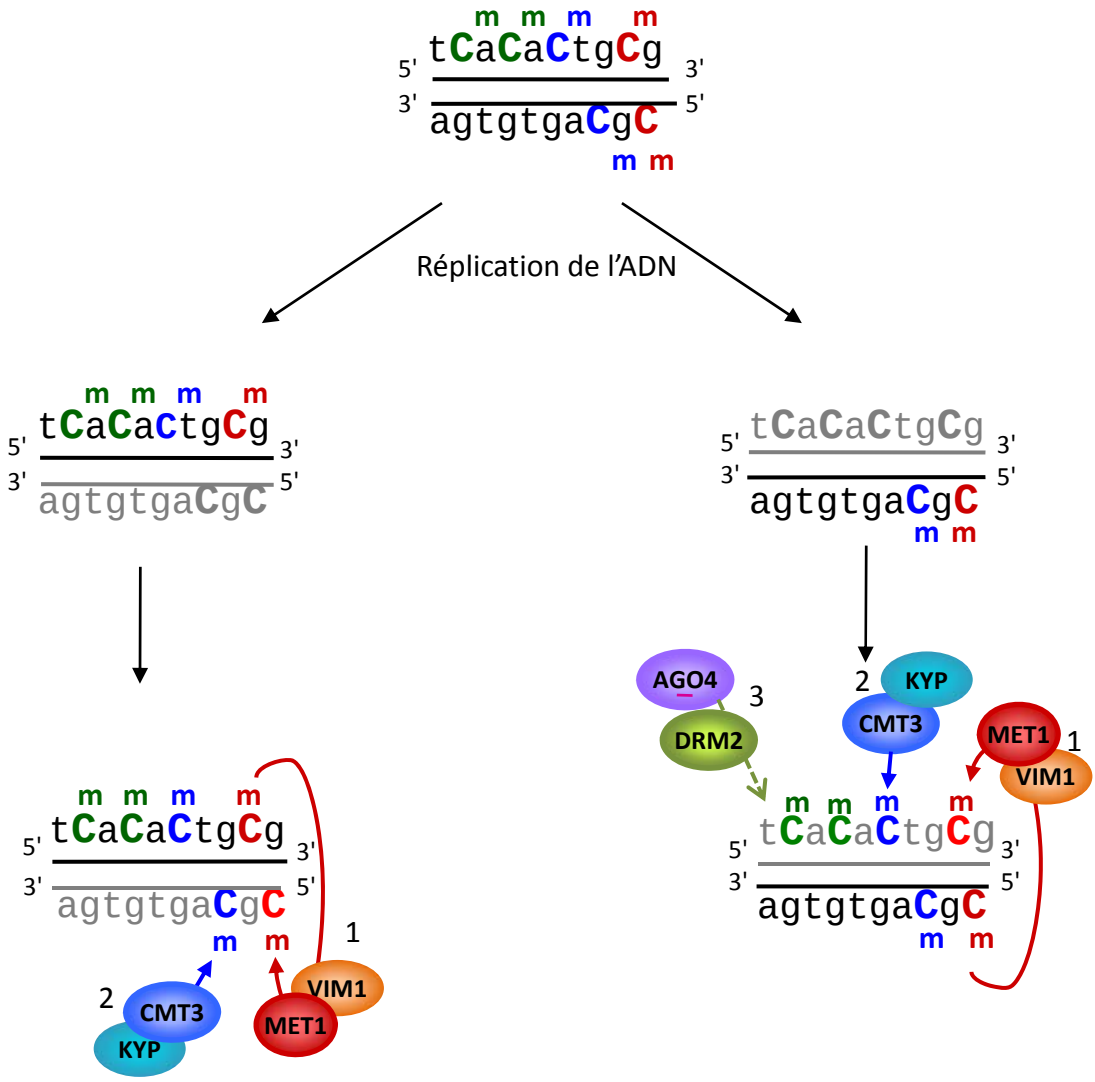


Figure 1.19 : Modèle du maintien de la méthylation de l'ADN sur les ET chez Arabidopsis.
Les point de 1, 2 et 3 sont détaillés figure 1.20. D'après Teixeira et al. 2010

and Comai 1998, Lindroth et al. 2001, Cokus et al. 2008) et requiert également l'action de SUVH4 (aussi appelée kryptonite, (KYP)), la principale histone méthyltransférase responsable de la diméthylation de la lysine 9 de l'histone H3 (H3K9me2) (Jackson et al. 2002, Malagnac et al. 2002). Une boucle de renforcement entre ces deux modifications est suggérée par le fait que le chromodomaine de CMT3 et le domaine SRA de KYP peuvent lier respectivement H3K9me2 et les CHG méthylés (Lindroth et al. 2004, Johnson et al. 2007, Inagaki et al. 2010) (fig. 1.19 et 1.20).

Enfin, la maintenance de la méthylation des cytosines en contexte asymétrique CHH est assurée par l'action constante de la machinerie du RNA-directed-DNA-Methylation (RdDM) qui rétablit *de novo* la méthylation après chaque cycle de réplication (Cao and Jacobsen 2002 a et b). Ce phénomène requiert une première étape de transcription par l'ARN POLYMERASE IV (POLIV), polymérase spécifique des plantes (Lahmy et al. 2010). Le transcrit ainsi formé sert de matrice à l'ARN polymérase ARN dépendante RDR2 (RNA DEPENDENT RNA POLYMERASE 2) qui génère l'ARN double brin. Des travaux récents suggèrent que la reconnaissance du transcrit par RDR2 soit permise par une interaction physique de cette dernière avec POLIV (Law et al. 2011). Ce long transcrit double brin est ensuite clivé par DICER LIKE 3 (DCL3) en petits ARN interférents (siARN) de 24nt, qui sont pris en charge sous forme simple brin dans un complexe contenant la protéine ARGONAUTE4 (AGO4), qui est dirigé au locus cible grâce à l'action de l'autre ARN polymérase spécifique des plantes, POLV, dont le rôle exact reste encore controversé (Wierzbicki et al. 2008, Wierzbicki et al. 2009). Notons ici qu'en plus d'AGO4, AGO6 et AGO9 jouent également un rôle dans la maintenance via le RdDM. Malgré une redondance partielle (Zheng et al. 2007), ces trois AGO présentent des divergences fonctionnelles (Vaucheret 2008) principalement dues à des différences de domaines d'expression (Eun et al. 2011, Havecker et al. 2010, Olmedo-Monfil et al. 2010). L'étape finale est le recrutement de la méthyltransférase DRM2 (DOMAIN REARRANGED METHYLTRANSFERASE 2) qui est l'homologue de la méthyltransférase *de novo* des mammifères Dnmt3 (Cao et al. 2000, Goll and Bestor 2005), et qui catalyse la méthylation dans tous les contextes nucléotidiques (Cao and Jacobsen 2002, Cao et al. 2003) (fig. 1.19 et 1.20).

Il est remarquable que ces trois machineries diffèrent par le taux de méthylation qu'elles produisent. De fait, la maintenance assurée par MET1 génère généralement un taux de

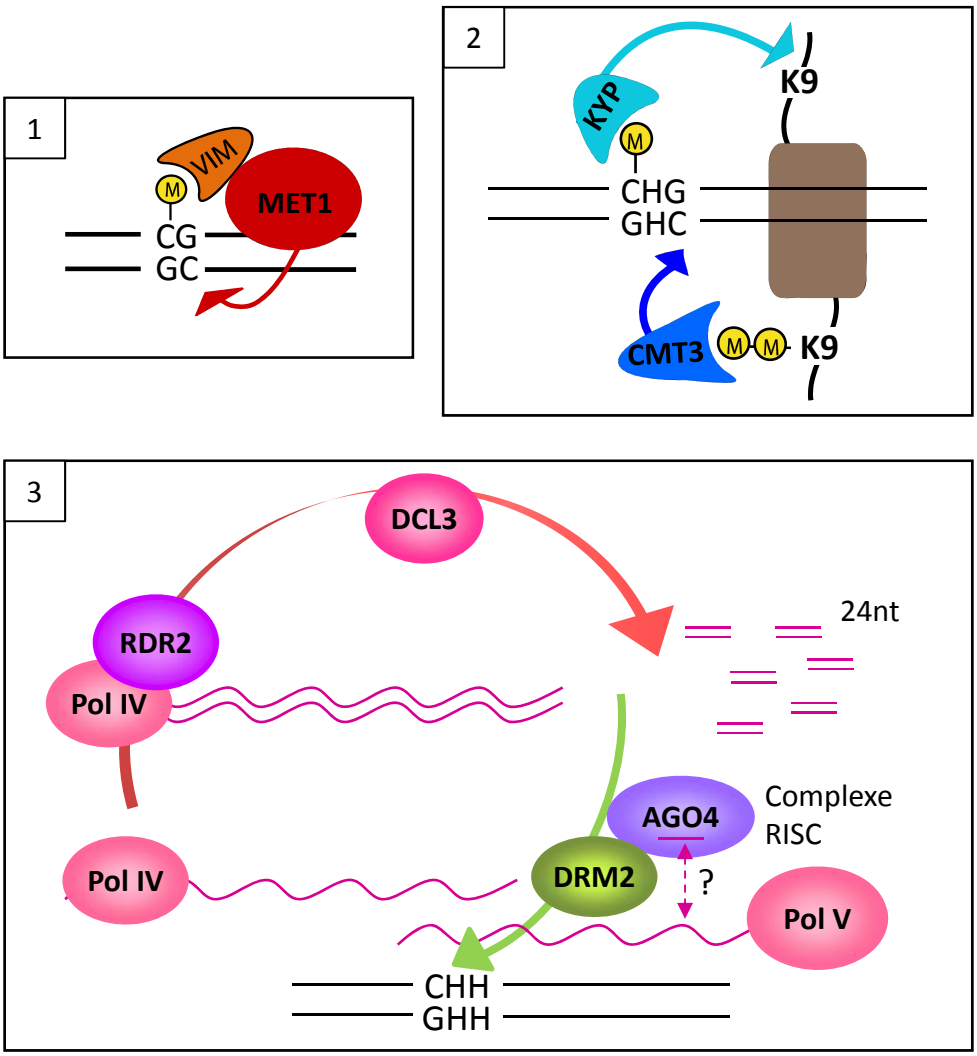


Figure 1.20 : Les trois voies de maintenance de la méthylation de l'ADN.
La numérotation correspond à celle de la figure 1.19:

1 : recrutement de MET1 par VIM1 sur les CG hémiméthylés

2 : boucle de renforcement entre H3K9me2 et la méthylation en contexte CHG via KYP et CMT3

3 : voie du RdDM

méthylation d'environs 80% à chaque site CG alors que l'action du RdDM atteint rarement un niveau supérieur à 40 % aux sites CHH et que la méthylation des sites CHG via CMT3 est intermédiaire (Lister et al. 2008, Cokus et al. 2008).

En plus des acteurs précédemment cités, la protéine remodeleuse de chromatine DECREASE IN DNA METHYLATION1 (DDM1), une ATPase à domaine SWI2/SNF2, joue un rôle majeur dans le maintien de la méthylation des séquences répétées (Vongs et al. 1993, Kakutani et al. 1995, Hirochika et al. 2000, Jeddeloh et al. 1999, Lippman et al. 2004, Teixeira et al. 2009). La fonction exacte de DDM1 reste inconnue même si une étude récente indique qu'elle servirait à faciliter l'accès des méthyltransférases de l'ADN aux régions de la chromatine associées à l'histone H1 (Zemach et al. 2013).

Les différentes machineries responsables de la maintenance de la méthylation de l'ADN sont partiellement interdépendantes les unes des autres. De fait, des mutations dans les différentes méthyltransférases de l'ADN n'engendrent pas uniquement des pertes de méthylation au niveau des cytosines qu'elles ciblent en théorie (Stroud et al. 2013). En outre une réduction de la production de petits ARN de 24nt a été observée dans les mutant *met1* et *ddm1* pour de nombreux locus (Lister et al. 2008 et données non publiées du laboratoire). Le premier mécanisme responsable de cette interdépendance a été mis en évidence très récemment. Il repose sur l'action de SHH1, une protéine capable de recruter la POLIV sur un sous-ensemble des cibles du RdDM en liant H3K9me2 (Law et al. 2011, Law et al. 2013).

1.3.2.2 Etablissement/Rétablissement de la méthylation de l'ADN

Si les mécanismes qui assurent la maintenance de la méthylation de l'ADN sont assez largement caractérisés, ceux menant à l'établissement *de novo* de la méthylation de l'ADN le sont moins. De fait, bien que de nombreux travaux, menés sur des transgènes ont mis en évidence le rôle prédominant de la machinerie du RdDM dans l'établissement de la méthylation de l'ADN chez les plantes (Chan et al. 2004, Greenberg et al. 2011), nous savons peu de choses sur la manière dont le RdDM reconnait ses cibles. Ainsi, les données génétiques suggèrent l'existence d'une étape antérieure permettant à POLIV de reconnaitre ses cibles. Se pourrait-il que le PTGS (Post transcriptional gene silencing, fig. 1.21), une autre voie de l'ARNi, soit impliqué? Plusieurs données militent en faveur de cette hypothèse,

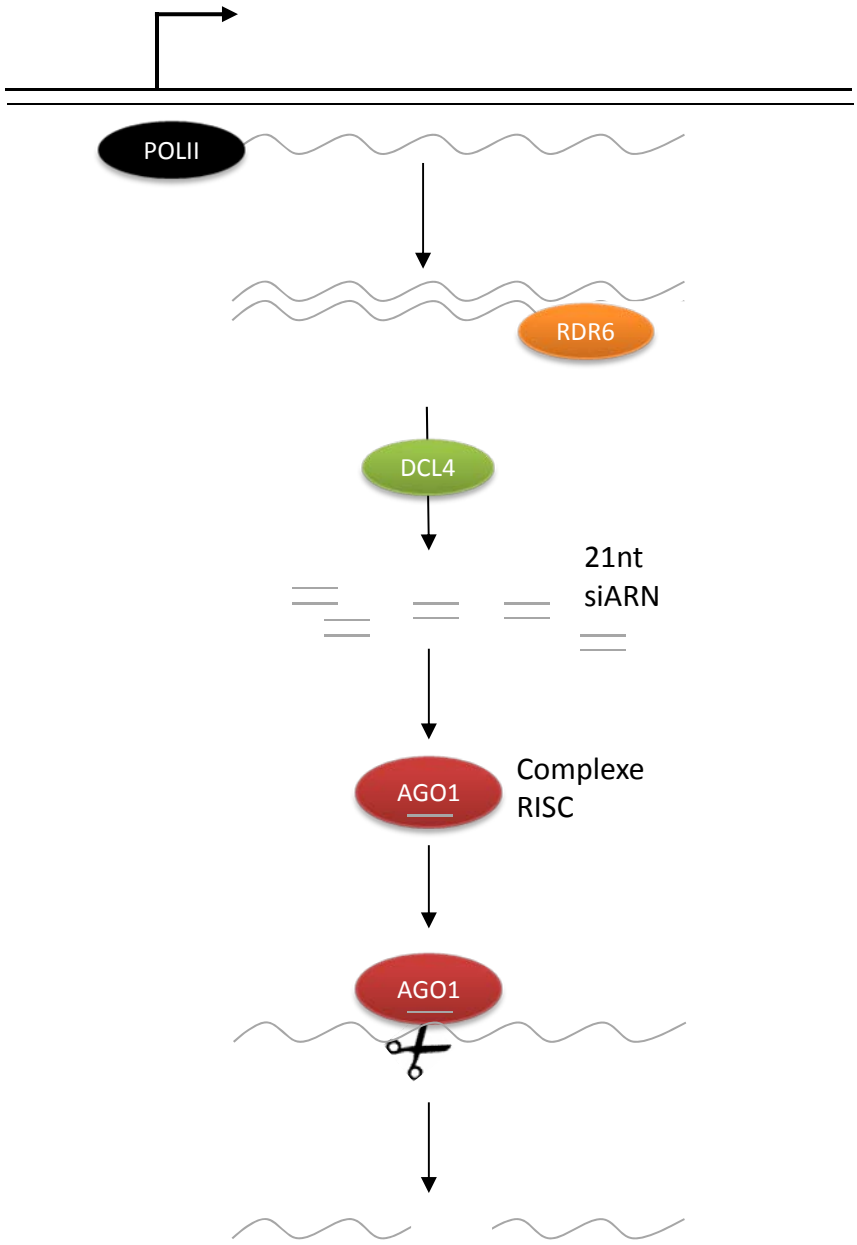


Figure 1.21 : Etapes principales de la voie du Sens-PTGS (SPTGS).
Adapté de Brodersen et al. 2006.

notamment le fait qu'il peut conduire à une méthylation de l'ADN indépendamment de l'action de RDR2 et POLIV (Baulcombe 1996, Mourrain et al. 2000, Elmayan et al. 1998, Jauvion et al. 2012). Il est donc tentant de supposer que dans certains cas, cette méthylation puisse permettre le recrutement de POLIV. Un cas de passage du PTGS vers le TGS a d'ailleurs été documenté sur un transgène dans des cellules en culture, mais la base moléculaire de ce « switch » n'a pas été démontrée (Fojtova et al. 2003).

L'établissement *de novo* de la méthylation de l'ADN par le RdDM, mis en évidence sur des transgènes, est très certainement responsable de la mise en place de la répression épigénétique sur les ET nouvellement insérés. En effet, nous savons déjà que le RdDM est essentiel à la reméthylation progressive d'un certain nombre d'ET suite à la perte de méthylation induite dans le mutant *ddm1* (fig. 1.22) (Teixeira et al. 2009). De plus, l'établissement de la méthylation par le RdDM a été mis en évidence pour le transposon MuDR du maïs. Ce transposon à ADN, très actif chez le maïs est très rapidement réprimé/méthylé lorsqu'il est mis en présence du locus Mu Killer (*MuK*). Le *MuK* est constitué d'une répétition inversée partielle d'un élément MuDR. Cette structure particulière engendre la production d'un ARN capable de s'apparier sur lui-même pour former une « épingle à cheveux » qui va être prise en charge par la machinerie du RdDM, produisant ainsi des siARN de 24nt spécifiques de MuDR et capables d'instruire en trans la répression des copies de MuDR (Slotkin et al. 2003, Slotkin et al. 2005).

1.3.2.3 Déméthylation de l'ADN

En l'absence d'activité d'établissement récurrent ou de maintien de la méthylation de l'ADN, celle-ci est inexorablement perdue au cours des réplications. En plus de cette déméthylation dite passive, il existe chez les mammifères et les plantes des mécanismes enzymatiques de déméthylation. On parle alors de déméthylation active. La première déméthylase identifiée chez les eucaryotes l'a été chez *Arabidopsis thaliana* (Choi et al. 2002). Il s'agit d'une ADN glycosylase, DEMETER (DME), qui excise les cytosines méthylées et qui serait préférentiellement exprimée dans la cellule centrale du gamétophyte femelle et l'albumen des graines, ainsi que dans la cellule végétative du pollen (Choi et al. 2002, Hsieh et al. 2009). Le génome d'Arabidopsis code trois autres déméthylases apparentées à DME : ROS1 (REPRESSOR OF SILENCING 1) ainsi que DML2 et 3 (DEMETER-LIKE 2 et 3) (Zhu 2009, Gehring

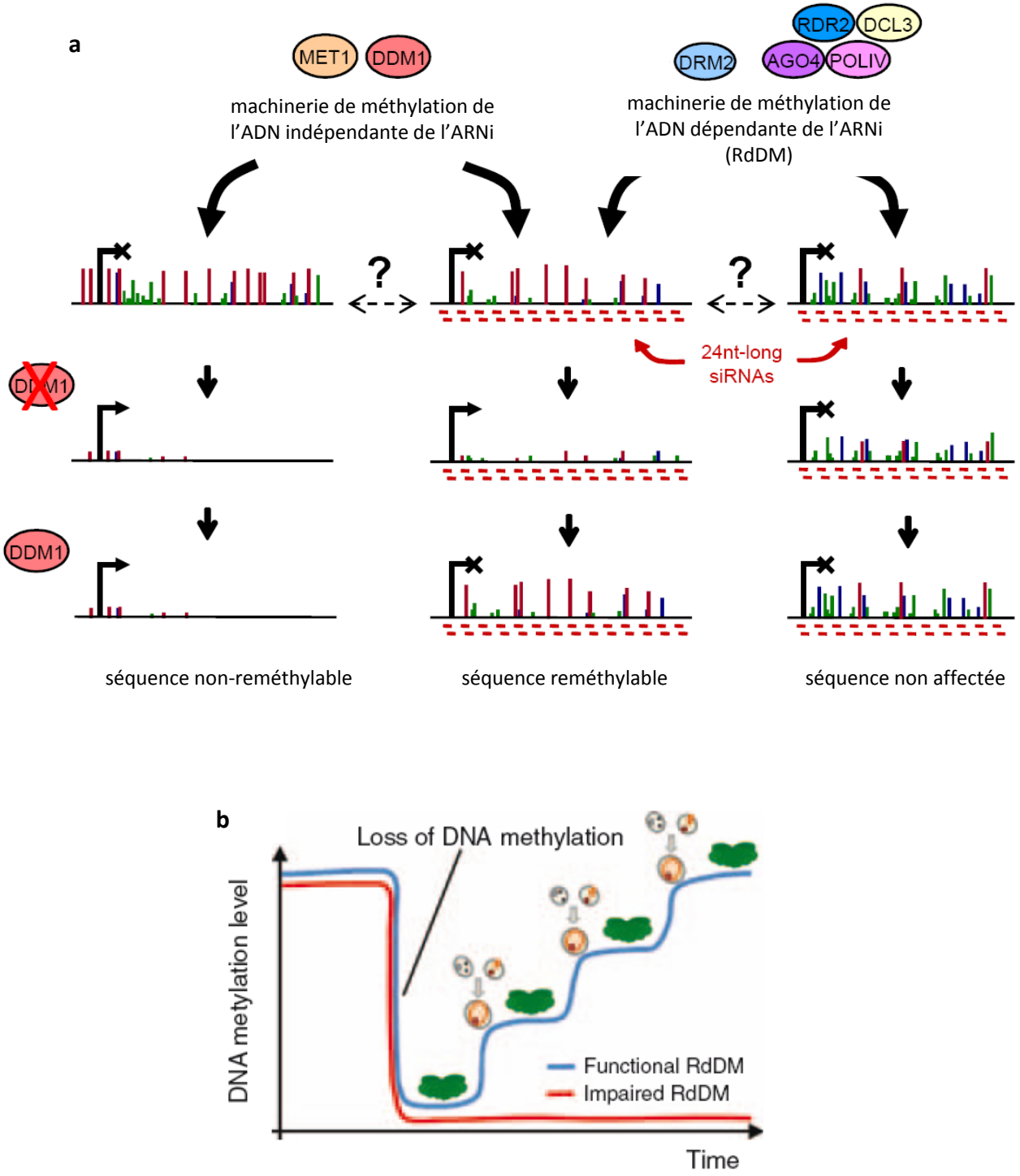


Figure 1.22 : Processus de restauration de la méthylation de l'ADN suite à une perte induite par une mutation dans le gène *DDM1.*

(**a**) Modèle du processus de reméthylation. Une mutation dans le gène *ddm1* engendre une perte drastique de la méthylation de l'ADN excepté sur les quelques séquences régulées quasi-exclusivement par le RdDM (séquences non affectées). Après ségrégation de la mutation *ddm1* par croisement contre une plante sauvage, certaines séquences sont capables de recouvrer un niveau sauvage de méthylation grâce à l'action du RdDM (séquences reméthylables). Les autres conservent un état hypométhylé stable (séquences non reméthylables). (**b**) Progressivité de la reméthylation au cours des générations. D'après Teixeira et al. 2009 et Teixeira et al. 2010.

et al. 2009). Bien que les mécanismes d'adressage des déméthylases aux locus qu'elles ciblent ne soient pas connus, des analyses génétiques suggèrent un lien avec le RdDM. Ainsi, l'hyperméthylation observée dans le triple mutant *ros1dml2dml3* est généralement associée à une sur-accumulation de siARN de 24nt et plusieurs suppresseurs de l'hyperméthylation induite par la mutation *ros1* sont des acteurs du RdDM (Zheng et al. 2007, Penterman et al. 2007, He et al. 2009). De plus il a été montré que ROS3, un acteur de la voie impliquant ROS1 est capable de se lier à des petits ARN *in vivo* (Zheng et al. 2008). Ces observations indiquent que, pour certains locus, les profils de méthylation résultent d'une action antagoniste entre le RdDM et la déméthylation (Teixeira and Colot 2010).

1.3.3 Autres marques et acteurs associés aux ET

Outre la méthylation de l'ADN, d'autres marques chromatiniennes sont associées aux ET chez Arabidopsis.

Evoquée précédemment, la diméthylation de la lysine 9 de l'histone H3 (H3K9me2) est une marque étroitement associée aux séquences répétées chez les plantes comme chez d'autres organismes, notamment ceux ne présentant pas de méthylation de l'ADN. Chez Arabidopsis néanmoins la maintenance de cette marque est interdépendante de celle de la méthylation de l'ADN car le domaine SRA de l'histone méthyltransférase SUVH4 (également connue sous le nom de KRYPTONITE (KYP)) peut se lier aux CHG méthylés (fig. 1.20). SUVH5 et SUVH6, deux autres histone méthyltransférases sont également impliquées dans la méthylation d'H3K9me2 étant donné que le triple mutant présente un phénotype moléculaire plus prononcé que le simple mutant *kyp,* mais leur recrutement semble être locus-spécifique (Ebbs and Bender 2006, Inagaki et al. 2010).

La monométhylation de la lysine 27 de l'Histone H3 (H3K27me1) est quant à elle catalysée par deux méthyltransférases d'histone ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) et ATXR6, mais leur adressage à l'hétérochromatine reste inconnu. Cette marque n'est pas requise pour la déposition d'H3K9me2 ni de la méthylation de l'ADN qui restent inchangées dans le double mutant ATXR5/6 (Jacob et al. 2009). Cependant, ce double mutant montre une réactivation transcriptionnelle de certains ET. L'un des rôles majeurs de

cette marque identifié à ce jour est d'empêcher une endoréplication excessive des régions hétérochromatiques (Jacob et al. 2010).

Enfin la monométhylation de la lysine 20 de l'histone H4 (H4K20me1) est associée avec les ET chez Arabidopsis (Roudier et al. 2011, Naumann et al. 2005). Peu de choses sont connues sur cette marque, cependant, sa localisation le long du génome d'Arabidopsis diffère drastiquement de celle des mammifères où elle est trouvée en 5' des gènes et associée à leur répression (Congdon et al. 2010).

En plus des marques chromatiniennes, certaines protéines ont été décrites comme impliquées dans le contrôle de la répression transcriptionnelle des ET. Parmi elles, MOM1 (Morpheus' Molecule 1) apparait comme l'une des plus énigmatiques étant donné que plusieurs travaux lui ont attribué des rôles et modes d'action diverses, voire contradictoires. Des mutations dans le gène *MOM1* ont pour conséquence une réactivation transcriptionnelle de certaines séquences répétées sans affecter la méthylation de l'ADN ni la compaction de la chromatine (Tariq et al. 2002, Scheid et al. 2002, Vaillant et al. 2006, Numa et al. 2010). Ses cibles présentent de la méthylation non CG et son action se ferait *via* une interaction avec POLV (Habu 2010, Yokthongwattana et al. 2010). Enfin, MOM1 participerait au maintien d'H3K9me2 sur certains locus (Numa et al. 2010). Récemment, il a été montré que les protéines MORC (appartenant à la famille protéique Microrchidia) sont également impliquées dans le maintien de la répression transcriptionnelle des ET (Moissiard et al. 2012, Brabbs et al. 2013). Cependant, ici encore, les analyses sont partiellement contradictoires quant au mode d'action de ces protéines. De fait, le double mutant *AtMORC1/6* présente une décompaction de l'hétérochromatine associée à la réactivation transcriptionnelle de rares ET et à aucun changement de méthylation de l'ADN ni d'H3K9me2 (Moissiard et al. 2012). Ce résultat suggère que l'état de compaction seul est déterminant pour assurer la répression efficace de certains ET. Cependant, il a également été montré que le simple mutant *AtMORC6* peut causer une perte stochastique de la répression associée à une réduction de la méthylation (Brabbs et al. 2013).

Pour finir, il est important d'évoquer le rôle des enzymes qui enlèvent les marques chromatiniennes associées avec l'activation transcriptionnelle et garantissent ainsi le maintien efficace des différentes marques répressives associées aux ET. Notamment des

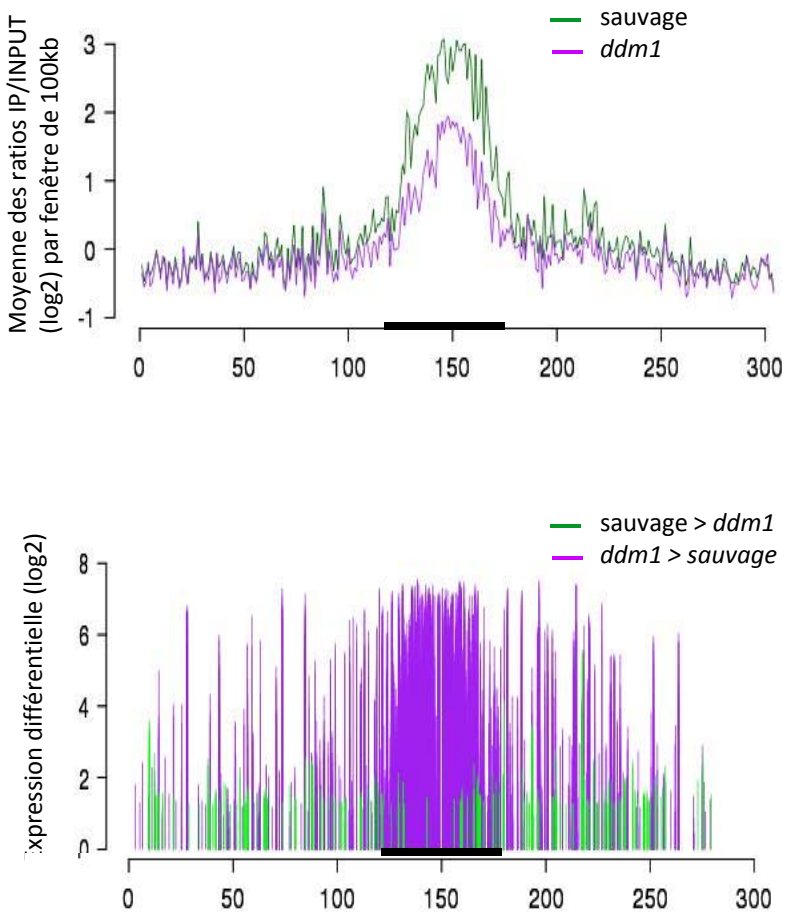


Figure 1.23 : La perte de méthylation induite par une mutation dans le gène *DDM1* cause la réactivation transcriptionnelle de très nombreux ET.

(**a**) Méthylation de l'ADN le long du chromosome 1 d' Arabidopsis (obtenue par MeDIP-chip) chez le sauvage et le mutant *ddm1*. (**b**) Expression différentielle (mesurée par ARN-chip) entre le sauvage et le mutant *ddm1.* La barre noire représente les deux régions péricentromériques.

mutations dans l'HISTONE DEACETYLASE6 (HDA6) impliquée dans la déacétylation d'H3K9, H3K14, et de la tétra-acétylation d'H4 (Earley et al. 2006) induisent, en plus d'un gain de marques actives, une réduction drastique d'H3K9me2 et de la méthylation de l'ADN sur certains locus (To et al. 2011, Lippman et al. 2003, Liu et al. 2012). Il est proposé qu'HDA6 travaille en coopération avec MET1 sur un sous-ensemble des séquences méthylées car l'interaction physique de ces deux protéines a été montrée *in vitro* et *in vivo* (To et al. 2011, Liu et al. 2012). Autre exemple notable, l'histone déméthylase JmJ14 qui catalyse la déméthylation d'H3K4me2 chez Arabidopsis (Lu et al. 2010) est nécessaire à la maintenance de la méthylation non-CG assurée par DRM2 (Deleris et al. 2010, Searle et al. 2010). Ces observations suggèrent qu'il existe une compétition permanente entre les voies activatrices et répressives de régulation de la chromatine.

1.3.4 Conséquences de la perte du contrôle épigénétique sur les ET

Les mutations affectant les gènes *MET1* et *DDM1* conduisent aux pertes de méthylation de l'ADN les plus drastiques (~70%) chez Arabidopsis. Ces pertes sévères de méthylation de l'ADN s'accompagnent d'une réduction d'H3K9me2 et d'une décondensation des chromocentres (Soppe et al. 2002). La conséquence de cette perturbation du contrôle épigénétique est la réactivation transcriptionnelle massive de plusieurs centaines d'ET de tous types (Lippman et al. 2004 Zemach et al. 2013) (fig. 1.23). Cette réactivation transcriptionnelle s'accompagne à certains locus de la production de siARN d'une taille de 21nt, ce qui suggère la mise en place d'un contrôle post-transcriptionnel pour certains des ET réactivés (Teixeira et al. 2009, Slotkin et al. 2009). Les mutants affectés dans la voie du RdDM, quant à eux, ne présentent que des changements mineurs de niveau de méthylation. Cependant, il semble que certains ET soient spécifiquement contrôlés par cette voie (Ito et al. 2011, Matsunaga et al. 2012).

Quoiqu'il en soit, il n'est pas clairement établi dans quelle mesure cette réactivation transcriptionnelle des ET se traduit par leur mobilisation. Si divers travaux ont permis de mettre en évidence la mobilisation de certains ET suite à des perte de méthylation chez différents mutants (Miura et al. 2001, Tsukahara et al. 2009 Singer et al. 2001 Johannes et al. 2009 Tsukahara et al. 2012 Ito et al. 2011, Matsunaga et al. 2012, Mirouze et al. 2009,

Reinders et al. 2009) la plupart d'entre eux ne s'intéressent qu'à une famille d'ET en particulier. A ce jour, une seule étude a été menée dans le but d'identifier de façon exhaustive les familles d'ET présentant des éléments mobiles chez le mutant *ddm1* (Tsukahara et al. 2009). Ces travaux ont notamment montré qu'un faible nombre de familles seulement présentait des éléments mobiles dans ce mutant. Néanmoins, il apparait qu'une famille de rétroélément à LTR COPIA (*ATCOPIA93*) est particulièrement active suite à une perte de méthylation de l'ADN car elle est responsable de phénomènes de « burst » de transposition. Cependant, ces analyses reposent sur des expériences de CGH (Comparative Genome Hybridization) et n'ont donc pas une résolution suffisante pour permettre l'identification du nombre et de l'identité des copies mobiles au sein d'une famille. De plus, bien que deux lignées mutantes aient été comparées dans cette étude, cela n'est pas assez pour estimer les différences de fréquences et dynamiques de mobilisation interindividus, autrement dit le caractère systématique versus stochastique de la mobilisation des différents ET.

1.4 Dynamique du contrôle des ET chez les plantes : au cours du développement et en réponse à l'environnement

1.4.1 Contrôle des ET au moment de la reproduction

Chez Arabidopsis, contrairement à ce qui est observé chez les mammifères, les profils de méthylation de l'ADN sont transmis au travers des générations plutôt que réinitialisés à chaque cycle de reproduction. Cette conclusion vient notamment de l'observation que les mutants affectés dans la voie du RdDM (méthylation *de novo*) gardent un niveau de méthylation quasi normal au travers des générations (Tran et al. 2005, Zhang et al. 2006, Penterman et al. 2007, Lister et al. 2008, Cokus et al. 2008). De plus, le fait que les pertes de méthylation induites par les mutations dans les gènes *MET1* et *DDM1* puissent perdurer au travers des générations pour de nombreux locus suggère qu'il n'existe pas un mécanisme capable de reprogrammer la totalité de l'épigénome à chaque cycle de reproduction (Kakutani et al. 1999). Néanmoins le processus de reméthylation mentionné plus haut, bien que restreint à un sous-ensemble des séquences méthylées du génome, illustre le rôle du RdDM dans le rétablissement des profils de méthylation au moment de la reproduction (Teixeira et al. 2009). De plus, plusieurs études récentes ont révélé l'existence d'une certaine dynamique des profils de méthylation de l'ADN lors de la reproduction (fig. 1.24).

Dans le gamétophyte femelle, on observe une déméthylation passive de la cellule centrale due à la répression de *MET1* par la voie RETINOBLASTOMA-RELATED 1 (RBR1) (Jullien et al. 2008), à laquelle s'ajoute une déméthylation active catalysée par la déméthylase DEMETER (DME). Cette hypométhylation qui perdure dans l'albumen est notamment à la base du phénomène d'empreinte parentale, qui chez les plantes est restreinte à cet organe nourricier (Hsieh et al. 2009, Gehring et al. 2009). Cette hypométhylation serait également à l'origine de la production de siARN de 24nt dans la cellule centrale et dans l'albumen (Mosher et al. 2009). Si le rôle de ces siARN n'est pas encore clair, il a été proposé que leur transport vers la cellule œuf et/ou l'embryon après fécondation joue un rôle de renforcement de la répression transcriptionnelle des ET à chaque début de cycle de vie. L'idée que le contrôle des ET puisse se faire grâce à des petits ARN mobiles est également soutenue par la mise en évidence que l'expression d'AGO9 dans les cellules sporophytiques

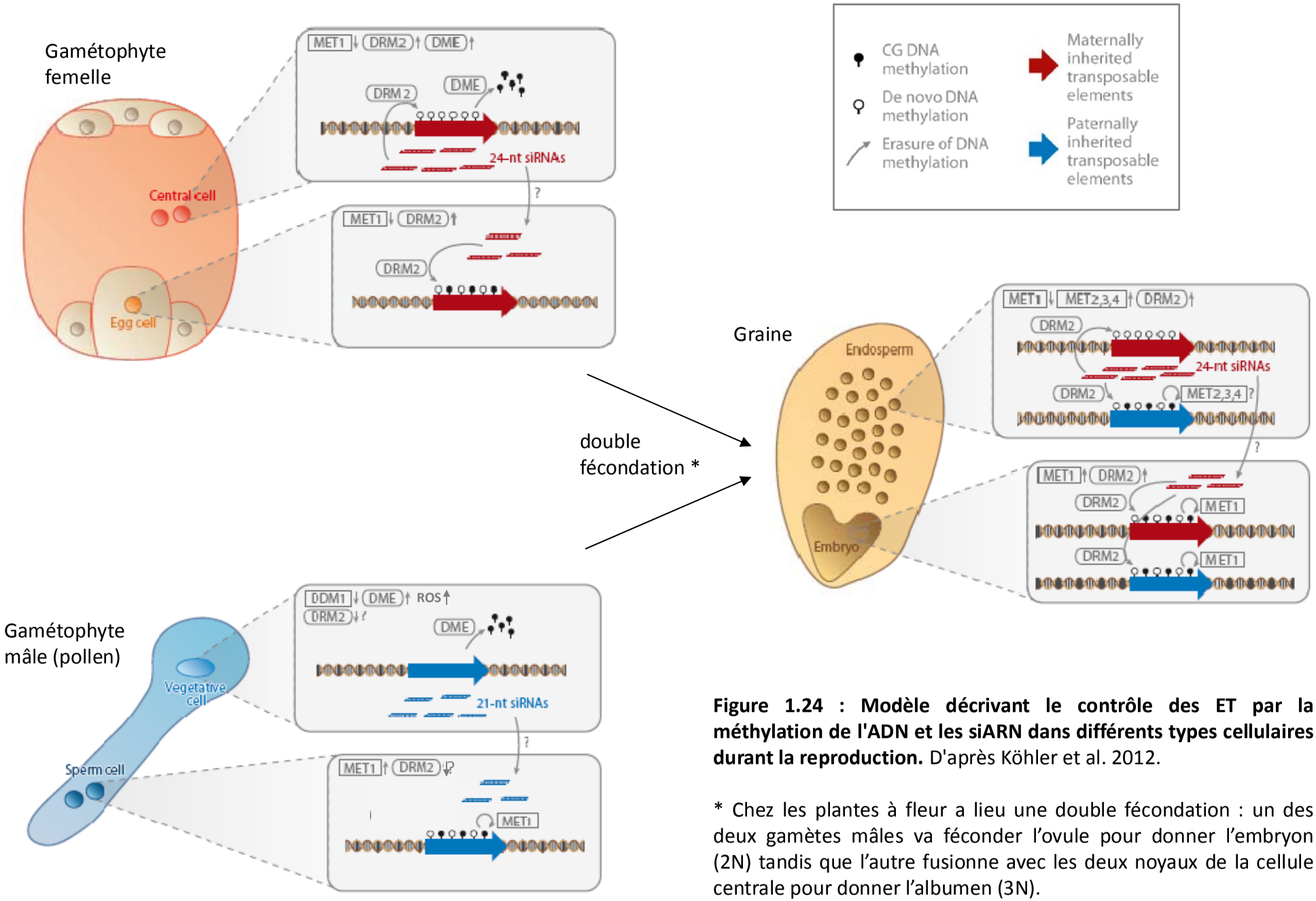


Figure 1.24 : Modèle décrivant le contrôle des ET par la méthylation de l'ADN et les siARN dans différents types cellulaires durant la reproduction. D'après Köhler et al. 2012.

* Chez les plantes à fleur a lieu une double fécondation : un des deux gamètes mâles va féconder l'ovule pour donner l'embryon (2N) tandis que l'autre fusionne avec les deux noyaux de la cellule centrale pour donner l'albumen (3N).

est nécessaire pour la répression des ET du gamète femelle dans un processus cellulaire non autonome. (Olmedo-Monfil et al. 2010, Durán-Figueroa and Vielle-Calzada 2010). Un phénomène analogue a été également observé pour AGO104 chez le maïs (Singh et al. 2011).

Du coté mâle, on observe également une déméthylation passive dans le noyau végétatif due à l'absence de MET1 et DDM1 (Jullien et al. 2008, Slotkin et al. 2009) à laquelle s'ajoute l'action de DME et ROS (Calarco et al. 2012). L'hypométhylation résultante serait associée à la réactivation transcriptionnelle des ET, qui engendre la production de siARN de 21nt (Slotkin et al. 2009), voire de 24nt. Ici encore il est proposé que ces siARN soient mobiles et impliqués dans la répression des ET dans le gamète mâle (Ibarra et al. 2012, Slotkin et al. 2009). Quoiqu'il en soit, il n'est pas clairement établi si cette répression est transcriptionnelle ou post transcriptionnelle.

Si ces évènements d'hypométhylation dans les cellules compagnes des gamètes semblent permettre un renforcement du contrôle de la répression des ET au moment de la reproduction, ils pourraient également être requis pour un phénomène de reprogrammation partiel de la méthylation de l'ADN sur les ET. De fait, dans les gamètes s'opère également une hypométhylation, bien que plus modérée que dans les cellules compagnes, due à une réduction de MET1 dans l'ovule et de DRM2 dans le gamète mâle (Calarco et al. 2012, Jullien et al. 2012). Cette perte partielle de méthylation dans les gamètes pourrait être rétablie grâce à l'action des siARN générés suite à l'hypométhylation observée dans les cellules compagnes et l'albumen (Calarco et al. 2012, Grimanelli and Roudier 2013).

1.4.2 Comportement des ET en réponses à des variations environnementales

Une des particularités des ET est leur capacité à « répondre au stress ». De fait, de nombreux travaux effectués chez les plantes ont mis en évidence une activation transcriptionnelle des ET en réponse à divers stress aussi bien biotiques (infection bactérienne (Pouteau et al. 1994, Grandbastien et al. 2005, Yu et al. 2013) ou virale (Buchmann et al. 2009) qu'abiotiques (stress salin (Naito et al. 2009), blessure (Mhiri et al. 1997), traitement au froid (Ivashuta et al. 2002, Naito et al. 2009), stress chaleur (Lang-Mladek et al. 2010, Pecinka et

al. 2010, Tittel-Elmer et al. 2010). La cause de cette réactivation transcriptionnelle demeure peu claire dans de nombreux cas. Cependant, certains stress semblent pouvoir engendrer l'expression spécifique de certains ET du fait de la présence de séquences de réponse au stress dans leurs promoteurs. C'est notamment le cas du rétrotransposon tnt1 du tabac qui contient dans sa LTR une région similaire à celle présente chez les gènes impliqués dans la réponse au stress (Mhiri et al. 1997, Grandbastien et al. 1994)

Chez Arabidopsis, le traitement à un stress chaleur induit l'activation transcriptionnelle de nombreux ET (Pecinka et al. 2010, Tittel-Elmer et al. 2010). Cependant, aucun changement de méthylation de l'ADN ni de marques d'histones n'a pu clairement être associé à cette activation. La seule cause potentielle identifiée est une réduction de l'occupation par les nucléosomes (Lang-Mladek et al. 2010, Pecinka et al. 2010) qui aurait pour conséquence une chromatine moins condensée. Cette hypothèse est d'ailleurs en accord avec la réactivation transcriptionelle de certains ET dans les mutants *AtMORC* qui présenteraient une chromatine décompactée mais pas de changement du niveau de méthylation ni de marques d'histones (Moissiard et al. 2012). Quoiqu'il en soit, l'induction transcriptionnelle des ET par la chaleur est transitoire et ne semble pas se traduire par leur mobilisation. La raison pour laquelle la mobilité des ET n'est pas observée dans ce cas est due (au moins pour certains d'entre eux) à l'action du RdDM. En effet, les études sur les rétroéléments LTR/COPIA *ONSEN* (famille *ATCOPIA78*) ont mis en évidence leur mobilisation uniquement dans les mutants affectés dans le RdDM et soumis à un stress chaleur (Ito et al. 2011, Matsunaga et al. 2012). Il semble en revanche que certains stress biotiques engendrent une perte de méthylation (Pavet et al. 2006, Yu et al. 2013) qui peut s'accompagner de la réactivation transcriptionnelle de certains ET (Yu et al. 2013).

Enfin, il a été montré que certaines conditions environnementales (comme une faible lumière) ainsi que certains stades développementaux (germination, transition florale) s'accompagnent de la décompaction des régions péricentromériques (Tessadori et al. 2007 Tessadori et al. 2009, van Zanten et al. 2010 van Zanten et al. 2011) et pourraient donc engendrer une réactivation des ET, mais cela reste encore à démontrer.

1.5 Rôle des ET dans la régulation génique

1.5.1 Implication des ET dans la régulation génique

Les ET, dès qu'ils transposent, sont de puissants mutagènes. Ils peuvent, par exemple, interrompre la séquence codante des gènes. Les insertions de ce type sont pour la plupart particulièrement délétères et sont donc rapidement purgées par la sélection naturelle. De plus, étant donné que les ET contiennent leur propres séquences promotrices et régulatrices, ils peuvent affecter les profils d'expression des gènes situés à proximité et ce de différentes manières. Par exemple, le gène *b1* du maïs, impliqué dans la synthèse d'une anthocyanine est normalement exprimé dans le « corps » de la plante. Néanmoins, deux variants de *b1* (*B-Peru et B-Bolivia*) présentent une expression dans la graine due à l'insertion d'un ET contenant des séquences suffisantes pour l'expression graine-spécifique (Selinger and Chandler 2001). De même, des études récentes réalisées chez le riz ont notamment montré que des insertions de MITE à moins de 100 bp du TSS de gènes peuvent conférer à ces derniers la capacité d'être induits par un stress salin et le froid (Naito et al. 2009). Enfin, les ET peuvent également produire des transcrits « outward reading » qui se prolongent en dehors de l'élément, dans le gène adjacent. Ces transcriptions intempestives peuvent activer les gènes adjacents comme dans le cas d'*hcf106* chez le maïs (décrit plus loin) ou à l'inverse bloquer leur expression comme dans le cas des rétroéléments à LTR *Dasheng* du riz dont l'expression est associée à la répression des gènes adjacents (Kashkush and Khasdan 2007).

1.5.2 Conséquences de la régulation épigénétique des ET sur les gènes situés à proximité

Un autre niveau de contrôle de l'expression des gènes par les ET provient du fait qu'ils sont la cible quasi-systématique des mécanismes de répression épigénétique, notamment la méthylation de l'ADN chez les plantes et les mammifères.

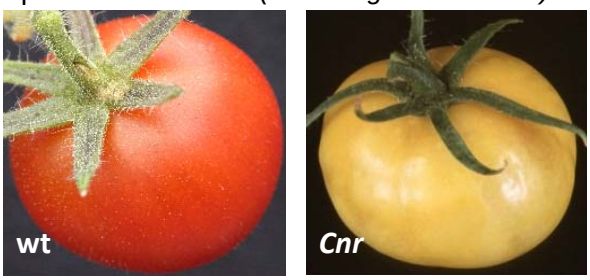
Des études récentes ont mis en évidence une corrélation négative chez *A. thaliana* entre le niveau d'expression des gènes et la densité d'ET méthylés (Hollister and Gaut 2009). De plus, des analyses basées sur des comparaisons entre *Arabidopsis thaliana* et *lyrata* à l'échelle du

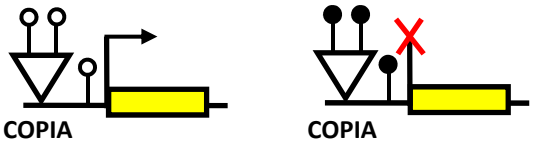
génome ont montré que les ET ciblés par les siARN de 24nt (donc par le RdDM) ont un impact plus fort sur l'expression génique que ceux qui ne sont pas associés avec des siARN de 24nt. En outre, les orthologues entre les deux espèces, qui diffèrent par la présence d'un ET à proximité, présentent des différences hautement significatives d'expression mais seulement si l'ET est associé avec des siARN de 24nt (Hollister et al. 2011). Cependant, ces effets ne sont détectables que si l'ET est situé à faible distance du gène (200 bp chez *A. thaliana* et 400 bp chez *A. lyrata*). Ces observations sont en accord avec le fait que la méthylation de l'ADN peut s'étendre sur une courte distance (environs 200 bp) à partir des séquences ciblées par le RdDM (Ahmed et al. 2011).

Il semble donc que l'état de méthylation des ET puisse jouer un rôle dans la régulation des gènes situés à proximité. Autrement dit, ces gènes vont potentiellement être induits/surexprimés dans les conditions développementales ou environnementales induisant l'hypométhylation des ET. Une fois encore, on peut évoquer ici les conditions de stress qui peuvent avoir pour conséquence l'hypométhylation de certains ET (Yu et al. 2013). Cependant le cas le mieux décrit et documenté est celui de l'empreinte parentale chez les plantes et Arabidopsis notamment. En effet, comme présenté précédemment, la chute d'expression de *MET1* et l'activation de *DME* dans la cellule centrale engendre la déméthylation des ET et cette perte de méthylation perdure dans l'albumen. Les copies paternelles de ces ET sont quant à elles méthylées. Ainsi, dans l'albumen les gènes dont l'expression est normalement réprimée à cause de la présence d'ET méthylés à proximité peuvent s'exprimer et ce, uniquement à partir de l'allèle maternel. La détection de séquences méthylées correspondant majoritairement à des ET (ou des fragments d'ET) à proximité des gènes qui présentent une expression albumen-spécifique a d'ailleurs permis d'identifier plusieurs nouveaux cas d'expression soumise à l'empreinte parentale (Gehring et al. 2009). Réciproquement, les gènes identifiés comme différentiellement exprimés selon leur origine parentale dans l'albumen sont souvent associés à des ET (Wolff et al. 2011). Cependant, à l'exemple de *MEDEA,* premier gène soumis à l'empreinte découvert chez Arabidopsis, l'association à des ET n'est pas systématique (Kinoshita et al. 1999 Spillane et al. 2004).

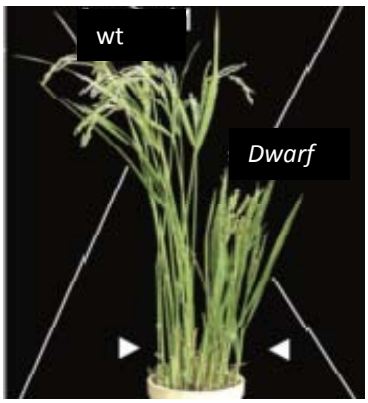
Enfin, un cas remarquable de l'impact de la méthylation des ET sur les gènes à proximité est la mise en évidence récente que la détermination du sexe chez le melon est contrôlée par la

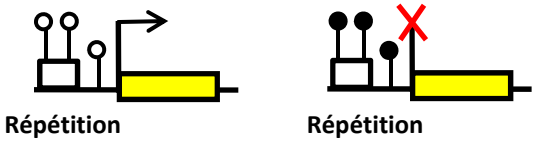
Epimutation *Cnr* *(Manning et al. 2006)*





Epimutation *Dwarf (Miura et al. 2009)*





Exemples d'instabilité



Figure 1.25 : Exemples d'épimutations causées par des variations de méthylation sur des ET (ou autres séquences répétées).

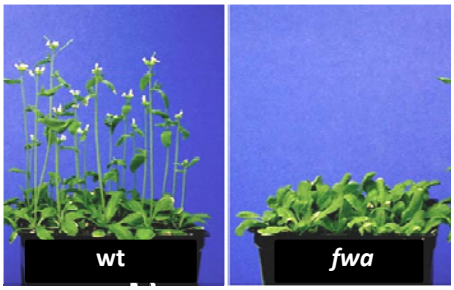
présence d'un transposon à ADN hAT en 3' du gène CmWIP1. La présence de l'ET méthylé entraine la méthylation stable du promoteur de CmWIP1 et son extinction transcriptionnelle. Ce gène étant nécessaire à l'induction des étamines, son extinction a pour conséquence la production de fleurs femelles exclusivement (Martin et al. 2009).

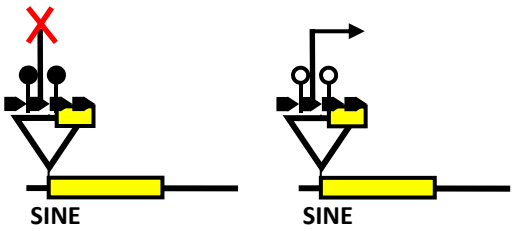
1.5.3 Les éléments transposables à l'origine de la création d'épimutations

Une épimutation est une variation héritable de l'expression d'un gène qui n'est pas due à un changement de la séquence nucléotidique. De par leur rôle potentiel dans la régulation des gènes situés à proximité, les ET sont une source privilégiée d'épimutations. De fait, parmi les quelques épimutations naturelles caractérisées (une dizaine seulement à ce jour), la majorité d'entre elles est causée par des variations de la méthylation de l'ADN sur des séquences d'ET situées à proximité du gène. Il est notable que les plantes fournissent le plus de cas d'épimutations, sans doute à cause de l'absence de reprogrammation massive de la méthylation de l'ADN dans l'embryon ou les cellules germinales. Dans les exemples de *cnr* et *dwarf* (fig. 1.25), l'épimutant présente un gain de méthylation sur la séquence d'ET située en 5' du gène ayant pour conséquence l'extinction de ce dernier (Manning et al. 2006, Miura et al. 2009). Un autre exemple bien caractérisé d'épimutation est celui d'*hcf106* chez le maïs. Ici, l'insertion d'un transposon à ADN Mu non méthylé en 5' du gène *hcf* cause l'extinction de ce dernier et se traduit par un phénotype albinos. Cependant lorsque Mu est méthylé il génère, à partir de son extrémité 3', un transcrit qui se prolonge dans le gène *hcf* et qui restaure le phénotype sauvage. Il est notable que l'apparition de clones révertants présentant Mu méthylé augmente au cours du développement de la plante. La formation des organes reproducteurs femelles chez le maïs étant plus précoce dans le développement que celle des organes mâles, l'allèle hypométhylé est plus fréquemment transmis par la femelle que par le mâle qui transmettra souvent l'allèle méthylé stablement (Martienssen et al. 1990, Barkan and Martienssen 1991).

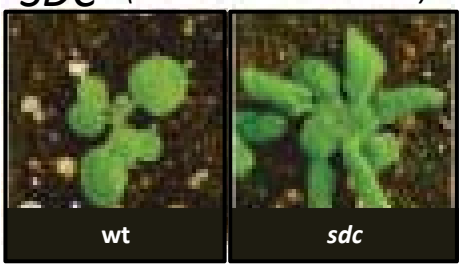
Bien que transmissibles, les épimutations sont en général instables et présentent des réversions plus ou moins élevées, au travers des générations comme au sein d'un même individu. Cette instabilité donne notamment lieu à des phénotypes de bigarrure (fig. 1.25) correspondant à des variations somatiques clonales de l'état de méthylation et d'expression.

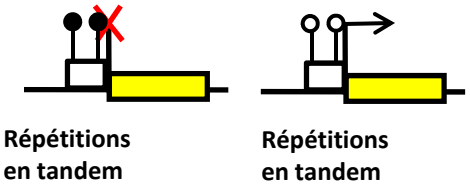
FWA *(Soppe et al. 2000)*





SDC *(Henderson et al. 2008)*





BNS *(Saze et al. 2007)*



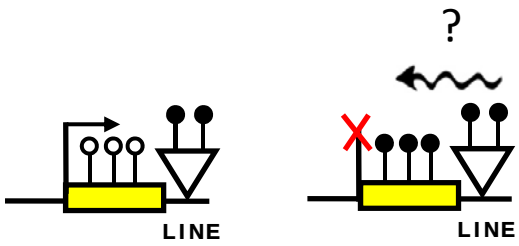


Figure 1.26 : Exemples de dérégulations géniques et leurs conséquences phénotypiques induites par des mutations affectant les voies de la méthylation de l'ADN chez Arabidopsis.

1.5.4 Impact de la perturbation du contrôle épigénétique des ET sur la régulation des gènes chez *Arabidopsis thaliana*

Compte tenu de l'impact que peuvent avoir les ET, sur l'expression des gènes situés à proximité, on peut s'attendre à ce que des perturbations du contrôle épigénétique des ET aient pour conséquence une dérégulation importante des gènes. Cependant, une mutation dans le gène *DDM1*, bien qu'entrainant une perte importante de la méthylation au niveau des ET, ainsi que leur réactivation transcriptionnelle, ne semble avoir que peu d'impact sur la régulation des gènes (Lippman et al. 2004, Vongs et al. 1993). De fait, le mutant *ddm1* ne présente pas de variation phénotypique majeure du moins dans les premières générations d'autofécondation. En revanche, après plusieurs générations d'autofécondation des altérations phénotypiques sporadiques mais sévères peuvent survenir (Kakutani 1997, Kakutani et al. 1996, Saze and Kakutani 2007, Soppe et al. 2000). Si certaines d'entre elles peuvent être dues à la mobilisation des éléments transposables, d'autres sont causées par des variations de la méthylation de l'ADN (fig. 1.26). C'est notamment le cas du gène *FWA* qui présente une séquence répétée en tandem dérivée d'un SINE qui couvre le promoteur et une partie du premier exon (Lippman et al. 2004). L'hypométhylation complète de cette séquence engendre l'expression ectopique du gène et se traduit par un phénotype caractéristique d'un retard important du temps de floraison. La perte complète de méthylation du locus *FWA* est très stable car elle peut être transmise au cours des générations même après restauration de la fonction DDM1 après croisement avec une plante sauvage (Kakutani 1997, Soppe et al. 2000). *FWA* est le premier cas d'épimutation (induite par une mutation) décrit chez Arabidopsis. D'autres épimutations ont depuis été identifiées dans les générations avancées du mutant *ddm1* dont *bonsai*. Etonnamment, cette épimutation est causée par l'hyperméthylation et non par l'hypométhylation du gène *BONSAI* (*BNS*) et est associée à son extinction. Comme son nom l'indique, *bonsai* produit un phénotype de petite taille. L'hyperméthylation du gène pourrait être causée par la réactivation transcriptionnelle d'un rétroélément LINE situé en 3' et dans l'orientation opposée du gène *BONZAI,* ce qui conduirait à la formation d'un ARN double brin, élément déclencheur du recrutement du RdDM (Saze and Kakutani 2007).

Enfin, le dernier exemple de gène identifié comme dérégulé par une perte de méthylation et présentant des conséquences phénotypiques majeures chez Arabidopsis est *SUPRESSOR OF*

DRM1 DRM2 CMT3 (*SDC*) identifié dans le triple mutant *drm1drm2cmt3*. La perte de méthylation au niveau de la répétition en tandem située en 5' du gène *SDC* a pour conséquence sa surexpression qui cause un phénotype de feuilles enroulées sur elles-mêmes. Cette dérégulation n'est cependant pas stable car cette séquence réacquiert immédiatement une méthylation sauvage après ségrégation des mutations (Henderson and Jacobsen 2008).

Ainsi donc, peu de gènes ont été identifiés (chez Arabidopsis comme chez les autres espèces) comme étant directement influencés par l'état de méthylation des ET ou autres séquences répétées situés dans leur voisinage. L'une des raisons possibles est que ces épimutations (naturelles ou induites) ont étés identifiées sur la base des phénotypes forts qu'elles causent. Néanmoins, il a été montré sur la population de lignées recombinantes épigénétiques (epiRIL) générée au laboratoire que des différences de profils de méthylation au niveau des séquences répétées entre les différentes lignées étaient associées à des variations héritables de traits quantitatifs (Johannes et al. 2009). Ces données suggèrent donc l'existence d'autres gènes contrôlés par la méthylation que les quelques uns déjà décrits.

1.6 Objectifs de la thèse

Dans le cadre de mon travail de thèse, j'ai dans un premier temps étudié le rôle de la méthylation de l'ADN dans le contrôle de la mobilisation des ET chez Arabidopsis. Pour ce faire, j'ai caractérisé les évènements de transposition ayant eu lieu dans la population de lignées recombinantes épigénétiques générée au laboratoire par re-séquençage de leur génome. A partir de ces données, j'ai également étudié la distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis.

Dans un second temps, je me suis intéressée à l'établissement de la répression épigénétique, et plus particulièrement à celui de la méthylation de l'ADN, sur les ET nouvellement insérés. Dans ce but, j'ai cherché à caractériser les modalités d'établissement de la méthylation de l'ADN sur les nouvelles insertions d'ET dans les epiRIL et j'ai mis au point un crible génétique afin d'identifier des acteurs impliqués dans ce processus.

Pour finir je me suis intéressée à l'influence directe des variations de méthylation des ET sur la régulation des gènes situés à proximité. J'ai donc étudié les variations d'expression de gènes situés à proximité des nouvelles insertions d'ET avant et après établissement du contrôle épigénétique. J'ai également participé à l'étude des conséquences de la perte de méthylation de l'ADN des ET sur l'expression des gènes adjacents.

2
Etude des conséquences d'une perte de méthylation de l'ADN sur la mobilisation des éléments transposables

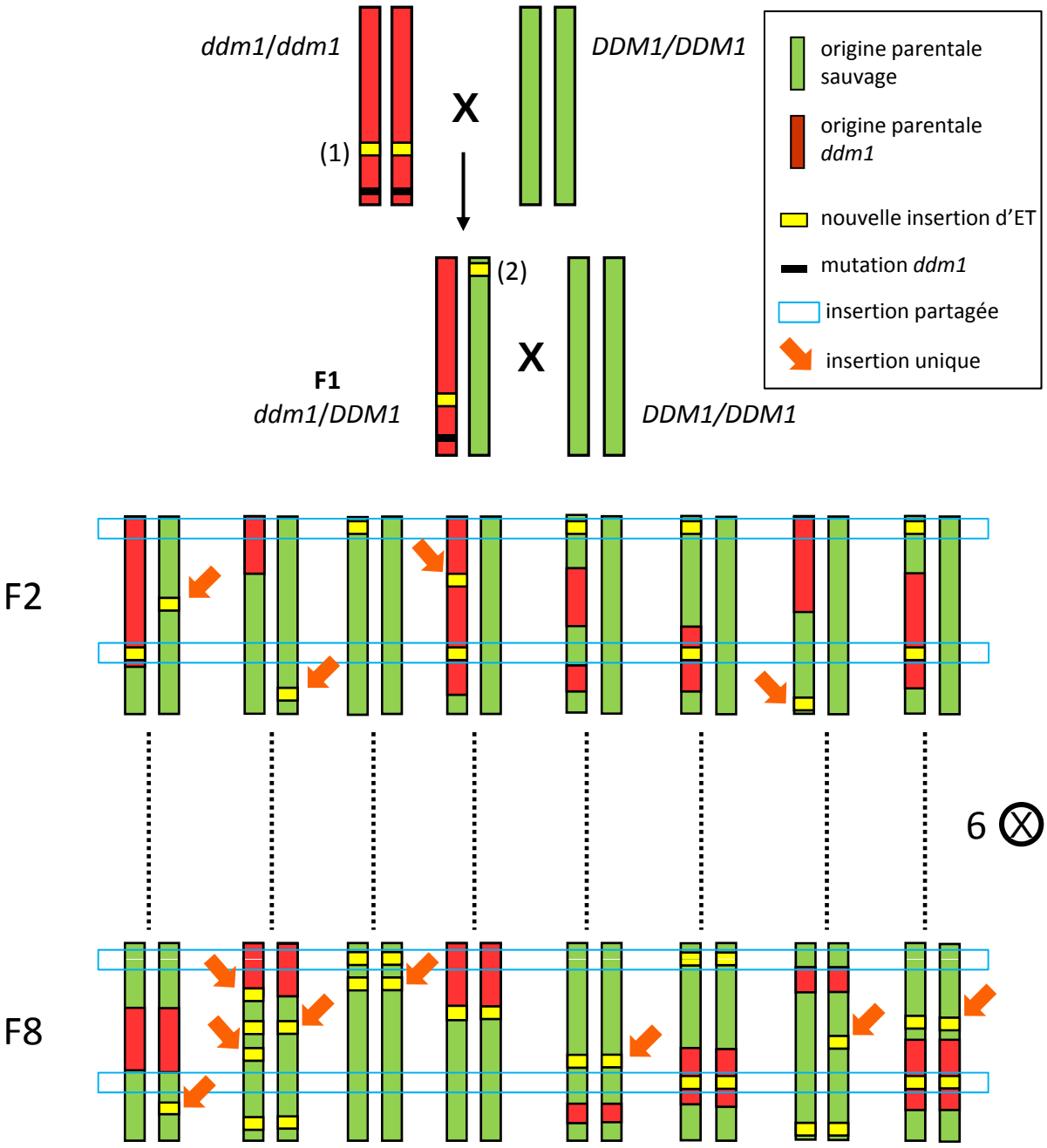


Figure 2.1 : Evènements de mobilisation potentiellement détectables dans les epiRIL.
La perte de méthylation induite dans le mutant *ddm1* cause la mobilisation de certains ET (1). Cette hypométhylation étant transmise à la descendance (du moins à certains locus), on peut s'attendre à de nouveaux évènements de transposition dans celle-ci. Alors que les nouvelles insertions produites dans le parent *ddm1* ou la F1 (1 et 2) vont ségréger dans la population d'epiRIL et vont donc être présentes dans plusieurs lignées (insertions partagées), celles produites à partir de la F2 seront propres à chaque lignée (insertions uniques). En F8, chaque lignée présente donc un jeu de nouvelles insertions, composé d'insertions partagées et d'insertions uniques.

Comme décrit dans l'introduction, une perte drastique de la méthylation de l'ADN, telle qu'induite par une mutation dans le gène *DDM1*, engendre une réactivation transcriptionnelle massive des ET. Cela dit, il n'est pas établi dans quelle mesure cette réactivation transcriptionnelle des ET se traduit par leur mobilisation. Afin de répondre à cette question ainsi que pour déterminer le profil d'insertion des ET, nous avons procédé au séquençage du génome de 53 epiRIL ainsi que des deux lignées parentales.

2.1 Identification des évènements de transposition ayant eu lieu lors de la génération des epiRIL

La population d'epiRIL est issue d'un croisement entre un parent sauvage et le mutant *ddm1-2* suivi d'un rétrocroisement avec le parent sauvage à la suite duquel seuls les individus homozygotes pour l'allèle sauvage du gène *DDM1* ont été sélectionnés et autofécondés sur 6 générations (fig. 2.1). Compte tenu du schéma de croisement, chaque point du génome est en moyenne d'origine sauvage dans 75% des lignées et hérité du parent *ddm1* dans 25% les lignées, sauf bien sûr à proximité du locus *DDM1*, systématiquement d'origine sauvage. Etant donné que l'hypométhylation induite par la mutation *ddm1* est transmise de façon stable pour de nombreux locus, les epiRIL présentent des profils de méthylation contrastés (fig. 2.1).

Afin d'identifier les évènements de transposition ayant eu lieu dans le mutant *ddm1* ou lors de la production des epiRIL (fig. 2.1) nous avons, en partenariat avec le Génoscope, réalisé le séquençage Illumina « paired-end » de banques « mate-pair » du génome de plus d'une cinquantaine de ces lignées ainsi que de deux individus sauvages et d'un mutant *ddm1* cousins des parents utilisés pour générer la population d'epiRIL.

Le séquençage paired-end, contrairement au séquençage de lectures uniques permet d'identifier des variations structurales en se basant sur la détection de paires dites discordantes, à savoir de paires dont les deux lectures ne sont pas positionnées à la bonne distance l'une de l'autre ou mal orientées par rapport au génome de référence. Le choix de banques mate-pair permet d'obtenir des lectures appariées situées physiquement à une

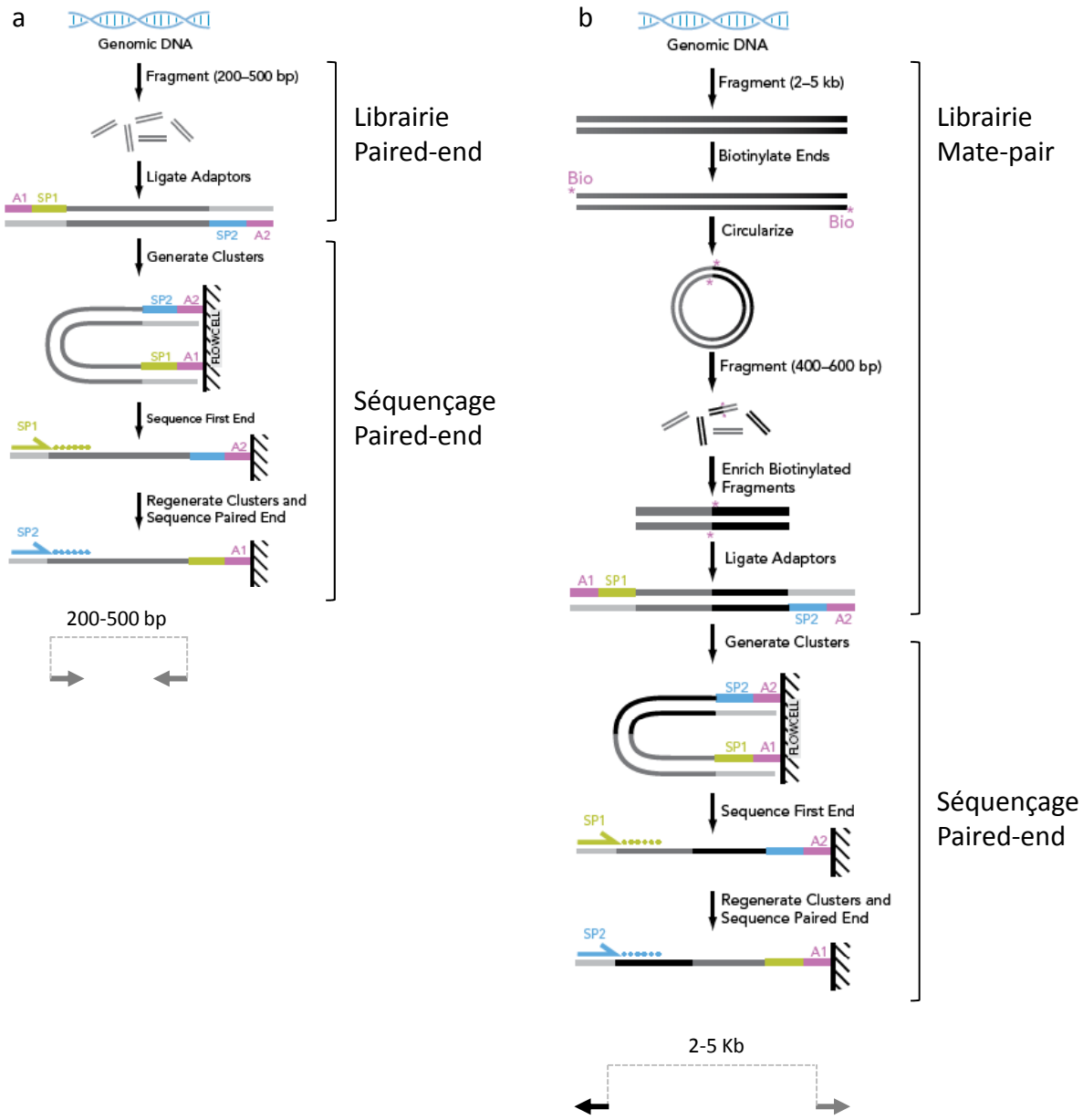


Figure 2.2 : Protocole de séquençage paired-end de banques paired-end (a) vs banques mate-pair (b).

grande distance l'une de l'autre (jusqu'à plusieurs kb), contrairement aux banques paired-end classiques qui couvrent de petites distances (500 bp maximum) (fig. 2.2). Le choix de banques mate-pair présente deux avantages majeurs : (i) elle permet d'augmenter le nombre de lectures qui soutiennent une variation structurale donnée et donc l'exhaustivité et la robustesse de la détection de ces évènements ; (ii) elle produit une plus grande couverture horizontale des variations structurales ce qui permet d'avoir des informations sur la quasi-totalité de la séquence insérée et notamment sa partie interne. Cette propriété est particulièrement importante dans le cadre de la détection de nouvelles insertions d'ET car l'une des difficultés majeures est l'identification précise du locus donneur. Avec l'approche que nous avons choisie, même un très faible niveau de polymorphisme, où qu'il soit localisé dans la séquence de l'ET, est suffisant pour discriminer entre plusieurs donneurs.

Afin d'optimiser la détection des nouvelles insertions d'ET basée sur des données de séquençage de banques mate-pair, j'ai participé à l'élaboration de TE-Tracker, un programme spécifiquement dédié à cette problématique. Ce travail est présenté dans un manuscrit en cours de finalisation, reproduit ci-après.

TE-Tracker: systematic identification of transposition events through whole-genome resequencing

Arthur Gilly[1,*], Mathilde Etcheverry[4,5,6,*], Mohammed-Amin Madoui[1,*], Julie Guy[1], Antoine Martin[4,5,6], Tony Heitkam[4], Karine Labadie[1], Jérémie Le Pen[4,5,6,†], Patrick Wincker[1,2,3], Vincent Colot[4,5,6,‡] and Jean-Marc Aury[1,‡]

[1]Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, 2 rue Gaston Crémieux, BP5706, 91057 Evry, France
[2]Centre National de Recherche Scientifique (CNRS), UMR 8030, CP5706, Evry, France
[3]Université d'Evry, UMR 8030, CP5706, Evry, France
[4]Institut de Biologie de l'Ecole Normale Supérieure, F-75230 Paris Cedex 05, France
[5]Centre National de la Recherche Scientifique (CNRS), UMR 8197, F-75230 Paris Cedex 05, France
[6]Institut national de la santé et de la recherche médicale (INSERM) U1024, F-75230 Paris Cedex 05, France

[*]These authors contributed equally
[†]Current address: Gurdon Institute and Department of Biochemistry, University of Cambridge, The Henry Wellcome Building of Cancer and Developmental Biology, Tennis Court Rd, Cambridge CB2 1QN, UK.
[‡]Corresponding authors

Abstract

The recent explosion in genome sequencing has confirmed that transposable elements (TEs) and their relics constitute the major fraction of genomic DNA in most known eukaryotes, including mammals and many plants (Lopez-Flores and Garrido-Ramos, 2010). Furthermore, resequencing experiments have provided compelling evidence that while TE mobilization is strongly repressed in most instances, this repression is not absolute and can be allievated in both pathological and normal situations (Huang et al., 2012).In particular, mobilization of specific TEs has been linked to diseases, including cancer, as well as to the normal development of the brain in humans (Lee et al., 2012; Hancks and Kazazian, 2012; Muotri et al., 2010). Until now, the identification of TE mobilization events using next generation sequencing (NGS) technologies has often relied on the use of general-purpose structural variant (SV) detection algorithms, which produce very large numbers of putative breakpoints and leave the calling step to the biologist. Here, we present a computational method for accurately detecting both the identity and destination of newly mobilized TEs in re-sequenced genomes. This method, called TE-Tracker, accepts as input the widely used BAM format, and does not rely on prior TE annotation but solely on the extraction of specific association patterns within paired-end alignments. We show that noise levels are orders of magnitude lower when using TE-Tracker instead of generic SV detection algorithms. We test TE-Tracker on a synthetic human genome and use it to provide a comprehensive view of transposition events induced by loss of DNA methylation in Arabidopsis.

Introduction

TEs and their abundant relics have been found in nearly all organisms and have been classified into several families based on sequence features and transposition mechanisms (Lopez-Flores and Garrido-Ramos, 2010). So-called DNA-transposons generally exhibit cut-and-paste transposition, while retrotransposons use an ARN intermediate and thus transpose using a copy-and-paste mechanism. Retro-elements are further divided into two subclasses, depending on the presence or absence of Long Terminal Repeats (LTR). The biological role of TEs has been the subject of great controversy, and although they had been assimilated to "selfish" or "junk" DNA for some time (Doolittle and Sapienza, 1980), they are now recognized as important factors in the evolution of genome structure and function (Hurst and Werren, 2001; Rebollo et al., 2012).

Indeed, it has been estimated that mobilization of LTR-retrotransposons is responsible for up to one tenth of spontaneous germline mutations (Maksakova et al., 2006) in laboratory mice. Similarly, mobilization of the human LINE1 (L1) non-LTR retrotransposon was found to account for 19% of the structural variation between individual genomes (Kidd et al., 2010), and has been linked to over a hundred human diseases (Hancks and Kazazian, 2012). TEs can also be mobilized in somatic cells and although this has been associated with pathological situations, such as cancer in humans (Lee et al., 2012; Iskow et al., 2010), there is now growing evidence that transposition does occur during normal development and cellular differentiation, notably in the brain (Thomas et al., 2012). Despite these observations, TE mobilization

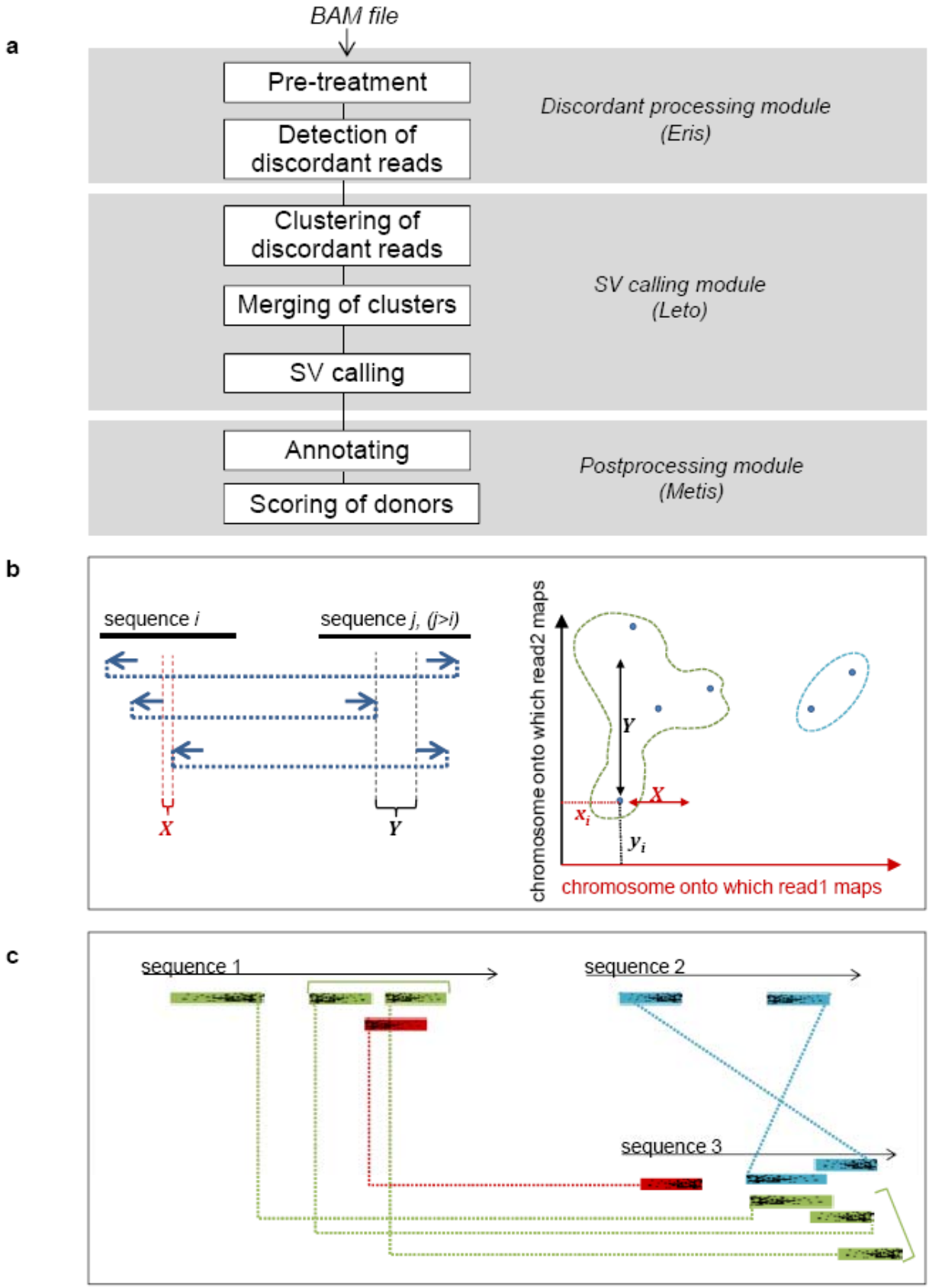


Figure 1 : TE-Tracker overview and main algorithms.

a. Main steps of the TE-Tracker pipeline. **b**. Principle of the single linkage clustering. Left: genome view of three anomalously spaced read pairs (blue arrows and lines). Discordant reads are ordered by increasing order on position chromosome and onset. Read pairs that are distant by no more than X on their ordered side and by no more than Y on their unordered one are clustered together. Right: graphical representation of the Single-Linkage Clustering (SLC) algorithm implemented in TE-Tracker. **c**. Cluster merging and calling. The bracketed green clusters overlap on sequence 3 and are close to each other on sequence 1 and will therefore be merged. However, if the red cluster is also present all green clusters will be rejected. Considering the set of blue and green clusters independently leads to multiple donors or acceptors.

remains a rare event, notably because of mechanisms, such as DNA methylation, that keep TEs in a transcriptionnally inactive form most of the time (Slotkin and Martienssen, 2007; Zamudio and Bourc'his, 2010). Additional control mechanisms must also exist, as transcriptional reactivation of TEs does not invariably lead to their mobilization. Yet, the extent to which TE mobilization is controlled post-transcriptionally is largely unknown.

With the advent of NGS technologies, it is now conceivable to resequence whole genomes in order to characterize TE mobilization systematically. However, this task is complicated by the inherently repeated, most often defective nature of TE sequences and by their frequent clustering. Perhaps as a result, few bioinformatic tools have been developed specifically for the detection of newly mobilized TEs in resequenced genomes and these tools often rely on prior annotation (Nellaker et al., 2012), or are restricted to the analysis of specific TE sequences (Lerat, 2010). Nonetheless, several studies have attempted to identify new TE insertions de novo using structural variant (SV) detection tools. Four broad types of such methods have been described over the past few years. They are based on the analysis of either (i) depth of coverage, (ii) split reads, or (iii) discordant paired reads, or else (iv) on de novo assembly (Alkan et al., 2011). Type (i) methods give a quantitative measure of the number of extra TE copies but do not provide information about their location. Type (ii) and (iii) methods identify one-sided events in the form of clusters of anomalously mapped reads, but they do not combine these one-sided events to produce bona fide TE insertions. In addition, the heavy computational burden of type (iv) methods as well as their poor performance with repeated sequences preclude their use for large-scale detection of new TE insertions (Li et al., 2011).
More recently, several programs have attempted to adopt an integrative approach by combining results from several methods (Jiang et al., 2012; Rausch et al., 2012; Sindi et al., 2012), but their signal-to-noise ratio is still typically low when considering specific types of structural variation (See Methods).Finally, none of these general-purpose tools can identify the donor TE or provide the internal sequence of transposed copies.

Here, we present TE-tracker, a new method dedicated to the systematic and robust identification of newly mobilized TEs based on whole-genome resequencing (Figure 1.a). Since TE-Tracker performs pairwise association of read mate groups that are anomalously spaced or oriented, discordant reads can be anchored on both sides of the recipient sequence until their mates overlap over the donor. Not only does this allow to detect new TE insertions de novo and without prior annotation, it also makes it possible to determine both the length and orientation of the newly inserted sequences, as well as the identity of the corresponding donor TEs.
We have used TE-Tracker to analyse the resequenced genome of Arabidopsis in the progeny of a mutant plant compromised for DNA methylation. Our findings indicate that despite extensive transcriptional reactivation of numerous TEs in the parental plant, only few of

these are mobilized, thus providing strong evidence of the importance of post-transcriptional mechanisms in preventing TE mobilization.

Results

The TE-Tracker pipeline

As with type (iii) SV detection methods (see Introduction), TE-Tracker starts with the analysis of the orientation and spacing of clustered (Figure 1.b) discordant paired reads to detect new insertions (Medvedev et al., 2009). Unlike these methods though, TE-Tracker aims to cover the entire length of the new TE insertion by searching at the new location for left and right-side clusters that overlap with each other (Figure 1.c). As a result, TE-Tracker can detect insertions and determine the donor copy for TEs that are up to 2L in length, where L is the mean size of DNA fragments used for sequencing. For example, in order to fully characterize the transposition landscape of Alu elements in the human genome ($\approx$ 300 bp), TE-Tracker would require a short fragment paired-end library of 150 bp mean length, whereas longer, recircularized fragments (such as mate-pairs) would have to be used for larger elements.

When more than two clusters satisfy the overlap condition at the same location, TE-Tracker attempts to consider them together, within some limits (Figure 1.c and see Methods), which allows orienting insertion events with respect to the donor TE sequence. The clustering algorithm has been optimized for speed and TE-Tracker was found to run between one and four order of magnitudes faster than three common pipelines (Zeitouni et al., 2010; Chen et al., 2009; Quinlan et al., 2010) that accept long fragment sequencing data as input (Supplementary Figure 1).

Typically, TE families contain mostly defective copies that are unable to be mobilized because of truncations or other mutations in their coding or regulatory sequences. Nonetheless, their potentially mobile copies are difficult to predict on the basis of sequence integrity alone, and there are no programs to date that attempt to identify those that transpose among potential candidates. Given that TE families may contain several mobile copies that differ from each other by a few sequence polymorphisms, we have included in TE-Tracker a donor-scoring feature, which selects within clusters those reads that contain discriminating polymorphisms (See Methods).

All of the features described above are implemented into a pipeline comprising three main programs (Figure 1.a), which allow pre- as well as post-processing steps such as filtering of the input alignment or annotation of the output result to be also performed (see Methods).

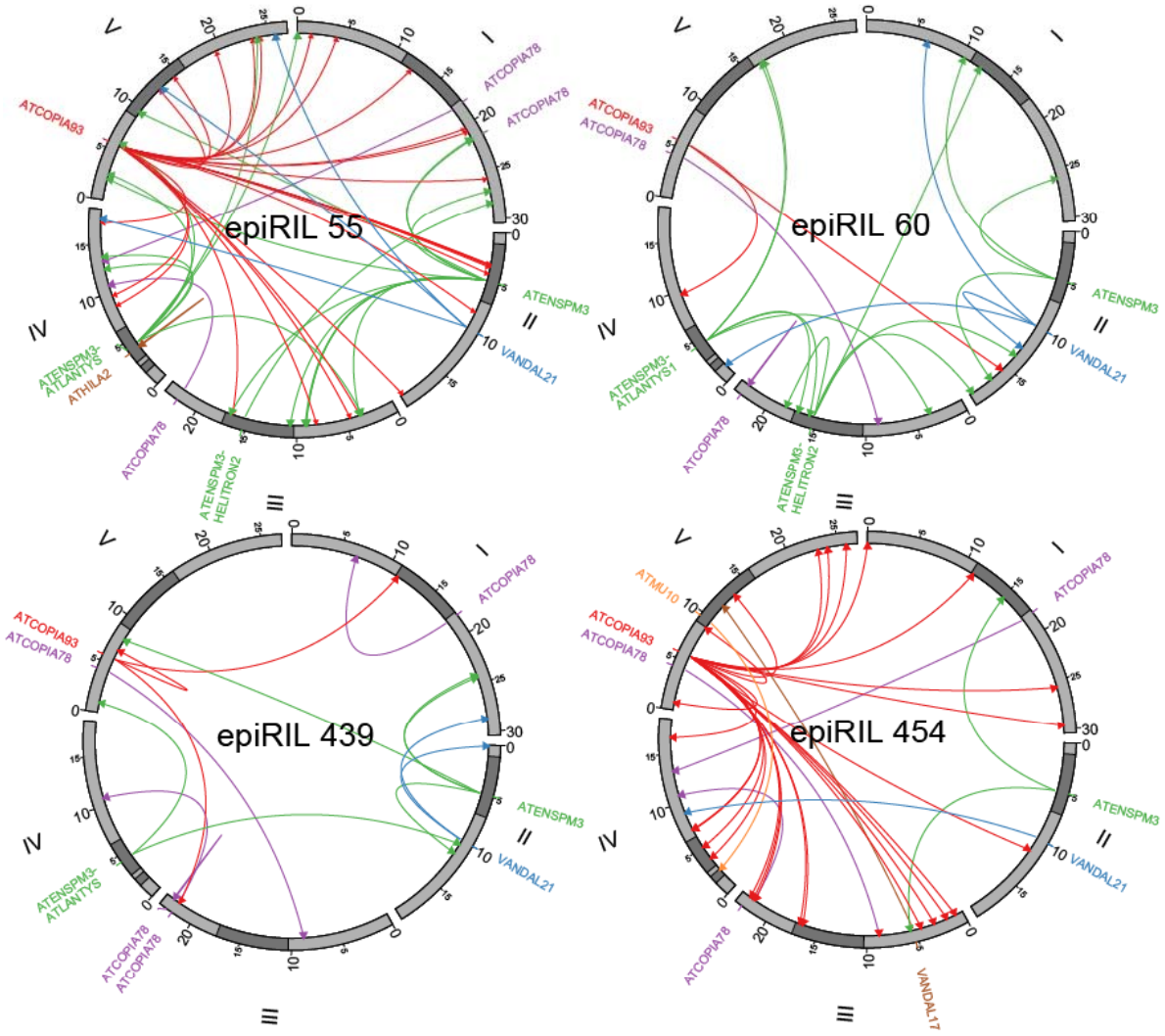


Figure 2 : Circos representation of new TE insertion events detected in four epiRILs . Exterior circle represents the five chromosomes of Arabidopsis with pericentromeric regions and heterochromatic knob on chromosome 4 in dark grey. Arrows link donor TEs with the new insertion sites. Only events mapped with no ambiguity (no multiple acceptor sites and no similarity with events detected in wt) are represented.

Assessment of TE-Tracker performance on simulated Arabidopsis and human chromosomes

We first tested TE-Tracker by simulating a mate-pair sequencing strategy for the complete Arabidopsis reference genome using an in-house paired read simulator (see Methods). Importantly, although the simulator took into account realistic rates and types of sequencing errors as well as of chimeric reads, this did not lead to any transposition event being detected in this un-rearranged version of the Arabidopsis genome, which suggests a null type-II error rate. When 80 random insertions involving two copies of the retro-element COPIA93 were simulated between a pair of Arabidopsis chromosomes and sequenced using the same simulator, 75 were detected, which gives a sensitivity of 95% (Supplementary Table 1). The remaining five insertions were located either in tandem or other nearby ($<$ 50 kb) repeat elements, or in sequence gaps.

We also performed similar tests on two human chromosomes, for which we simulated the mobilization of two, 6 kb-long L1-type elements that differ by 124 nucleotides and that have been described as active in the human brain (Baillie et al., 2011). Of the 20 random insertions generated (with random donor), 17 were detected (Supplementary Table 2), and the three remaining ones were inserted in sequence gaps. Furthermore, only one L1 donor was misattributed in this set, despite there being about one hundred distinct, potentially mobile full length L1 on these two chromosomes.

Application of TE-tracker to the exploration of the transposition landscape in Arabidopsis

We then applied TE-Tracker to the identification of novel TE insertions in a set of four Arabidopsis epigenetic recombinant inbred lines (Johannes et al., 2009) derived from a cross between a wild type (*wt*) plant and a plant mutated for the *DECREASE IN DNA METHYLATION 1* (*DDM1*) gene. *ddm1* mutant plants in which DNA methylation as well as transcriptional silencing of TEs is severely compromised (Vongs et al., 1993), inducing their transcriptional reactivation and thus potentially leading to their re-mobilization (Lippman et al., 2004; Singer et al., 2001; Miura et al., 2001; Tsukahara et al., 2009). The four epiRILs together with two *wt* lines were sequenced using Illumina mate-pair libraries (around 5.5 kb mean length), which should therefore enable the detection of new insertions for almost all of the TEs that are potentially active in the genome, as over 90% of all full-length annotated Arabidopsis TEs are less than 11 kb long (Buisine et al., 2008; Ahmed et al., 2011). Effective sequencing coverage (after alignment) ranged from 17X to 37X (Supplementary Table 3). Results are illustrated in Figure 2 and summarized in Supplementary Tables 4-9. TE-Tracker reported a total of 220 distinct insertions among the four epiRILs, of which 162 matched annotated TE sequences (Supplementary Tables 4-9, column "Donor annotation"). The vast

majority (136) of these insertions were not detected in the two *wt* parental lines, as expected if most transposition events occurred in the *ddm1* parental line or during the propagation of the epiRILs. To validate these results, a random set of 69 potentially novel insertions as well as three insertions also shared with the two *wt* parents were tested by PCR. In all 72 cases, the presence of the insertion could be confirmed (Supplementary Table 10), which provides further evidence of the high specificity of TE-Tracker. Furthermore, sequencing of 26 PCR products corresponding to new insertions was used to evaluate the performance of TE-tracker in identifying donor TEs. In all but one case, the donor scoring module was able to identify the correct TE donor sequence. Also, sequencing of both ends of 12 new insertions confirmed the presence of a target site duplication in each case, as expected for true transposition events (Supplementary Figure 2, Supplementary Table 10). Among these, we validated several insertions involving composite sequences that were not previously annotated as full-length TE units, namely fragments of ATENSPM3 and ATLANTYS1 and fragments of ATENSPM3 and HELITRON2 (Supplementary Figure 3).

Of the 136 distinct novel TE insertions identified, six were shared among the four epiRILs (Supplementary Tables 4-9). This proportion is lower than that expected (exact one-sided binomial test, p-value= 2.137e-11) if all insertions had occurred in the *ddm1* parental line used to establish the epiRIL population (Johannes et al., 2009)and suggests therefore that TE mobilization could also occur in subsequent generations in at least some cases. Furthermore, transposition in the epiRILs concerns only a small number of TE families (Supplementary Tables 4-9), which is consistent with a previous report of TE mobilization in *ddm1* (Tsukahara et al., 2009). These findings, together with the fact that most TE sequences are transcriptionally reactivated in *ddm1* (Lippman et al., 2004), suggest therefore an important role of postranscriptional mechanisms in preventing TE mobilization in Arabidopsis. Our analysis indicates in addition that mobilization, when it occurs, often concerns only one of the potentially mobile TE members of a given family. For instance, despite there being two highly similar copies of the LTR retroelement family COPIA93, only one appears to be mobile in the genome of the Columbia accession, as was previously reported (Mirouze et al., 2009). However, there are exceptions to this rule, as exemplified by the fact that several members of the LTR retroelement family COPIA78, which is closely related to COPIA93, have been mobilized. As many of these new COPIA78 insertions are shared among at least two of the epiRILs, transposition is likely to have taken place in most cases in the parental *ddm1* line or in the F1, which contradicts a previous claim that COPIA78 cannot transpose in this mutant background (Ito et al., 2011). Furthermore, in the case of COPIA78 insertions, the donor scoring feature often yielded two potential donors with similar high scores. Detailed analysis of one such COPIA78 insertion showed the existence of distinct sequential blocks corresponding to either donor. This is in agreement with previous reports indicating that similarly to what is seen in viruses (Goodrich and Duesberg, 1990), two RNA intermediates

matching distinct LTR-TE family members could be encapsidated together. As a result, TE sequences could undergo recombination by template switching during cDNA synthesis (Jordan and McDonald, 1998), thus leading to the insertion of a composite sequence presenting block-wise similarity to both of the parent elements (Supplementary Figure 4). However, we cannot rule out in the case of the COPIA78 element that the new insertion corresponds in fact to to the mobilization of still non-assembled COPIA78 donor elements. Resequencing also revealed that the DNA transposon VANDAL21 inserts preferentially close to the transcription start site of genes and in the same orientation as these (9 /12 instances and 8 /9 instances, respectively), likely indicating a role for transcription initiation of the target locus in the insertion of this particular TE family. Finally, we note that overall, new TE insertions are spread across the entire genome (Figure 2), which contrasts with the pericentromeric localization of most TE sequences present in the Arabidopsis genome. This provide therefore strong evidence that purifying selection plays an important role in eliminating insertions that occur within the gene-rich regions of Arabidopsis chromosomes.

Discussion

We have presented here a program, TE-Tracker, that accurately detects both the source and destination of novel transposition events in resequenced genomes. Since TE-Tracker only relies on the detection and clustering of discordant paired reads and not on TE annotation, it is generic and enables to track any mobilized TE, irrespective of its identity. With appropriate parameterization, TE-Tracker is also able to detect all potential donor sequences for a given insertion, and by discriminating reads that map better to a particular donor, it can attribute the correct one among them if they differ by at least one nucleotide. Furthermore, TE-Tracker is faster and produces significantly less noise than common SV detection programs (see Methods), therefore allowing the researcher to focus exclusively and exhaustively on TE mobilization events in a re-sequenced genome.

Clustering algorithms such as TE-Tracker are generally memory-intensive when run over a large number of points; in particular, it is known that the optimal performance of the single-linkage algorithm used in TE-Tracker is $\mathcal{O}(n^2)$, where n the number of points (Sibson, 1973). In an omics context, this will result in increased computational load proportional to the number of discordant reads, either because of larger genomes or higher sequencing depth. For TE-Tracker, we choose to favor speed at the expense of memory use. For performance optimization, we developed a seed-type heuristic that reduces the amount of pairs in memory to a fraction of the total number (see Methods). Furthermore, at any given time, read mate mappings that belong to different pairs of chromosomes and are mapped in a specific orientation are considered independently and sequentially, which implies that performance of TE-Tracker will not depend on overall genome size/sequencing depth but on the average

sequence size/sequencing depth for individual chromosomes. Hence, discordant reads are subdivided in up to $4 \times \left(\!\!\binom{k}{2}\!\!\right)$ chunks where k is the number of chromosomes. This is why performance evaluation for a pair of two human chromosomes can be considered to reflect performance over the human genome as a whole. For an organism with very large chromosomes, it would be possible to reduce concerns for memory load by adding a further optimization step, for example one that would perform two-dimensional Fast Fournier Transform for spatial density estimation of read pair mappings. This would further scale down the size of the individual proximity matrices considered by the program, however during our tests, we never encountered a case where such a procedure was necessary.

Acceptor sites are generally defined by TE-Tracker within a few kb, depending on the size of the mobilized TE but in no case can be less than a few dozens of bp, because of split reads at the precise breakpoint, which are often left unmapped by a general-purpose aligner such as BWA. These "blind spots" can nevertheless be overcome by performing an additional split-aligning step within the insertion interval defined by TE-Tracker (e.g. using BWA's Smith-Waterman implementation) and searching for appropriate motifs in read sequences. This search algorithm was not included in the standard distribution of TE-Tracker due to its low (combinatorial) performance, but could easily be added so as to match the base-pair precision provided by integrative structural variant detection tools such as PRISM.

To conclude, we show that TE-tracker, when combined with a relevant sequencing protocol, provides an efficient and reliable method for detecting new transposition events in re-sequenced genomes. Indeed TE-Tracker stands in marked contrast to general-purpose structural variant detection tools, which typically generated lists of thousands of clusters and/or breakpoints, which the biologist needs to sieve through. Furthermore, none of these tools provide information on donor TE sequences. We believe that the development of biologically constrained programs (BCPs) such as TE-Tracker, which aim at providing robust answers to specific questions, will greatly help overcome the noise issues that typically plague the analysis of short read sequencing data.

Methods

Sequencing

DNA was extracted from seedlings grown under long-day conditions, using DNeasy Qiagen kits. About 10 microgram of genomic DNA were sonicated separately to a 4-6 Kb size range using the E210 covaris instrument (Covaris, Inc., USA). Libraries were prepared following Illumina's protocol (Illumina Mate Pair library kit). Briefly, fragments were end-repaired and biotin labeled. A size selection of fragments with length of interest (around 5Kb) was performed. DNA were then circularized and linear, non-circularized DNA were eliminated by digestion. Circularized DNA were fragmented to 300-700-bp size range using covaris E210. Biotinylated DNA were purified, end-repaired, then 3'-adenylated, and Illumina adapters were added. DNA fragments were PCR-amplified using Illumina adapter-specific primers. Finally, the PCR amplified libraries (350-650 bp) were size-selected. Libraries were then quantified using a Qubit Fluorometer (Life technologies) and libraries profiles were evaluated using an Agilent 2100 bioanalyzer (Agilent Technologies, USA). Each library was sequenced using 100 base-length read chemistry in a paired-end flow cell on the Illumina GAIIx (2 lanes) or HiSeq2000 (1 lane) (Illumina, USA).

Read Mapping

Reads were then mapped with BWA v. 0.6.1 (Li and Durbin, 2009), using the parameters `-R 10000 -l 35 -O 11`, onto the TAIR10 reference sequence (Lamesch et al., 2012). Reads hanging over chromosome ends were removed using `picard CleanSam`, duplicate pairs were removed using `picard MarkDuplicates` (http://picard.sourceforge.net). Supplementary Table 3 displays mean vertical coverage, as well as the portion of the genome covered by less than 10 reads, in percent, for each of the lines.

TEtracker algorithm

The `TE-Tracker` algorithm comprises three Perl modules. The three modules are named `eris`, `leto` and `metis` and deal respectively with preprocessing, clustering of discordant pairs and postprocessing (annotation and scoring of results). The core of the pipeline lies in the `leto` module, which is itself but a wrapper around the `slclust` program, written in C++ using the `boost` library (Siek et al., 2000) for better performance.
The pipeline's design was user-driven and centered around modularity. It is therefore possible to replace the modules with custom ones, provided the command-line argument set is consistent with other elements of the pipeline. Conversely, and under the same conditions, it is possible to use another clustering program in lieu of `slclust`, all it takes is to change one single path in the general configuration file.

TE-Tracker will use that file (`SV.conf`) to determine all parameters to be used during execution, as well as the number of times that each step needs to be performed.
In the following sections we will describe the workings of the pipeline's most important steps.

Preprocessing

Each alignment file was preprocessed using the `eris` module. Parameters used were `-treat_bam input:0-1`, which removed all mappings whose best match contained more than 1 mismatch.
The use of this parameter was driven by our concern for noise reduction and was justified by the average high coverage depth of all our lines. Testing showed that what would later be verified as being noise was significantly reduced using this procedure.

Discordant pair detection

Let a mapped read $r(c, o, l)$ be defined by its chromosome c, its orientation o ($+$ or $-$) and its localisation l on c, and a read pair $p(r_a, r_b)$, the couple of two reads with $l_a < l_b$ and $d = l_b - l_a$ the distance between two reads of the same pair if $c_a = c_b$. Let $P_i = p_{i_1}, .., p_{i_n}$ the set of the n different mapping possibilities for the pair i. From the P_i we calculate the median M , the median absolute deviation (MAD) and upper and lower limits of the distances from the d_i , respectively d_{inf} , d_{sup} with $d_{inf} = M - 3.MAD$ and $d_{sup} = M + 3.MAD$. For a large insert library, a pair mapping $p_i(r_a, r_b) \in Pi$ is a proper mapping if $c_a = c_b$, $(o_a, o_b) = (-, +)$ and $d_{inf} < l_b - l_a < d_{sup}$. If such mapping do not exist in P_i at least once, the pair is considered as discordant and its mapping possibilities are classified following their mapping signatures. For each mapping possibility of one read pair $p_i \in P_i$, we have:

$$\text{if } c_a = c_b \begin{cases} \text{if } (o_a, o_b) = (-, +) & : \begin{cases} p_i \in Del \text{ if } d_i > d_{sup} \\ p_i \in Ins \text{ if } d_i < d_{inf} \end{cases} \\ \text{if } (o_a, o_b) = (+, -) & : p_i \in Dup \\ \text{if } (o_a, o_b) \in \{(-, -); (+, +)\} & : p_i \in Inv \end{cases}$$

$$\text{if } c_a \neq c_b : p_i \in Trans$$

Del, Ins, Dup, Inv, $Trans$, being sets of discordant read pairs having respectively a deletion, insertion, duplication, inversion and translocation signatures. For an organism with m chromosomes, we split the sets Del, Ins, Dup and Inv into n subsets (i.e. Del_1, ... ,Del_m, Ins_1, ... ,Ins_m, ...) and the $Trans$ set into $4 \times A_n^2$ subsets (i.e. $Trans_{(1,+);(2,+)}$, $Trans_{(1,+);(2,-)}$, $Trans_{(1,-);(2,+)}$, $Trans_{(1,-);(2,-)}$, ..., $Trans_{(m-1,-);(m,-)}$).

Single-linkage clustering and merging

We also calculated upper and lower limits of the depth of coverage c_{inf} , c_{sup} using M and MAD. Pairs whose reads both map in a genomic region with a very high coverage depth (typically $> 1000\times$ and containing repeated elements and low complexity sequences) are discarded from the discordant pair set. Discordant reads are sorted by l_a and clustered using single linkage clustering. For each subset, we built $G = (V, E)$, an undirected graph where nodes V are pairs $(p_i, p_j) \in E$ if:

$$|l_{i_a} - l_{j_a}| < \frac{4s}{c_{inf}} \tag{1}$$

and

$$|l_{i_b} - l_{j_b}| < d_{sup} - d_{inf} \tag{2}$$

with l_{i_a} and l_{j_a} respectively the position of the reads a of the pair i and the reads a of the pair j, and s the read size. Two pairs are linked if the distance between the two reads r_{j_a}, r_{j_a} is smaller than expected relative to the coverage depth variation (1) and the fragment size variation (2).
The single linkage processus starts from a single read pair, the seed. It tries to link it to the next available pairs; when linking is not possible anymore the last linked pair is used as seed, which is helpful in terms of CPU and memory usage. Nearby clusters with identical signatures are merged, this cluster extension allows no penalization due to low covered regions that may interrupt the linking process. After merging, the clusters are filtered by their size, rejecting those larger than d_{sup}.

Calling

Intra and interchromosomal translocations are called by searching overlapping clusters of different orientation(Figure 1.c) at the donor location that allows detection of translocation up to $2 \times M + 6 \times MAD$ bp. A deletion pattern cluster overlapping a duplication pattern cluster is needed to call an intrachromosomal translocation in the sense donor orientation. If a deletion pattern cluster does not overlap any duplication pattern cluster and if its size is over d_{inf} , this cluster supports a deletion. An inversion pattern cluster overlapping a inversion pattern cluster of the opposite orientation is needed to call an antisense intrachromosomal translocation.

Donor scoring

When several donors are available for one putative insertion site, the `leto` module outputs them all and groups them by an unique "acceptor ID". It is possible to assign a donor score to each using the `metis` module. Discordant reads anchoring at the acceptor site on one side,

and at every successive potential donor on the other, are extracted from the discordant BAM file generated by the `eris` module with parameters `-discordant=sorted`. For each donor, the program goes through the multiple mappings of each read, and looks for mates on the donor-side that map significantly better on a particular donor than on all the others. The count of these mates constitutes the score. The idea behind this is that multiple copies of a single TE in the genome, although they share most of their sequence, diverge by at least some SNPs. If this is true, then for each such insertion there will be discordant reads spanning these discriminating positions. With lenient mapping parameters such as ours, they will map on other copies of the original donor element, however their score will be lower there because of the SNPs they contain. Thus, a common score of 0 means that there is complete sequence identity between all possible donors.

TE-Tracker Parameterization

Most of TE-Tracker's parameters are self-explanatory and concern only the pipelining features of the program. The only two parameters that significantly influence output are X and Y, involved in the central Single-Linkage clustering algorithm.

There are two ways to deal with these parameters. By default, TE-Tracker produces a point estimate of the optimal parameters through statistical inference; however it is also possible to specify them manually.

Optimal parameters are evaluated as follows: since X is dependent on the coverage distribution, and Y on the insert size distribution, they have to be estimated in quite a different way. Estimation of Y is straightforward: it corresponds to a sufficiently large quantile of the empirical insert size distribution.

Similar distribution-based estimation for X lead to systematic underestimation. Thus another method was implemented, based on Let (C_n) be the random variable that gives coverage depth for each position n in the genome.

We model (C_n) as a two-state Markov chain, where one state corresponds to low-coverage (vertical coverage 0 or 1) regions and the second to normally covered ones. Under this model, the probability of a gap of size k is given by the probability of a k-sized null path: $p_{l_k} = \pi_{h \to l} \prod_{i=1}^{k} p_{ll}$ where $\pi_{h \to l}$ is the transition probability from the normal state to the low-coverage state and p_{ll} is the probability of staying in the low-coverage state. Both of these values can be empirically estimated by traversing a "pileup" file built from the alignment file.

We then calculate $p_l(k)$ for gaps of increasing size until a k_0 is encountered with probability $p_l(k_0)$ below a certain threshold. We then assign $X = k_0$, since k_0 will correspond to the maximum gap size that can be expected to be found under this model with a probability below the specified threshold. Typically, we found that satisfactory values of X were obtained with a very stringent treshold, for example when $p_l(k)$ was so small that it became indiscernible from 0 under double floating point precision.

Running TE-Tracker

In order to maximize sensitivity, we ran the `leto` module on a regular grid of increasing X and Y parameters (see below) and pooled the results by traversing the grid. A step size of 50 was chosen for X, while a step size of 300 was chosen for Y. X ranged from 50 to 2000, Y from 300 to 6000, which amounts to a total of 800 runs of the `leto` module. For this, we needed only one run of the prefiltering `eris` module, which allowed us to complete each grid in about a day's time.

The high number of points on the grid allowed us to draw a three-dimensional surface onto which the number of events outputted for each run was shown (Supplementary Figure 5). We noticed a clear maximum in all of the lines, and in every occasion, the point estimation of the optimal (X, Y) parameters was found to be quite close to it.

For our traversing program, we first looked for the said maximum, then we built a dictionary of donors from the insertions found in it. Then, we went through every run's output file and added data to the dictionary: we performed the cartesian product of the dictionary and the output file and added events that did not overlap either on donor or acceptor site. This allows to build a comprehensive landscape of all insertions that appear at least once on the grid. When an event was found in several points of the grid, the one supported by the most reads was kept, which ensures that the optimal parameters for each insertion were used for each line of the final output file. Looking for the maximum first allows to reduce the computational impact of such a combinatorial algorithm.

Performance evaluation

Comparison with other software

In order to assess its performance, we compared the events produced by TE-Tracker to those outputted by other software. However, this procedure proved difficult because all existing programs display only one edge of a mobilized transposon, with no or wrongful event size information. In addition, several others such as SHORE(Ossowski et al., 2008) do not assemble *INV*, *INS*, *DEL* and *DUP* events into putative two-sided events, which makes it difficult to quantify how much of an event the program has really detected.

Additionally, few existing programs were documented to handle large insert size libraries such as those produced by our mate-pair sequencing approach. We therefore performed our tests on two sequencing runs of the same epi-RIL line, the first one using a mate-pair preparation (median insert size 5kb), the second a more usual paired-end one with median insert size of 500bp.

Events involving COPIA93 and ATENSMP3, that had been PCR-validated and for which the donor has been clearly identified by biological means, were used in this benchmark. Seventeen COPIA93-events and 6 ATENSPM3-events in total had been validated on this particular line. Results are presented in Supplementary Tables 11 and 12. For programs other than TE-Tracker, a positive find was called whenever the program's output contained a cluster that overlapped both event's extremities.

Paired-end tests

For the paired-end run, we tested the popular SV-detection programs BreakDancer Max(Chen et al., 2009), SHORE(Ossowski et al., 2008), PRISM(Jiang et al., 2012) and SVDetect(Zeitouni et al., 2010). However, SVDetect was dismissed further on due to its long run-time in a single-core setting, as well as SHORE due to its ambiguous results and lack of support.

Mate-pair tests

Only BreakDancer(Chen et al., 2009), Hydra(Quinlan et al., 2010) and SVDetect (Zeitouni et al., 2010) claimed to be able to detect structural events based on large-fragment sequencing. For all these programs, we also assessed computational performance, for which we compared both CPU time and maximum memory use. These results are presented in Supplementary Figure 1.

Simulation

As mentioned in the main text, we also performed tests on simulated mate-pair sequencing runs to assess theoretical sensitivity and specificity. For this, we used the custom-built `Sphinx` script that generates artificial transpositions in a reference sequence. It is able to generate sense and antisense transpositions, as well as cut-paste, copy-paste and deletion events.

This tampered reference sequence is then supplied as an argument to the `SimSeqG.pl` script, which simulates mate-pair sequencing according to the features found in an existing alignment. It first scans an existing BAM file for aligned pairs and considers their fragment size and quality.

Using this data, it then builds a constant-frequency bin histogram for both the quality distribution of each base and the fragment size distribution for paired-end oriented mates and mate-pair oriented ones.

`SimSeqG.pl` has been designed according to the BCP principle and aims to mirror the sequencing process as closely as possible:

- A fragment size is sampled from the empirical distribution;
- a location is randomly selected in the genome and a sequence corresponding to the sampled fragment length is extracted starting at this position;
- the fragment is circularized and a random splice length is chosen from the pair-end fragment length distribution;
- a splice start is randomly chosen around the circularization point and the sequence is extracted from the circularized fragment;
- both ends of the subfragment are extracted and sequenced: for each base, the program will draw a quality corresponding to its position from the empirical quality distribution. Then, it will produce a sequencing error at that position with probability given by the base quality.
- once in a while, at a rate defined from the BAM learning set, a configuration leading to the production of a parasitic paired-end fragment is produced and the result is sequenced in a similar way;
- reads and qualities are then written in FASTQ format.

PCR Validation

A list of primers used for the validation of detected insertions is provided in Supplementary Table 13.

Set operations

Every set operation on shared insertions were performed using overlapping acceptor regions as the main criterion. The program used for intersecting regions was `bedops`(Neph et al., 2012).

Data Access

TE-Tracker is free for academic noncommercial use and can be downloaded from http://www.genoscope.cns.fr/
All additional programs as well as sequenced and simulated FASTQ files are also available at

this address.

Acknowledgments

This work was supported by the French National Agency for Research (ANR; Project Acronym EPIMOBILE).

Disclosure Declaration

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Figure Legends

Figure 1 TE-Tracker overview and main algorithms – **a.** Main steps of the TE-Tracker pipeline. **b.** Principle of the single linkage clustering. Left: genome view of three anomalously spaced read pairs (blue arrows and lines). Discordant reads are ordered by increasing order on position chromosome and offset. Read pairs that are distant by no more than X on their ordered side and by no more than Y on their unordered one are clustered together. Right: graphical representation of the Single-Linkage Clustering (SLC) algorithm implemented in TE-Tracker. **c.** Cluster merging and calling. The bracketed green clusters overlap on sequence 3 and are close to each other on sequence 1 and will therefore be merged. However, if the red cluster is also present all green clusters will be rejected. Considering the set of blue and green clusters independently leads to multiple donors or acceptors.

Figure 2 Circos representation of new TE insertion events detected in four epiRILs – Exterior circle represents the five chromosomes of Arabidopsis with pericentromeric regions and heterochromatic knob on chromosome 4 in dark grey. Arrows link donor TEs with the new insertion sites. Only events mapped with no ambiguity (no multiple acceptor sites and no similarity with events detected in wt) are represented.

References

Ahmed I, Sarazin A, Bowler C, Colot V, and Quesneville H. 2011. Genome-wide evidence for local dna methylation spreading from small rna-targeted sequences in arabidopsis. *Nucleic Acids Res* **39**: 6919–31. Ahmed, Ikhlak Sarazin, Alexis Bowler, Chris Colot, Vincent Quesneville, Hadi England Nucleic Acids Res. 2011 Sep 1;39(16):6919-31. doi: 10.1093/nar/gkr324. Epub 2011 May 17.

Alkan C, Coe BP, and Eichler EE. 2011. Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**: 363–76. Alkan, Can Coe, Bradley P Eichler, Evan E Canadian Institutes of Health Research/Canada Howard Hughes Medical Institute/ England Nat Rev Genet. 2011 May;12(5):363-76. doi: 10.1038/nrg2958. Epub 2011 Mar 1.

Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al.. 2011. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**: 534–7. Baillie, J Kenneth Barnett, Mark W Upton, Kyle R Gerhardt, Daniel J Richmond, Todd A De Sapio, Fioravante Brennan, Paul M Rizzu, Patrizia Smith, Sarah Fell, Mark Talbot, Richard T Gustincich, Stefano Freeman, Thomas C Mattick, John S Hume, David A Heutink, Peter Carninci, Piero Jeddeloh, Jeffrey A Faulkner, Geoffrey J 090385/Wellcome Trust/United Kingdom 090385/Z/09/Z/Wellcome Trust/United Kingdom BB/H005935/1/Biotechnology and Biological Sciences Research Council/United Kingdom England Nature. 2011 Oct 30;479(7374):534-7. doi: 10.1038/nature10531.

Buisine N, Quesneville H, and Colot V. 2008. Improved detection and annotation of transposable elements in sequenced genomes using multiple reference sequence sets. *Genomics* **91**: 467–75. Buisine, Nicolas Quesneville, Hadi Colot, Vincent Genomics. 2008 May;91(5):467-75. doi: 10.1016/j.ygeno.2008.01.005. Epub 2008 Mar 14.

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, et al.. 2009. Breakdancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* **6**: 677–81. Chen, Ken Wallis, John W McLellan, Michael D Larson, David E Kalicki, Joelle M Pohl, Craig S McGrath, Sean D Wendl, Michael C Zhang, Qunyuan Locke, Devin P Shi, Xiaoqi Fulton, Robert S Ley, Timothy J Wilson, Richard K Ding, Li Mardis, Elaine R HG003079/HG/NHGRI NIH HHS/ P01 CA101937/CA/NCI NIH HHS/ Nat Methods. 2009 Sep;6(9):677-81. doi: 10.1038/nmeth.1363. Epub 2009 Aug 9.

Doolittle WF and Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**: 601–3. Doolittle, W F Sapienza, C ENGLAND Nature. 1980 Apr 17;284(5757):601-3.

Goodrich DW and Duesberg PH. 1990. Retroviral recombination during reverse transcription. *Proc Natl Acad Sci U S A* **87**: 2052–6. Goodrich, D W Duesberg, P H 5-R35-CA39915-04/CA/NCI NIH HHS/ Proc Natl Acad Sci U S A. 1990 Mar;87(6):2052-6.

Hancks DC and Kazazian H H J. 2012. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22**: 191–203. Hancks, Dustin C Kazazian, Haig H Jr Research Support, N.I.H., Extramural Review England Current opinion in genetics & development

Curr Opin Genet Dev. 2012 Jun;22(3):191-203. doi: 10.1016/j.gde.2012.02.006. Epub 2012 Mar 8.

Huang CR, Burns KH, and Boeke JD. 2012. Active transposition in genomes. *Annual Review of Genetics* **46**: 651–75. Huang, Cheng Ran Lisa Burns, Kathleen H Boeke, Jef D P30 CA006973/CA/NCI NIH HHS/ Annu Rev Genet. 2012;46:651-75. doi: 10.1146/annurev-genet-110711-155616.

Hurst GD and Werren JH. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* **2**: 597–606. Hurst, G D Werren, J H England Nat Rev Genet. 2001 Aug;2(8):597-606.

Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, and Devine SE. 2010. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141**: 1253–61. Iskow, Rebecca C McCabe, Michael T Mills, Ryan E Torene, Spencer Pittard, W Stephen Neuwald, Andrew F Van Meir, Erwin G Vertino, Paula M Devine, Scott E 1R01CA132065/CA/NCI NIH HHS/ 5P01CA116676/CA/NCI NIH HHS/ F32HG004207/HG/NHGRI NIH HHS/ R01 CA077337-10/CA/NCI NIH HHS/ R01 CA132065-01A2/CA/NCI NIH HHS/ R01 CA132065-02/CA/NCI NIH HHS/ R01 CA132065-03/CA/NCI NIH HHS/ R01 HG002898/HG/NHGRI NIH HHS/ R01 HG002898-06/HG/NHGRI NIH HHS/ R01CA086335/CA/NCI NIH HHS/ R01CA116804/CA/NCI NIH HHS/ R01GM078541/GM/NIGMS NIH HHS/ R01HG002898/HG/NHGRI NIH HHS/ Cell. 2010 Jun 25;141(7):1253-61. doi: 10.1016/j.cell.2010.05.020.

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, and Paszkowski J. 2011. An sirna pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**: 115–9. Ito, Hidetaka Gaubert, Herve Bucher, Etienne Mirouze, Marie Vaillant, Isabelle Paszkowski, Jerzy England Nature. 2011 Apr 7;472(7341):115-9. doi: 10.1038/nature09861. Epub 2011 Mar 13.

Jiang Y, Wang Y, and Brudno M. 2012. Prism: pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics* **28**: 2576–83. Jiang, Yue Wang, Yadong Brudno, Michael Canadian Institutes of Health Research/Canada England Oxford, England Bioinformatics. 2012 Oct 15;28(20):2576-83. doi: 10.1093/bioinformatics/bts484. Epub 2012 Jul 31.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P, et al.. 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS genetics* **5**: e1000530. Johannes, Frank Porcher, Emmanuelle Teixeira, Felipe K Saliba-Colombani, Vera Simon, Matthieu Agier,

Nicolas Bulski, Agnes Albuisson, Juliette Heredia, Fabiana Audigier, Pascal Bouchez, David Dillmann, Christine Guerche, Philippe Hospital, Frederic Colot, Vincent PLoS Genet. 2009 Jun;5(6):e1000530. doi: 10.1371/journal.pgen.1000530. Epub 2009 Jun 26.

Jordan IK and McDonald JF. 1998. Evidence for the role of recombination in the regulatory evolution of saccharomyces cerevisiae ty elements. *J Mol Evol* **47**: 14–20. Jordan, I K McDonald, J F J Mol Evol. 1998 Jul;47(1):14-20.

Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, Kallicki J, Kaul R, Wilson RK, and Eichler EE. 2010. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* **143**: 837–47. Kidd, Jeffrey M Graves, Tina Newman, Tera L Fulton, Robert Hayden, Hillary S Malig, Maika Kallicki, Joelle Kaul, Rajinder Wilson, Richard K Eichler, Evan E HG004120/HG/NHGRI NIH HHS/United States P01 HG004120-03/HG/NHGRI NIH HHS/United States Howard Hughes Medical Institute/United States Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. United States Cell Cell. 2010 Nov 24;143(5):837-47. doi: 10.1016/j.cell.2010.10.027.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al.. 2012. The arabidopsis information resource (tair): improved gene annotation and new tools. *Nucleic Acids Res* **40**: D1202–10. Lamesch, Philippe Berardini, Tanya Z Li, Donghui Swarbreck, David Wilks, Christopher Sasidharan, Rajkumar Muller, Robert Dreher, Kate Alexander, Debbie L Garcia-Hernandez, Margarita Karthikeyan, Athikkattuvalasu S Lee, Cynthia H Nelson, William D Ploetz, Larry Singh, Shanker Wensel, April Huala, Eva 5P41HG002273-09/HG/NHGRI NIH HHS/ England Nucleic Acids Res. 2012 Jan;40(Database issue):D1202-10. doi: 10.1093/nar/gkr1090. Epub 2011 Dec 2.

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette L J r, Lohr JG, Harris CC, Ding L, Wilson RK, et al.. 2012. Landscape of somatic retrotransposition in human cancers. *Science* **337**: 967–71. Lee, Eunjung Iskow, Rebecca Yang, Lixing Gokcumen, Omer Haseley, Psalm Luquette, Lovelace J 3rd Lohr, Jens G Harris, Christopher C Ding, Li Wilson, Richard K Wheeler, David A Gibbs, Richard A Kucherlapati, Raju Lee, Charles Kharchenko, Peter V Park, Peter J Cancer Genome Atlas Research Network F32 AG039979/AG/NIA NIH HHS/United States F32AG039979/AG/NIA NIH HHS/United States K25AG037596/AG/NIA NIH HHS/United States R01 GM082798/GM/NIGMS NIH HHS/United States R01GM082798/GM/NIGMS NIH HHS/United States RC1HG005482/HG/NHGRI NIH HHS/United States U01 HG005209/HG/NHGRI NIH HHS/United States U01 HG005725/HG/NHGRI NIH HHS/United States U01HG005209/HG/NHGRI NIH HHS/United States U01HG005725/HG/NHGRI NIH HHS/United States U24 CA144025/CA/NCI NIH HHS/United States U24CA144025/CA/NCI NIH HHS/United

States Research Support, N.I.H., Extramural United States Science (New York, N.Y.) Science. 2012 Aug 24;337(6097):967-71. doi: 10.1126/science.1222077. Epub 2012 Jun 28.

Lerat E. 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity* **104**: 520–33. Lerat, E England Heredity (Edinb). 2010 Jun;104(6):520-33. doi: 10.1038/hdy.2009.165. Epub 2009 Nov 25.

Li H and Durbin R. 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**: 1754–60. Li, Heng Durbin, Richard 077192/Z/05/Z/Wellcome Trust/United Kingdom England Oxford, England Bioinformatics. 2009 Jul 15;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. Epub 2009 May 18.

Li Y, Zheng H, Luo R, Wu H, Zhu H, Li R, Cao H, Wu B, Huang S, Shao H, et al.. 2011. Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. *Nat Biotechnol* **29**: 723–30. Li, Yingrui Zheng, Hancheng Luo, Ruibang Wu, Honglong Zhu, Hongmei Li, Ruiqiang Cao, Hongzhi Wu, Boxin Huang, Shujia Shao, Haojing Ma, Hanzhou Zhang, Fan Feng, Shuijian Zhang, Wei Du, Hongli Tian, Geng Li, Jingxiang Zhang, Xiuqing Li, Songgang Bolund, Lars Kristiansen, Karsten de Smith, Adam J Blakemore, Alexandra I F Coin, Lachlan J M Yang, Huanming Wang, Jian Wang, Jun Nat Biotechnol. 2011 Jul 24;29(8):723-30. doi: 10.1038/nbt.1904.

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al.. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471–6. Lippman, Zachary Gendrel, Anne-Valerie Black, Michael Vaughn, Matthew W Dedhia, Neilay McCombie, W Richard Lavine, Kimberly Mittal, Vivek May, Bruce Kasschau, Kristin D Carrington, James C Doerge, Rebecca W Colot, Vincent Martienssen, Rob Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. England Nature Nature. 2004 Jul 22;430(6998):471-6.

Lopez-Flores I and Garrido-Ramos MA. 2010. The repetitive dna content of eukaryotic genomes. *Genome dynamics* **7**: 1–28. Lopez-Flores, I Garrido-Ramos, M A Switzerland Genome Dyn. 2010;7:1-28. doi: 10.1159/000337118. Epub 2012 Jun 25.

Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, and Mager DL. 2006. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet* **2**: e2. Maksakova, Irina A Romanish, Mark T Gagnier, Liane Dunn, Catherine A van de Lagemaat, Louie N Mager, Dixie L PLoS Genet. 2006 Jan;2(1):e2.

Medvedev P, Stanciu M, and Brudno M. 2009. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods* **6**: S13–20. Medvedev,

Paul Stanciu, Monica Brudno, Michael Nat Methods. 2009 Nov;6(11 Suppl):S13-20. doi: 10.1038/nmeth.1374.

Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, and Mathieu O. 2009. Selective epigenetic control of retrotransposition in arabidopsis. *Nature* **461**: 427–30. Mirouze, Marie Reinders, Jon Bucher, Etienne Nishimura, Taisuke Schneeberger, Korbinian Ossowski, Stephan Cao, Jun Weigel, Detlef Paszkowski, Jerzy Mathieu, Olivier England Nature. 2009 Sep 17;461(7262):427-30. doi: 10.1038/nature08328. Epub 2009 Sep 6.

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, and Kakutani T. 2001. Mobilization of transposons by a mutation abolishing full dna methylation in arabidopsis. *Nature* **411**: 212–4. Miura, A Yonebayashi, S Watanabe, K Toyama, T Shimada, H Kakutani, T England Nature. 2001 May 10;411(6834):212-4.

Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, and Gage FH. 2010. L1 retrotransposition in neurons is modulated by mecp2. *Nature* **468**: 443–6. Muotri, Alysson R Marchetto, Maria C N Coufal, Nicole G Oefner, Ruth Yeo, Gene Nakashima, Kinichi Gage, Fred H 1-DP2-OD006495-01/OD/NIH HHS/United States DP2 OD006495/OD/NIH HHS/United States DP2 OD006495-01/OD/NIH HHS/United States R01 MH088485-03/MH/NIMH NIH HHS/United States R01MH088485/MH/NIMH NIH HHS/United States Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England Nature Nature. 2010 Nov 18;468(7322):443-6. doi: 10.1038/nature09544.

Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, and Ponting CP. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol* **13**: R45. Nellaker, Christoffer Keane, Thomas M Yalcin, Binnaz Wong, Kim Agam, Avigail Belgard, T Grant Flint, Jonathan Adams, David J Frankel, Wayne N Ponting, Chris P R01 NS031348/NS/NINDS NIH HHS/ Medical Research Council/United Kingdom Wellcome Trust/United Kingdom Cancer Research UK/United Kingdom England Genome Biol. 2012 Jun 15;13(6):R45. doi: 10.1186/gb-2012-13-6-r45.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al.. 2012. Bedops: high-performance genomic feature operations. *Bioinformatics* **28**: 1919–20. Neph, Shane Kuehn, M Scott Reynolds, Alex P Haugen, Eric Thurman, Robert E Johnson, Audra K Rynes, Eric Maurano, Matthew T Vierstra, Jeff Thomas, Sean Sandstrom, Richard Humbert, Richard Stamatoyannopoulos, John A 1U54HG004592/HG/NHGRI NIH HHS/ 5U01ES017156/ES/NIEHS NIH HHS/ England Oxford, England Bioinformatics. 2012 Jul 15;28(14):1919-20. doi: 10.1093/bioinformatics/bts277. Epub 2012 May 9.

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, and Weigel D. 2008. Sequencing of natural strains of arabidopsis thaliana with short reads. *Genome Res* **18**: 2024–33. Ossowski, Stephan Schneeberger, Korbinian Clark, Richard M Lanz, Christa Warthmann, Norman Weigel, Detlef Genome Res. 2008 Dec;18(12):2024-33. doi: 10.1101/gr.080200.108. Epub 2008 Sep 25.

Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, Mell JC, and Hall IM. 2010. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Res* **20**: 623–35. Quinlan, Aaron R Clark, Royden A Sokolova, Svetlana Leibowitz, Mitchell L Zhang, Yujun Hurles, Matthew E Mell, Joshua C Hall, Ira M 1F32HG005197-01/HG/NHGRI NIH HHS/ DP2OD006493-01/OD/NIH HHS/ Genome Res. 2010 May;20(5):623-35. doi: 10.1101/gr.102970.109. Epub 2010 Mar 22.

Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, and Korbel JO. 2012. Delly: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333–i339. Rausch, Tobias Zichner, Thomas Schlattl, Andreas Stutz, Adrian M Benes, Vladimir Korbel, Jan O England Oxford, England Bioinformatics. 2012 Sep 15;28(18):i333-i339. doi: 10.1093/bioinformatics/bts378.

Rebollo R, Romanish MT, and Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annual Review of Genetics* **46**: 21–42. Rebollo, Rita Romanish, Mark T Mager, Dixie L Canadian Institutes of Health Research/-Canada Annu Rev Genet. 2012;46:21-42. doi: 10.1146/annurev-genet-110711-155621. Epub 2012 Aug 16.

Sibson R. 1973. Slink - optimally efficient algorithm for single-link cluster method. *Computer Journal* **16**: 30–34. O9837 Times Cited:152 Cited References Count:10.

Siek J, Lee L, and Lumsdaine A. 2000. Boost graph library. http://www.boost.org/libs/graph/.

Sindi SS, Onal S, Peng LC, Wu HT, and Raphael BJ. 2012. An integrative probabilistic model for identification of structural variation in sequencing data. *Genome Biol* **13**: R22. Sindi, Suzanne S Onal, Selim Peng, Luke C Wu, Hsin-Ta Raphael, Benjamin J R01 HG005690/HG/NHGRI NIH HHS/ R01 HG5690/HG/NHGRI NIH HHS/ England Genome Biol. 2012;13(3):R22. doi: 10.1186/gb-2012-13-3-r22.

Singer T, Yordan C, and Martienssen RA. 2001. Robertson's mutator transposons in a. thaliana are regulated by the chromatin-remodeling gene decrease in dna methylation (ddm1). *Genes & development* **15**: 591–602. Singer, T Yordan, C Martienssen, R A Genes Dev. 2001 Mar 1;15(5):591-602.

Slotkin RK and Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272–85. Slotkin, R Keith Martienssen, Robert R01 GM067014-01A1/GM/NIGMS NIH HHS/United States Research Support, N.I.H., Extramural Review England Nature reviews. Genetics Nat Rev Genet. 2007 Apr;8(4):272-85.

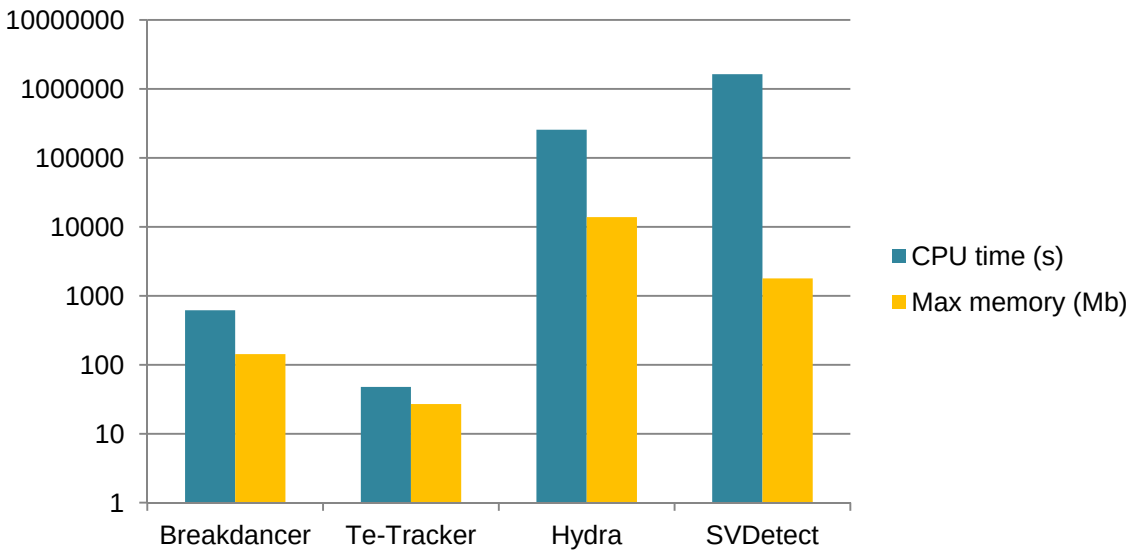
Thomas CA, Paquola AC, and Muotri AR. 2012. Line-1 retrotransposition in the nervous system. *Annual review of cell and developmental biology* **28**: 555–73. Thomas, Charles A Paquola, Apua C M Muotri, Alysson R 1-DP2-OD006495-01/OD/NIH HHS/ 1F31NS076198-01A1/NS/NINDS NIH HHS/ R01 MH094753-01/MH/NIMH NIH HHS/ Annu Rev Cell Dev Biol. 2012;28:555-73. doi: 10.1146/annurev-cellbio-101011-155822.

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, and Kakutani T. 2009. Bursts of retrotransposition reproduced in arabidopsis. *Nature* **461**: 423–6. Tsukahara, Sayuri Kobayashi, Akie Kawabe, Akira Mathieu, Olivier Miura, Asuka Kakutani, Tetsuji England Nature. 2009 Sep 17;461(7262):423-6. doi: 10.1038/nature08351. Epub 2009 Sep 6.

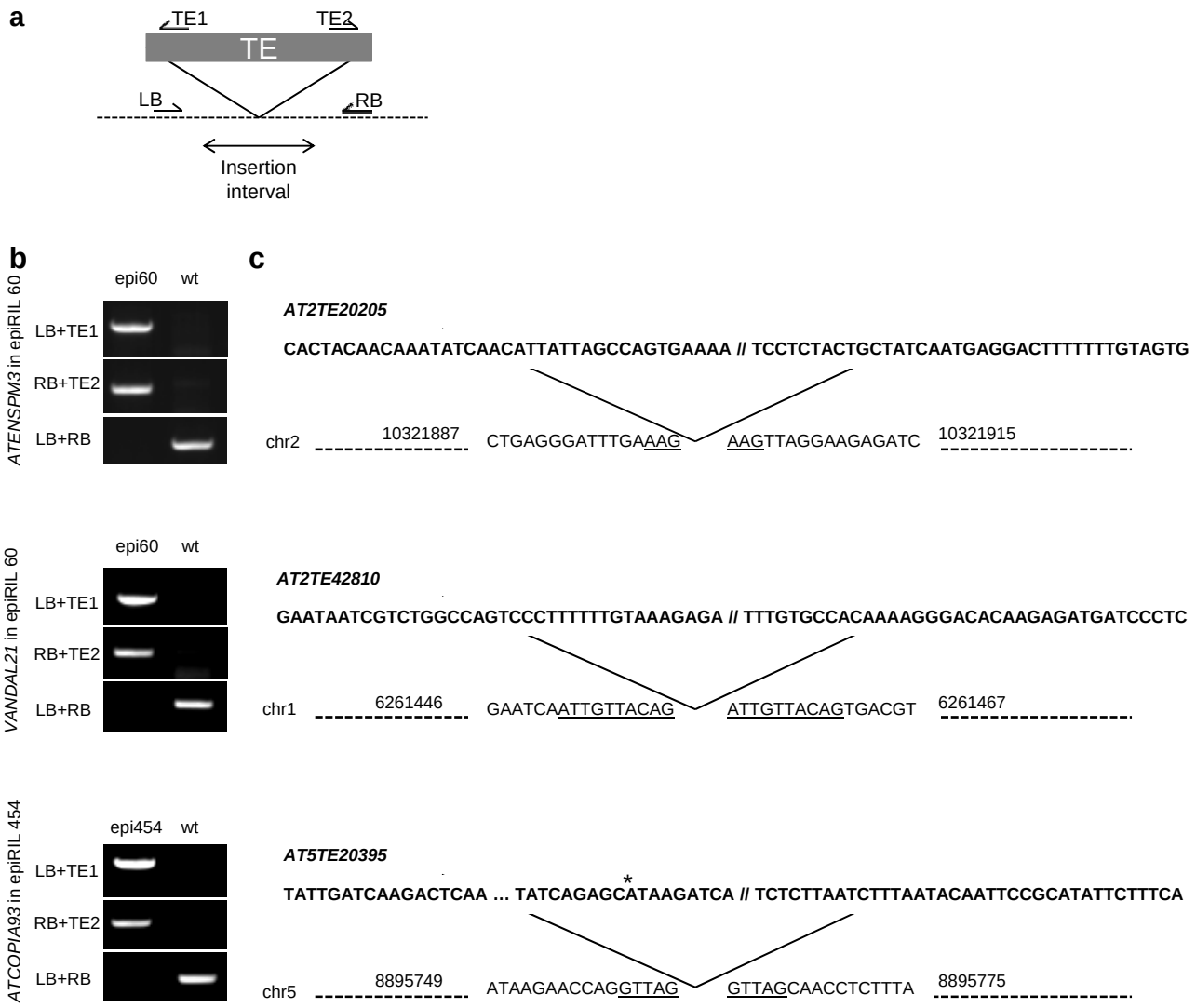
Vongs A, Kakutani T, Martienssen RA, and Richards EJ. 1993. Arabidopsis thaliana dna methylation mutants. *Science* **260**: 1926–8. Vongs, A Kakutani, T Martienssen, R A Richards, E J New York, N.Y. Science. 1993 Jun 25;260(5116):1926-8.

Zamudio N and Bourc'his D. 2010. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity* **105**: 92–104. Zamudio, N Bourc'his, D England Heredity (Edinb). 2010 Jul;105(1):92-104. doi: 10.1038/hdy.2010.53. Epub 2010 May 5.

Zeitouni B, Boeva V, Janoueix-Lerosey I, Loeillet S, Legoix-ne P, Nicolas A, Delattre O, and Barillot E. 2010. Svdetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* **26**: 1895–6. Zeitouni, Bruno Boeva, Valentina Janoueix-Lerosey, Isabelle Loeillet, Sophie Legoix-ne, Patricia Nicolas, Alain Delattre, Olivier Barillot, Emmanuel England Oxford, England Bioinformatics. 2010 Aug 1;26(15):1895-6. doi: 10.1093/bioinformatics/btq293.



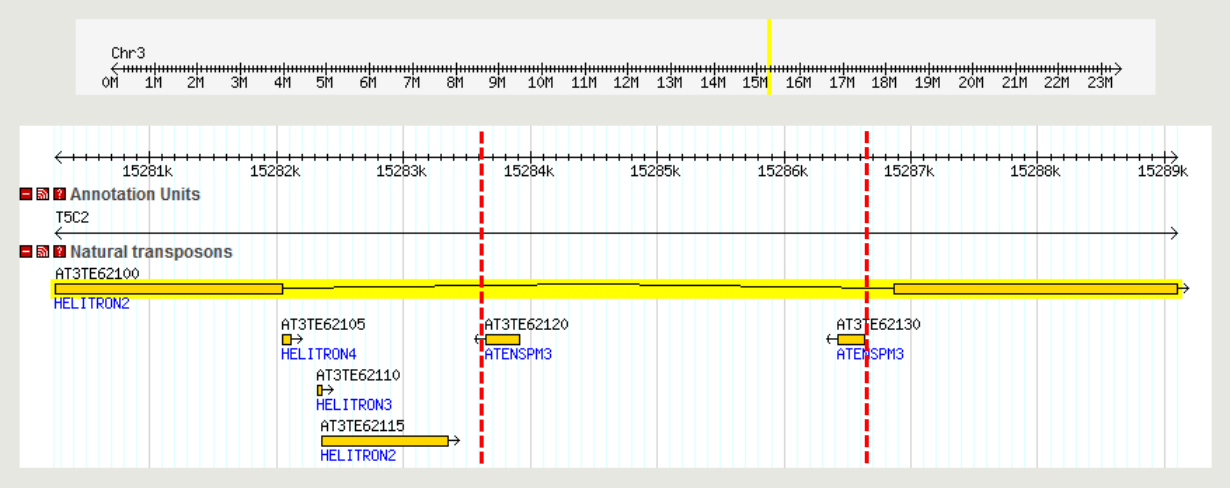
Supplementary figure 1 : Comparative evaluation of TE-Tracker time and memory usage. Both variables were evaluated using Platform Computing® Load Sharing Facility. Time is displayed in seconds, peak memory usage in megabytes, both displayed in log-scale.



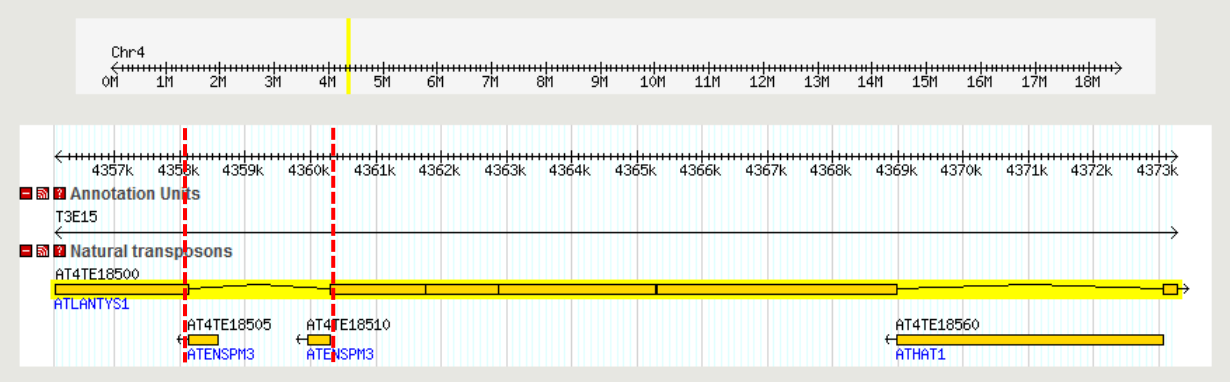
Supplementary figure 2 : Validation of new TE insertions.

a. Scheme of the PCR validation procedure, arrows represent primers. **b.** PCR results for 16 three new insertions identied in two epiRILs and involving three distinct donors : *ATENSPM3* (*AT2TE20205*), *VANDAL21* (*AT2TE42810*) and *ATCOPIA93* (*AT5TE20395*). **c.** Sequencing results of the PCR products that correspond to the borders of new insertions. TE sequence is represented in bold and target site duplications generated by TE insertion are underlined. The star highlights the single polymorphism located in the extremity of the TE and that discriminates between the two ATCOPIA93 potential donors.

a *ATENSPM3,ATENSPM3,HELITRON2*



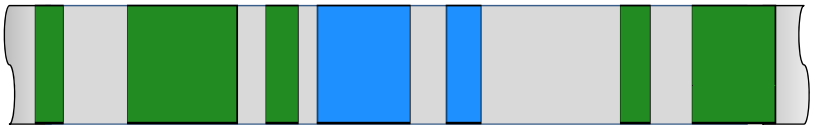
b *ATENSPM3,ATLANTYS1,ATENSPM3*



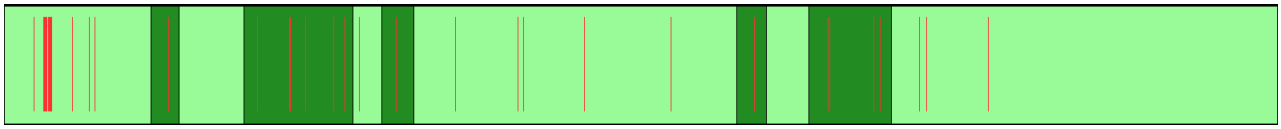
Supplementary figure 3 : **TAIR (Lamesch et al., 2012) Gbrowse display of composite elements detected by TE-Tracker.**

a. Element composed of two sequences annotated as ATENSPM3 and one annotated as HELITRON2. **b.** Element composed of two sequences annotated as ATENSPM3 and one annotated as ATLANTYS1. Red dotted lines indicate the boundaries of the mobile sequence as detected by TE-Tracker.

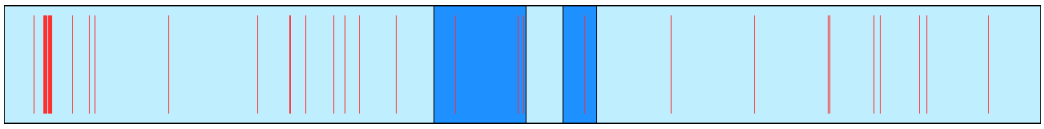
Putative composite sequence



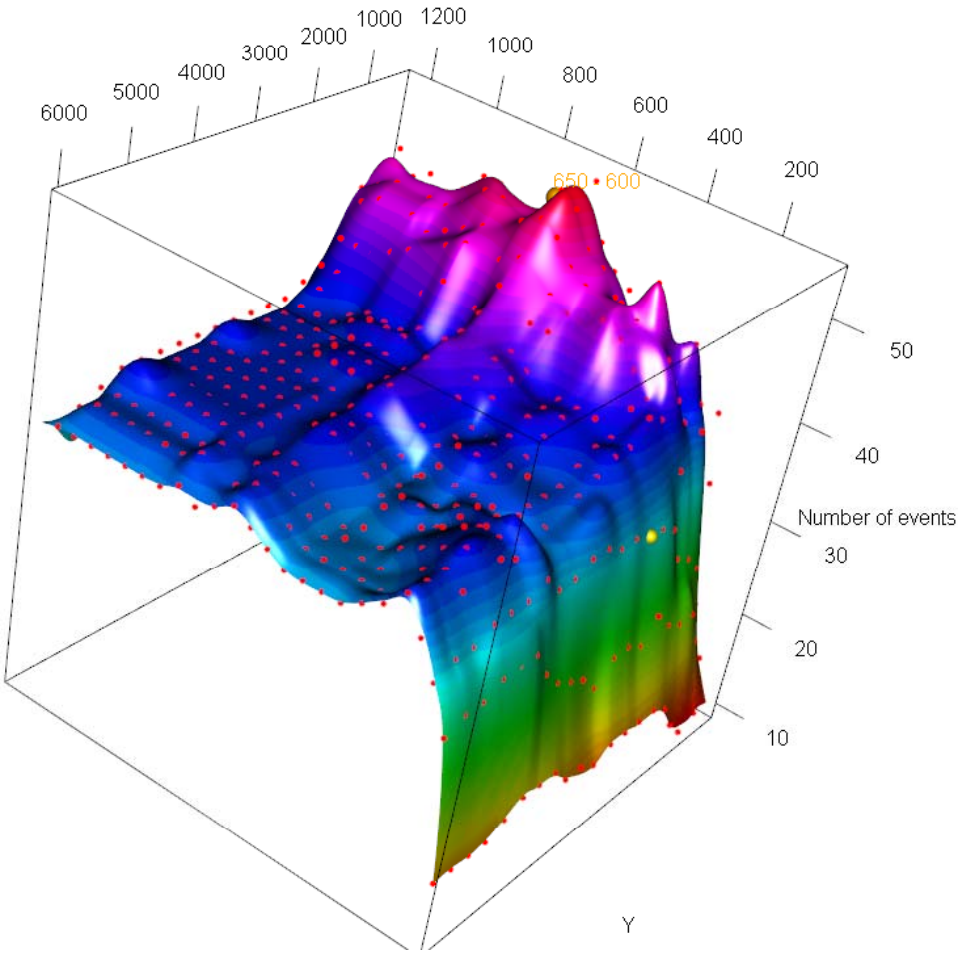
AT5TE15240



AT3TE89830



Supplementary Figure 4 : **Multiple-origin unique regions for donor suggest a recombination-like event during transposition.**
Clusters of reads mapping to acceptor region chr3:8763824..8765272 of epiRIL 454 have their mates mapping on several genomic regions corresponding to ATCOPIA78 sequences, which indicates multiple potential donors. The donor scoring feature extracted mates that mapped significantly better on one copy than on all others; this yielded not one, but two candidates. Piling up the said reads led to the configuration above. In light blue and green are the two copies, aligned unto each other. Red ticks indicate variants (SNP and indels) between the two copies. In dark blue and green are the regions delimited by specific reads for each copy, respectively. On top is the putative inserted sequence with blocks coming from either copy. Ticks not included in darker regions represent non-unique variations that are therefore also found when aligning to other *ATCOPIA78* sequences.



Supplementary figure 5 : **Number of events on two-dimensional grid for epiRIL 439.** Example of surface generated by a 800-point grid in X-Y space-Surface was loess-interpolated, coordinates of the maximum are displayed in orange, Markov-chain point estimate is displayed by the orange sphere near the peak, distribution estimate is displayed by a yellow sphere.

	Acceptor stire		Donor			
ID	chr.	position	chr.	start	stop	detected
1	chr1	248294	chr5	5630011	5634975	x
2	chr1	623731	chr5	5630011	5634975	x
3	chr1	770974	chr5	5630011	5634975	x
4	chr1	1071330	chr5	5630011	5634975	x
5	chr1	2037961	chr5	5630011	5634975	x
6	chr1	2257882	chr5	5630011	5634975	x
7	chr1	2481930	chr5	5630011	5634975	x
8	chr1	2871763	chr5	5630011	5634975	x
9	chr1	3340610	chr5	5630011	5634975	x
10	chr1	5016892	chr5	5630011	5634975	x
11	chr1	5046091	chr5	5630011	5634975	x
12	chr1	5250696	chr5	5630011	5634975	x
13	chr1	6416036	chr5	5630011	5634975	x
14	chr1	6715956	chr5	5630011	5634975	x
15	chr1	6912218	chr5	5630011	5634975	x
16	chr1	7982407	chr5	5630011	5634975	x
17	chr1	8603037	chr5	5630011	5634975	x
18	chr1	9735777	chr5	5630011	5634975	x
19	chr1	10284309	chr5	5630011	5634975	
20	chr1	10671044	chr5	5630011	5634975	x
21	chr1	12351929	chr5	5630011	5634975	x
22	chr1	13861366	chr5	5630011	5634975	x
23	chr1	14152163	chr5	5630011	5634975	x
24	chr1	16793854	chr5	5630011	5634975	x
25	chr1	17841838	chr5	5630011	5634975	x
26	chr1	19477905	chr5	5630011	5634975	x
27	chr1	20198447	chr5	5630011	5634975	x
28	chr1	20795441	chr5	5630011	5634975	x
29	chr1	20917772	chr5	5630011	5634975	x
30	chr1	22738695	chr5	5630011	5634975	x
31	chr1	22788970	chr5	5630011	5634975	x
32	chr1	23436613	chr5	5630011	5634975	x
33	chr1	24714782	chr5	5630011	5634975	x
34	chr1	25429340	chr5	5630011	5634975	x
35	chr1	25902534	chr5	5630011	5634975	x
36	chr1	27314695	chr5	5630011	5634975	x
37	chr1	27552632	chr5	5630011	5634975	x
38	chr1	27828210	chr5	5630011	5634975	x
39	chr1	29071608	chr5	5630011	5634975	x
40	chr5	1505621	chr5	5630011	5634975	x

	Acceptor stire		Donor			
ID	chr.	position	chr.	start	stop	detected
41	chr5	1674163	chr5	5630011	5634975	x
42	chr5	2174400	chr5	5630011	5634975	x
43	chr5	3304484	chr5	5630011	5634975	x
44	chr5	3910297	chr5	5630011	5634975	x
45	chr5	4230314	chr5	5630011	5634975	x
46	chr5	6104389	chr5	5630011	5634975	x
47	chr5	6854645	chr5	5630011	5634975	x
48	chr5	6906526	chr5	5630011	5634975	x
49	chr5	7812336	chr5	5630011	5634975	x
50	chr5	9091588	chr5	5630011	5634975	x
51	chr5	9698725	chr5	5630011	5634975	x
52	chr5	10365579	chr5	5630011	5634975	x
53	chr5	10661178	chr5	5630011	5634975	x
54	chr5	11826789	chr5	5630011	5634975	
55	chr5	12309261	chr5	5630011	5634975	x
56	chr5	13315486	chr5	5630011	5634975	x
57	chr5	13727742	chr5	5630011	5634975	x
58	chr5	14112116	chr5	5630011	5634975	x
59	chr5	14356539	chr5	5630011	5634975	
60	chr5	14432804	chr5	5630011	5634975	
61	chr5	15059811	chr5	5630011	5634975	x
62	chr5	15108704	chr5	5630011	5634975	x
63	chr5	16366871	chr5	5630011	5634975	x
64	chr5	17754143	chr5	5630011	5634975	x
65	chr5	18929477	chr5	5630011	5634975	x
66	chr5	19550283	chr5	5630011	5634975	x
67	chr5	19717309	chr5	5630011	5634975	x
68	chr5	21710387	chr5	5630011	5634975	x
69	chr5	22466087	chr5	5630011	5634975	x
70	chr5	22514806	chr5	5630011	5634975	x
71	chr5	23367420	chr5	5630011	5634975	x
72	chr5	23406966	chr5	5630011	5634975	x
73	chr5	23418427	chr5	5630011	5634975	x
74	chr5	23445199	chr5	5630011	5634975	x
75	chr5	24842981	chr5	5630011	5634975	x
76	chr5	24946033	chr5	5630011	5634975	x
77	chr5	25260678	chr5	5630011	5634975	x
78	chr5	25589670	chr5	5630011	5634975	x
79	chr5	26965212	chr5	5630011	5634975	x

Supplementary Table 1: Synthetic TE insertions in Arabidopsis

acceptor site		donor		
chromosome	position	chromosome	start	stop
chr19	3943363	chr12	101539821	101545842
chr19	8369980	chr12	75268648	75274681
chr19	9250040	chr12	75268648	75274681
chr19	18299455	chr12	101539821	101545842
chr19	31836999	chr12	101539821	101545842
chr19	32550986	chr12	75268648	75274681
chr19	37320442	chr12	101539821	101545842
chr19	38820257	chr12	101539821	101545842
chr19	40646550	chr12	75268648	75274681
chr19	42456336	chr12	75268648	75274681
chr19	42984248	chr12	101539821	101545842
chr19	45046186	chr12	101539821	101545842
chr19	52203029	chr12	101539821	101545842
chr19	52246544	chr12	75268648	75274681
chr19	54702746	chr12	101539821	101545842
chr19	55041491	chr12	75268648	75274681
chr19	58771303	chr12	75268648	75274681

Supplementary Table 2: Synthetic TE insertions in Human

line	coverage	*p(c)* < 10 (%)	sequencing
55	24.64	2.61	GAIIx
60	27.55	1.70	GAIIx
454	17.68	11.39	HiSeq2000
439	32.71	0.62	HiSeq2000
wt1	29.54	0.84	GAIIx
wt2	37.27	0.48	HiSeq2000

Supplementary Table 3: Sequencing and alignment properties

Supplementary Table 4-9 :
Ces tables sont présentées dans l'annexe 1 du manuscrit de thèse

	TE	validation of the presence	validation of the donnor
present in epiRILs only	*ATENSPM3*	13/13	9/9
	VANDAL21	5/5	3/3
	ATCOPIA93	42/42	14/14
	ATCOPIA13	1/1	-
	ATCOPIA78	4/4	2/3
	ATENSPM3;ATLANTYS1;ATENSPM3	2/2	ND
	ATENSPM3;ATENSPM3;HELITRON	3/3	-
present in wt	*ATGP2*	1/1	ND
	ATHILA6A	1/1	ND
	ATCOPIA78	1/1	ND
TOTAL		72/72	25/26

Supplementary Table 10: Summary of PCR validation of new insertions

	Paired-End		Mate-Pair		
	True Positives(/17)	Total Positives	True Positives	Total Positives	Correctly Scored
TE-Tracker	-	-	100%	352	17
BreakDancerMax***	17%	1600	0%	2859	-
PRISM**	47%	9641	-	-	-
SHORE			-	-	-
SVDetect*	?	?	94%	125929	-
Hydra	-	-	82%	42521	-

Supplementary Table 11: Comparative sensitivity evaluation for common structural variation detection tools (for *ATCOPIA93* insertions)

	Paired-End		Mate-Pair		
	True Positives(/6)	Total Positives	True Positives	Total Positives	Correctly Scored
TE-Tracker	-	-	100%	176	6
BreakDancerMax***	100%	1600	0%	2859	-
PRISM**	100%	9641	-	-	-
SHORE			-	-	-
SVDetect*	?	?	100%	125929	-
Hydra	-	-	100%	42521	-

Supplementary Table 12: Comparative sensitivity evaluation for common structural variation detection tools (for *ATENSPM3* insertions)

*We did not manage to get SVDetect to produce results on paired-end data (program crashed or run did not finish).

** Results filtered for events supported by more than 10 reads

*** For BreakDancer, we allowed for an extra 500 base pairs left or right.

A blank cell means the test was not performed, a question mark means the program did not terminate, a hyphen indicates an unsupported function

primers flanking the acceptor sites				
primer name	primer sequence	acceptor site		
		chr	start	stop
55_COPIA13_i1_F	CCCACATCCAAAATCCCTTT	chr3	17341732	17341847
55_COPIA13_i1_R	TGGTGCAGAATATGGAATGG			
55_COPIA78_i1_F	ATCATCACCGTGGTCCTTGT	chr4	10739659	10739988
55_COPIA78_i1_R	GACCACTGAAATCGGTCCAC			
55_COPIA78_i2_F	TGATATGTGTGGCTCTTGTTCA	chr4	13252525	13253326
55_COPIA78_i2_R	CCTCGAAAATTTAGGGCAGA			
55_COPIA93_i1_F	GTAAGTGCATTTGATGGATTCG	chr1	3990167	3990279
55_COPIA93_i1_R	CCATGAATCGGATTTTAAACTGT			
55_COPIA93_i2_F	ACCATCACCAAACTCAACTGG	chr1	12169447	12169578
55_COPIA93_i2_R	TGAATCTCCGGATCGCTTAC			
55_COPIA93_i3_F	ATTGATTTCTTTTGGCAGTGACT	chr1	20679800	20679956
55_COPIA93_i3_R	CTACATGAATCCCCACCACA			
55_COPIA93_i4_F	GCTGATCTGCAGCTTATTTCCT	chr1	25815758	25815886
55_COPIA93_i4_R	CTGGACATGCAGCTTAACAAAG			
55_COPIA93_i5_F	GTAGTGAGTCTTGTGAGCAGAGGT	chr2	3215890	3216071
55_COPIA93_i5_R	CCATAAAGATACTGACGCCACA			
55_COPIA93_i6_F	CCGGTAGCTCTACTAAGCGAAG	chr2	3275722	3275965
55_COPIA93_i6_R	CTATGTGAGGTGGGGAAAGGT			
55_COPIA93_i7_F	ATGGTGCCTTGATCTTCAGC	chr2	3505788	3506083
55_COPIA93_i7_R	AGAGAGTCCTCACCTAACAAGCA			
55_COPIA93_i8_F	AAGCTCTGAACCATCCATGC	chr2	4171056	4171205
55_COPIA93_i8_R	AAAGATCCCCTTATCCTTGGAA			
55_COPIA93_i9_F	TCTTCAGCAGTAACCACACGA	chr2	8272168	8272305
55_COPIA93_i9_R	TTGGATGGGGGAGATAAGAAC			
55_COPIA93_10_F	AGGACCATCTGATGACTTTATTCT	chr2	19475806	19475903
55_COPIA93_10_R	TACGCCGGATCCTCCTAGT			
55_COPIA93_i11_F	AAACCCCTAAAACCCTTCTTCT	chr3	4589949	4590097
55_COPIA93_i11_R	GTGAAATGGCTCATGGTTAAAG			
55_COPIA93_i12_F	GAGGAGATACGGGAAATGTGA	chr3	8036011	8036108
55_COPIA93_i12_R	GGATCACACATGCTTGAAAAGA			
55_COPIA93_i13_F	GGACTTGAGAGGGCAAATAGTG	chr4	8561572	8561742
55_COPIA93_i13_R	CTCTAAATTCAACCTCACCGAAC			
55_COPIA93_i14_F	ACCCGTAAGACTTTGCGGTA	chr5	12603975	12604164
55_COPIA93_i14_R	CCATCGAGTTAACGGCATTC			
55_COPIA93_i15_F	TCCGACTTGCTGGGTTTACT	chr5	14550492	14550576
55_COPIA93_i15_R	ATGAGATGGTCTTTCGCTACG			
55_COPIA93_i16_F	AGAAGGCCATGTCTTGATCCT	chr5	19519546	19519749
55_COPIA93_i16_R	TGGTATCCACGTTGGAAGAAG			
55_SPM3_i1_F	CACACCTTCACCCATGTGAT	chr1	21486789	21486913
55_SPM3_i1_R	CAACCCTTTACCAGGGTACCTA			
55_SPM3_i2_F	TGTGTTCAAGGGACAAAAACAG	chr1	27101182	27103667
55_SPM3_i2_R	TACGTCCCGTTTCTTTGGTC			
55_SPM3_i3_F	GATCTTATGTTACGTCAAAGCCAAC	chr3	9043438	9043531
55_SPM3_i3_R	TGGAGTCCTGCCTTTATTCTTG			
55_SPM3_i4_F	CCAAACCTATGTGTTGACGAAG	chr3	10647578	10647846
55_SPM3_i4_R	GGGAAAGTCGACAGCGATAG			
55_SPM3_i5_F	GGAAAGCCAAGATTTTGATGAC	chr3	16618765	16619206
55_SPM3_i5_R	CAACACCATTCATGACAAAACC			
55_SPM3_i6_F	TTTCTCAGGCAGGAAGAGACA	chr5	24028648	24028773
55_SPM3_i6_R	GCGATCAAGCTCTCATCAAAC			
55_COPIA78_i1_F	TGCAGATCGTCAACGTTTGT	chr3	8764557	8764584
55_COPIA78_i1_R	TTCAGCACGCCACTACATTC			
55_COPIA78_i2_F	TTTTCCCAATCCAAAAATCA	chr3	22615358	22615562
55_COPIA78_i2_R	TTAAATGATGACGCGGAAAG			
60_COPIA93_i1_F	ATGCTGCAACCTCTGAACCT	chr2	15677093	15677345
60_COPIA93_i1_R	TCTCATGTGCTCTTGGTTGG			
60_SPM3_i1_F	ACTAGATTTCGCTGTCGTTCC	chr1	25956040	25956309
60_SPM3_i1_R	AGCGAAAATCAATAGCGTGA			
60_SPM3_i2_F	CAATTTCGACACCTTTCTCG	chr2	10321820	10321904
60_SPM3_i2_R	AAGCAGAAGAAACCCATCATC			
60_SPM3_i3_F	CAGAAACTCCCCTTTTCAGG	chr2	17097260	17097372
60_SPM3_i3_R	AATTTGCATTTGTCGTCAGG			
60_SPM3HELI_i1_F	CCAAACAACAAACCAAGTAATCG	chr1	12178270	12181765
60_SPM3HELI_i1_R	TGGAATCTTAGCTTGTTTTGG			
60_SPM3HELI_i2_F	TGTCTGGCATTACAACATTTAAAA	chr3	3464127	3467214
60_SPM3HELI_i2_R	CACTGATGATTCCGGGAAGA			
60_SPM3HELI_i3_F	ACGGGCTCTTGGTGTATCTG	chr3	18274281	18277593
60_SPM3HELI_i3_R	GGATCCTTAATGTAAGCTTTTCTGG			
60_SPM3ATLA_i1_F	CGAATATCGATGGACTAACAATACC	chr3	3464127	3467214
60_SPM3ATLA_i1_R	TCGTGTTTACGGAGAACGTG			

Supplementary Table 13 (part1) : primer sequences for PCR validation of new TE insertions

primers flanking the acceptor sites				
primer name	primer sequenC e	acceptor site		
		chr	start	stop
60_SPM3ATLA_i2_F	CGTCGAGGCTTGAGAGAAGT	chr3	15773104	15778796
60_SPM3ATLA_i2_R	TGTCCCACCCATCTCTCTTT			
60_VANDAL21_i1_F	CTAAGACACCAATCCGGTGA	chr1	6261367	6261460
60_VANDAL21_i1_R	CGGCTGAGTTAGTTGTGGAG			
60_VANDAL21_i2_F	TTTGGAGTTGACGGAAGATG	chr2	12660632	12660713
60_VANDAL21_i2_R	GCCGTGTGACGTTAGTAGGA			
60_VANDAL21_i3_F	AATCAATCCAGCGAAGAAGAC	chr4	1151516	1151621
60_VANDAL21_i3_R	TTCGTCGGTAAATGAACCAG			
439_COPIA93_i1_F	ATCAACAAATAGATATCTTTGCATTTT	chr1	11211722	11211794
439_COPIA93_i1_R	TTCTTTTCGGATTCTGTGTGG			
439_COPIA93_i2_F	ATGGTTCGATTGTGGAGGAA	chr3	21970792	21970868
439_COPIA93_i2_R	CAACATTCTAGGGTTTTACCTTGC			
439_COPIA93_i3_F	TCTCCCACCCAAAATCTCAG	chr5	6591834	6592026
439_COPIA93_i3_R	GCGTGTTGTTTTGCAGTGAT			
439_SPM3_i1_F	AAAAGATCAACATGGTCACTCG	chr1	24091117	24091194
439_SPM3_i2_F	TGGACTTCGACGTTAGATCTTG	chr1	24334475	24334558
439_SPM3_i3_F	GCTGAGAATGATGTTTGGTTG	chr2	11505759	11505852
439_SPM3_i4_F	AACGTCAAATACATTCGCCA	chr5	7815404	7815488
439_VANDAL21_i1_F	CGATTTGACCCCAAAACAAA	chr1	28631078	28631250
439_VANDAL21_i2_F	TCCAGAGCTTCCCGTCAGTTT	chr2	79850	80054
454_COPIA93_i1_F	CTACTGGCGTGGATGCTTTT	chr1	3115	3249
454_COPIA93_i2_F	TATGACTGGGAAGCACGATG	chr1	11191545	11191795
454_COPIA93_i3_F	TTCTTCCAAAATTTGATTACTTGC	chr1	25653718	25654122
454_COPIA93_i4_F	CCCAAGTGTAGTCCAAAACG	chr1	29616254	29616578
454_COPIA93_i5_F	TGTACAAAAGCTAGGTAGGAGGATG	chr2	11496506	11496664
454_COPIA93_i6_F	TTTTGTGATATTTCTGCGTAGATTG	chr3	807190	807542
454_COPIA93_i7_F	TTGGAATTCGATTCAACTAGAGG	chr3	1737581	1737875
454_COPIA93_i8_F	TGTGCGTAGCTTTTCCTTGA	chr3	3083438	3083594
454_COPIA93_i9_F	GCCTTAATGAATGGCCCAATA	chr3	4492935	4493027
454_COPIA93_10_F	GACCCCAACCAATCAATCAC	chr3	16580047	16580428
454_COPIA93_i11_F	AAAATGTCACTTATCATTGCAAGC	chr3	16894460	16894609
454_COPIA93_i12_F	CGCTTGCTAGGCAACCTGTA	chr3	21980027	21980291
454_COPIA93_i13_F	GAAGGGGAAACCCTATGGAA	chr4	3651904	3652219
454_COPIA93_i14_F	TTATCTTTAATGTGACGATTGCTTT	chr4	5006317	5006790
454_COPIA93_i15_F	ACGCATCATCAGGCTTCTCT	chr4	7003389	7003463
454_COPIA93_i16_F	CCAACGTGGGTTAGATCCAT	chr4	7083314	7083461
454_COPIA93_i17_F	CTTGAGGGGCTGTCTCTGAT	chr4	16682304	16682584
454_COPIA93_i18_F	GGATTCTGCTTTTAGACCATCAA	chr5	729472	729675
454_COPIA93_i19_F	CGGTGTTTGTAGTGGGAACC	chr5	8895666	8895869
454_COPIA93_i19_R	GCAAAAACATTCGTACATTTCG			
454_COPIA93_i20_F	TCTCTTGTCTTGTGCGACTCA	chr5	22921922	22922268
454_COPIA93_i21_F	TGACAAGAAAACCACAAGACAAA	chr5	23913821	23913992
454_COPIA93_i22_F	GCTACGCTGACTGGGCTTCT	chr5	25834575	25834843
wt_COPIA78_i1_F	TGTGTCGTTGAAACACATCC	chr4	9320384	9320617
wt_COPIA78_i1_R	ACAATCCGTACCAAGCGAAC			
wt_ATGP2_i1_F	TAAACCCGTAAGGCGCAAG	chr3	14200805	14201497
wt_ATHILA6A_i1_R	ATTCAAAAACCGGAGAAGCA	chr3	5285507	5285698

primers within donor TEs	
VANDAL21_LB_R	CTTGCAGGAGGAAAACG
VANDAL21_RB_F	TAGGCGGCATGACTGATTTC
COPIA93_LB_R	CAAGCACAAACGGACTGATG
COPIA93_RB_F	TTGCATAAGTCTTGCGCTTG
SPM3_LB_R	TAAGTGTTGGCGCTGAAGTG
SPM3_RB_F	AAAAGTGTGGAGGCCATCAG
ATHILA6A_LB_R	GCAAAGTGTGCAAGAATGATG
ATGP2_LB_R	GATGCGGACTGCCTAAAGTC
ATGP2_RB_F	GGGAAAAGCAAGGGATTTGT
COPIA13_LB_R	GTGCGAATCCAACCCACTAT
COPIA13_RB_F	CCTGGTCAAGGCATTTTGTT
SPM3HELI_LB_R	GGGCATAACATAGCGTTTATGA
SPM3HELI_RB_F	TCGAAGTATTGGCGGAATAAGT
SPM3ATLA_LB_R	CCATAACAGGGGCATAAGACA
SPM3ATLA_RB_L	TTTTAGCTTCCCGCTTAGTGA
COPIA78_LB_R	TGAGAGGGGGAGGAGGTATT
COPIA78_RB_F	AATACCTCCTCCCCCTCTCA
COPIA78(1kb)_LB_R	CTTGTAATGACCCAAGAAGTTGCTCTATTG
COPIA78(1kb)_RB_F	GGCTTACATTATTCAACTACTAGTGATTACAAGC

Supplementary Table 13 (part2) : primer sequences for PCR validation of new TE insertions

2.2 Résultats du séquençage du génome des epiRIL : identité des ET mobilisés, fréquence et dynamique de mobilisation

A ce jour, les génomes de 53 epiRIL ainsi que d'un individu *ddm1* et de deux individus sauvages ont été séquencés et analysés avec TE-Tracker. Les séquençages ont été effectués sur des ADN extraits de lots de plantes de la génération F9 (au moins 12 plantes par lot), ce qui doit permettre l'identification des insertions communes à toutes les plantes au sein du lot, mais pas celles propres à chacune d'entre elles. Ainsi donc, en séquençant un lot de plantes F9, ce sont les insertions présentes dans la plante mère unique F8 que l'on détecte.

Les listes de nouvelles insertions pour chacune des epiRIL ainsi que pour le mutant *ddm1* que j'ai utilisées pour l'ensemble des analyses rapportées ci-après sont placées dans l'annexe 2. Elles correspondent aux listes obtenues à partir de TE-tracker auxquelles ont été soustraites les insertions également détectées dans les plantes sauvages ainsi que les évènements ambigus (voir Matériels et Méthodes).

2.2.1 Analyses globales de la mobilisation des ET dans les epiRIL et le mutant *ddm1*

Le séquençage du génome du mutant *ddm1* n'a révélé que quatre évènements de transposition faisant intervenir quatre familles d'ET (fig. 2.3), la famille de transposons à ADN de type EN/SPM *ATENSPM3* et trois familles de rétroéléments à LTR, *ATCOPIA21*, *ATCOPIA57* et *ATCOPIA78*. Les familles *ATENSPM3* et *ATCOPIA21* avaient précédemment été décrites comme mobiles dans le mutant *ddm1* (Miura et al. 2001, Tsukahara et al. 2009). En revanche, *ATCOPIA57* n'a jamais été décrit auparavant comme étant mobile. La famille *ATCOPIA78,* qui regroupe les ET également connus sous le nom d'*ONSEN,* a fait l'objet de plusieurs études récentes qui ont montré la mobilité des *ONSEN* chez les mutants affectés dans la voie du RdDM soumis à des conditions de stress chaleur (Ito et al. 2011). Cependant, jamais leur mobilité n'avait pu être mise en évidence dans le mutant *ddm1*. Le faible nombre d'insertions identifiées ici mais également l'absence de nouvelles insertions d'ET connus pour être fortement mobiles dans ce fond génétique mutant (notamment le rétroélément *ATCOPIA93* et le transposon à ADN *VANDAL21*) peuvent paraître contradictoires au regard

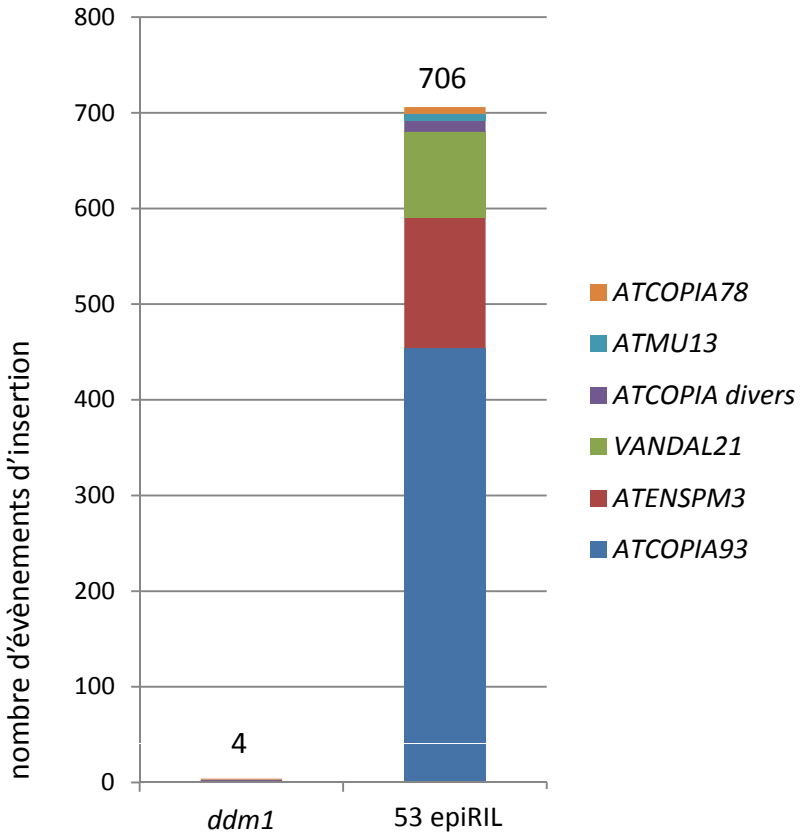


Figure 2.3 : Nombre d'évènements d'insertion par famille d'ET mobilisés dans le mutant *ddm1* et les epiRIL.

de travaux antérieurs (Tsukahara et al. 2009). Cependant, les études précédentes portaient sur des plantes *ddm1* issues de huit autofécondations successives contre quatre seulement dans notre cas, ce qui pourrait expliquer le plus faible nombre d'insertions détecté par nous.

Le séquençage du génome des 53 epiRIL a permis de mettre en évidence un total de 706 évènements d'insertion d'ET (fig. 2.3). Le contraste important entre ce nombre et celui des nouvelles insertions détectées dans le mutant *ddm1* est un premier indicateur de l'activité des ET dans les epiRIL, malgré la restauration de la fonction DDM1 dès la F1 (*ddm1* étant une mutation récessive, Vongs et al. 1993). Cependant, seules 13 familles d'ET sont concernées : neuf familles de rétroéléments à LTR de la superfamille COPIA (*ATCOPIA13*, ATCOPIA20, *ATCOPIA21, ATCOPIA31, ATCOPIA51, ATCOPIA63, ATCOPIA78, ATCOPIA93 et ATRE1*), deux familles de transposons à ADN, l'une de type Mu (*VANDAL21* et *ATMU*) et l'autre de type EN/SPM (*ATENSPM3*). Le nombre de familles mobilisées est donc faible au regard du nombre total de familles que compte le génome d'Arabidopsis (>300) (Buisine et al. 2008, Ahmed et al. 2011) et surtout en comparaison avec la quantité de famille réactivées transcriptionnellement suite à la perte de méthylation induite par la mutation *ddm1* (Lippman et al. 2004, Zemach et al. 2013). Plusieurs hypothèses permettent d'expliquer cet apparent paradoxe. La plus triviale est que, bien qu'exprimée, une importante proportion des ET est constituée d'éléments dégénérés, présentant des ORF mutées, tronquées voire absentes et sont, par conséquent, incapables de transposer. Au laboratoire, nous avons ainsi déterminé, sur la base de plusieurs critères de séquence, que plus de la moitié des familles d'ET du génome d'Arabidopsis était composée de copies sans capacité de transposition (données non montrées). La deuxième hypothèse est l'existence de mécanismes contrôlant l'activité des ET à un niveau post-transcriptionnel. En effet la répression post-transcriptionnelle voire post- traductionnelle a été mise en évidence pour plusieurs ET. On peut citer notamment l'effet dose décrit pour les transposons à ADN *Ac* du maïs qui arrêtent de transposer malgré la production de leur transposase (Scofield et al. 1993, Kunze et al. 1993) ou encore la régulation des rétroéléments à LTR de type COPIA Ty1 chez *S. cerevisae* par des protéines codées par le génome hôte qui seraient impliquées dans la dégradation de l'ADNc (Lesage and Todeschini 2005). Enfin, un des aspects cruciaux est la régulation tissu-spécifique de l'expression des ET. De par notre schéma expérimental nous ne nous intéressons qu'aux évènements d'insertion qui ont été transmis au travers des générations,

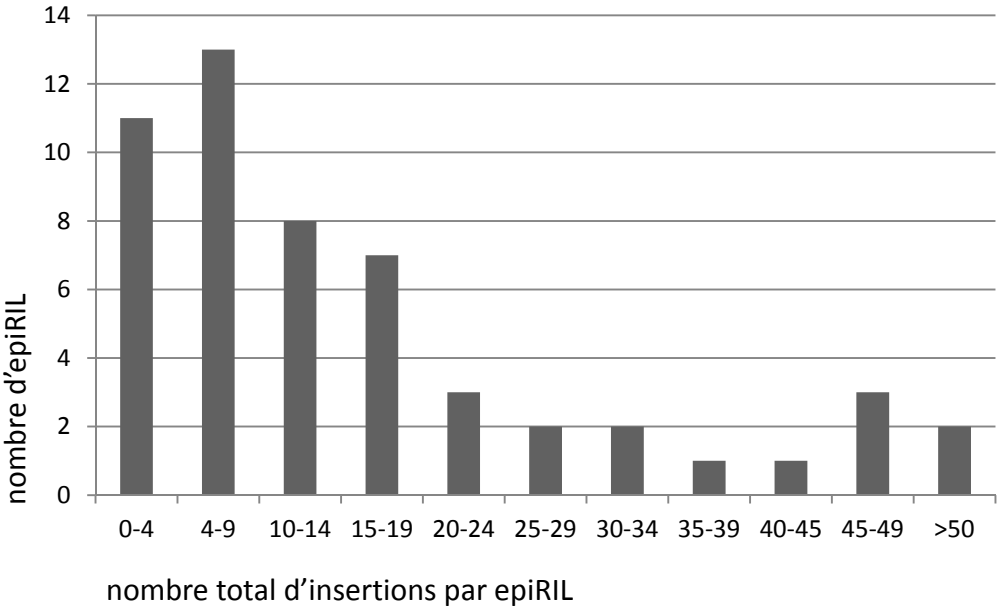


Figure 2.4 : Distribution du nombre de nouvelles insertions dans les epiRIL.

ce qui implique que les insertions ont eu lieu dans des cellules qui « contribuent » à la génération suivante. Or, les analyses transcriptionnelles dont nous disposons ont été réalisées sur plantules entières majoritairement composées de cellules somatiques. Il est possible que certains ET transcriptionnellement actifs dans les tissus somatiques soient réprimés passivement (absence de facteurs de transcription) ou activement (via le TGS ou le PTGS par exemple) dans certains tissus, notamment dans les organes reproducteurs. A ce sujet, il faut noter que, comme décrit dans l'introduction, la régulation des séquences répétées semble différer dans les gamètes et les tissus somatiques.

Il est notable que la majorité des familles présentant des éléments mobiles appartiennent à la superfamille de rétroéléments à LTR COPIA qui s'est amplifiée et diversifiée récemment dans le génome d'Arabidopsis (Peterson-Burch et al. 2004).

Nos résultats montrent également que le nombre d'insertions diffère dramatiquement entre les différentes familles pour lesquelles nous avons mis en évidence une mobilisation. De fait, la très vaste majorité (>95%) des évènements de transposition identifiés est due à la mobilisation d'ET appartenant à trois familles seulement (*ATCOPIA93, ATENSPM3* et *VANDAL21*). Cette disparité est particulièrement grande pour les 9 familles de LTR-COPIA mobiles puisque seul *ATCOPIA93* présente une fréquence de mobilisation élevée, contribuant à 65% de l'ensemble des évènements de transposition détectés dans les epiRIL. Deux raisons peuvent être évoquées ici : (i) le nombre de copies mobiles par famille est très variable, (ii) certaines familles présentent des copies extrêmement mobiles. L'identification des locus donneurs permettra de trancher entre ces hypothèses (voir plus loin).

Enfin, le nombre de nouvelles insertions par lignée est hautement variable entre les epiRIL allant de 0 à 87 (fig. 2.4). L'existence de lignées présentant un grand nombre de nouvelles insertions suggère des phénomènes de « burst » qui concernent principalement *ATCOPIA93* (cf tables en annexe 2). On notera cependant que la lignée 122 qui possède le plus de nouvelles insertions (87) le doit presqu'autant à *ATENSPM3* (32 insertions) qu'à *ATCOPIA93* (46 insertions).

Rétroéléments			
famille	donneurs	RT	LTR
ATCOPIA93	AT5TE20395 (évadé)	oui	oui
	AT1TE41580 (attrapé)	oui	oui
ATCOPIA78	AT1TE59755	oui	oui
	AT3TE89830	oui	oui
	AT5TE15240	oui	oui
	AT3TE92525	oui	oui
ATCOPIA21	AT5TE65370	oui	oui
ATCOPIA20	AT2TE13385	oui	oui
ATCOPIA31	AT1TE38210	oui	oui
ATCOPIA63	AT1TE57025	non	oui
	AT5TE33540	non	oui
ATCOPIA13	AT2TE23850+AT2TE23855	oui	oui
ATRE1	AT1TE72060	oui	oui
ATCOPIA51	AT1TE36030+AT1TE36035	oui	oui
Transposons à ADN			
famille	donneurs	Transposase	TIR
ATENSPM3	AT2TE20205	?	oui
	AT3TE62120+AT3TE62130+ AT3TE62100 (composite1)	non	oui
	AT4TE18510+AT4TE18500+ AT4TE18505 (composite2)	non	oui
VANDAL21	AT2TE42810	oui	oui
ATMU13	AT4TE23205+AT4TE23185 +AT4TE23200+AT4TE23195 +AT4TE23190	oui	oui

Figure 2.5 : Copies donneuses identifiées par TE-tracker.

2.2.2 Identification, caractérisation et propriétés des locus donneurs

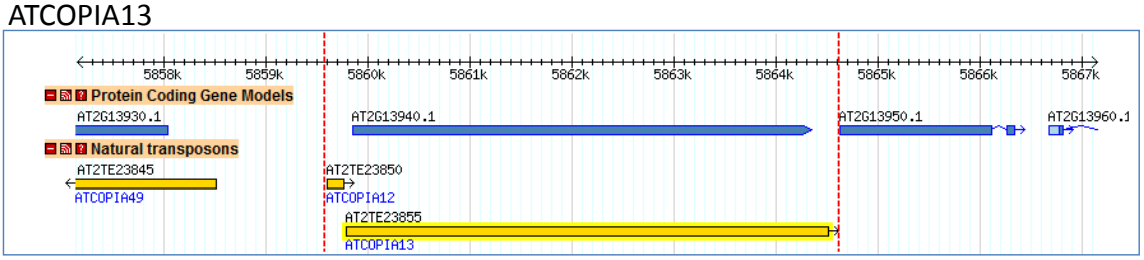
L'approche par séquençage de banques mate-pair associée au module d'attribution de score de TE-tracker permet de déterminer précisément le locus donneur pour la majorité des insertions détectées. Le nombre et l'identité des donneurs mobilisés dans les epiRIL sont résumés dans la table fig. 2.5.

Au total, nous avons identifié 19 locus donneurs avec une bonne confiance dans les epiRIL (score élevé et /ou validation par PCR-reséquençage) et leur nombre varie de 1 à 4 selon les familles. On peut donc conclure ici que les différences drastiques de fréquence de mobilisation observées entre familles ne sont pas dues à des variations majeures du nombre de copies mobilisées.

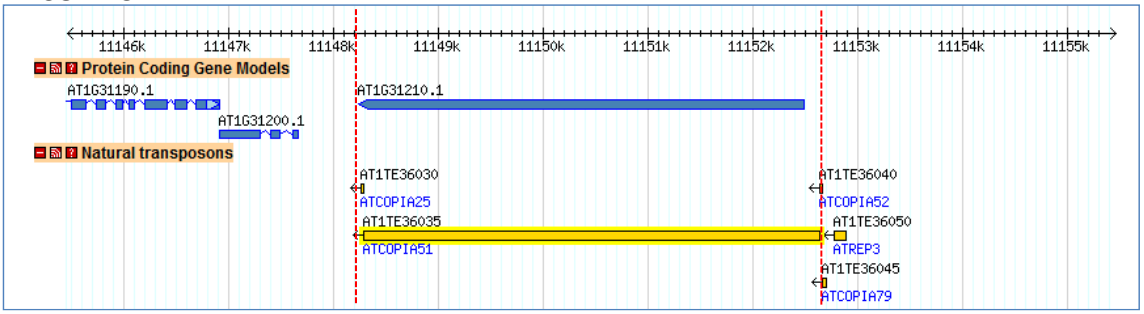
Parmi les différents locus donneurs identifiés, certains couvrent plusieurs annotations d'ET (d'après l'annotation de Buisine et al. 2008) et/ou sont plus étendus que celles-ci. Ces unités « composites » détectées comme mobiles et validées par PCR sont représentées fig. 2.6. J'ai pu identifier des structures de LTR et TIR pour les deux composites de rétroéléments à LTR et les trois composites de transposons à ADN respectivement en réalisant des BLAST des extrémités. Enfin, les deux composites *ATENSPM3* contiennent chacun un gène (*AT3G43340* et *AT4G07526)* qui n'est pas annoté comme un gène d'ET et des analyses de BLASTp et PSI-BLAST n'ont pas permis de mettre en évidence un quelconque lien entre les protéines putatives codées par ces gènes et des protéines caractéristiques d'ET connus. Cependant, très peu d'informations sur ces gènes sont disponibles et si la protéine putative codée par *AT3G43340* présente un domaine pseudouridine synthase, celle codée par *AT4G07526* ne présente aucun domaine ni fonction connus.

Afin de déterminer l'autonomie potentielle des différents locus donneurs, je me suis appuyée sur les résultats d'une étude réalisée au laboratoire basée sur la détection de RT et LTR (pour les rétroéléments) ou de transposase et TIR (pour les transposons à ADN) chez tous les ET annotés du génome d'Arabidopsis (fig. 2.5). Parmi les 14 LTR-rétroéléments identifiés, tous présentent des LTR de séquences quasi-identiques et 12 présentent une ORF codant une RT à priori intacte. Les deux autres copies donneuses appartiennent à la famille *ATCOPIA63*, et leur mobilisation résulte vraisemblablement de l'utilisation d'une RT codée soit par un autre élément de la même famille (au moins un *ATCOPIA63* code une RT intacte

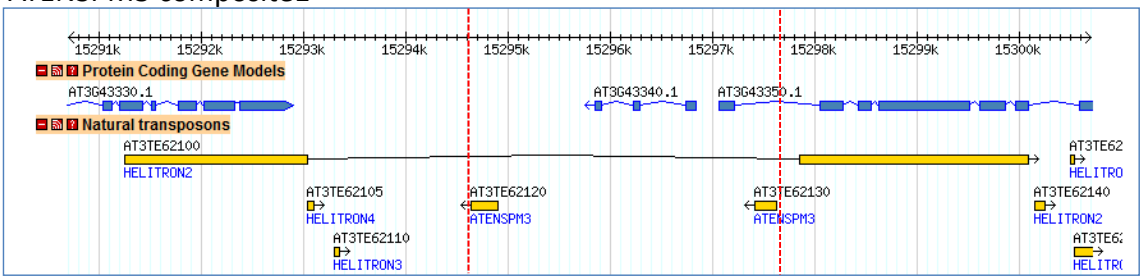
ATCOPIA13



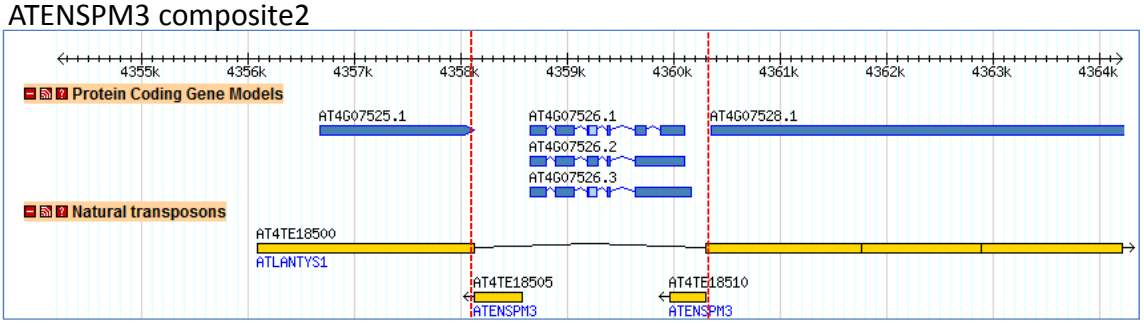
ATCOPIA51



ATENSPM3 composite1



ATENSPM3 composite2



ATMU13

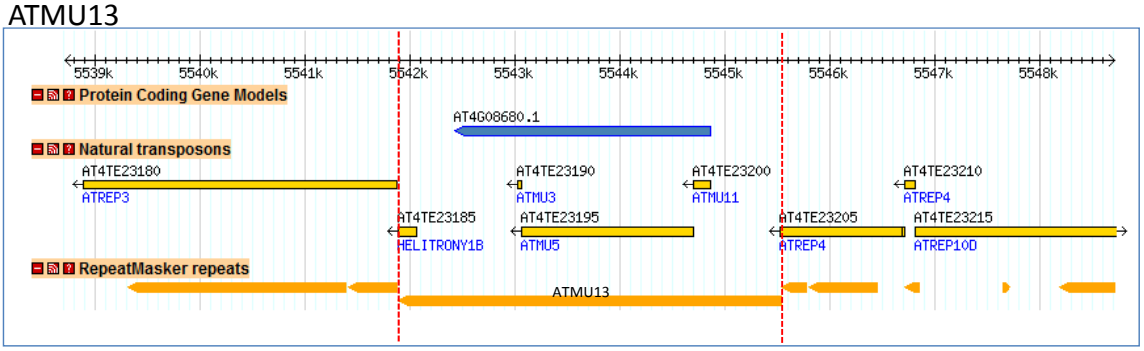


Figure 2.6 : Vues Gbrowse des cinq locus donneurs « composites ». Les régions encadrées par des pointillés rouges correspondent aux unités mobiles.

Remarque : L'annotation établie par « RepeatMasker » a été rajoutée pour le dernier composite car elle correspond exactement à la séquence mobilisée. Pour les autres composites, cette annotation n'apportait pas plus d'information que « Natural transposon ».

selon nos analyses), soit par celle d'un ET appartenant à une autre famille LTR-COPIA. Pour les transposons à ADN, tous les donneurs présentent des TIR mais seuls les locus donneurs de *VANDAL21* et *ATMU13* présentent une ORF continue. Le donneur *ATENSPM3* (locus *AT2TE20205)* présente lui une ORF interrompue ce qui suggère la présence d'introns. Cette observation est en accord avec les travaux effectués sur les éléments Spm du maïs qui avaient mis en évidence un épissage alternatif donnant lieu à plusieurs protéines (Masson et al. 1989). Enfin, il est probable que les deux composites *ATENSPM3* soient mobilisés en trans.

D'autre part, comme cela est décrit dans le manuscrit présenté ci-dessus, certaines nouvelles insertions de LTR-COPIA sont attribuées à deux donneurs avec un score élevé. Une analyse fine des lectures correspondant à certaines de ces insertions a permis de mettre en évidence leur structure chimérique (cf. figure supplémentaire 4 du manuscrit Gilly et al.). Ce résultat n'est pas inattendu car nous savons que lors de la rétrotransposition des rétroéléments à LTR deux ARN correspondant à des locus distincts peuvent être encapsidés ensemble et la RT peut changer de matrice lors de la transcription inverse.

Dans un premier temps, j'ai comparé la liste des locus donneurs dans les epiRIL avec la liste des copies potentiellement mobiles pour chacune des familles mobilisées en me focalisant sur les copies potentiellement autonomes (d'après les analyses réalisées au laboratoire non montrées). Pour les rétroéléments qui présentaient des erreurs d'annotation (*ATCOPIA13* et *ATCOPIA51*) j'ai étendu la recherche de LTR en dehors de l'annotation pour la détermination des locus potentiellement mobiles (fig. 2.7). Pour la majorité des familles, il existe plus d'éléments potentiellement mobiles autonomes que ceux qui transposent dans les epiRIL et les raisons de cette différence restent énigmatiques. Le cas le plus frappant est celui d'*ATCOPIA93*, qui présente trois copies potentiellement mobiles dont deux d'entre elles sont identiques à plus de 99% et sont connues sous les noms d'*EVADE* et *ATTRAPE*. Bien que ces deux copies soient mobiles dans les epiRIL, *ATTRAPE* n'a été attribué qu'à un seul évènement de transposition contre plusieurs centaines pour *EVADE*. C'est d'ailleurs la première fois que la mobilisation d'*ATTRAPE* a été détectée car l'étude de la mobilisation des éléments *ATCOPIA93* dans divers mutants n'avait mis en évidence que la mobilisation d'*EVADE* (Mirouze et al. 2009).

Rétroéléments		
famille	Copies mobilisées autonomes (LTR+RT)	Copies potentiellement mobiles et autonomes (RT+LTR)
ATCOPIA93	2	3
ATCOPIA78	4	8
ATCOPIA21	1	1
ATCOPIA20	1	1
ATCOPIA31	1	2
ATCOPIA63	0	1
ATCOPIA13	1	4
ATRE1	1	2
ATCOPIA51	1	4
Transposons à ADN		
famille	Copies mobilisées autonomes (transposase+TIR)	Copies potentiellement mobiles et autonomes (transposase+TIR)
VANDAL21	1	6
ATENSPM3	?	?
ATMU13	1	?

Figure 2.7 : Nombre de copies mobiles autonomes et de copies apparentées potentiellement mobiles et autonomes.

	FAMILLE	TOTAL	UNIQUES	PARTAGEES
Rétroélements	ATCOPIA93	454	452	2
	ATCOPIA78	7	0	7
	ATCOPIA63	4	4	0
	ATCOPIA21	2	2	0
	ATCOPIA13	2	2	0
	ATRE1	1	1	0
	ATCOPIA20	1	0	1
	ATCOPIA31	1	1	0
	ATCOPIA51	1	1	0
Transposons à ADN	ATENSPM3	136	125	11
	VANDAL21	90	89	1
	ATMU13	8	8	0
TOTAL		707	685	22

Nombre faible d'insertions, toutes partagées
Nombre élevé d'insertions, la plupart uniques
Nombre faible d'insertions, toutes uniques

Figure 2.8 : Dynamique de mobilisation des différentes familles d'ET dans les epiRIL.

2.2.3 Dynamique de transposition des ET dans les epiRIL

Afin d'estimer plus finement la dynamique de mobilisation pour les différentes familles d'ET, j'ai déterminé pour chacune d'entre elles le nombre d'insertions partagées (qui ont donc eu lieu soit dans le mutant *ddm1* soit dans la plante F1 unique à partir de laquelle a été réalisé le croisement en retour à l'origine de chacune des epiRIL), ainsi que le nombre d'insertions uniques (fig. 2.8). Comme attendu au vu du faible nombre d'insertions détectées dans le mutant *ddm1*, la très grande majorité des insertions identifiées dans les epiRIL est spécifique à chacune des lignées. Ceci permet de conclure définitivement que les ET sont toujours actifs dans les epiRIL après ségrégation de la mutation *ddm1*.

Néanmoins, la dynamique de mobilisation diffère selon les familles. On peut distinguer : (1) les familles d'ET présentant un petit nombre d'insertions toutes partagées (*ATCOPIA78* et *ATCOPIA20*) indiquant une mobilisation exclusivement dans *ddm1* ou la F1, (2) les familles présentant un nombre élevé d'insertions dont une majorité sont non partagées entre epiRIL (*ATCOPIA93*, *ATENSPM3* et *VANDAL21*) ce qui témoigne d'une mobilisation dans *ddm1* et /ou la F1 et dans les épiRIL, (3) les familles d'ET ayant un faible nombre d'insertions toutes non partagées, révélant une mobilisation exclusivement dans les epiRIL. Parmi elles, certaines présentent des nouvelles insertions dans plusieurs epiRIL (*ATMU13*, *ATCOPIA21* et *ATCOPIA12*) et d'autres dans une seule (*ATCOPIA63*, *ATCOPIA51*, *ATCOPIA31* et *ATRE1*).

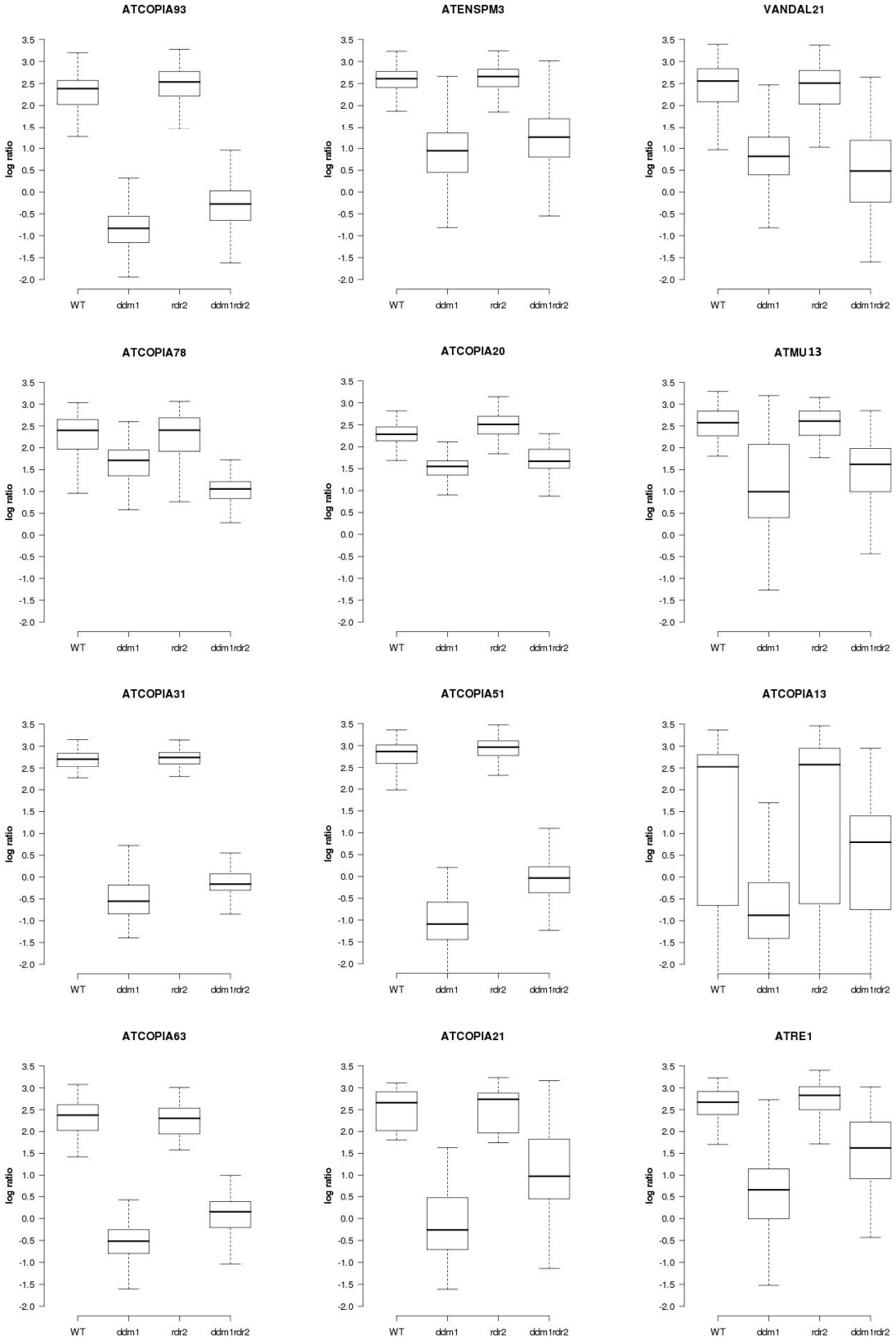


Figure 2.9 : Distribution chez le sauvage et les mutants *ddm1*, *rdr2* et *ddm1rdr2,* du niveau de méthylation (MeDIP-chip) rapporté par les sondes couvrant les séquences des différentes familles d'ET pour lesquelles de nouvelles insertions ont été détectées.

2.2.4 Etude comparative des différentes familles comptant des ET mobiles dans les epiRIL

Afin de mieux comprendre les différences de fréquence et de dynamique de mobilisation des ET appartenant aux différentes familles dans les epiRIL, j'ai étudié l'implication des différentes voies de la méthylation de l'ADN (dépendante de DDM1 et RdDM) dans le contrôle de ces ET. Pour ce faire, j'ai réalisé des expériences d'immunoprécipitation des cytosines méthylées suivies d'hybridation sur puce à ADN (MeDIP-chip) chez le sauvage et le mutant *ddm1* ainsi que chez le mutant *rdr2* et le double mutant *ddm1rdr2*. J'ai également tiré avantage de données de séquençage profond de petits ARN dans ces mêmes fonds génétiques ainsi que du transcriptome de *ddm1* (ARN-chip) disponibles au laboratoire.

Les analyses présentées ci-dessous ont été effectuées à l'échelle de la famille d'ET car les techniques de MeDIP-chip et d'ARN-chip ne permettent pas, dans la plupart des cas, d'obtenir des données spécifiques pour chaque copie au sein d'une famille à cause des trop grandes similarités de séquence.

Comme attendu, toutes les familles comptant des ET mobiles dans les epiRIL présentent une perte de méthylation dans le mutant *ddm1* en comparaison avec le sauvage (fig. 2.9). Cependant, l'intensité de cette perte varie entre les familles. Notamment, les familles *ATCOPIA78* et *ATCOPIA20* qui n'ont engendré que des insertions partagées présentent une perte de méthylation moindre dans le mutant *ddm1* que toutes les autres familles mobilisées. Cette observation indique donc l'existence de mécanismes capables d'assurer le maintien de la méthylation en l'absence de l'action de DDM1 pour ces familles. Il faut néanmoins considérer ici que ce type d'analyses ne permet pas de déterminer si une perte partielle de méthylation observée à l'échelle d'une famille d'ET est due à la perte complète au niveau de certains membres de la famille alors que les autres membres gardent un niveau élevé de méthylation ou si cela traduit une perte partielle concertée au niveau de tous les membres.

Les résultats de l'analyse de méthylation dans le mutant *rdr2* montrent que la compromission du RdDM seul n'engendre pas de perte de méthylation globale pour aucune des familles. Il faut noter ici que le MeDIP-chip ne permet pas de mettre en évidence une perte subtile de méthylation. Or, la voie du RdDM est responsable du maintien de la méthylation en contexte CHH qui est de faible intensité (40% maximum des cytosines sont

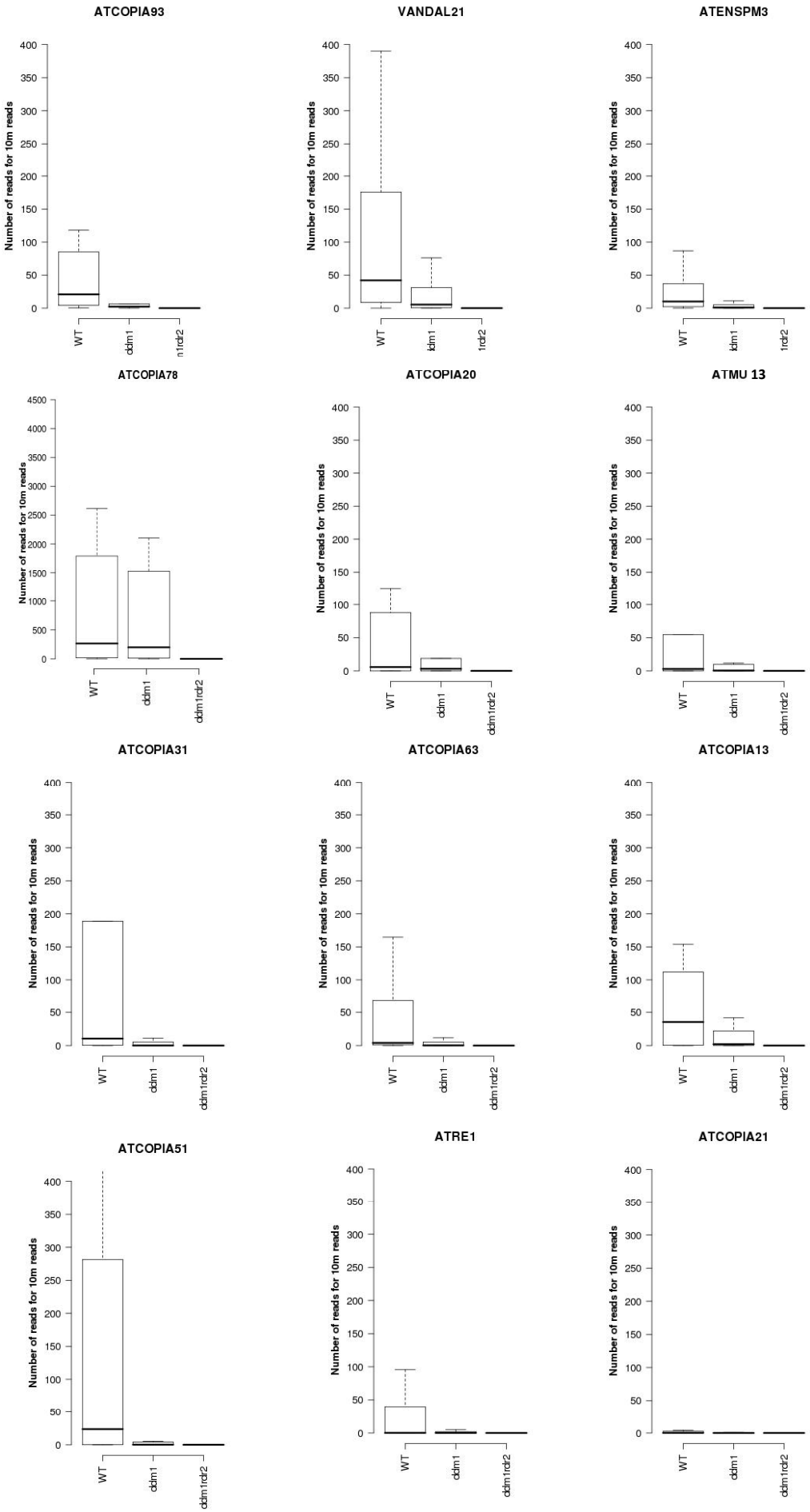


Figure 2.10 : Abondance chez le sauvage et les mutants *ddm1* et *ddm1rdr2* des siARN de 24nt ciblant les différentes familles d'ET pour lesquelles de nouvelles insertions ont été détectées. A noter que l'échelle des ordonnées est différente pour *ATCOPIA78.*

méthylées à une position donnée). Cependant, l'importance du RdDM dans le contrôle de la méthylation de la famille *ATCOPIA78* est révélée dans le double mutant *ddm1rdr2* qui présente, pour ces séquences, une perte de méthylation plus importante que dans le simple mutant *ddm1*. Pour cette famille, il est donc clair que c'est l'action du RdDM qui lui permet de garder un certain niveau de méthylation dans le mutant *ddm1*. En revanche, *ATCOPIA20* ne présente pas une perte de méthylation plus importante dans le double mutant *ddm1rdr2*, ce qui suggère que le maintien d'un niveau de méthylation relativement élevé dans *ddm1* pour les membres de cette famille n'est pas dû à l'action du RdDM. Enfin, pour la majorité des familles, le niveau de méthylation dans le double mutant *ddm1rdr2* bien que très faible, apparait légèrement supérieur à celui mesuré dans le simple mutant *ddm1*. Ceci est probablement dû au fait que le double mutant *ddm1rdr2* n'a subi que deux cycles d'autofécondation successives alors que le simple mutant *ddm1* en a subi quatre, or l'hypométhylation induite par la mutation *ddm1* a tendance à s'aggraver au cours des générations.

Les données issues du séquençage profond des petits ARN sont en accord avec celles concernant la méthylation de l'ADN (fig 2.10). *ATCOPIA78* est dix fois plus ciblée en moyenne par les siARN de 24nt que les autres familles considérées ici. De plus, elle est la seule famille à conserver quasi totalement ses siARN dans le mutant *ddm1* alors que les autres familles les perdent drastiquement voire totalement. En revanche, et comme attendu, on observe une perte totale des siARN de 24nt ciblant les *ATCOPIA78* dans le double mutant *ddm1rdr2*. Ces observations expliquent donc pourquoi *ATCOPIA78* est la seule famille à présenter une perte de méthylation plus importante dans *ddm1rdr2* que dans *ddm1* et pourquoi la perte de méthylation dans *ddm1* est aussi forte pour les autres familles.

Enfin, l'analyse différentielle d'expression par ARN-chip dans le mutant *ddm1* révèle des comportements contrastés entre les différentes familles comptant des d'ET mobilisés (fig. 2.11). Aucune corrélation ne peut être mise en évidence entre le niveau d'activation transcriptionnelle de ces familles dans *ddm1* et les fréquences et dynamiques de mobilisation. Une fois encore, il faut prendre en considération que les expériences d'ARN-chip sont effectuées sur plantules entières, or c'est l'expression dans les tissus qui donneront lieu à la génération suivante qui importe. Si par exemple les évènements de

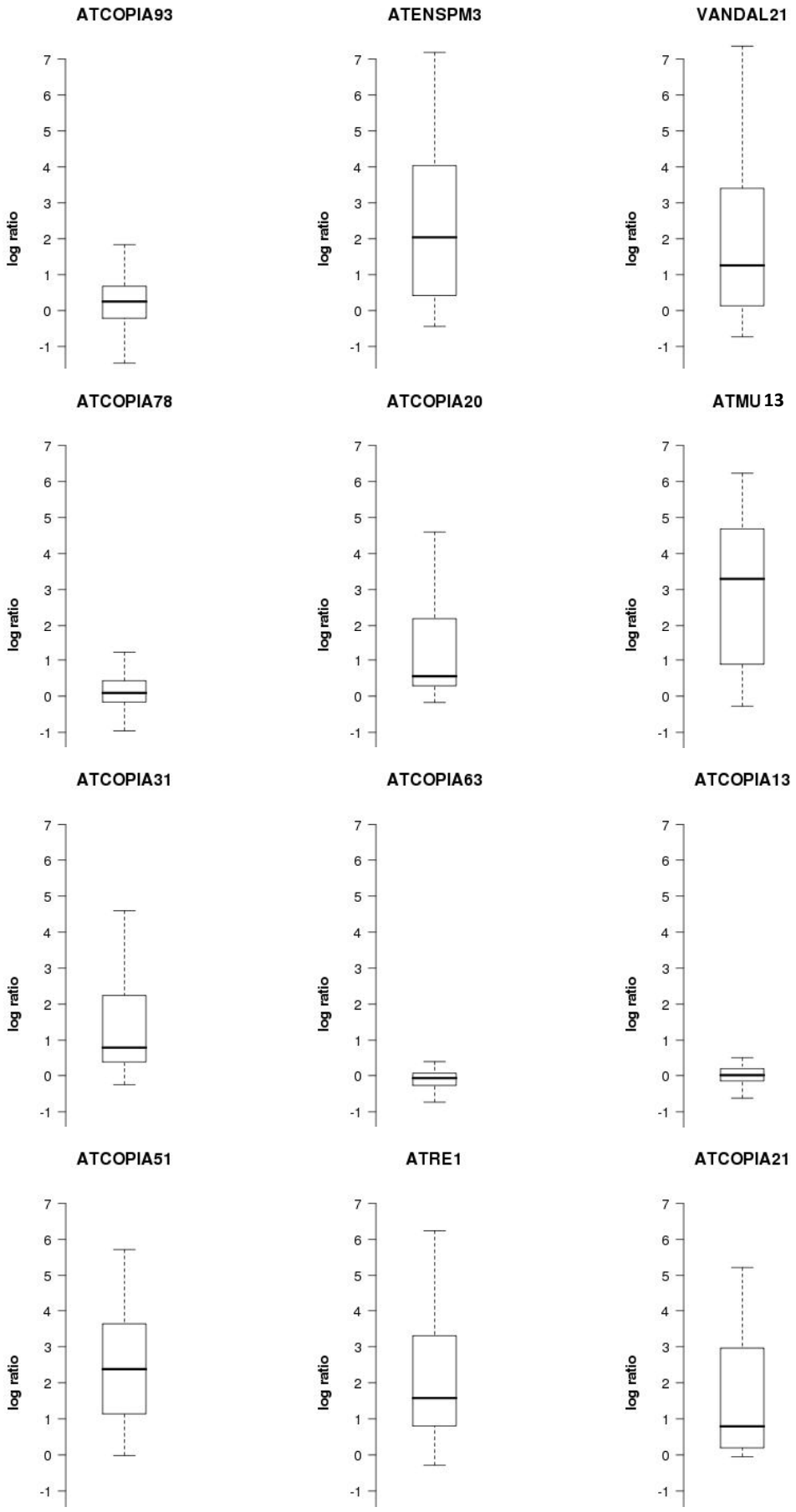


Figure 2.11 : Distribution du niveau d'expression différentielle (ARN-chip) rapportée par les sondes couvrant les séquences des différentes familles d'ET pour lesquelles des nouvelles insertions ont été détectées entre *ddm1* et sauvage.

mobilisation que nous avons détectés sont dus à une activation spécifique dans les gamètes ou les inflorescences, elle ne sera pas détectée ici.

Dans l'ensemble, les analyses de la méthylation de l'ADN, de l'expression et de l'abondance en siARN de 24nt dans les différents fonds mutants ne permettent pas d'expliquer les différences drastiques de fréquence et dynamique de mobilisation observées entre les différentes familles d'ET. Néanmoins, les familles responsables d'insertions toutes partagées ne présentent qu'une perte modérée de méthylation dans le mutant *ddm1* en comparaison avec les autres familles. Pour le cas d'*ATCOPIA78*, cela s'explique clairement par un contrôle de la méthylation par la voie du RdDM. Cette action forte du RdDM sur cette famille pourrait également expliquer le fait que la mobilisation ne soit possible que dans *ddm1* ou dans la F1. De fait, un ciblage fort par la machinerie du RdDM chez le sauvage et qui perdure chez le mutant *ddm1* est la signature caractéristique du phénomène de reméthylation décrit dans l'introduction. Il est donc probable que les copies résidentes d'*ATCOPIA78* soient efficacement reméthylées donc réinactivées après restauration de la fonction DDM1, mais également que les nouvelles insertions soient ciblées par ce mécanisme et donc rapidement réprimées également. Ce point sera développé plus en détail dans la partie 3.

Enfin, s'il est vrai que les familles d'ET qui comptent les copies les plus mobiles montrent une forte perte de méthylation dans *ddm1*, cela est vrai aussi pour la majorité de celles présentant un faible nombre d'insertions toutes non partagées. Concernant ces dernières, une mobilisation restreinte à un très faible nombre de lignées pourrait être due à un défaut occasionnel de reméthylation dans ces lignées spécifiquement. De fait, le mécanisme de reméthylation présente une pénétrance variable selon les séquences. Afin de tester cette hypothèse, j'ai évalué l'état de méthylation d'un de ces éléments, *AT2TE23855 (ATCOPIA13),* dans les epiRIL à partir des données de MeDIP-chip disponibles au laboratoire. J'ai dû me restreindre à cet élément car c'est le seul qui présente des sondes suffisamment spécifiques sur la puce à ADN utilisée pour le MeDIP-chip. Bien que n'ayant transposé que dans deux epiRIL, *AT2TE23855* est dans un état hypométhylé dans plusieurs autres lignées, et cela dans les proportions attendues dans le cas de la ségrégation mendélienne de locus stablement hypométhylés (dans 25% des epiRIL). L'absence de transposition dans ces dernières lignées reste donc énigmatique mais pourrait être liée à la ségrégation de facteurs agissant en trans.

	ATCOPIA93 (AT5TE20395)			VANDAL21 (AT2TE42810)			ATENSPM3 (AT2TE20205)		
epiRIL	ORIGINE PARENTALE	INS. PARTAGEE	INS. UNIQUE	ORIGINE PARENTALE	INS. PARTAGEE	INS. UNIQUE	ORIGINE PARENTALE	INS. PARTAGEE	INS. UNIQUE
8	ND		2	WT		3	WT	2	
14	WT			DDM1		1	DDM1		
24	WT			WT			WT		
36	WT		4	DDM1		5	WT		
46	WT			WT			WT	1	
52	WT			WT	1	1	WT	1	1
53	WT			WT			WT		
55	DDM1	1	24	WT		5	DDM1		11
60	WT	1	1	DDM1	1	4	WT		3
69	WT			WT		4	WT		3
70	ND		3	WT		2	WT		
92	DDM1	1	5	WT	1	5	DDM1		
95	DDM1	1	22	WT	1		WT		8
98	DDM1	1	8	DDM1		5	DDM1	2	
99	WT			WT			WT		
108	WT		2	WT	1		WT	1	
118	WT			WT			WT		
122	DDM1	1	46	ND	1	1	DDM1		32
137	WT			ND			WT		3
144	WT			WT		3	WT		2
150	WT		1	WT		1	WT		
159	WT		22	WT		2	WT		1
166	WT			WT			WT		
193	DDM1	1	23	ND		5	DDM1		
218	WT			DDM1	1		DDM1		
222	WT			WT			WT	1	
252	WT			WT		2	WT		1
260	ND			WT		8	WT		
276	WT	1	3	ND		5	WT	1	
277	WT			WT			WT		4
340	WT			DDM1		2	WT	1	
344	WT			WT			WT		
356	DDM1		40	WT			WT	3	
362	ND	2	37	ND		2	ND		
363	ND		12	WT			WT	1	
366	DDM1		15	WT			WT	1	5
368	WT			WT			WT		
371	WT	1	1	WT	1	7	WT	1	
375	WT	1	9	WT			WT		
377	ND		41	ND	1	2	ND		
393	DDM1		10	WT			WT		
408	DDM1		7	WT			WT		8
421	DDM1		18	DDM1		3	DDM1		
425	DDM1		42	WT		3	WT	1	
437	DDM1		10	WT			WT		
439	ND		3	DDM1		2	WT		4
454	DDM1		24	WT	1	1	WT	2	
458	WT	1	5	DDM1		2	WT		
466	WT			WT			DDM1		
467	WT	1	1	WT			WT		
480	WT			WT			WT		
503	WT	1	11	WT		2	WT		3
559	WT			WT		1	WT		

Figure 2.12 : Tableau récapitulatif des insertions d'*AT5TE20395*, *AT2TE42810* et *AT2TE20205* et origine parentale du locus donneur dans les epiRIL.

2.2.5 Mobilisation des ET en fonction de l'origine parentale du locus donneur et de la présence d'insertions partagées.

Pour les trois ET potentiellement autonomes les plus mobiles *AT5TE20395* (*ATCOPIA93*), *AT2TE20205* (*ATENSPM3*) et *AT2TE42810* (*VANDAL21*) j'ai cherché à déterminer leur « potentiel de mobilisation », autrement dit à établir dans quelle mesure la présence d'une copie potentiellement active dans les epiRILs (locus donneur hérité du parent *ddm1* et /ou insertions partagées héritées du parent *ddm1* ou de la plante F1) se traduit par la mobilisation de ces ET. Pour ce faire, j'ai mis en relation de façon systématique l'origine parentale du locus donneur et la présence d'insertions partagées avec la présence d'insertions uniques pour chacune des epiRIL. Pour établir l'origine parentale des locus donneurs je me suis appuyée sur les cartes de recombinaison des epiRILs établies au laboratoire (annexe 3 Colome-Tatche et al. 2012). Les résultats de cette analyse sont regroupés figure 2.12.

- *ATCOPIA93*

Il apparaît très clairement que dans le cas d'*ATCOPIA93*, la présence du locus donneur dans un intervalle d'origine *ddm1* (donc théoriquement hypométhylé) et/ou la présence de l'insertion partagée est systématiquement associée à la présence d'insertions uniques. Cependant, des insertions uniques sont également détectées dans des epiRIL ne présentant pas ces caractéristiques. Cette dernière observation pourrait refléter soit une mobilisation dans les premières générations des epiRIL avant fixation à l'état homozygote de la copie *AT5TE20395* d'origine sauvage (fig. 2.13 (1)), soit l'héritage à l'état hémizygote de la copie partagée pendant quelques générations avant élimination par ségrégation (fig. 2.13 (2)). Dans les deux cas, la présence de copies actives a pu engendrer l'apparition de nouvelles insertions uniques. Ces hypothèses pourront être testées au moyen d'études verticales au sein de chaque lignée en y analysant à chaque génération l'état de méthylation du locus donneur et la présence de l'insertion partagée.

- *VANDAL21*

Dans le cas de *VANDAL21* la tendance concernant l'origine parentale du donneur est la même que celle observée pour *ATCOPIA93*, à savoir la présence quasi-systématique (8 cas

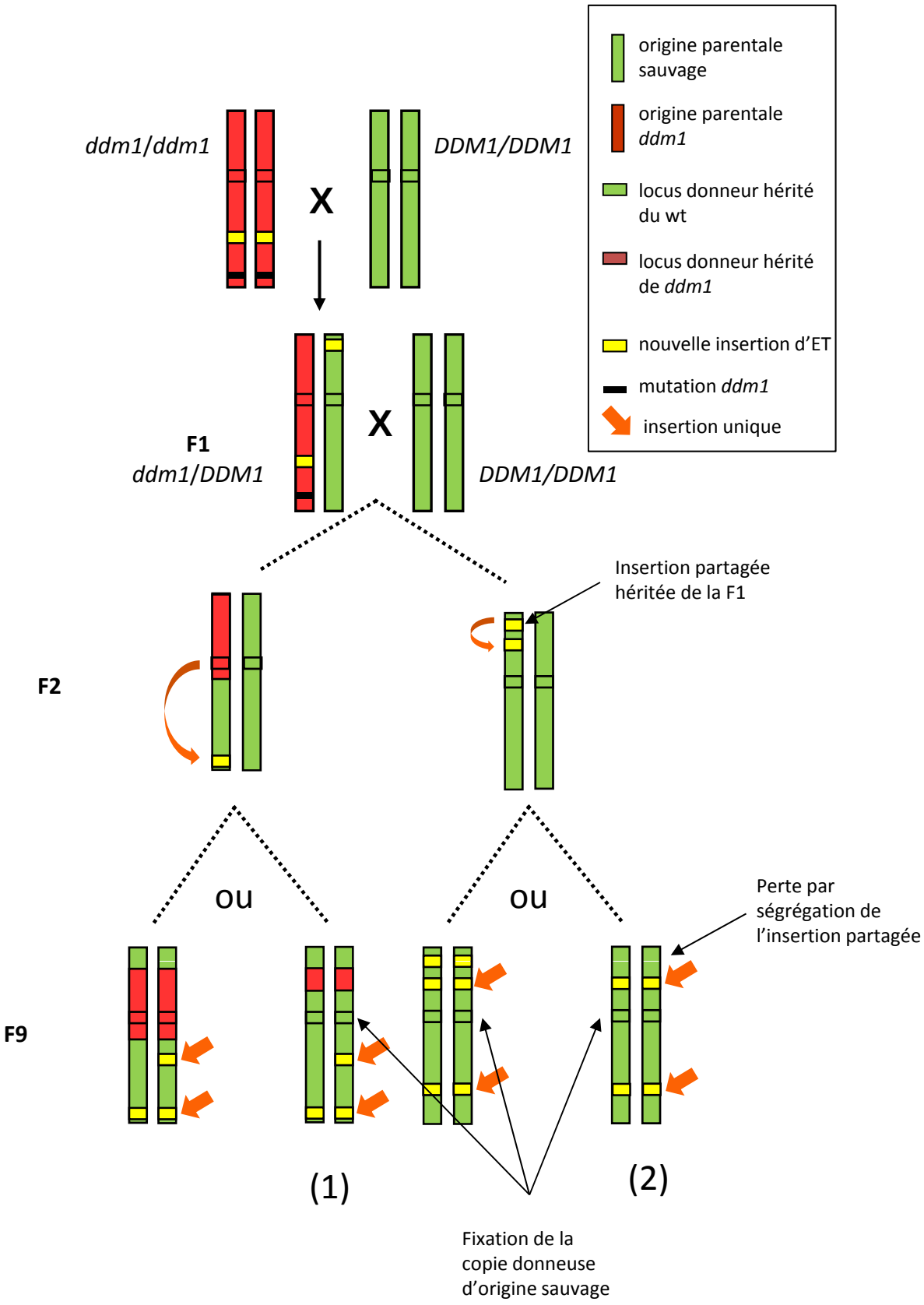


Figure 2.13 : Origines possibles des insertions uniques dans les epiRIL possédant un locus donneur hérité du parent sauvage et aucune insertion partagée (cas 1 et 2).

sur 9) de nouvelles insertions uniques quand le donneur est hérité de *ddm1*. En revanche, il n'en va pas de même pour le cas de la présence de l'insertion partagée qui n'est pas systématiquement associée à celle d'insertions uniques.

- *ATENSPM3*

Enfin, dans le cas d'*ATENSPM3*, la tendance est complètement inversée car l'origine *ddm1* du locus donneur est corrélée négativement avec la présence de nouvelles insertions uniques (seulement deux cas sur neuf). L'hypothèse la plus triviale pouvant expliquer cette observation est directement liée au mode de transposition d'*ATENSPM3* par un mécanisme de couper/coller. De fait, si la mobilisation d'un transposon à ADN est suivie de la réparation du site donneur par « non-homologous end joining » et que la nouvelle insertion n'est pas fixée dans la descendance, alors on a disparition de l'ET dans la lignée. Afin de tester cette hypothèse j'ai réalisé des PCR pour la détection de la présence d'*AT2TE20205* au locus donneur dans les epiRIL (fig. 2.14). *AT2TE20205* est systématiquement absent au locus donneur lorsqu'il a été hérité du parent *ddm1,* ce qui traduit une excision vraisemblablement ancestrale et héritée du parent *ddm1*. Il est également remarquable que la présence d'une ou plusieurs insertions partagées est négativement corrélée avec la présence de nouvelles insertions uniques dans le cas d'*ATENSPM3*. S'il est vrai que la mobilisation des SPM se traduit le plus souvent par un site donneur vide après excision, la mobilisation des insertions partagées dans chaque lignée qui produit des insertions uniques doit souvent être associée avec la « perte » de ces insertions partagées. Cette hypothèse peut être testée en recherchant dans chaque lignée présentant des insertions uniques d'*ATENSPM3* la présence d'empreintes attestant de la présence antérieure des insertions partagées. De fait, j'ai déjà mis en évidence un premier cas dans l'epiRIL 439 qui présente la signature d'excision d'une des insertions partagées (fig. 2.15). Quoiqu'il en soit, cela n'explique pas la raison pour laquelle tant pour *VANDAL21* qu'*ATENSPM3* les insertions partagées sont mobilisées dans certaines lignées mais pas dans d'autres.

En résumé il apparaît donc que la présence d'une ou plusieurs copies potentiellement mobiles d'*AT5TE20395* est systématiquement associée avec sa prolifération dans chacune des epiRIL bien qu'à des fréquences très variées (le nombre d'insertions uniques est très variable selon les lignées). En revanche, ce n'est pas le cas pour *AT2TE20205* et *AT2TE42810*

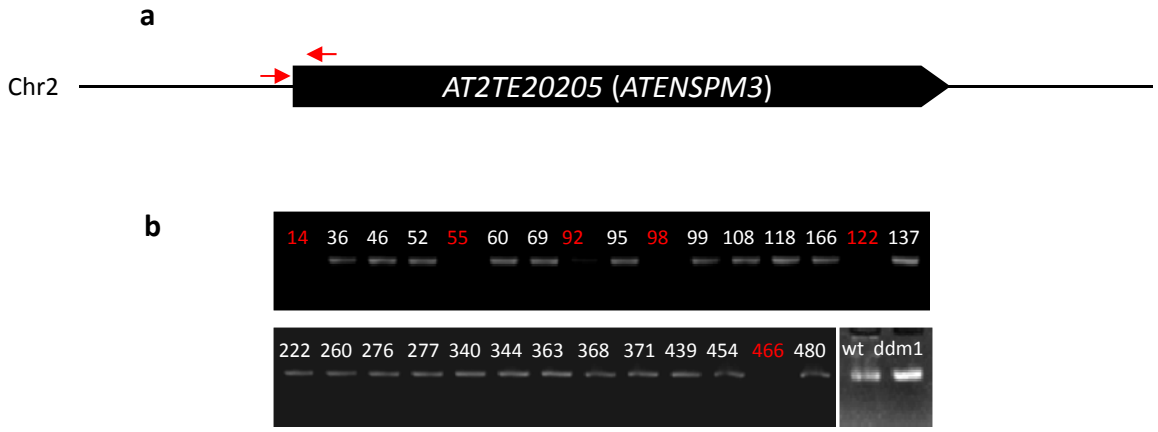
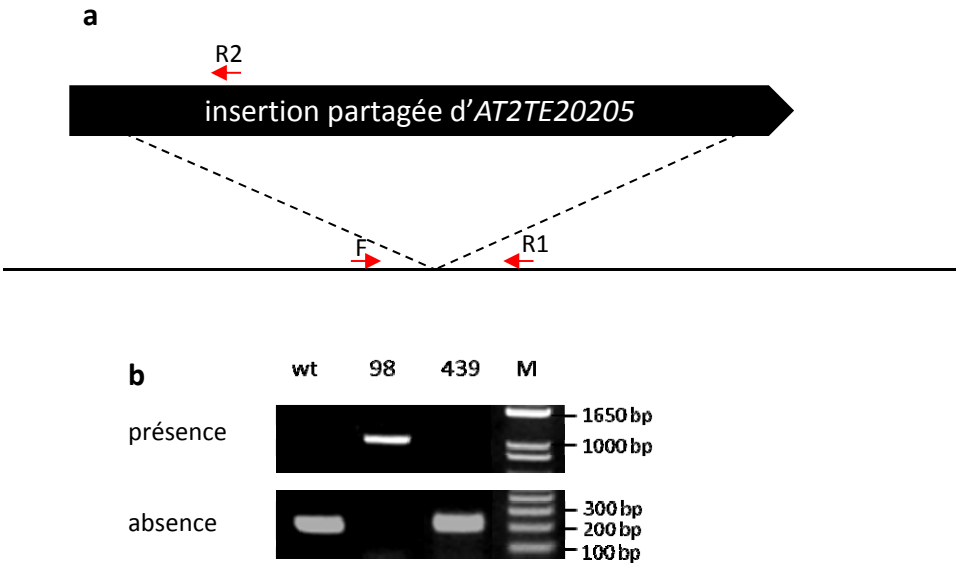


Figure 2.14 : Test de la présence d'*AT2TE20205* (*ATENSPM3*) au locus donneur.
(**a**) Représentation schématique du positionnement des amorces de PCR utilisées pour tester la présence d'*AT2TE20205* au locus donneur (l'échelle n'est pas respectée). (**b**) PCR montrant la présence/absence d'*AT2TE20205* au locus donneur. Les lignées en rouge ont le locus donneur dans un intervalle hérité du parent *ddm1*.



c chr4 : 16482535-16482686

```
ref  CTAGTGCGAATCCAGTCTGTCAGACAATGAGTAAACGAGTGACATGTCATCTCACGCATA
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
439  CTAGTGCGAATCCAGTCTGTCAGACAATGAGTAAACGAGTGACATGTCATCTCACGCATA

ref  CTTATTCCCTGAGAAGA----TAATTATTAACATTAACATCGAATGAATTTAAAAATAAG
     |||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
439  CTTATTCCCTGAGAAGATAATTAATTATTAACATTAACATCGAATGAATTTAAAAATAAG

ref  TATTACAAGTACCAAATAGTATTAAGCAGAATTTAT
     |||||||||||||||||||||||||||||||||||
439  TATTACAAGTACCAAATAGTATTAAGCAGAATTTAT
```

Figure 2.15 : Excision d'une insertion partagée d'*AT2TE20205* (*ATENSPM3*) dans l'epiRIL 439.
(**a**) Représentation schématique du positionnement des amorces PCR utilisées pour montrer l'absence (F+R1) ou la présence (F+R2) d'une insertion partagée d'*AT2TE20205* (l'échelle n'est pas respectée). (**b**) PCR montrant la présence *vs* absence de l'insertion partagée dans le sauvage, l'epiRIL 439 et l'epiRIL 98 qui seule présente l'insertion partagée. (**c**) Résultat du séquençage du produit de PCR correspondant à l'absence de l'insertion dans l'epiRIL 439 aligné sur le génome de référence. Les nucléotides en bleu correspondent à l'empreinte laissée par l'excision d'*AT2TE20205*.

car certaines lignées présentant des insertions partagées de ces ET ne possèdent aucune insertion unique. Cela peut indiquer que ces insertions partagées ne sont plus actives dans ces lignées.

2.2.6 Etude de l'activation en trans des transposons à ADN *ATENSPM3* et *VANDAL21*

L'absence de la mobilisation du locus donneur *AT2TE20205* lorsqu'il est hérité du sauvage alors que de nouvelles insertions, potentiellement actives, sont présentes peut sembler contradictoire avec des travaux effectués sur les éléments Spm du maïs qui mettaient en évidence la réactivation de copies silencieuses (alors appelées cryptiques) lorsque mises en présence de copies actives (Fedoroff 1989). Cette réactivation passait notamment par une étape de déméthylation de la copie cryptique grâce à l'action de la protéine TNPA codée par la copie active. Il est possible que les éléments *ATENSPM3* d'Arabidopsis ne présentent pas cette propriété, ce qui expliquerait l'absence de mobilisation de la copie donneuse héritée du parent sauvage. Afin de tester cette hypothèse, j'ai mesuré par McrBC-qPCR le niveau de méthylation de la copie au site donneur lorsqu'il est hérité du parent sauvage dans des epiRIL présentant ou non des nouvelles insertions d'*AT2TE20205* ainsi que chez la plante sauvage et le mutant *ddm1* (fig. 2.16). Plusieurs epiRIL, bien qu'ayant hérité le locus donneur du parent sauvage, présentent une hypométhylation drastique de son extrémité 5'. Cette hypométhylation n'est observée que pour des epiRIL présentant des nouvelles insertions d'*At2TE20205,* ce qui suggère qu'elle est bien causée par ces nouvelles copies. Cependant, toutes les epiRIL présentant des nouvelles insertions ne montrent pas l'hypométhylation de l'extrémité 5', ce qui peut signifier soit que ce phénomène n'a pas une pénétrance complète soit que, dans ces lignées, les nouvelles insertions ne sont pas/plus actives (voir partie 3). Notons qu'il n'y a pas de lien apparent entre l'hypométhylation du locus donneur et le nombre ou l'ancienneté (unique vs partagée) des nouvelles insertions (fig. 2.12). Quoiqu'il en soit, il semble que cette hypométhylation induite en trans ne se traduise pas par la mobilisation car aucune excision du locus donneur *AT2TE20205* hérité du sauvage n'a été observée.

J'ai étendu cette analyse à *VANDAL21* bien que ce phénomène d'hypométhylation en trans n'est jamais été mis en évidence pour les éléments de type Mu. Les résultats de cette

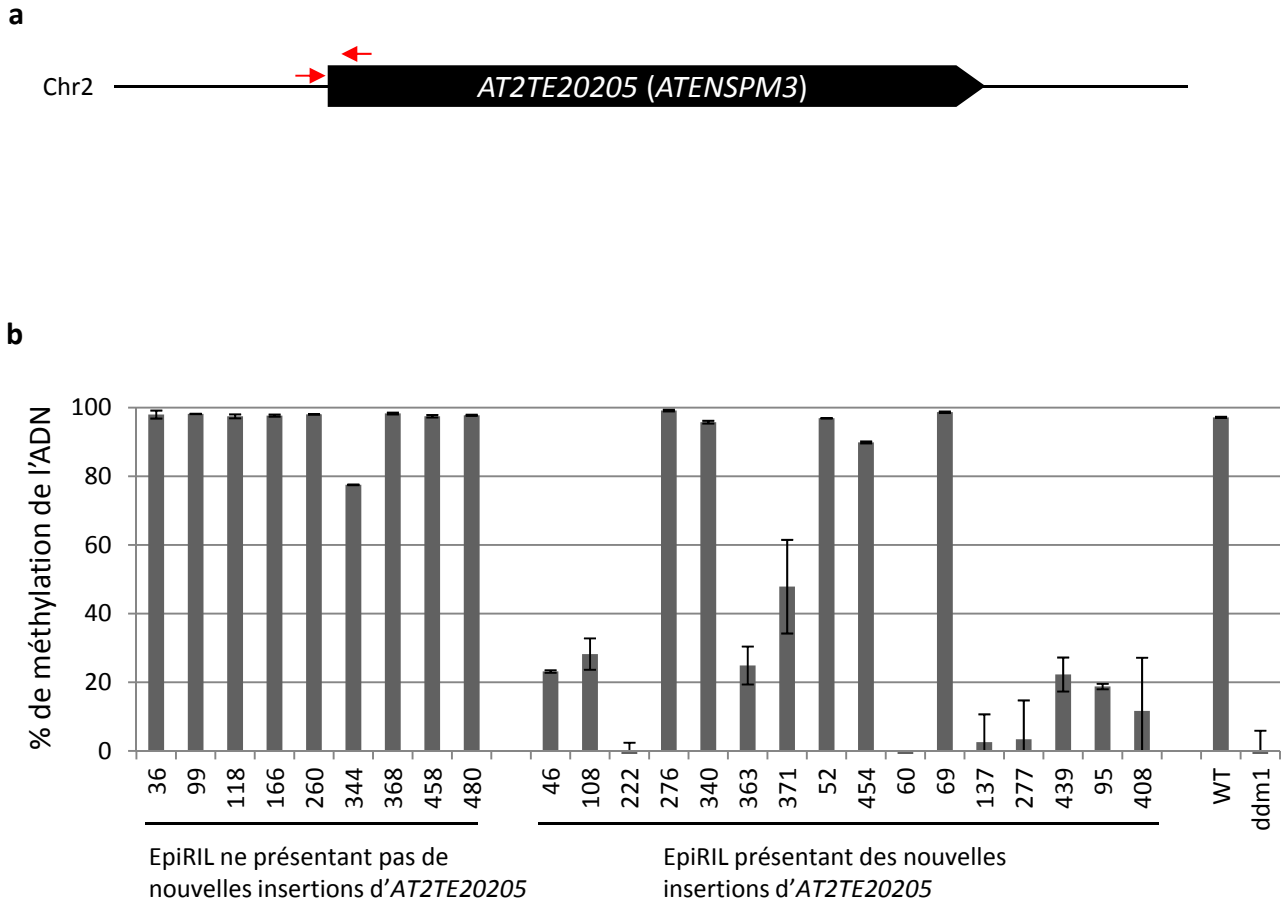


Figure 2.16 : Les nouvelles insertions d'*AT2TE20205* (*ATENSPM3*) induisent l'hypométhylation du locus donneur.

(**a**) Représentation schématique du positionnement des amorces de qPCR utilisées pour mesurer le niveau de méthylation en 5' du locus donneur *AT2TE20205* (les échelles ne sont pas respectées). (**b**) % de méthylation (McrBC-qPCR) de l'extrémité 5' du locus donneur en fonction de la présence ou non de nouvelles insertions. Les epiRIL présentant des nouvelles insertions sont ordonnées par ordre croissant de celles-ci. Les barres d'erreur représentent l'écart type entre deux répliques techniques.

analyse montrent clairement l'hypométhylation systématique du locus donneur hérité du parent sauvage lorsque de nouvelles insertions sont présentes (fig. 2.17). Ces analyses ont été valorisées dans une publication réalisée dans le cadre d'une collaboration avec l'équipe du Dr. T Kakutani (voir en annexe 4 : Fu et al, 2013) qui a notamment montré que cette hypométhylation est induite en trans par une protéine codée par *VANDAL21* et qu'elle est transitoire car le locus reméthyle après ségrégation de la copie active. Ce phénomène engendre la réactivation transcriptionnelle de la copie cryptique et est systématiquement associé à son excision dans les tissus somatiques (mais probablement dans une petite fraction de cellules car elle n'a pu être mise en évidence que par « nested PCR ») (fig. 2.17). En revanche la transmission germinale de l'excision du locus donneur n'a été observée que dans deux epiRIL (fig. 2.17). Notons ici que l'aspect systématique du phénomène d'hypométhylation en trans du locus donneur hérité du sauvage lorsque de nouvelles copies sont présentes suggère qu'au moins une des nouvelles insertions est active dans ces lignées.

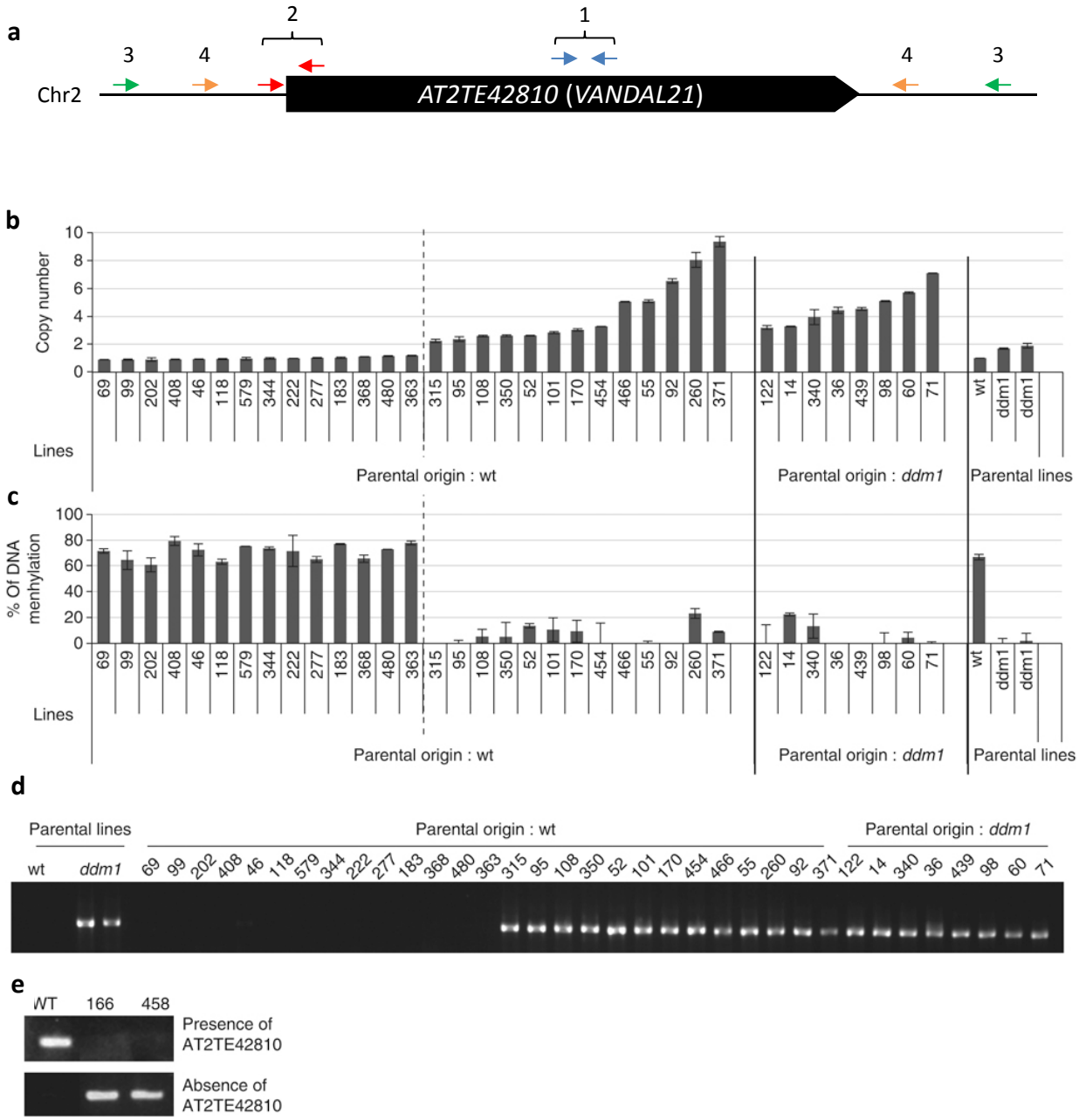


Figure 2.17 : Les nouvelles insertions d'*AT2TE42810* (*VANDAL21*) induisent l'hypométhylation du locus donneur ainsi que son excision.

(**a**) Représentation schématique du positionnement des amorces de PCR et qPCR sur le locus donneur *AT2TE42810* (les échelles ne sont pas respectées). (**b**) Estimation du nombre de copies d'*AT2TE42810* dans chacune des epiRIL estimé par qPCR en utilisant les amorces 1. (**c**) % de méthylation à l'extrémité 5' du locus donneur estimé par McrBC-qPCR en utilisant les amorces 2. Les barres d'erreur représentent l'écart type entre deux répliques techniques. (**d**) Analyse de l'excision somatique d'*AT2TE42810* par « nested » PCR en utilisant les amorces 3 puis 4. (**e**) Transmission germinale de l'allèle excisé dans deux epiRIL détectée par PCR classique en utilisant les amorces 2 pour la présence et 3 pour l'absence. L'origine parentale du locus est sauvage pour l'epiRIL 166 et *ddm1* pour 458.

2.3 Distribution des nouvelles insertions d'ET

Chez Arabidopsis, les ET sont principalement localisés au niveau des régions péricentromériques, ce qui pourrait résulter soit d'un ciblage préférentiel soit de leur élimination rapide après insertion le long des bras chromosomiques.

Diverses études portant sur le ciblage de plusieurs ET ont révélé qu'il était essentiellement dicté par des propriétés de la transposase/intégrase et qu'il peut se faire à plusieurs niveaux. Tout d'abord, comme bon nombre d'endonucléases, ces enzymes peuvent reconnaitre spécifiquement certaines séquences d'ADN. Ensuite, certaines transposases/intégrases contiennent un domaine d'interaction avec des modifications d'histones. Cela a été montré pour certains chromovirus (rétroéléments à LTR GIPSY) qui présentent un chromodomaine reconnaissant la marque H3K9me2 engendrant leur insertion dans l'hétérochromatine (Gao et al. 2008). Enfin, la transposase/intégrase peut interagir avec des protéines associées spécifiquement à l'ADN ou la chromatine comme cela a été très bien décrit pour les rétroéléménts à LTR Ty chez *Saccharomyces cerevisiae*. De fait, le ciblage des éléments Ty3 (rétroélément à LTR GYPSY) au niveau des gènes d'ARNt se ferait d'une manière analogue au recrutement de POL III à ces locus, par une interaction de l'intégrase avec TFIIIC (Aye et al. 2001, Lesage and Todeschini 2005). De même, le ciblage de Ty5 (rétroélément à LTR COPIA) dans l'hétérochromatine serait dû à l'interaction de l'intégrase avec la protéine hétérochromatique SIR4 (Xie et al. 2001). Il est important de noter que ces spécificités ne sont pas caractéristiques des différentes superfamilles d'ET car des ciblages différents impliquant des mécanismes divers ont été identifiés pour des ET d'une même superfamille et ce, au sein d'une même espèce ou entre deux espèces apparentées. C'est notamment le cas des éléments appartenant aux familles de rétroéléments à LTR COPIA Ty1, Ty2, Ty4, et Ty5 chez la levure *S. cerevisiae* (Lesage and Todeschini 2005) ou encore des deux rétroéléments à LTR COPIA apparentés *Tal1* et *EVADE* (*ATCOPIA93*) respectivement présents chez *Arabidopsis lyrata* et *Arabidopsis thaliana* (Tsukahara et al. 2012).

D'autres paramètres peuvent également affecter la localisation des nouvelles insertions d'ET. On peut mentionner notamment les cas du transposon à ADN *Ac* du maïs et de l'élément P de la drosophile qui tendent à s'insérer à proximité du locus donneur par des

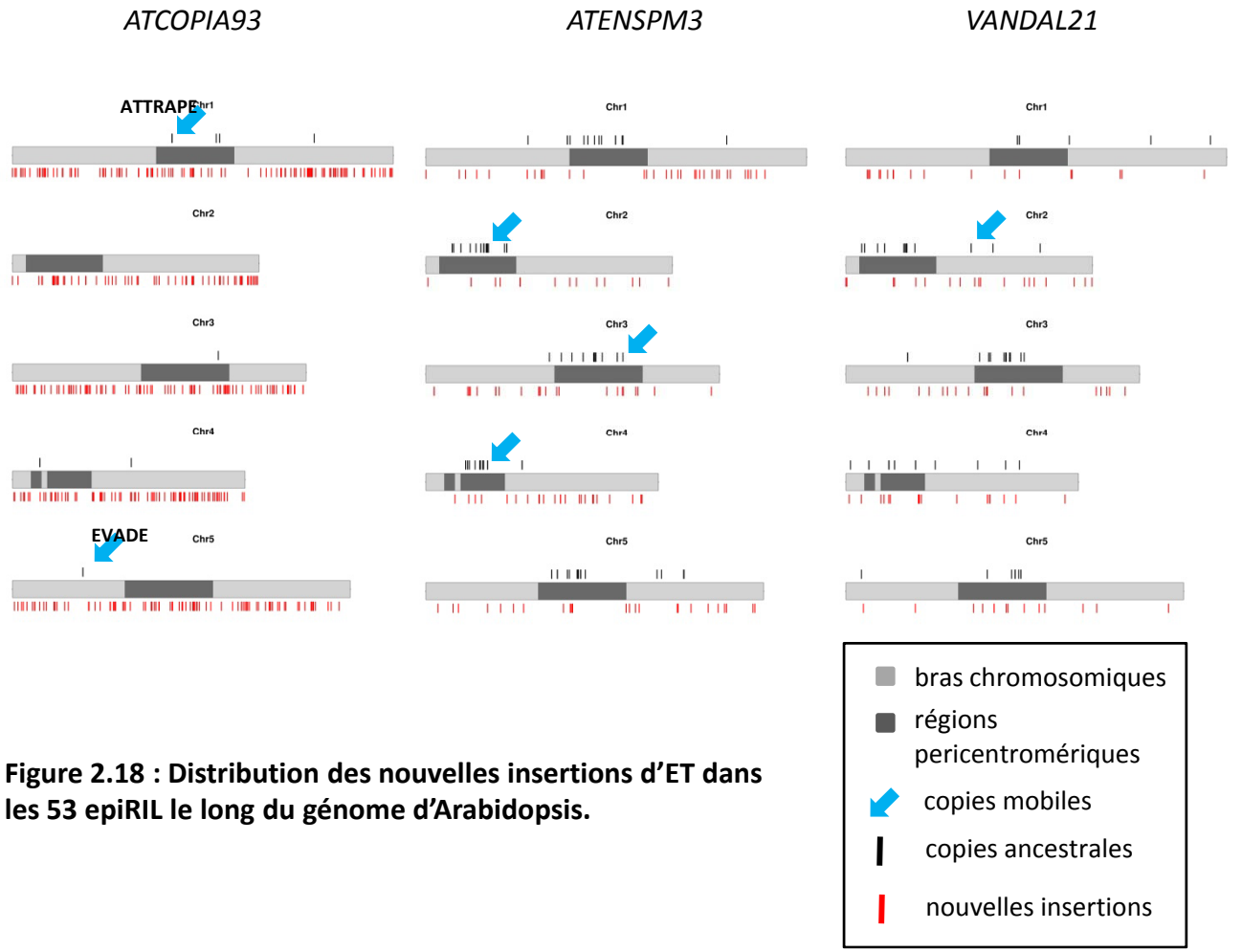


Figure 2.18 : Distribution des nouvelles insertions d'ET dans les 53 epiRIL le long du génome d'Arabidopsis.

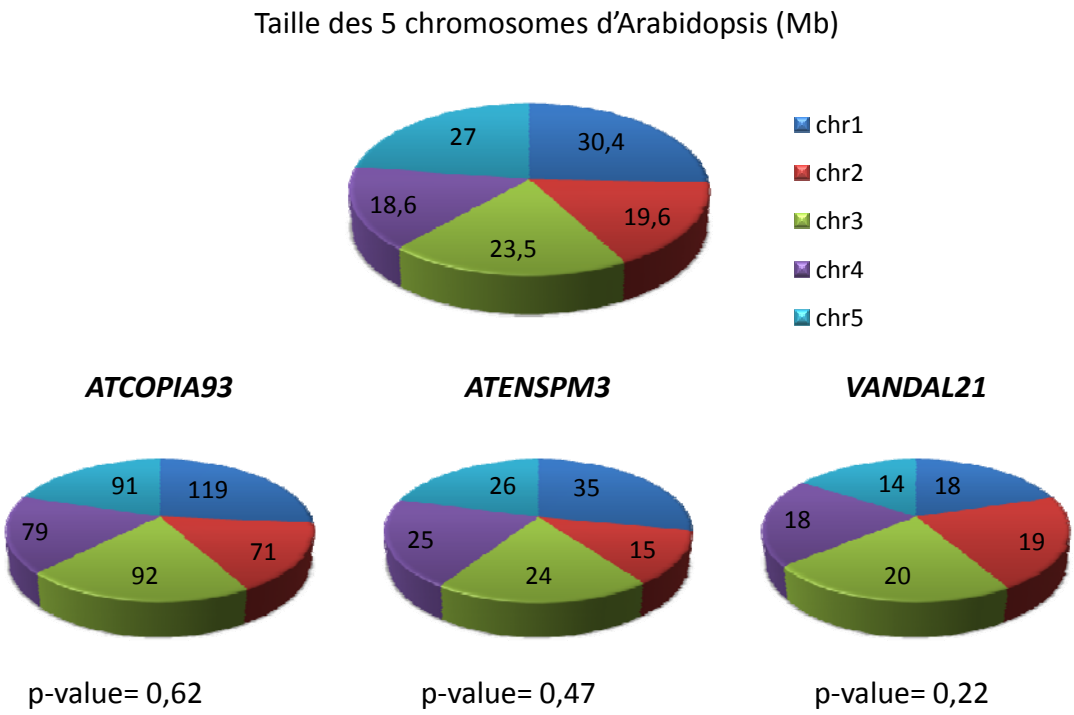


Figure 2.19 : Distribution des nouvelles insertions d'ET sur les chromosomes d'Arabidopsis.
Les p-value ont été obtenues par un test de χ^2.

mécanismes encore méconnus mais basés sur le couplage de la transposition de ses éléments avec la réplication (Greenblatt 1984, Dooner and Belachew 1989).

L'analyse des sites d'insertion des nouvelles copies d'ET dans les epiRIL doit permettre de discriminer entre les deux hypothèses concernant la distribution des ET dans le génome d'Arabidopsis car les évènements d'insertions sont identifiés juste après qu'ils ont eu lieu et donc bien avant que la sélection naturelle ou la dérive génétique ne puissent conduire à leur élimination. De fait, seuls les évènements de transposition causant la létalité car invalidant des gènes essentiels à la survie échappent à notre analyse.

2.3.1 Distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis

J'ai analysé les distributions des nouvelles insertions d'ET le long des chromosomes d'Arabidopsis pour les trois familles les plus mobiles dans les epiRIL en ne prenant en compte que les insertions uniques afin de pouvoir déterminer si elles ont lieu préférentiellement dans des intervalles d'origine *ddm1* ou sauvage (fig. 2.18).

Dans les trois cas, les nouvelles insertions sont réparties sur les 5 chromosomes sans préférence significative pour celui ou ceux sur lesquels se trouvent les locus des donneurs (fig. 2.18 et 2.19). De plus, la distribution des nouvelles insertions n'est pas significativement différente de celle attendue théoriquement sous hypothèse de hasard entre les régions péricentromériques et les bras chromosomiques (fig. 2.20) (d'après la définition des régions péricentromériques établie dans Bernatavichute et al. 2008). Cependant, bien que non statistiquement significatives, des tendances se dessinent pour *VANDAL21* et *ATENSPM3* qui présentent une proportion plus importante de nouvelles insertions respectivement dans les régions péricentromériques et les bras chromosomiques. Enfin, il apparaît que la distribution des nouvelles insertions contraste dramatiquement de celle des copies résidentes de chacune des familles, qui sont majoritairement trouvées dans les régions péricentromériques (bien que, pour *ATCOPIA93*, le trop faible nombre de copies résidentes ne permettent pas de réaliser de test statistique) (fig. 2.20).

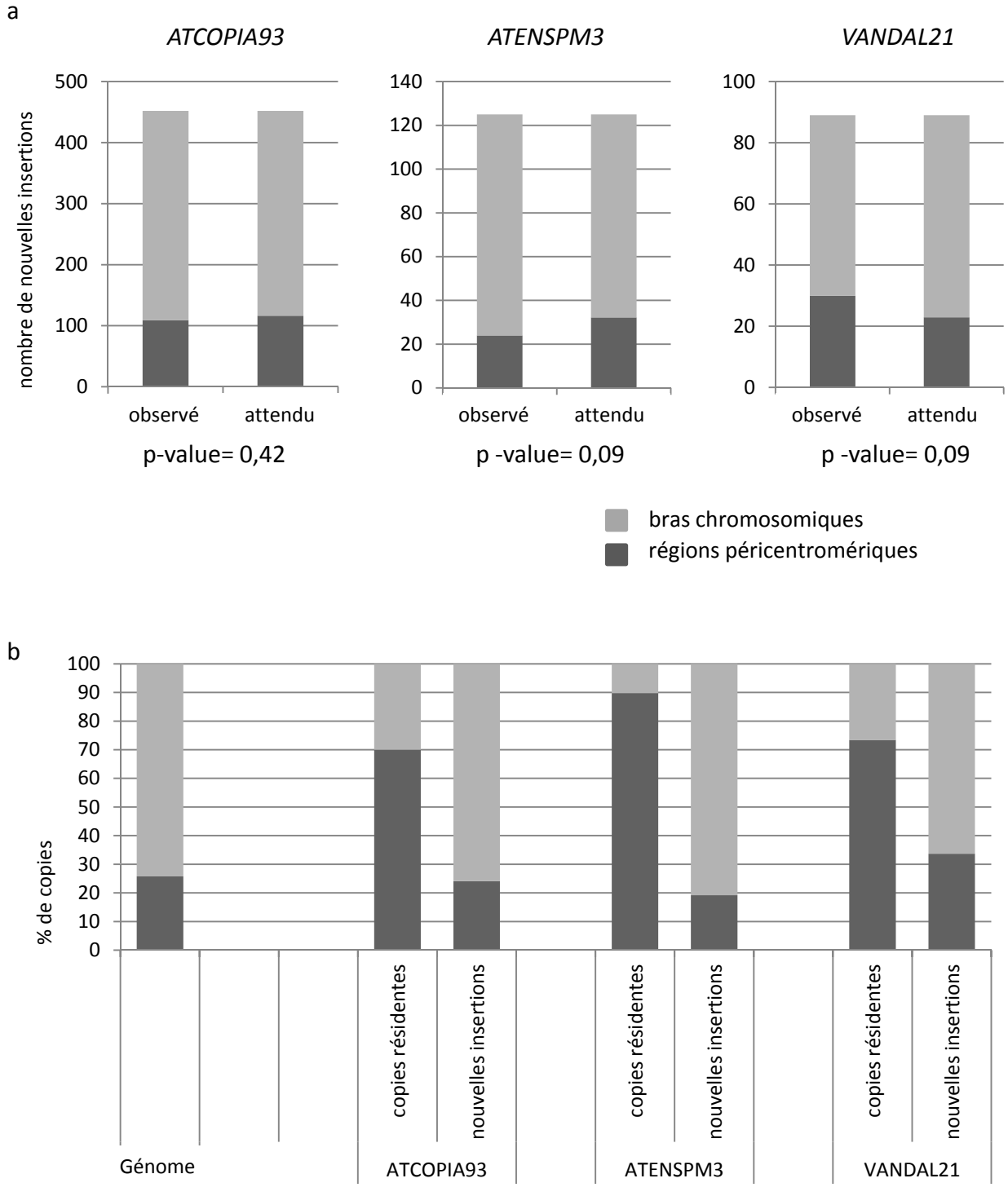


Figure 2.20 : Distribution des nouvelles insertions d'ET entre régions péricentromériques et bras chromosomiques par rapport à une distribution attendue sous hypothèse de hasard (a) et par rapport à la distribution des copies résidentes (b).
Les p-value ont été obtenues par un test de χ^2.

2.3.2 Distribution des nouvelles insertions d'ET en fonction des états chromatiniens

Comme décrit dans l'introduction, quatre états chromatiniens distincts ont été décrits chez Arabidopsis. Ces états sont caractérisés par des combinaisons différentes de marques chromatiniennes et sont associés à des activités transcriptionnelles contrastées. J'ai donc cherché à déterminer si les nouvelles insertions d'ET présentaient une préférence pour certains états. Il est essentiel de noter ici que la carte génomique des états chromatiniens établie au laboratoire est basée sur des expériences réalisées sur des plantules entières qui sont composées quasi-exclusivement de cellules somatiques différenciées. Bien que la dynamique chromatinienne au cours du développement soit encore peu connue, il est probable que la distribution des états chromatiniens le long du génome d'Arabidopsis soit quelque peu différente dans les cellules contribuant à la génération suivante (zygote, cellules souches du méristème apical, gamètes…). De fait, nous savons déjà que des variations plus ou moins importantes de la méthylation de l'ADN se produisent lors de la reproduction (voir introduction).

Comme décrit dans l'introduction, les états chromatiniens forment des domaines de petite taille dans la majorité des cas. J'ai donc restreint mon analyse aux insertions cartographiées dans des intervalles accepteurs d'une taille maximale de 1kb afin de limiter les situations de recouvrement de plusieurs domaines. Pour les trois familles d'ET analysées, la distribution des nouvelles insertions en fonction des états chromatiniens est significativement différente du hasard. De plus, trois profils de distribution très distincts voire contrastés se dessinent pour les trois familles (fig. 2.21).

Dans le cas d'*ATCOPIA93*, les insertions dans les états chromatiniens CS2 et CS4 sont surreprésentées, tandis que celles dans le CS1 sont sous représentées. Les insertions dans le CS3 sont quant à elles observées dans la proportion attendue théoriquement sous une hypothèse de hasard. Cependant, dans les epiRIL, les régions héritées du parent *ddm1* présentent des altérations majeures du CS3 à savoir une perte des marques répressives 5meC et H3K9me2 ainsi qu'un gain de marques associées à la transcription comme H3K4me2. Afin d'affiner mon analyse, j'ai donc déterminé l'origine parentale des intervalles accepteurs localisés dans des domaines définis comme CS3 chez la plante sauvage. Le résultat de cette analyse révèle un biais très marqué pour les intervalles chromosomiques

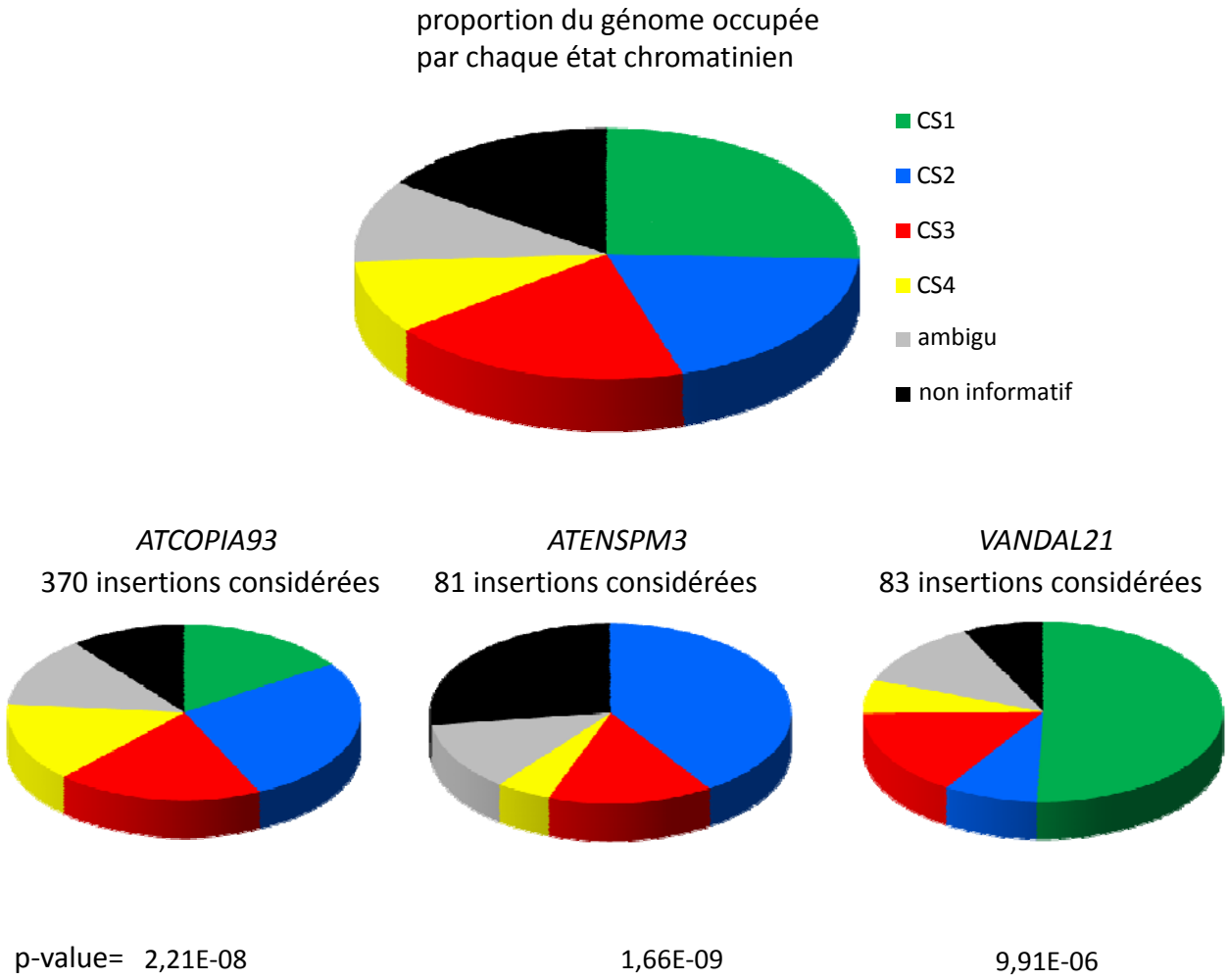
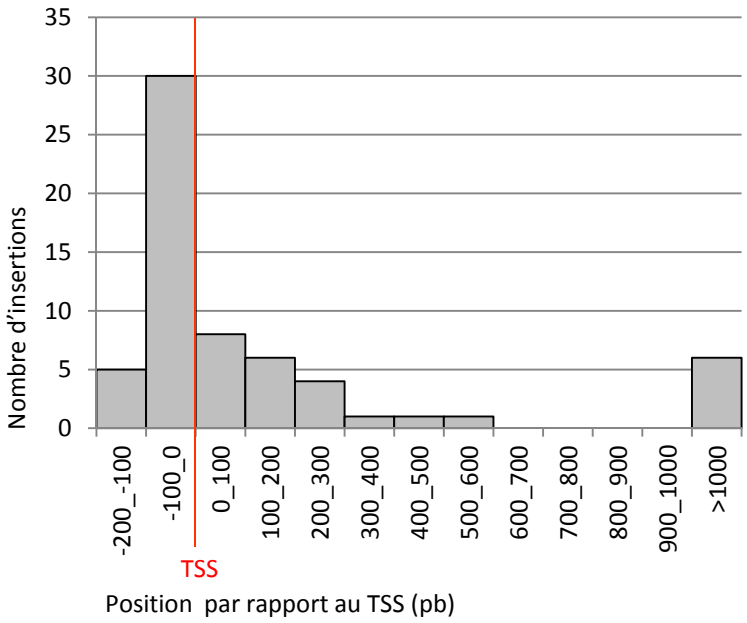


Figure 2.21 : Proportion des nouvelles insertions d'ET dans chacun des état chromatiniens.
En plus des quatre états chromatiniens définis par Roudier et al, 2011 décrits dans l'introduction, j'ai également représenté sur cette figure l'état qualifié d'ambigu (en gris) qui correspond aux sondes de la puce à ADN associées à des marques caractéristiques de plusieurs états distincts ainsi que les régions du génome qui correspondent aux sondes ne rapportant aucun signal pour les marques testées (en noir). Les p-value ont été obtenues par un test de χ^2.

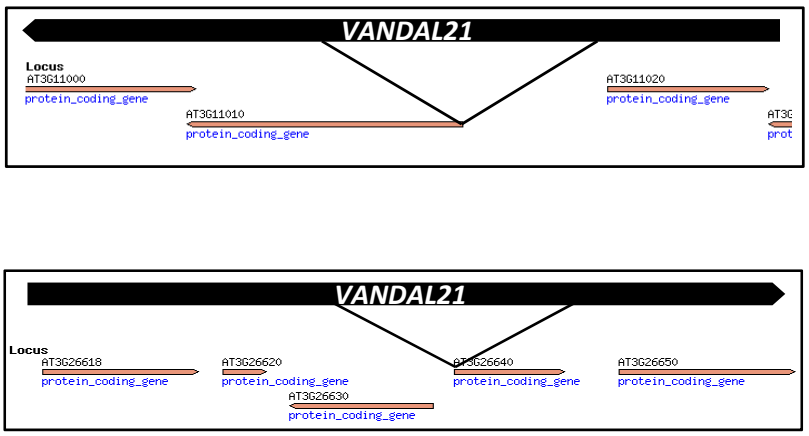
d'origine *ddm1* (47/56 insertions considérées contre 14/56 sous hypothèse de hasard p-value = 2,3E-24 avec un test de χ2). Pour résumer, *ATCOPIA93* montre donc une distribution préférentielle des nouvelles insertions dans le CS2, le CS4 et le CS3 dérivé de *ddm1*. La caractéristique principale partagée par ces trois états est l'absence de méthylation de l'ADN. Cela coïncide avec le déficit en insertion dans le CS1 car la méthylation de l'ADN est aussi une des marques assez fréquemment associée avec cet état. Afin de confirmer cette hypothèse j'ai analysé directement l'état de méthylation (à partir de données de MeDIP-chip) des sites accepteurs des nouvelles insertions d'*ATCOPIA93* en prenant en compte leur origine parentale. Trois cas de figure sont à prendre en considération pour les intervalles accepteurs qui peuvent être : (i) des régions qui ne sont pas ou peu méthylées dans le sauvage et dans le mutant *ddm1*, (ii) des régions qui sont méthylées dans le sauvage mais qui perdent leur méthylation dans le mutant *ddm1* et (iii) des régions qui sont méthylées fortement dans le sauvage et dans le mutant *ddm1*. Il apparait que 85% des insertions d'*ATCOPIA93* ont lieu dans des intervalles de catégorie (i) 15% dans des intervalles de type (ii) et 0% dans des intervalles de types (iii). Cependant, lorsque les insertions sont localisées dans des régions de type (ii), ces régions sont très majoritairement hérités de *ddm1* (22/25 insertions considérées contre 6,25/25 sous hypothèse de hasard, p-value= 3,5E-13 avec un test de χ2). Considérés ensemble, ces résultats confirment l'insertion préférentielle d'*ATCOPIA93* dans les régions non méthylées. On peut alors envisager que, tout comme certaines enzymes de restriction, l'intégrase des *ATCOPIA93* soit sensible à la méthylation de l'ADN. D'autres hypothèses sont bien sûr possibles, notamment une préférence pour d'autres marques chromatiniennes dont la distribution est corrélée négativement avec celle de la méthylation de l'ADN, comme par exemple le variant d'histone H2AZ (Zilberman et al. 2008).

Les nouvelles insertions d'*ATENSPM3* présentent une distribution encore plus contrastée. Les insertions dans le CS1 sont absentes tandis que celles dans le CS2 sont fortement surreprésentées. Le trop faible effectif d'insertions dans le CS3 ne permet pas de déterminer s'il existe un biais dans l'origine parentale des sites accepteurs. A ce stade de l'analyse, il est difficile d'expliquer ce profil de distribution en ne considérant qu'un seul paramètre. Le biais en faveur du CS2 pourrait être dû à un ciblage spécifique des insertions vers ce type chromatinien par exemple grâce à l'interaction de la transposase avec une marque

a



b



c

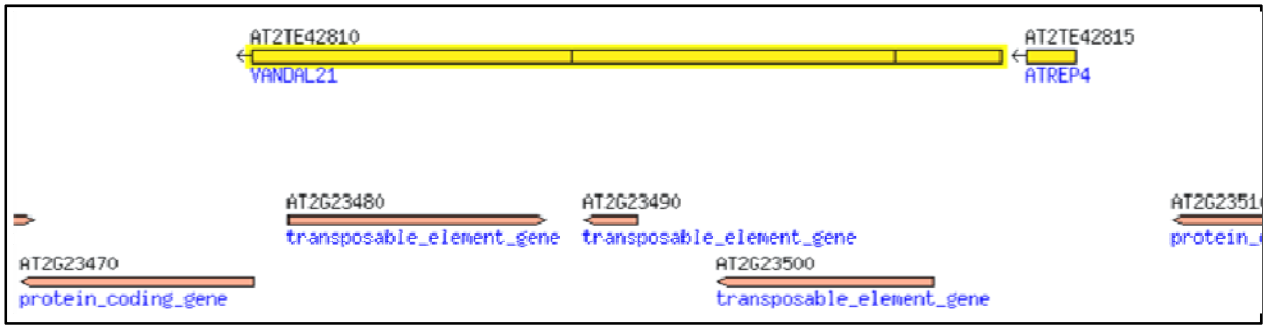


Figure 2.22 : Distribution préférentielle des nouvelles insertions de *VANDAL21* à proximité du TSS des gènes.
(**a**) Distribution des nouvelles insertions de *VANDAL21* par rapport au TSS des gènes. (**b**) Deux exemples de nouvelles insertions de *VANDAL21* à proximité du TSS des gènes et dans la même orientation que le gène. (**c**) Vue Gbrowse du locus donneur *AT2TE42810*.

chromatinienne caractéristique de cet état : H3K27me3. Cependant, cette hypothèse n'explique pas l'absence totale d'insertions dans le CS1 uniquement. Cette absence suggère quant à elle des insertions préférentielles en dehors de la chromatine transcriptionnellement active. Cela n'explique pas alors le biais vers CS2 plutôt que vers CS3 ou CS4 qui sont également très peu actifs transcriptionnellement.

Enfin, la distribution des nouvelles insertions de *VANDAL21* montre une préférence très nette pour le CS1 au détriment des autres catégories qui sont toutes également sous-représentées. Si le faible nombre d'insertions dans le CS3 ne permet pas ici non plus de tester statistiquement la présence d'un biais dans l'origine parentale des sites accepteurs, on notera cependant que 9/12 sont d'origine *ddm1* contre ¼ attendu. Ces deux observations suggèrent un biais d'insertion préférentiel de *VANDAL21* dans les régions chromatiniennes transcriptionnellement actives. Ce résultat est en contradiction avec l'observation précédente qui montrait, pour les nouvelles insertions de *VANDAL21*, une tendance (non significative) vers les régions péricentromériques, théoriquement riches en domaines CS3. On pourrait alors faire l'hypothèse d'une régulation à deux niveaux : une attraction au niveau des régions péricentromériques au sein desquelles s'effectuerait un ciblage spécifique dans les régions transcriptionnellement actives. En effet, bien que composées majoritairement d'ET, les régions péricentromériques contiennent également des gènes.

2.3.3 Distribution des nouvelles insertions d'ET par rapport aux gènes

L'analyse de la localisation des nouvelles insertions d'ET par rapport aux gènes est importante non seulement pour ajouter à la compréhension du ciblage éventuel des insertions mais également pour explorer l'impact de ces nouvelles insertions sur la régulation de l'expression des gènes au sein desquels ou à proximité desquels elles ont eu lieu.

En me basant sur le même jeu d'insertions que pour la distribution en fonction des états chromatiniens (insertions cartographiées dans des intervalles < 1Kb), j'ai recherché combien d'entre elles sont situées dans un gène ou à une distance maximale de 200pb d'un gène. Les proportions ainsi obtenues sont respectivement de 70%, 55% et 77% pour *ATCOPIA93*,

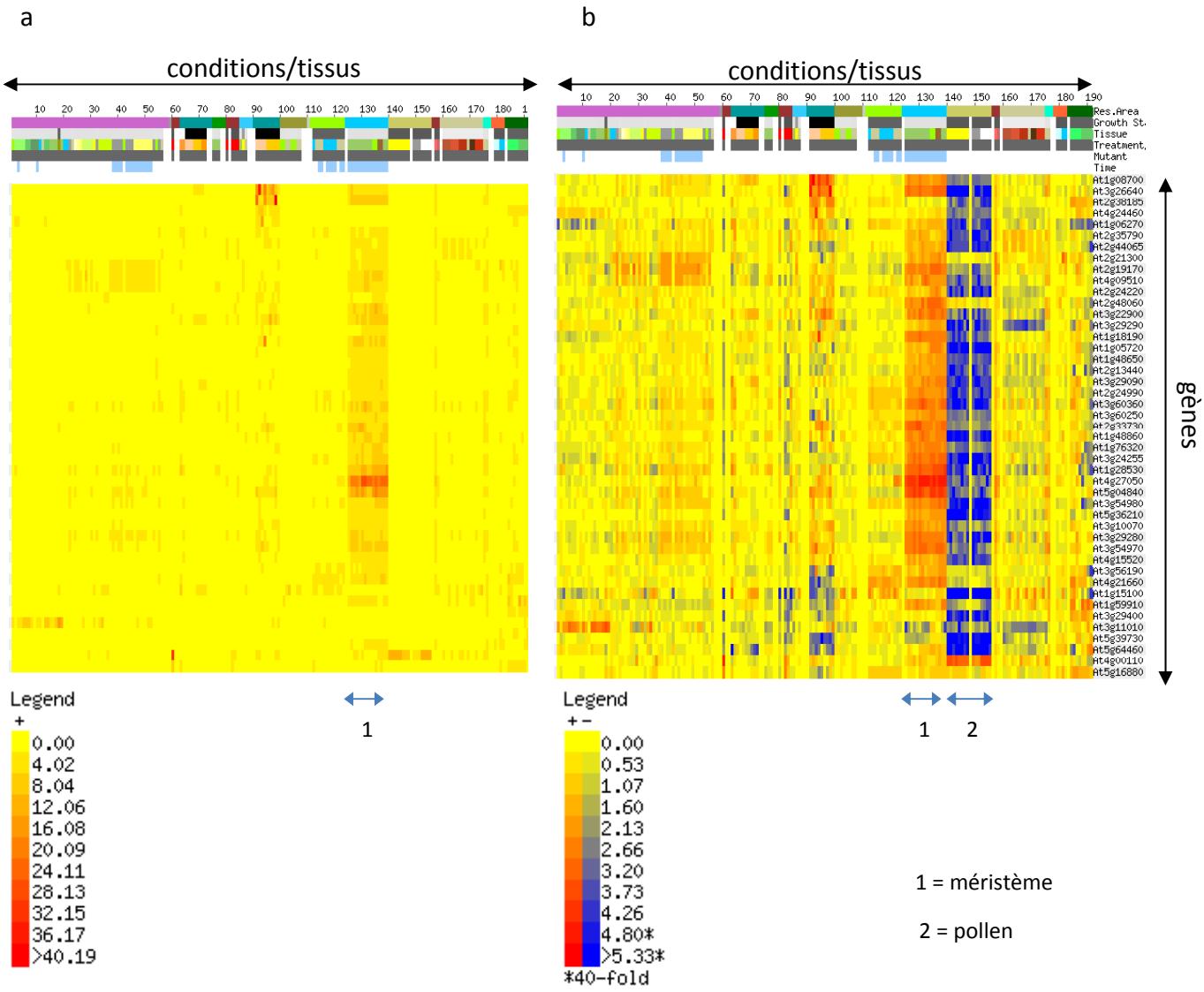


Figure 2.23 : e-northern des gènes ciblés par les nouvelles insertions de *VANDAL21*. Figure établie à partir de données de transcriptomes publiques (http://bar.utoronto.ca). (**a**) Représentation de l'expression en « fold change » par rapport aux conditions contrôles. (**b**) Représentation de l'expression transformée en «log(fold change) ».
Remarque : l'absence d'expression dans le pollen n'est pas spécifique de ces gènes car des tirages aléatoires de gènes présentaient des profils similaires (données non montrées). Ce n'est pas le cas en revanche pour l'expression méristématique.

ATENSPM3 et *VANDAL21* (les résultats de cette analyse sont listés dans l'annexe 5). Afin de déterminer si ces proportions sont significativement différentes du hasard nous avons réalisé des simulations aléatoires (bootstrap) de listes d'intervalles (1000 listes pour chaque ET). Il apparaît que le nombre d'insertions de *VANDAL21* au sein ou à proximité d'au moins un gène est significativement supérieur au hasard (p-value = 0,03) et non significativement différentes du hasard pour les insertions d'*ATCOPIA93* et *ATENSPM3* (p-value = 0,16 et 0,84 respectivement).

Afin d'affiner mes analyses quant à la position de ces insertions par rapport aux gènes, je me suis ensuite focalisée sur les insertions dont la localisation à proximité des gènes peut être déterminée encore plus précisément, dans un intervalle de 200pb. Ainsi, j'ai concentré mon analyse sur les insertions située à <400pb d'un gène (taille de l'intervalle accepteur ajoutée à la distance maximum par rapport au gène). J'ai cherché à déterminer si ces insertions présentaient un biais pour les extrémités 5' ou 3' des gènes. Les insertions de *VANDAL21* sont majoritairement localisées dans la région 5' des gènes en amont du TSS (fig. 2.22a). De plus, ces insertions proches du TSS sont en grande majorité orientées dans le sens du gène ce qui suggère un lien entre les mécanismes d'insertion et de transcription (fig. 2.22b). Il est notable que le locus donneur *VANDAL21* (*AT2TE42810*) montre également ces caractéristiques (fig. 2.22c). Ces mêmes observation ont été faites par nos collaborateurs (Fu et al. 2013).

J'ai ensuite cherché à déterminer si ces gènes présentaient des profils d'expression particuliers en me basant sur des données publiques d'expression. Pour les insertions d'*ATENSPM3* et *ATCOPIA93* aucun profil d'expression significatif n'a pu être mis en évidence (données non montrées). En revanche, les gènes présentant des nouvelles insertions de *VANDAL21* à proximité du TSS tendent à montrer une expression méristématique (fig. 2.23). Ce profil constitue un indicateur du tissu dans lequel *VANDAL21* a transposé dans les epiRIL. En effet, si, comme le suggèrent nos données, *VANDAL21* s'insert préférentiellement près des gènes transcriptionnellement actifs et que les gènes ciblés sont, pour la plupart actifs principalement dans le méristème apical, cela suggère que *VANDAL21* est mobilisé majoritairement dans les cellules souches du méristème apical plutôt que dans les gamétophytes ou l'embryon précoce. Bien sûr, cette analyse ne nous renseigne pas sur la

possibilité d'évènements de transposition dans d'autres types cellulaires ou organes ne contribuant pas à la génération suivante.

2.4 Discussion

2.4.1 Remarques générales sur l'approche expérimentale

Bien que les conséquences de la perte du contrôle épigénétique sur la mobilisation des ET ont déjà été étudiées à plusieurs reprises, l'approche décrite ici est originale par de nombreux aspects. C'est en effet la première fois que la mobilisation est déterminée à l'échelle du génome pour tous les ET, sans apriori d'annotation et ce à l'échelle d'une population. Le séquençage paired-end de librairies mate-pair associé à la méthode d'analyse que nous avons développée permet l'identification robuste des évènements d'insertion et, dans la majorité des cas, de la copie donneuse. Cependant, notre méthode ne permet pas à ce jour d'identifier les évènements d'insertion d'une taille supérieure à 11kb (cf. manuscrit Gilly et al.). Bien que peu nombreux au sein du génome d'Arabidopsis, certains ET >11 kb ont été identifiés comme potentiellement mobiles sur la base de critères de séquence. L'une des améliorations prévue pour notre méthode d'analyse est l'ajout d'un module pouvant identifier de tels évènements. La détection des insertions sans *a priori* sur les annotations a permis également de mettre en évidence la mobilisation de structures complexes composites de plusieurs annotations d'ET. Si à ce stade de l'analyse je ne me suis intéressée qu'aux évènements d'insertion dont le donneur recouvrait au moins partiellement une annotation d'ET, il est important de souligner que des évènements d'insertions n'impliquant aucun ET annoté ont aussi été détectés. S'il s'agit souvent de séquences centromériques imbriquées les unes dans les autres et qui constituent probablement des faux positifs, une analyse approfondie des séquences sans annotation pourrait néanmoins permettre l'identification éventuelle de nouveaux ET.

2.4.2. Eléments mobilisés suite à la perte du contrôle épigénétique

Nous avons identifié un petit nombre de familles présentant des ET mobiles dans le mutant *ddm1* et dans les epiRIL, ce qui est en accord avec des études antérieures réalisées sur ce mutant (Tsukahara et al. 2009). Cependant, nos travaux ont pu mettre en évidence la mobilisation d'éléments appartenant à des familles jamais identifiées comme mobiles chez Arabidopsis ou alors, comme dans le cas des *ATCOPIA78*, mobiles dans des conditions différentes des nôtres. Si on ajoute à cette observation que peu de copies seulement au sein de chacune de ces familles sont mobiles, il apparaît que très peu d'ET sont mobilisés suite à la perturbation du contrôle épigénétique induite par la mutation *ddm1*. Ce nombre paraît d'autant plus faible si on le compare à la réactivation transcriptionnelle massive de très nombreux ET dans le mutant *ddm1 (Lippman et al. 2004, Zemach et al. 2013)*. Plusieurs aspects dont certains ont été évoqués précédemment sont à prendre en compte ici pour tenter d'expliquer ces apparentes contradictions.

Le premier aspect concerne l'activation transcriptionnelle de copies défectueuses, incapables de transposer. Plus de la moitié des familles d'ET d'Arabidopsis serait composée de telles copies. Cependant, il n'est pas aisé de déterminer quelles copies d'ET sont potentiellement mobiles ou non en s'appuyant uniquement sur des critères de séquence. Ceux choisis au laboratoire (notamment présence LTR + RT pour les rétroéléments et de TIR + transposase pour les transposons à ADN) ne sont pas très strictes. On ne peut donc ne pas exclure que le nombre de familles sans aucun membre mobilisable soit encore plus élevé.

L'autre aspect essentiel à considérer est que la mobilisation au travers des générations nécessite l'expression dans des cellules qui participent à la génération suivante comme les cellules souches du méristème apical, les gamètes ou le zygote. L'activation transcriptionnelle des ET dans le mutant *ddm1* a été évaluée sur des plantules entières qui sont majoritairement composées de cellules somatiques et il est envisageable, voire probable que les profils d'expression des ET soient différents dans les cellules contribuant à la génération suivante. Cette hypothèse est d'ailleurs en accord avec les observations qui suggèrent que le contrôle des ET est renforcé au moment de la reproduction grâce à des mécanismes faisant intervenir différentes voies de l'ARNi. Afin de tester cette hypothèse, il

faudrait comparer les profils d'expression (par hybridation *in situ* par exemple) de plusieurs ET mobiles ou seulement potentiellement mobiles.

Cependant, ces différentes hypothèses ne permettent pas d'expliquer le cas très particulier des ET de la famille *ATCOPIA93 EVADE* et *ATTRAPE*. Comme décrit précédemment, ces deux ET présentent une très grande similarité de séquence (>99%), des LTR intègres et identiques ainsi que des ORF continues en apparence intactes. Cependant seul *EVADE* est fortement mobilisé suite à une perte de méthylation de l'ADN (Mirouze et al. 2009 et nos travaux). En outre, des travaux réalisés en collaboration avec l'équipe d'Olivier Voinnet (présentés dans la partie 3) indiquent que seul *EVADE* est exprimé dans le mutant *met1*, et on peut supposer qu'il en est de même dans le mutant *ddm1* et les epiRIL. Or, des travaux récents indiquent que certains ET sont marqués par H3K27me3 (marque caractéristique de la répression par le complexe PRC2) dans le mutant *met1* (Deleris et al. 2012) et que *EVADE* et *ATTRAPE* en font partie. Cependant, les analyses ne permettent pas de discriminer entre ces deux locus. L'absence de réactivation transcriptionnelle d'*ATTRAPE* dans *met1* (et probablement dans *ddm1*) pourrait donc résulter de son ciblage spécifique en fond mutant par le complexe PRC2. Des expériences de CHIP-qPCR utilisant des amorces spécifiques d'*ATTRAPE* ou d'*EVADE* permettraient de tester cette hypothèse.

Quoiqu'il en soit, le décalage qui existe entre le nombre d'ET réactivés transcriptionnellement dans le mutant *ddm1* et le nombre d'ET mobiles n'était pas inattendu. De fait, les mutants *ddm1* sont viables et fertiles et peuvent être propagés sur au moins huit générations d'autofécondation, ce qui est manifestement incompatible avec une mobilisation intempestive de plusieurs centaines d'ET.

2.4.3. Distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis

L'étude de la distribution des ET le long du génome d'Arabidopsis a permis de mettre en évidence des préférences d'insertions spécifiques de chacune des familles étudiées.

J'ai montré tout d'abord que les ET appartenant aux familles *ATCOPIA93*, *VANDAL21* et *ATENSPM3* ne présentent pas de biais d'insertion à l'échelle du génome entre les régions

péricentromériques et bras chromosomiques, en contradiction avec la distribution des copies résidentes.

Si des travaux antérieurs avaient montré une absence apparente de biais concernant les insertions d'*ATCOPIA93* et *ATENSPM3* le long du génome d'Arabidopsis (Miura et al. 2001, Tsukahara et al. 2012) l'étude de la distribution des nouvelles insertions en fonction des états chromatiniens a permis de mettre en évidence certains biais. Dans le cas d'*ATCOPIA93* qui semble s'insérer préférentiellement dans les régions non méthylées du génome, il serait intéressant de tester si l'intégrase présente une affinité différente pour l'ADN méthylé et non méthylé.

Le ciblage préférentiel de *VANDAL21* dans le CS1 et à proximité du TSS des gènes est en accord avec les études réalisées sur d'autres éléments de la superfamille Mu chez le riz et le maïs (Naito et al. 2009, Liu et al. 2009). Il semble donc que les éléments Mu, bien que très diversifiés (Wicker et al. 2007) présentent les mêmes spécificités de site d'insertion chez les différents organismes étudiés. Cependant, c'est la première fois à ma connaissance qu'une préférence d'orientation dans le sens du gène a été mise en évidence pour ce type d'éléments (voir également en annexe 4 l'article Fu et al, 2013).

Considérées ensemble, ces différentes observations suggèrent que la suraccumulation des ET (du moins de ceux considérés ici) dans les régions péricentromériques du génome d'*Arabidopsis thaliana* ne résulte pas d'un biais préférentiel d'insertion mais plutôt d'une élimination des insertions situées le long les bras chromosomiques. La première raison évoquée est la contre sélection des insertions ayant un effet délétère sur les gènes au sein ou à proximité desquels elles ont lieu. Le génome d'Arabidopsis étant très compact, une insertion sur un bras chromosomique a statistiquement de grandes chances de tomber dans ou à proximité immédiate d'un gène. L'autre force pouvant entrainer l'élimination des ET dans les bras chromosomiques est la recombinaison homologue. Chez Arabidopsis, il a été montré que la recombinaison homologue est très fortement inhibée dans les régions péricentromériques (Giraut et al. 2011). Il est donc possible que des insertions, n'ayant potentiellement aucun effet délétère, soient cependant plus fréquemment éliminées si elles sont localisées dans les bras chromosomiques, plus recombinogènes, que celles localisées dans les régions péricentromériques par dérive génétique.

3
Etude de l'établissement du contrôle épigénétique sur les éléments transposables

Si les mécanismes responsables du maintien du contrôle épigénétique et tout particulièrement de la méthylation de l'ADN sur les éléments transposables commencent à être bien connus, il n'en est pas de même pour ceux concernant l'établissement *de novo* de ce contrôle. En effet, bien que les expériences menées pour comprendre l'extinction inopinée de l'expression de transgènes ont montré l'implication de la machinerie du RdDM pour l'établissement de la méthylation de l'ADN, les étapes précoces de ce processus et les conditions de son ciblage spécifique sur les copies d'ET nouvellement insérées demeurent largement hypothétiques. Je me suis donc intéressée à l'établissement du contrôle épigénétique sur les ET nouvellement insérés dans les epiRIL et j'ai également mis au point un crible génétique visant à identifier de nouveaux acteurs impliqués dans l'établissement de la méthylation de l'ADN.

3.1 Etude de l'établissement de la méthylation de l'ADN sur les copies d'ET nouvellement insérées

3.1.1 Etude comparative de deux rétroéléments à LTR de type COPIA : *ATCOPIA78* et *ATCOPIA93*

Dans la partie 2, nous avons noté la dynamique de mobilisation drastiquement opposée entre les ET qui composent les familles *ATCOPIA93* et *ATCOPIA78*, qui appartiennent pourtant à la même superfamille. De fait, l'élément de la famille *ATCOPIA93, EVADE,* est très actif dans les epiRIL car il est responsable de plusieurs centaines d'évènements d'insertion, dont la très vaste majorité est propre à chaque lignée. En revanche, toutes les insertions des éléments *ATCOPIA78* (*ONSEN*) sont partagées par plusieurs epiRIL, ce qui démontre un arrêt de la mobilisation à partir de la F2, soit dès le rétablissement de la fonction DDM1. Il semble donc qu'un mécanisme de répression cible très précocement les *ATCOPIA78* (copies résidentes et nouvelles insertions) et plus tardivement, voire pas du tout *ATCOPIA93*. Je me suis donc attachée à caractériser plus avant ces deux familles d'ET afin de mieux comprendre leur dynamique de mobilisation et surtout, les modes d'établissement du contrôle épigénétique sur les copies nouvellement insérées.

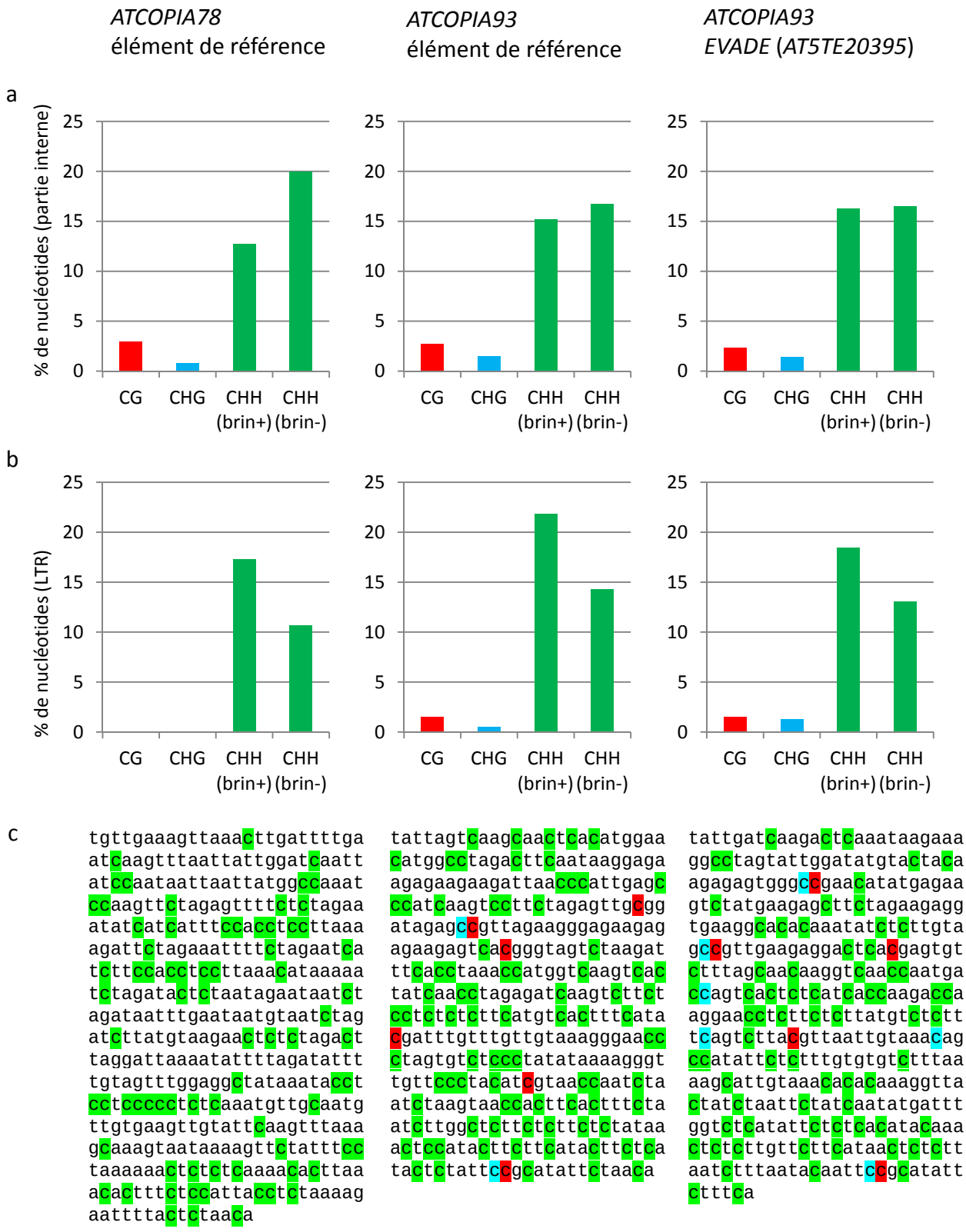


Figure 3.1 : Contexte nucléotidique des cytosines présentes dans ATCOPIA78 et *ATCOPIA93*.

(**a**) et (**b**) Proportion des cytosines dans les différents contextes nucléotidiques dans la région interne des éléments et des LTR . (**c**) Séquences des LTR.

Ces deux familles d'éléments transposables ont déjà fait l'objet de plusieurs études. La mobilisation d'*EVADE* (*ATCOPIA93*) a été identifiée dans les epiRIL *met1*, les doubles mutants *met1kyp et met1nrpd2a* (NRPD2a étant la sous-unité commune à POLIV et POLV), les epiRIL dérivées du mutant *met1* (Mirouze et al. 2009) ainsi que dans des générations avancées de *ddm1* (Tsukahara et al. 2009). Les éléments *ONSEN (ATCOPIA78)* en revanche ont jusqu'à présent été décrits comme mobiles uniquement dans des mutants affectés dans la voie du RdDM soumis à un stress chaleur. Ces observations ainsi que celles faites dans la partie 2 sur le comportement des *ATCOPIA93* et des *ATCOPIA78* chez le sauvage et chez les mutants affectant les différentes voies de la méthylation démontrent que le contrôle d'*ATCOPIA93* dépend quasi exclusivement de la voie de méthylation dite de maintenance (DDM1 et MET1) alors que celui des *ATCOPIA78* dépend de l'action conjointe de DDM1 et du RdDM.

Si une différence si drastique dans les voies de contrôle de la méthylation de l'ADN sur deux familles d'ET appartenant pourtant à la même superfamille peut sembler surprenante, une partie de l'explication est fournie par l'examen de leur séquence nucléotidique. J'ai analysé les contextes nucléotidiques des cytosines (CG, CHG et CHH) pour ces deux familles d'ET en me basant sur leur séquence consensus (établies par Repbase update (Jurka 2000)). J'ai également étudié directement le locus *EVADE* (*AT5TE20395*) car sa séquence est assez divergente de la séquence consensus des *ATCOPIA93*. Dans la région interne, les éléments de référence des deux familles ainsi qu'*EVADE* présentent des cytosines dans les trois contextes nucléotidiques CG, CHG et CHH en proportions équivalentes (fig. 3.1a). En revanche, la situation est tout autre pour les LTR, puisque que celles d'*ATCOPIA93* (consensus et *EVADE*) présentent des cytosines dans les trois contextes nucléotidiques alors que les cytosines des LTR d'*ATCOPIA78* sont exclusivement en contexte asymétrique CHH (fig.3.1b et c). Seul le RdDM peut donc instruire à chaque cycle de division la méthylation des LTR d'*ATCOPIA78*, néanmoins aidé semble-t-il par l'action de DDM1, et maintenir ainsi leur répression transcriptionnelle. A l'inverse, la présence de cytosines en contexte symétrique pour les LTR d'*ATCOPIA93* explique que la méthylation de celles-ci puisse être assurée de façon prépondérante par la machinerie de maintenance.

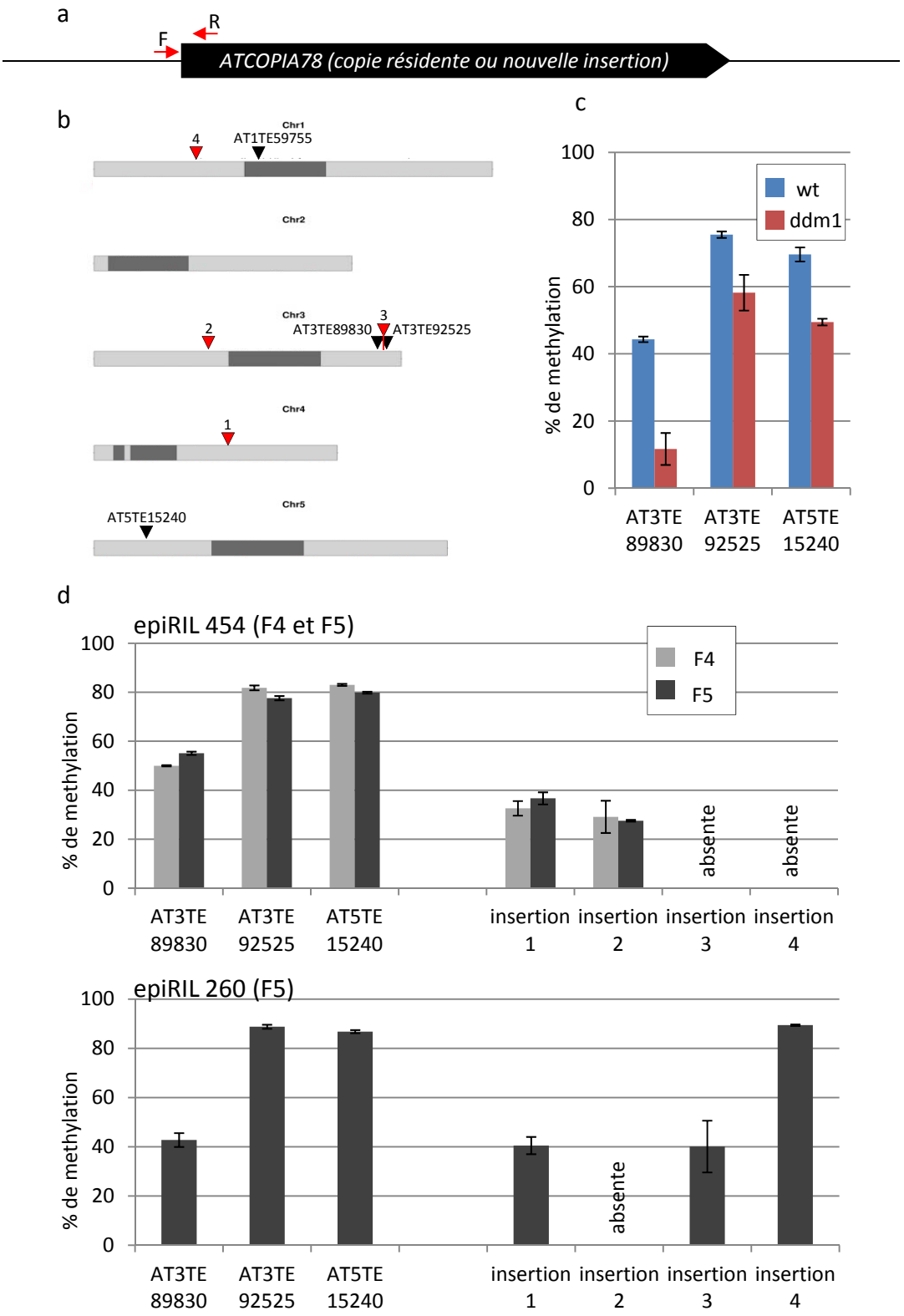


Figure 3.2 : Etablissement/Rétablissement du contrôle épigénétique sur les copies nouvellement insérées/résidentes de *ATCOPIA78*.

(**a**) Représentation schématique illustrant le positionnement des amorces utilisées pour analyser, par McrBC-qPCR, la méthylation de l'extrémité 5' des copies résidentes et des nouvelles insertions de *ATCOPIA78*. L'amorce « Forward » (F) est spécifique de la région flanquant chaque insertion alors que l'amorce « reverse » (R) localisée dans l'élément est la même pour toutes les copies testées. (**b**) Position des copies donneuses (triangles noirs) et des nouvelles insertions partagées d'*ATCOPIA78* dans les epiRIL (triangles rouges) le long du génome d'Arabidopsis. (**c**) % de méthylation par McrBC-qPCR de trois des copies donneuses chez la plante sauvage et le mutant *ddm1.* (**d)** % de méthylation par McrBC-qPCR des copies donneuses ainsi que de quatre nouvelles insertions partagées dans les générations précoces des epiRILs 454 et 260. Les barres d'erreur représentent l'écart type entre deux répliques techniques.

- *ATCOPIA78*

Comme suggéré dans la partie 2, la forte production de siARN de 24nt qui ciblent les *ATCOPIA78* en contexte sauvage et également dans le mutant *ddm1* doit pouvoir conférer aux copies résidentes la capacité d'être reméthylées et donc ré-inactivées suite à la restauration de la fonction DDM1. Dès lors, si une hypométhylation partielle induite par la mutation *ddm1* permet la mobilisation d'*ATCOPIA78*, celle-ci ne peut perdurer dans les epiRIL. Quant aux nouvelles copies, qui *a priori* doivent s'insérer dans un état complètement hypométhylé sauf à supposer que les ADNc sont cibles de la méthylation, elles pourraient théoriquement être la source de nouveaux évènements de mobilisation. Cependant, le fait qu'aucune insertion propre à une lignée n'ait été détectée suggère l'établissement rapide du contrôle épigénétique également sur les nouvelles insertions d'*ATCOPIA78*. Ce contrôle pourrait être soit transcriptionnel (TGS via l'établissement de la méthylation par le mécanisme de RdDM) soit post-transcripionnel (PTGS). Etant donné que cette famille d'ET est très fortement ciblée par le RdDM, on peut faire l'hypothèse d'un ciblage quasi-immédiat de ces nouvelles copies par la machinerie du RdDM. J'ai donc réalisé des expériences de McrBC-qPCR pour déterminer le niveau de méthylation de la LTR 5' de trois copies donneuses chez la plante sauvage et le mutant *ddm1* ainsi que de quatre copies nouvellement insérées d'*ATCOPIA78*, dans les générations précoces des epiRIL 260 et 454 (fig. 3.2).

En contexte sauvage, les trois copies donneuses présentent un taux variable de méthylation compris entre 40 et 80% selon les locus (fig. 2c). En accord avec les analyses effectuées à partir de données de MeDIP chip à l'échelle de la famille (partie 2), la perte de méthylation induite par la mutation *ddm1* n'est pas complète sur les trois locus. Il apparait que seul *AT3TE89830*, qui présente un niveau de méthylation relativement faible chez le sauvage perd complètement sa méthylation dans le mutant *ddm1* alors qu'*AT3TE92525* et *AT5TE15240* ne sont que très modérément affectés. Dans les epiRIL 260 et 454, *AT3TE89830* et *AT3TE92525* sont hérités du parent sauvage et présentent donc, comme attendu, des niveaux de méthylation comparables à ceux de l'individu sauvage (fig. 3.2 c et d). *AT5TE15240* a été hérité du parent *ddm1* dans ces deux epiRIL mais présente également un niveau élevé de méthylation comparable à celui observé chez l'individu sauvage (fig. 3.2d). Cette observation soutient donc l'hypothèse selon laquelle les copies résidentes

d'*ATCOPIA78* sont des séquences reméthylables. Concernant les nouvelles insertions, la méthylation de l'ADN est détectée sur chacune d'entre elle, bien qu'à des niveaux différents et parfois faibles dans les générations F4 et F5 de l'epiRIL 454 et la F5 de l'epiRIL 260 (fig. 3.2d). Dans l'epiRIL 454, aucune augmentation de la méthylation n'est observable entre la F4 et la F5 pour les deux nouvelles insertions étudiées, ce qui suggère que le niveau de méthylation atteint est définitif. Une analyse portant sur les générations plus avancées permettrait de confirmer ou non ce point. On peut également noter que l'insertion partagée par les deux epiRIL (insertion n°1) présente un niveau de méthylation équivalent dans les deux lignées. Le fait que plusieurs nouvelles insertions au sein d'une même lignée, ici la 260, montrent des niveaux de méthylation différents est en accord avec ce qui est observé pour les copies donneuses, qui elles aussi possèdent des niveaux de méthylation distincts. Il serait intéressant d'étudier s'il existe une corrélation entre le profil de méthylation d'une nouvelle insertion et celui de la copie donneuse correspondant. Cette analyse nécessite cependant de réassembler *de novo* toutes les nouvelles insertions qui sont en général des composites de plusieurs donneurs. Enfin, même si le taux de méthylation de certaines nouvelles insertions est très faible, il semble en revanche suffisant pour une extinction efficace car l'arrêt de la mobilisation s'est fait au plus tard en F2. On ne peut cependant exclure que l'arrêt de la mobilisation de certaines copies résulte pour tout ou partie de mécanismes de contrôle post-transcriptionnel.

L'analyse de générations encore plus précoces (notamment la F2) sera nécessaire afin de déterminer les premières étapes de l'établissement du contrôle épigénétique. Des techniques d'analyse de la méthylation plus précises et sensibles (comme le séquençage bisulfite) devront également être utilisées pour l'étude précise des profils de méthylation des nouvelles insertions afin notamment de pouvoir les comparer entre elles et avec les copies donneuses.

- *ATCOPIA93*

Si l'établissement du contrôle épigénétique sur les nouvelles insertions des ET de la famille *ATCOPIA78* semble rapide, le cas de l'élément *ATCOPIA93 EVADE* est manifestement plus complexe. Dans le cadre d'une collaboration avec l'équipe d'Olivier Voinnet (ETH Zurich), je me suis intéressée aux modalités conduisant à l'arrêt de la mobilisation d'*EVADE* (*ATCOPIA93*). Les résultats de ces analyses sont résumés plus bas et décrits dans l'article qui suit. Ma contribution à ce travail concerne l'étude des epiRIL 454 et 439 (Figure 3d ; Figure 4b, c, e, f, g, h ; Figure 6e, f, g, h et Figure supplémentaire 7b, c, d) qui vient en complément de l'étude de l'epiRIL 15 dérivée du mutant *met1* (Reinders et al. 2009). [Remarque : les données de la figure 6 concernant les conséquences du gain de méthylation au niveau des copies nouvellement insérées sur la régulation des gènes situés à proximité seront présentées et discutées dans la partie 4 de la thèse]. L'expression d'*EVADE* causée par son hypométhylation s'accompagne de la production de siARN de 21nt dépendant de RDR6, signature du PTGS, qui correspondent à la région 3' du gène GAG. L'augmentation du nombre de copies au cours des générations est associée à une augmentation du taux de transcrits et des siARN de 21nt, jusqu'à l'obtention d'un nombre de copies limite (2N=40). Une fois ce nombre de copies atteint, on observe systématiquement l'arrêt de la mobilisation, une chute abrupte du niveau d'ARN messagers et l'établissement rapide et concerté de la méthylation de l'ADN par le RdDM sur les LTR de toutes les copies d'*EVADE*. L'établissement de la méthylation au niveau des LTR est précédé par la méthylation de la région GAG qui était ciblée par les siARN de 21nt et qui, une fois le nombre limite de copies atteint, produit également des siARN de 24nt qui conduisent la méthylation. La cause la plus probable de cette production de siARN de 24nt est une saturation de la machinerie du PTGS due à une quantité trop importante du double brin généré par RDR6 et la reconnaissance de cet ARN double brin par DCL3 en plus de DCL1. Quoiqu'il en soit, on observe à la génération suivante la chute d'expression d'*EVADE*, l'arrêt de la production de siARN de 21nt et l'établissement de la méthylation sur les LTR (associé à la production de siARN de 24nt ciblant ces LTR).

Reconstructing *de novo* silencing of an active plant retrotransposon

Arturo Marí-Ordóñez[1,2], Antonin Marchais[1,5], Mathilde Etcheverry[3,5], Antoine Martin[3,4], Vincent Colot[3] & Olivier Voinnet[1,2]

© 2013 Nature America, Inc. All rights reserved.

Transposable elements (TEs) contribute to genome size, organization and evolution. In plants, their activity is primarily controlled by transcriptional gene silencing (TGS), usually investigated at steady states, reflecting how long-established silent conditions are maintained, faithfully reiterated or temporarily modified. How active, invasive TEs are detected and silenced *de novo* in plants remains largely unknown. Using inbred lineages of hybrid *Arabidopsis thaliana* epigenomes combining wild-type and mutant chromosomes, we have deciphered the sequence of physiological and molecular events underlying the *de novo* invasion, proliferation and eventual demise of the single-copy endogenous retrotransposon *Evadé (EVD)*. We show how this reconstructed TE burst causes widespread genome diversification and *de novo* epiallelism that could serve as sources for selectable and potentially adaptive traits.

TEs abound in plant genomes, contributing to their size and organization via successive proliferation-deletion 'bursts' (refs. 1,2) that generate functional variation upon which selection may operate[3,4]. TEs can also influence the chromatin state of neighboring genes, often in a developmental or stress-responsive manner[5], through cellular pathways that restrict gene activity primarily via promoter DNA methylation and TGS. Several mechanisms account for TGS in *A. thaliana*. In DNA methylation, 5-methylcytosine (5mC) maintenance by MET1 and CMT3 copies CG and CHG methylated patterns to daughter DNA strands[6]; *de novo* methylation at CG, CHG and CHH sites (where H is A, T and/or C) by contrast, relies on RNA-directed DNA methylation (RdDM)[7]. RdDM entails transcription of TE loci by the plant-specific RNA polymerase IV (Pol IV); the resulting RNA is copied by RNA-dependent RNA polymerase 2 (RDR2) into double-stranded RNA (dsRNA) and processed by Dicer-like 3 (DCL3) into 24-nt small interfering RNAs (siRNAs). Upon loading into the silencing effector ARGONAUTE 4 (AGO4), siRNAs guide *de novo* methyltransferases by recruiting AGO4 to scaffold RNA produced by the distinct RNA polymerase V (Pol V). The methylated DNA is thought to further attract RNA Pol IV at TEs in a self-reinforcing mechanism believed to operate mainly in *cis*[7]. DDM1, a SWI/SNF chromatin remodeler required for maintenance, facilitates access of DNA methyltransferases to heterochromatin[8].

Plant TE silencing has been mostly studied at steady states, reflecting how long-established silent conditions are maintained (for example, through MET1 and DDM1) or faithfully reiterated (for example, through RdDM). Such pre-established states may be temporarily reversed by stress or during development, correlating with the production of TE-derived 21-nt siRNAs that possibly reinforce TE suppression via post-transcriptional gene silencing (PTGS or RNA interference, RNAi)[2,9–12]. However, these studies have focused on evolutionarily ancient, strongly proliferated and degenerated transposon families, which are highly similar in sequence and are mostly clustered in centromeric and pericentromeric regions, acting as potential reservoirs for homology-based silencing[13,14]. Consequently, how new TEs with no homology to host genomes are detected and silenced *de novo* remains largely unknown, as are the genomic and epigenomic effects of their proliferation.

Ideally, investigating this process entails the reconstruction of *de novo* TE invasion-silencing events. Use of TE-based transgenes[15] for this end has generated confounding results, given the intrinsic susceptibility of transgenes to RNAi and TGS, artificial transposition patterns and lack of evolutionary context. Outbreeding can circumvent these caveats[16] but is limited, owing to the scarce genetic resources available in wild accessions. Endogenous transposition bursts were reconstructed in *A. thaliana met1* and *ddm1* single mutants or by their combination with RdDM mutants[17,18]. Reactivated TEs included members of the evolutionarily young, low-copy *ATCOPIA93* family of long-terminal repeat (LTR) retrotransposons, which proliferate by reverse transcription of RNA intermediates. Studying these bursts provided insights into the requirements and genomic consequences of *de novo* TE invasion. However, transgenerational studies, needed to explore the proliferation of reactivated TEs, are hindered by the developmental defects and sterility of inbred *met1* and *ddm1*

[1]Swiss Federal Institute of Technology (ETH-Z), Department of Biology, Zurich, Switzerland. [2]Life Science Zurich Graduate School, Molecular Life Science Program, University of Zurich, Zurich, Switzerland. [3]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche (UMR) 8197 and Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, Paris, France. [4]Present address: Institut de Biologie Intégrative des Plantes (IBIP), Biochimie et Physiologie Moléculaire des Plantes, UMR 5004 CNRS, Institut National de la Recherche Agronomique (INRA), SupAgro, UM2, Montpellier, France. [5]These authors contributed equally to this work. Correspondence should be addressed to O.V. (voinneto@ethz.ch).

Received 11 April; accepted 20 June; published online 14 July 2013; doi:10.1038/ng.2703

single-mutant plants. The continued presence of these mutations also intrinsically precludes analyses of *de novo* TE silencing following their reactivation[19].

Recently, populations of epigenetic recombinant inbred lines (epiRILs) were derived from crosses between wild-type and *met1* or *ddm1* homozygous *A. thaliana*[20,21]. Inbreeding epiRILs by single-seed descent in the presence of wild-type copies of *MET1* or *DDM1* generated collections of lines with mosaic epigenomes consisting of wild-type and methylation-depleted segments. We reasoned that epiRILs might provide a unique resource to explore transgenerationally the full biology of reactivated low-copy TEs without the complications noted above.

RESULTS

De novo genome invasion by a single-copy LTR retroelement

In *met1-3* inbred plants, the mutant used to generate the original epiRIL population, loss of DNA methylation allowed mobilization of an intact *ATCOPIA93* family representative, *EVD*[18], defining only two single-insertion loci in the *A. thaliana* Col-0 genome: *AT5G17125*, a euchromatic locus corresponding to *EVD* itself, and the pericentromeric, heterochromatic locus *AT1G34967*, termed *Attrapé* (*ATR*). Several *met1* epiRIL lines showed *EVD* mobilization[18], and we confirmed that, in epiRIL line 15 (epi15), *EVD* reactivation correlated with the unmethylated status of its 5′ LTR, inherited from the *met1* parent. By contrast, the *ATR* 5′ LTR showed wild-type DNA methylation levels in epi15; moreover, only *EVD* transcripts were detected in both the *met1* parent and in epi15 (**Fig. 1a**,**b** and **Supplementary Fig. 1**), suggesting that MET1-independent mechanisms restrict *ATR* expression. *EVD* mobilization was then followed over several generations of epi15 inbreeding, starting from the eighth generation (F_8) at which epiRILs were available[20,21]. Southern blot analysis showed a considerable increase in *EVD* copy number and accumulation of extrachromosomal DNA between generations F_8 and F_{11} (**Fig. 1c**). Thus, by overcoming the complications seen in *met1* and *ddm1* single-mutant inbred plants and circumventing transgenesis, the epi15 lineage provided the closest approximation to a *de novo* genomic invasion event and was ideally suited to study the proliferation, detection and eventual demise of the single euchromatic *EVD* copy of *A. thaliana* Col-0.

EVD transcription occurs in the single L2 cell layer

Despite the increasing *EVD* copy number, morphological defects were infrequent in epi15, as reported for other *met1* epiRILs (**Fig. 1d**,**e**)[18].

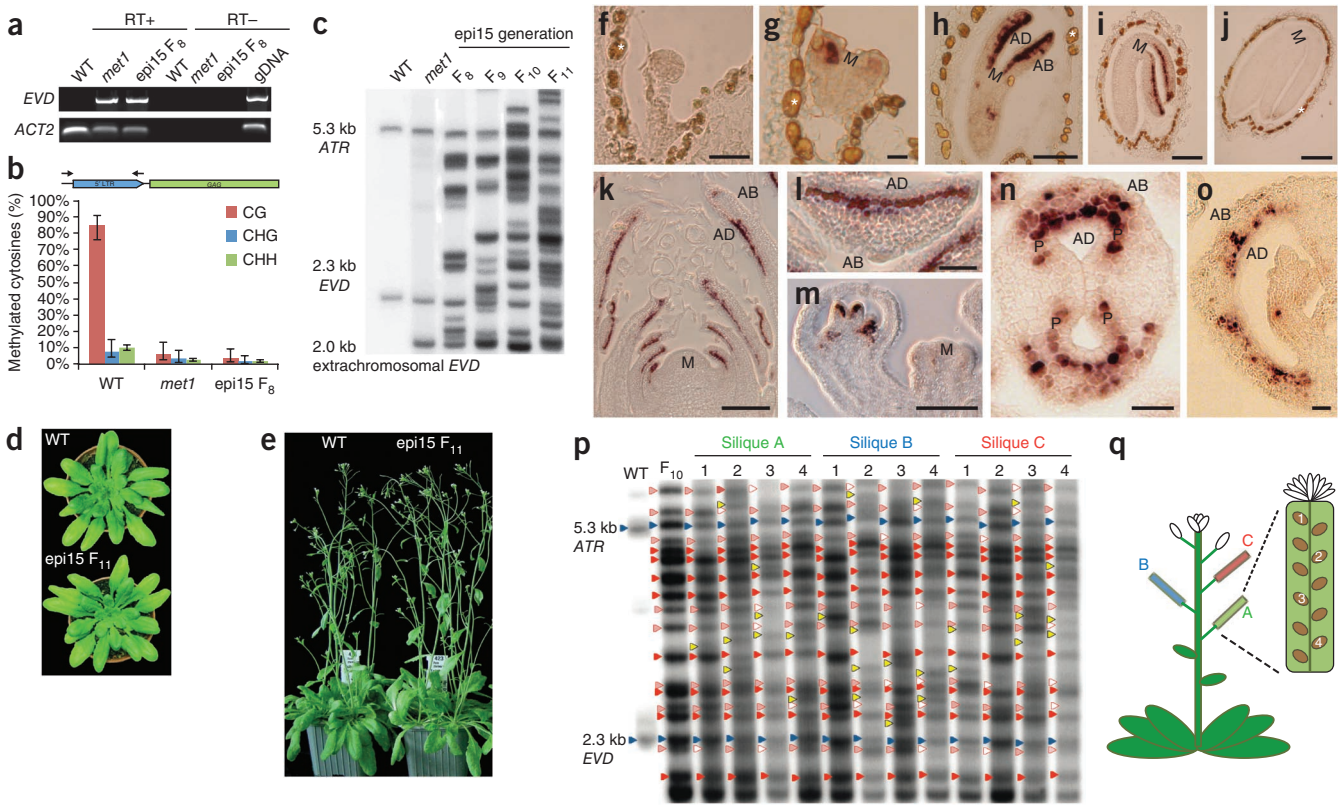


Figure 1 *EVD* transcript accumulation pattern. (**a**) RT-PCR analysis of *EVD* RNA levels in wild-type, *met1* and epi15 F_8 plants, normalized to *ACTIN2* (*ACT2*) levels. WT, wild type; RT, reverse transcriptase; gDNA, genomic DNA. (**b**) Bisulfite sequencing–based methylation analysis at CG, CHG and CHH sites of the *EVD* 5′ LTR in wild-type, *met1* and epi15 F_8 plants (arrows indicate the location of primers used for PCR amplification after bisulfite treatment of DNA; details in **Supplementary Fig. 1** and **Supplementary Table 1**; error bars, 95% confidence Wilson score intervals). (**c**) Southern blot analysis of *EVD* transposition in epi15 generations. (**d**,**e**) Wild-type and epi15 F_{11} plants during vegetative (**d**) or reproductive (**e**) growth. (**f**–**o**) *In situ* hybridization of *EVD* RNA in an F_{10} plant. (**f**–**j**) Longitudinal sections of embryos at globular (**f**), heart (**g**), mid-torpedo (**h**), late-torpedo (**i**) and mature green (**j**) stages. (**k**) Longitudinal section of the SAM. (**l**) Transversal section of a young leaf. (**m**) Longitudinal section of an inflorescence meristem and floral primordium at floral stage 7. (**n**) Gynoecium transversal section at floral stage 8. (**o**) Longitudinal section of ovule primordia at floral stage 12. M, meristem; AD, adaxial; AB, abaxial; P, placenta; nonspecific labeling of the seed coat is marked (asterisk). Scale bars, 20 μm for **f**,**g**,**l**, 100 μm for **h**–**k**, 50 μm for **m** and 10 μm for **n**,**o**. (**p**) Southern blot analysis of *EVD* transposition in siblings of a single F_{10} plant. Blue arrows, original *EVD* and *ATR* insertions; red arrows, bands present in both the parental plant and progeny; pale red arrows, bands present in the parental plant segregating in the progeny; yellow arrows, bands present in the progeny but absent in the parental plant. (**q**) Schematic of organs used in **p**. A–C depict individual fruits (siliques).

© 2013 Nature America, Inc. All rights reserved.

To gain insight into the *EVD* life cycle in its host, we investigated *EVD* RNA accumulation patterns in epi15 F_{10} individuals using *in situ* hybridization (**Fig. 1f–o**). Undetectable in globular-stage embryos (**Fig. 1f**), *EVD* RNA was restricted to the adaxial subepidermal (L2) layer of cotyledons at all subsequent stages of embryonic development (**Fig. 1g–i**) until dormant seed formation occurred (**Fig. 1j**). L2-specific *EVD* RNA was also detected during postembryonic growth, mostly in young developing leaves (**Fig. 1k,l**), but remained undetectable in the shoot apical meristem (SAM) during embryonic (**Fig. 1g–j**) and adult (**Fig. 1k–o**) growth. The SAM, set early in embryogenesis (**Fig. 1f,g**), continuously generates stem-cell niches from which all aerial organs derive. Floral meristems also lacked *EVD* RNA (**Fig. 1m**), but this RNA accumulated in the L2 of young gynoecia—the modified foliar organs bearing the ovules[22,23] (**Fig. 1m,n**). In particular, *EVD* RNA was detected in the four placental regions (**Fig. 1n**), from each of which one L2 cell differentiates into a megaspore mother cell undergoing meiosis to form the haploid gametophyte[24] (**Supplementary Fig. 2**). *EVD* RNA was, however, undetectable later in developing ovules (**Fig. 1o**). Unlike in wild-type plants with no hybridization signals, these patterns were seen identically in two other independent *met1* epiRILs and were recapitulated by a transcriptional reporter fusion to the *EVD* 5′ LTR promoter (**Supplementary Fig. 3**). L2-specific *EVD* transcription and meristem exclusion thus explained why somatic retrotransposition seemed developmentally inconsequential to the epi15 lineage. The prefertilization transience of *EVD* germline expression also likely delineated a discrete spatiotemporal window for transmitting new insertions to progenies. Because each gametophyte originates from a distinct placental L2 cell, individual seeds of the same fruits (siliques) were predicted to display distinct numbers and genomic patterns of *EVD* insertion. This prediction was confirmed by Southern blot analyses in F_{10} plants, unraveling an extensive genetic diversification process (**Fig. 1p,q**).

Antiviral PTGS forms the first host defense against *EVD*

Correlating with increased numbers of *EVD* insertions and higher *EVD* expression, *EVD*-derived small RNA (sRNA) levels increased between generations F_8 and F_{11} in epi15 plants and remained below detection in wild-type plants (**Fig. 2a–d**). *EVD*-derived sRNAs were in both sense and antisense orientations, were predominantly 21–22 nt long and were mostly derived from the 3′ end of *gag* (3′ *gag*) in a symmetrical pattern evoking processing of a long double-stranded RNA (dsRNA) (**Fig. 2c,d**)[25,26]. However, *EVD* RNA secondary structure prediction was not able to identify such a precursor, suggesting that it was perhaps a reverse transcription or retrotransposition byproduct. This idea was tested with an *EVD* transgene in which the ubiquitous *35S* promoter (*35S:EVD*) replaced the LTR promoter required for reverse transcription and retrotransposition. In five of ten individual *35S:EVD* T_1 transformants (wild-type Col-0), *EVD* expression correlated with the production of siRNAs of 21–22 nt in length specifically from 3′ *gag*, as was seen in epi15 plants between generations F_8 and F_{11} (**Fig. 2e**, lanes 1–3). This high incidence of PTGS and congruence of sRNA patterns suggested that long dsRNA synthesis was intrinsic to *EVD* transcription, probably involving a host-encoded RNA-dependent RNA polymerase (RDR) protein. RDR6 was an obvious candidate, as all of its known functions are coupled to DCL4 and DCL2 activities, which produce 21-nt and 22-nt siRNAs, respectively[27]. Introducing *35S:EVD* into the *rdr6-12* null background indeed markedly decreased 3′ *gag* siRNA accumulation (**Fig. 2e**, lanes 4 and 5); the remaining primary siRNAs were eliminated in *dcl2/dcl4* double mutants (**Fig. 2e**, lanes 5–7). Therefore, *EVD* was subjected to the same DCL4-, DCL2- and RDR6-dependent siRNA pathway that targets viruses for PTGS[28,29].

The Gag nucleocapsid protein protects *EVD* against PTGS

Although abundant *EVD*-derived siRNAs were efficiently loaded into their cognate effectors, AGO1 and AGO2, from generation F_8 onward (**Fig. 2f**), *EVD* expression and mobilization kept increasing from F_8 to F_{11}, as if *EVD* resisted PTGS. Successful viral infections in plants, in spite of the presence of abundant virus-derived siRNAs, rely on the production of *trans*-acting viral suppressors of RNA silencing (VSRs)[30]. To test whether *EVD* encodes a VSR, we subjected *35S:EVD* to a silencing suppression assay in *Nicotiana benthamiana*[31] in which a transiently expressed transgene encoding green fluorescent protein (*GFP*) is both a trigger and a target of PTGS, which is achieved within 4 d. Coexpression of P19, a VSR that binds 21-nt siRNAs, strongly enhances *GFP* mRNA accumulation in this assay, unlike coexpression of mutant $P19^m$ that is deficient in siRNA binding (**Fig. 2g**, lanes 1 and 2, and **Supplementary Fig. 4**). *GFP* mRNA levels remained low upon coexpression of *35S:EVD* (**Fig. 2g**, lane 3), despite the production of cognate *EVD* polyprotein products (**Supplementary Fig. 5**), arguing against an *EVD*-encoded *trans*-acting VSR.

Notably, *EVD* transcripts were still detected 4 d and up to 8 d after infection, despite high siRNA accumulation, whereas *GFP* mRNA levels were below detection at these time points, even if coexpressed with P19 (**Fig. 2g**, lane 3 and **Supplementary Fig. 4b**). Our search for an *EVD*-encoded *cis*-acting factor possibly explaining this resistance of *EVD* mRNAs to PTGS focused on Gag, which contributes the structural proteins of virus-like particles (VLPs) required for coating retrovirus and/or retroelement RNA[32]. In *Saccharomyces cerevisiae*, VLP coating protects Ty3 LTR retrotransposon RNA from endonucleases via the conserved zinc-finger CCHC motif ($CX_2CX_4HX_4C$, where X is any amino acid) of the Gag nucleocapsid RNA-binding domain (RBD)[33]. The *EVD*-encoded Gag CCHC motif was readily identified by sequence alignment (**Supplementary Fig. 5e**)[34], and the underlying sequence was removed from *35S:EVD* by an in-frame deletion, leading to $35S{:}EVD^{\Delta RBD}$ (**Fig. 2g**). Having confirmed cognate $EVD^{\Delta RBD}$ polyprotein processing and $Gag^{\Delta RBD}$ stability (**Supplementary Fig. 5b,c**), $35S{:}EVD^{\Delta RBD}$ was introduced into plants with or without expression of P19 or non-functional $P19^m$. Without P19, $EVD^{\Delta RBD}$ mRNA levels were strongly reduced compared to those of wild-type *EVD* that remained unaltered irrespective of P19 or $P19^m$ (**Fig. 2h**, lanes 3 and 4 and lanes 5 and 6). In contrast, mRNA from $EVD^{\Delta RBD}$ reached wild-type levels with expression of P19 but not $P19^m$ (**Fig. 2h**, lanes 7 and 8). Therefore, the Gag nucleocapsid RBD enables *cis* protection of *EVD* RNA against PTGS; this property probably facilitated *EVD* proliferation over multiple generations in the epi15 lineage, despite the strong accumulation of 21- to 22-nt siRNAs resulting from transcription of new insertions.

Effective *EVD* silencing occurs at a threshold of ~40 copies

Transgenerational investigation of the biology of *EVD* in the inflorescences of bulked individuals showed that *EVD* RNA levels and copy numbers reached their peaks at generations F_{11} and F_{12}, respectively, coinciding with the highest 3′ *gag*–derived siRNA levels (**Fig. 3a,b**). *EVD* RNA levels declined beyond generation F_{11}, until their pronounced loss in generation F_{14}, a loss that was not attributable to reduced numbers of *EVD* insertions, as a plateau of ~40 copies per genome was attained from generation F_{12} onward (**Fig. 3a,b**). F_{14} plants, however, had strongly reduced 3′ *gag*–derived siRNA levels (**Fig. 3b** and **Supplementary Fig. 6a**). Because the accumulation of these siRNAs requires *EVD* transcription by RNA polymerase II (Pol II), this suggested that *EVD* had undergone TGS in generation F_{14}; accordingly, F_{14} coincided with the onset of LTR-derived 24-nt siRNAs (**Fig. 3b,c**). Initially undetectable

© 2013 Nature America, Inc. All rights reserved.

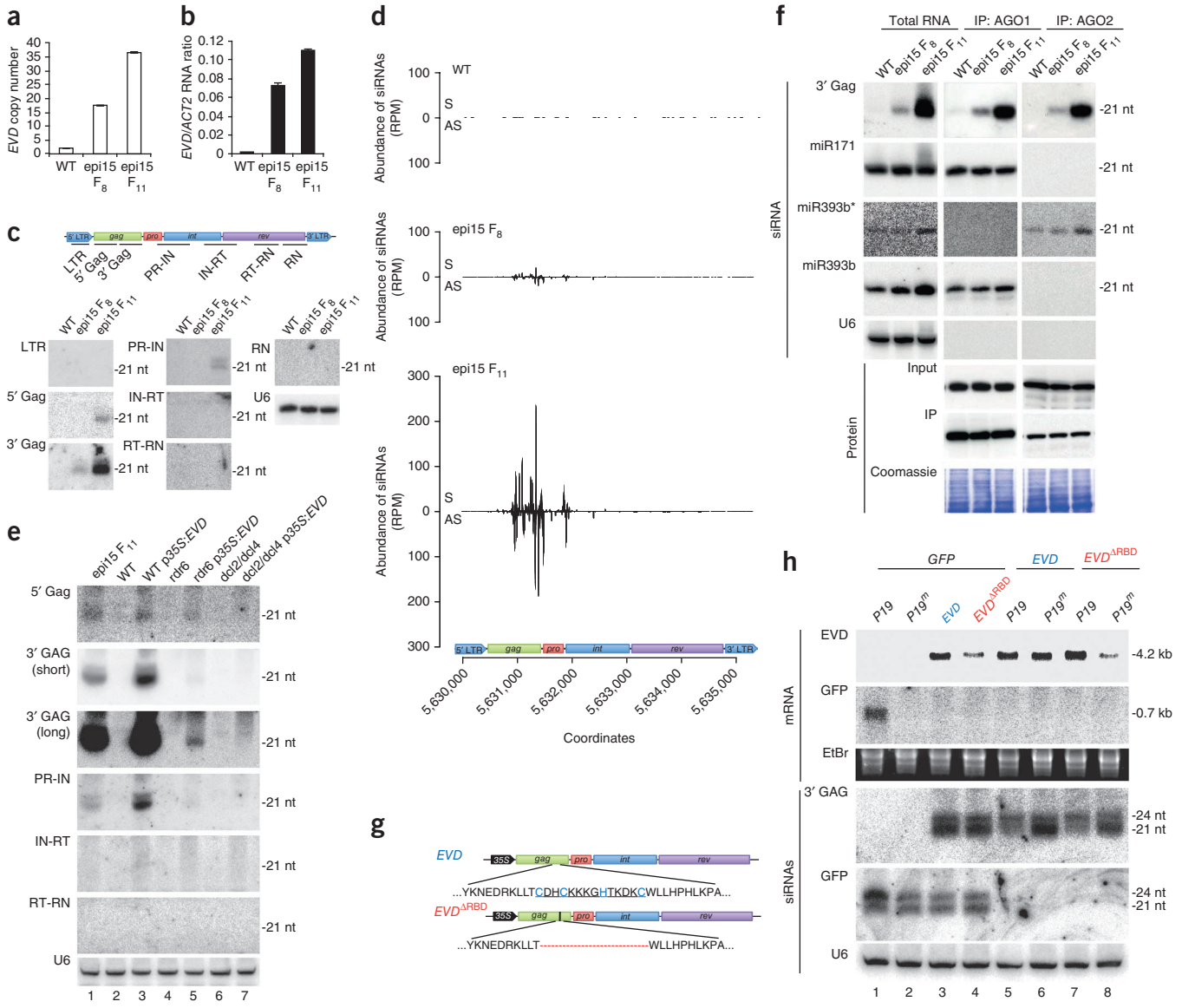


Figure 2 Antiviral PTGS is activated against and suppressed by *EVD*. (**a**) *EVD* copy number analysis by quantitative RT-PCR (qRT-PCR) in wild-type, epi15 F_8 and epi15 F_{11} plants. (**b**) qRT-PCR analysis of *EVD* RNA levels in wild-type, epi15 F_8 and epi15 F_{11} plants, normalized to *ACT2* levels. Error bars, s.d. in **a**,**b**. (**c**) RNA blot analysis of *EVD* siRNAs in wild-type, epi15 F_8 and epi15 F_{11} plants. Distinct probes mapping *EVD* to the genome were used. U6, RNA loading control. (**d**) Mapping of *EVD* 21-nt and 22-nt deep sequencing siRNA reads (same RNA as in **c**) onto the *EVD* genome. The *y* axis shows normalized reads per million sequences (RPM), and the *x* axis shows chromosome 5 coordinates at the *EVD* locus. S, sense; AS, antisense. (**e**) RNA blot analysis of *EVD* siRNA levels of *35S:EVD* in wild-type, *rdr6* or *dcl2/dcl4* background. epi15 F_{11} RNA was used as a reference; the same probe set was used as in **c**. U6, RNA loading control. (**f**) RNA blot analysis of *EVD* siRNAs in AGO1 and AGO2 immunoprecipitates (IP) using the 3′ Gag probe as in **c**. miR171 (AGO1), miR393b (AGO1) and miR393b* (AGO2) were probed to validate specific sRNA enrichment in cognate AGOs. U6, control for RNA carryover from cleared immunoprecipitation extracts. Protein blot analysis of input and immunoprecipitation fractions is shown, along with Coomassie staining of total protein. (**g**) Schematics of *35S:EVD* and *35S:EVD*$^{\Delta RBD}$. The encoded CCHC motif is highlighted in blue. (**h**) RNA blot analyses 4 d after infection in *N. benthamiana* leaves. Lanes 1–4, *GFP* silencing suppression assay; coexpression of *35S:GFP* with *35S:P19* or *35S:P19*^{m} provides positive and negative controls, respectively (**Supplementary Fig. 2**). Lanes 5–8, *EVD* and *EVD*$^{\Delta RBD}$ mRNA and siRNA levels with or without coexpression of *35S:P19* or *35S:P19*^{m}. EtBr, ethidium bromide staining of rRNA; U6, sRNA loading control.

between generations F_8 and F_{11}, these species also accumulated in F_{12} and F_{13}, albeit less than in F_{14}, coinciding with reduced 3′ *gag*–derived siRNAs (**Fig. 3b**) and *EVD* insertions reaching ~40 copies per genome.

To address whether this sudden TGS was epi15 specific or instead reflected a bona fide feature of *EVD* biology, we used two separate epiRILs in the *ddm1* background[21] in which *EVD* mobilization was confirmed. In both lines epi454 and epi439, with 33 and 7 *EVD* copies in generation F_8, respectively (**Supplementary Fig. 6b**), abrupt *EVD* silencing coincided with insertions reaching ~40 copies per genome, even though a dissimilar number of generations (7 and 17, respectively) was required to attain this figure in each line (**Fig. 3d** and **Supplementary Fig. 7**). Therefore, potent TGS of *EVD* occurs at a fixed copy number threshold (~40), coinciding with the production of LTR-derived 24-nt siRNAs and concurrent loss of all 3′ *gag*-derived siRNA classes.

© 2013 Nature America, Inc. All rights reserved.

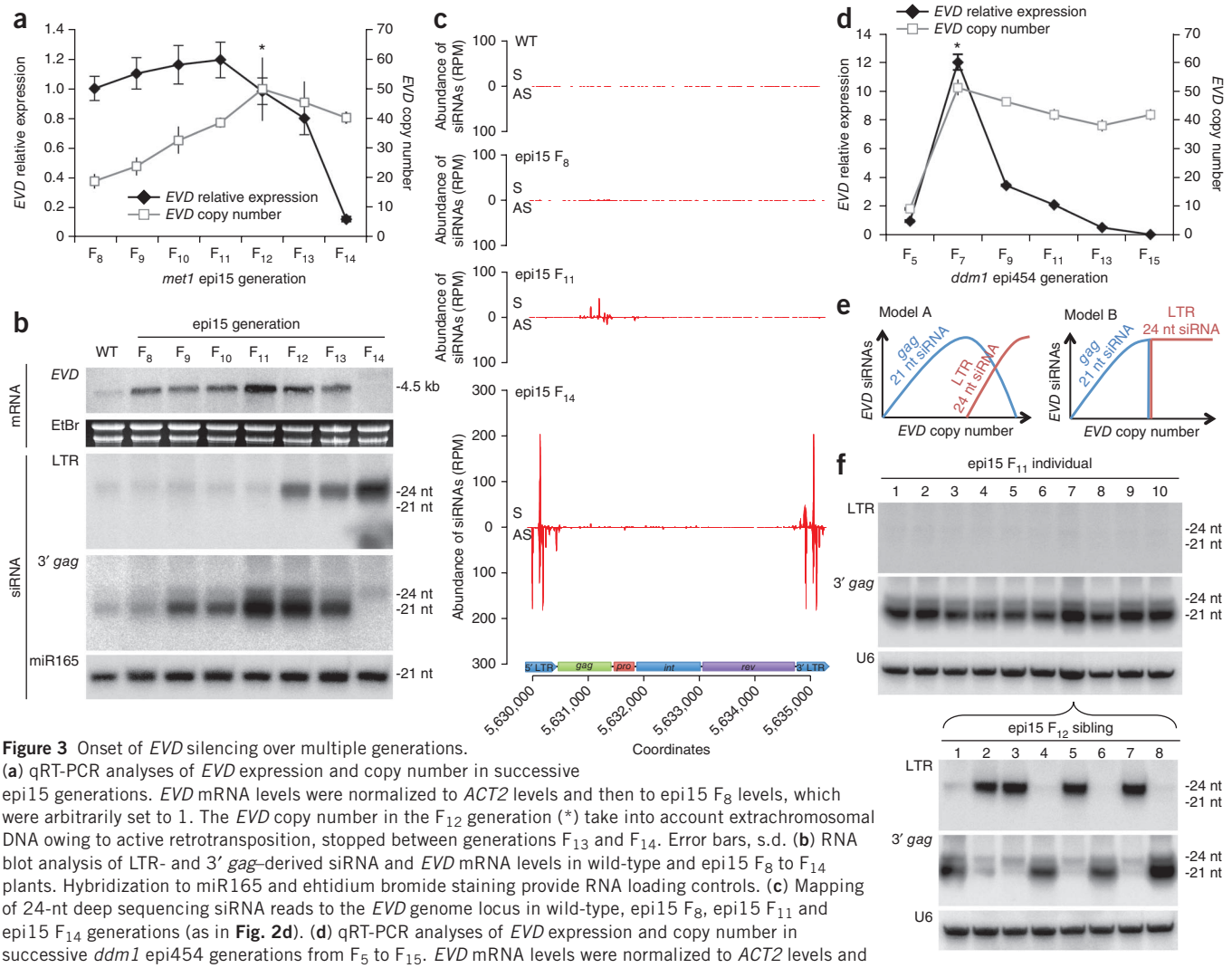


Figure 3 Onset of *EVD* silencing over multiple generations. (**a**) qRT-PCR analyses of *EVD* expression and copy number in successive epi15 generations. *EVD* mRNA levels were normalized to *ACT2* levels and then to epi15 F_8 levels, which were arbitrarily set to 1. The *EVD* copy number in the F_{12} generation (*) take into account extrachromosomal DNA owing to active retrotransposition, stopped between generations F_{13} and F_{14}. Error bars, s.d. (**b**) RNA blot analysis of LTR- and 3′ *gag*–derived siRNA and *EVD* mRNA levels in wild-type and epi15 F_8 to F_{14} plants. Hybridization to miR165 and ehtidium bromide staining provide RNA loading controls. (**c**) Mapping of 24-nt deep sequencing siRNA reads to the *EVD* genome locus in wild-type, epi15 F_8, epi15 F_{11} and epi15 F_{14} generations (as in **Fig. 2d**). (**d**) qRT-PCR analyses of *EVD* expression and copy number in successive *ddm1* epi454 generations from F_5 to F_{15}. *EVD* mRNA levels were normalized to *ACT2* levels and then to epi454 F_5 levels, which were arbitrarily set to 1. The *EVD* copy number in the F_7 generation (*) take into account extrachromosomal DNA owing to active retrotransposition. Error bars, s.d. (**e**) Two alternative models for the induction of LTR-derived 24-nt siRNAs. (**f**) RNA blot analysis of LTR- and 3′ *gag*–derived siRNAs in ten individual epi15 F_{11} plants and eight F_{12} siblings from the single F_{11} plant 7. U6, RNA loading control.

© 2013 Nature America, Inc. All rights reserved.

Two mutually exclusive *EVD* silencing states coexist in epiRILs

Our results suggest that distinct siRNA species coexisted in later epi15 generations (**Fig. 3b**), as if a progressive transgenerational increase in *EVD* copy number had instigated an accumulation of 24-nt LTR-derived siRNAs, ultimately triggering TGS and loss of 3′ *gag*–derived siRNAs in generation F_{14} (**Fig. 3e**, model A). An alternative scenario existed, however, in which occurrence of 21- to 22-nt 3′ *gag*–derived siRNAs and of 24-nt LTR-derived siRNAs represented either one of two mutually exclusive states (**Fig. 3e**, model B), reflecting differences in *EVD* copy number between individuals from the same generations (**Fig. 1q**). In most generations, including from F_{11} to F_{13}, these distinct states would have been confounded by analysis of bulked individuals. This idea was confirmed by analyzing single F_{12} siblings derived from a unique F_{11} parent (**Fig. 3f**), validating model B. Therefore, the sharp onset of silencing at F_{14} probably reflected the fact that a vast majority of plants had by this generation acquired an *EVD* copy number of ~40, which presumably underpins a threshold of overall *EVD* RNA expression above which TGS is triggered.

Many *EVD* copies concurrently undergo LTR methylation and TGS

Bisulfite DNA sequencing of *AT5G17125*, the original Col-0 *EVD* insertion, showed the coincidence of 24-nt siRNAs with LTR methylation in F_{14} but not F_8–F_{11} plants, which was even higher than in wild type and in the CHH context, the RdDM hallmark (**Fig. 4a** and **Supplementary Fig. 7a**). To test whether LTR *de novo* methylation affected all or only some *EVD* copies and correlated with a copy number of ~40, we exploited whole-genome resequencing data available for epiRILs 454 and 439 (A. Gilly, M.E., M.-A. Madoui, J. Guy, A. Martin *et al.*, unpublished data). This allowed McrBC- and quantitative PCR (qPCR)-based DNA methylation profiling of specific *EVD* insertions that were randomly spawned and genetically fixed from the epi454 and epi439 F_8 generations (ten and three insertions, respectively; **Fig. 4b** and **Supplementary Fig. 7b**).

In epi454 F_9 plants, in which ~40 *EVD* copies had been attained, 2 of 4 plants tested (plants 1 and 3) had low *EVD* RNA levels and exclusively LTR-derived 24-nt siRNAs (**Fig. 4c**,**d**), correlating with LTR methylation at all *EVD* insertions inspected (**Fig. 4e**,**f** (replicates 1 and 3)); in contrast, the two other F_9 plants (plants 2 and 4) accumulated only

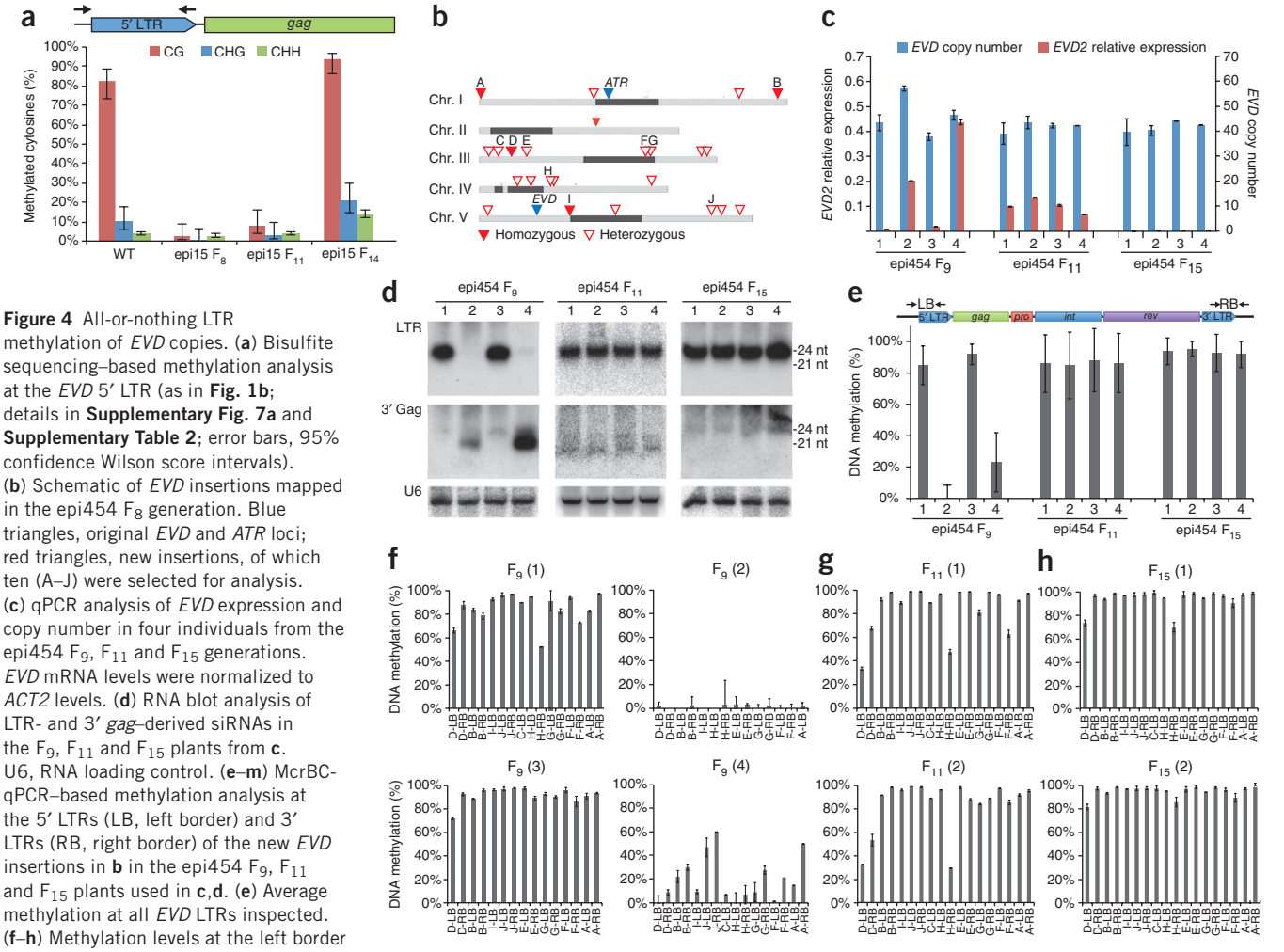


Figure 4 All-or-nothing LTR methylation of *EVD* copies. (**a**) Bisulfite sequencing–based methylation analysis at the *EVD* 5′ LTR (as in **Fig. 1b**; details in **Supplementary Fig. 7a** and **Supplementary Table 2**; error bars, 95% confidence Wilson score intervals). (**b**) Schematic of *EVD* insertions mapped in the epi454 F_8 generation. Blue triangles, original *EVD* and *ATR* loci; red triangles, new insertions, of which ten (A–J) were selected for analysis. (**c**) qPCR analysis of *EVD* expression and copy number in four individuals from the epi454 F_9, F_{11} and F_{15} generations. *EVD* mRNA levels were normalized to *ACT2* levels. (**d**) RNA blot analysis of LTR- and 3′ *gag*–derived siRNAs in the F_9, F_{11} and F_{15} plants from **c**. U6, RNA loading control. (**e**–**m**) McrBC-qPCR–based methylation analysis at the 5′ LTRs (LB, left border) and 3′ LTRs (RB, right border) of the new *EVD* insertions in **b** in the epi454 F_9, F_{11} and F_{15} plants used in **c**,**d**. (**e**) Average methylation at all *EVD* LTRs inspected. (**f**–**h**) Methylation levels at the left border (5′ LTR) and right border (3′ LTR) of specific *EVD* insertions at generations F_9 (4 replicate experiments, **f**), F_{11} (2 replicate experiments, **g**) and F_{15} (2 replicate experiments, **h**). Error bars, s.d. in **c**,**e**–**h**.

© 2013 Nature America, Inc. All rights reserved.

3′ *gag*–derived siRNAs, had higher *EVD* expression (**Fig. 4c**,**d**) and had low or no LTR methylation (**Fig. 4e**,**f** (replicates 2 and 4)). In the F_{11} and F_{15} generations, which had low or undetectable *EVD* RNA levels, all examined plants contained exclusively LTR-derived 24-nt siRNAs (**Fig. 4c**,**d**), with LTR methylation seen at all *EVD* copies inspected (**Fig. 4e**,**g**,**h**). Analyses of epi439 provided notably similar results (**Supplementary Fig. 7b**–**d**), which, combined with those shown in **Figure 3**, indicate that *EVD* silencing entails a PTGS-to-TGS shift, whereby many *EVD* copies accumulated over generations simultaneously undergo an 'all-or-nothing' methylation process mediated by LTR-derived 24-nt siRNAs.

3′ *gag* DNA methylation precedes the PTGS-to-TGS shift

We wanted to determine what mechanism could connect the *EVD* copy number of ~40 to the PTGS-to-TGS shift in *EVD* silencing. We first considered that the gain in *EVD* copies across generations had favored the emergence, by position effects or rearrangements, of *EVD* loci producing disproportionate amounts of 24-nt siRNAs with potent *trans* TGS properties. This idea, however, was ruled out in segregation analyses of an F_{14} backcross (**Supplementary Fig. 8a**). Further mining of deep sequencing data showed that, in the F_{11} generation but not in the F_8 generation, 3′ *gag*–derived 21- to 22-nt siRNAs were mixed with a much smaller proportion of 24-nt siRNAs, whose levels were strongly reduced in the F_{14} generation when TGS of *EVD* had been achieved (**Fig. 3c**). Thus, as for the 21- to 22-nt species, 3′ *gag*–derived 24-nt siRNA accumulation was positively correlated with increased *EVD* copy number and Pol II–dependent *EVD* transcript levels. In the F_{11} generation, these siRNAs were detected in immunoprecipitates of AGO4, the cognate effector for 24-nt siRNAs; this signal had dropped by the F_{14} generation, and an LTR-derived 24-nt siRNA signal was detected instead, correlating with loss of Pol II–dependent *EVD* transcription and TGS (**Fig. 5a**).

The processing patterns of 21- to 22-nt and 24-nt 3′ *gag*–derived siRNAs in the F_{11} generation were notably superimposable (**Fig. 5b**). The most abundant 24-nt siRNAs indeed invariably corresponded with versions of the 21-nt species that were 2 to 3 nt longer (**Fig. 5b** and **Supplementary Fig. 8b**), implying that these species are processed from the same RDR6-dependent dsRNA precursor, albeit by distinct DCL proteins. Because amplified 24-nt siRNAs are usually produced by DCL3 from RDR2-generated but not from RDR6-generated dsRNA, we first considered whether DCL3 levels had increased between the F_8 and F_{11} generations, potentially granting DCL3 access to RDR6 products. Protein blot analyses disproved this idea, however, also confirming unaltered DCL4 levels (**Fig. 5c**). Therefore, the most parsimonious explanation is that DCL3 could effectively access *EVD*-derived dsRNA in the F_{11} generation owing to the exceptionally large increase

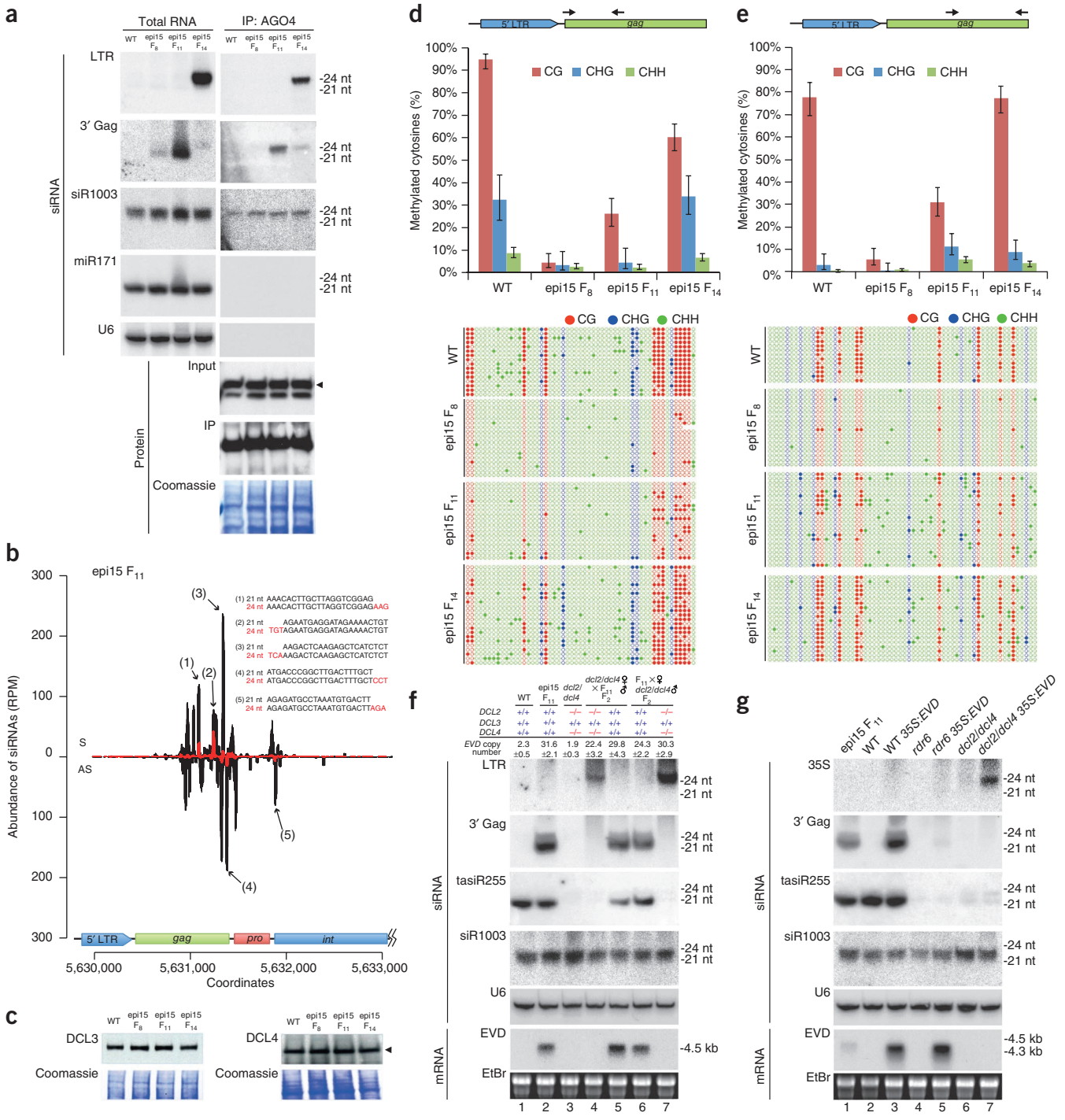


Figure 5 RDR6-DCL3–dependent RdDM triggers *EVD* TGS. (**a**) RNA analysis of LTR- and 3′ *gag*–derived siRNAs in AGO4 immunoprecipitations. siR1003 (AGO4) and miR171 (AGO1) were probed to validate specific sRNA enrichment with AGO4. U6, control for RNA carryover from cleared immunoprecipitation extracts. Protein blot analysis of input and immunoprecipitation fractions is shown (the arrowhead indicates the AGO4 band in input samples), as is Coomassie staining of input protein. (**b**) Comapping of deep sequencing siRNA reads of 21–22 nt (black) and 24 nt (red) in length in epi15 F_{11} plants (as in **Figs. 2d** and **3c**). Inset, sequence alignment of the five most abundant 24-nt siRNAs (peaks 1–5) with overlapping 21-nt siRNA reads. (**c**) Protein blot analysis of endogenous DCL3 and DCL4 in wild-type and epi15 F_8, epi15 F_{11} and epi15 F_{14} plants. Coomassie staining was performed as in **a**. (**d**,**e**) Top, bisulfite methylation analysis at *EVD* 5′ (**d**) and 3′ (**e**) *gag* regions (**Supplementary Table 3**; error bars, 95% confidence Wilson score intervals). Primers could not differentiate *EVD* from *ATR*; thus, wild-type values are for both elements and include new *EVD* insertions in epi15. Bottom, dot-plot methylation pattern of individual clones. Filled circles, methylated cytosine; empty circles, unmethylated cytosine. (**f**) RNA analysis of *EVD* siRNA and mRNA levels in segregating populations of an epi15 F_{11} cross to the *dcl2/dcl4* double mutant, using female (♀) or male (♂) progenitors. Mutant versus wild-type status of *DCL4* and *DCL2* is indicated alongside the average copy number (*EVD* copy number ± s.d.) displayed by pools of two plants with the same genotype. (**g**) RNA blot analysis of *35S*- and 3′ *gag*–derived *EVD* siRNA and *EVD* mRNA levels in *35S:EVD* plants in the wild-type, *rdr6* or *dcl2/dcl4* background. tasiR255 and siRNA1003 detection was used to validate cognate mutant backgrounds.

© 2013 Nature America, Inc. All rights reserved.

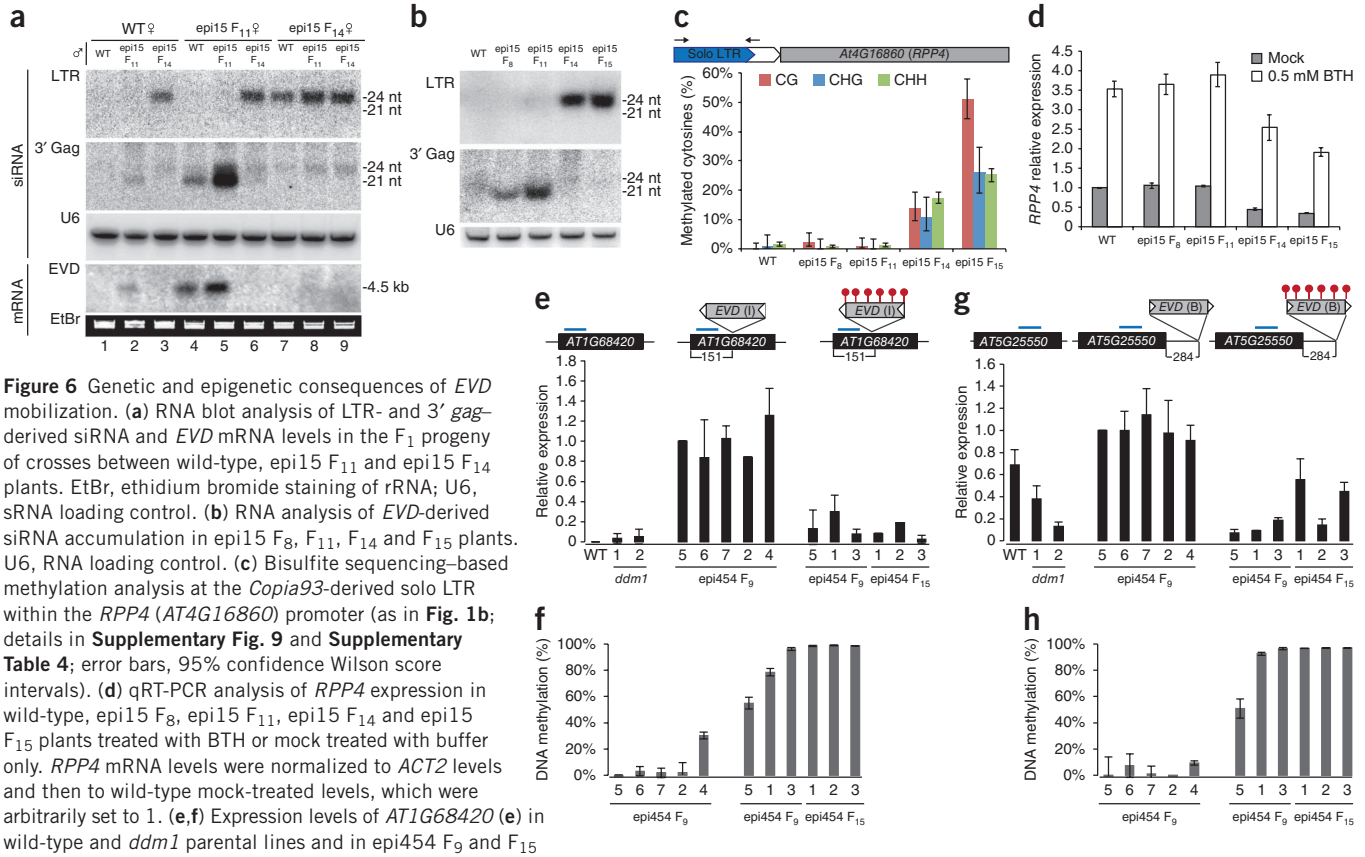


Figure 6 Genetic and epigenetic consequences of *EVD* mobilization. (**a**) RNA blot analysis of LTR- and 3′ *gag*–derived siRNA and *EVD* mRNA levels in the F_1 progeny of crosses between wild-type, epi15 F_{11} and epi15 F_{14} plants. EtBr, ethidium bromide staining of rRNA; U6, sRNA loading control. (**b**) RNA analysis of *EVD*-derived siRNA accumulation in epi15 F_8, F_{11}, F_{14} and F_{15} plants. U6, RNA loading control. (**c**) Bisulfite sequencing–based methylation analysis at the *Copia93*-derived solo LTR within the *RPP4* (*AT4G16860*) promoter (as in **Fig. 1b**; details in **Supplementary Fig. 9** and **Supplementary Table 4**; error bars, 95% confidence Wilson score intervals). (**d**) qRT-PCR analysis of *RPP4* expression in wild-type, epi15 F_8, epi15 F_{11}, epi15 F_{14} and epi15 F_{15} plants treated with BTH or mock treated with buffer only. *RPP4* mRNA levels were normalized to *ACT2* levels and then to wild-type mock-treated levels, which were arbitrarily set to 1. (**e**,**f**) Expression levels of *AT1G68420* (**e**) in wild-type and *ddm1* parental lines and in epi454 F_9 and F_{15} plants carrying the new insertion I (**Fig. 4d**) in the unmethylated or methylated state as assessed by McrBC-qPCR (as in **Fig. 4e**) (**f**). (**g**,**h**) Expression (**g**) and methylation (**h**) were determined as in **e**,**f** but with *AT5G25550*, carrying the 3′ proximal *EVD* insertion B (**Fig. 4b**). *AT1G68420* and *AT5G25550* mRNA levels were normalized to those of *AT5G13440*, *AT2G36060* and *AT4G29130*. Blue lines, probes for qRT-PCR. Error bars, s.d. in **d**–**h**.

in the concentration of this molecule, which likely overwhelmed the normal DCL4 and DCL2 processing machineries. This increase in concentration can be inferred from a matching, pronounced increase in 3′ *gag*–derived siRNA levels between the F_8 and F_{11} generations (**Figs. 2d** and **5a**). As anticipated from their loading onto AGO4, 3′ *gag*–derived 24-nt siRNAs induced *de novo* methylation of 3′ *gag* DNA in all cytosine contexts in F_{11} plants; also as expected, this methylation was absent in F_8 plants (**Fig. 5d**,**e**). Therefore, coinciding with the *EVD* copy number threshold of ~40, PTGS of *EVD* had acquired a DNA-level component via *de novo* 3′ *gag* DNA methylation.

3′-to-5′ spreading of methylation initiates TGS of *EVD*

RDR6-DCL3–dependent 3′ *gag de novo* methylation in the F_{11} generation and beyond probably constituted a key epigenetic trigger for the eventual demise of *EVD*. This hypothesis predicted that introducing a low *EVD* copy number into the *dcl2/dcl4* double-mutant background would accelerate TGS onset by enabling RDR6-dependent dsRNA to be processed solely by DCL3, as predicted from the hierarchical interactions linking siRNA-generating DCL proteins[25,28,35–38]. We used plants in which an introgressed *dcl2/dcl4* double mutation and F_{11}-derived active *EVD* insertions were allowed to segregate. Either wild-type plants or double homozygotes mutant for both *dcl4* and *dcl2* were selected on the basis of an *EVD* insertion number ranging from 20–30 copies per genome; pools of 2 plants with the same genotype were then analyzed. As expected, all plants with the wild-type background accumulated *EVD* RNA and 3′ *gag*–derived siRNAs exclusively (**Fig. 5f**, lanes 5 and 6). Notably, neither 3′ *gag*–derived siRNAs nor *EVD* RNA accumulated in plants with the *dcl2/dcl4* double-mutant genotype, which by contrast had only LTR-derived 24-nt siRNAs (**Fig. 5f**, lanes 4 and 7). Therefore, allowing the sole action of DCL3 on *EVD*-derived dsRNA was sufficient to trigger rapid and potent TGS of a low number of *EVD* copies.

A second anticipated result implicating 3′ *gag*–derived siRNAs as having a key role in TGS onset was the existence of a 3′ *gag*—to–5′ LTR methylation spreading process, which was indeed supported by a gain in DNA methylation in all cytosine contexts in the 5′ *gag* region in the F_{14} generation but not in the F_{11} generation (**Fig. 5d**). If 3′-to-5′ methylation spreading were sufficient for TGS initiation, it would be predicted to impair *EVD* transcription, regardless of the nature of its regulatory sequences, an idea that was tested with the *35S*:*EVD* lines (**Fig. 2e**). Indeed, *EVD* transcription was abolished upon introduction of *35S*:*EVD* in the *dcl2/dcl4* double mutant but not in the *rdr6* single mutant; this finding coincided with abundant amounts of *35S* promoter–derived 24-nt siRNAs triggering transgene TGS (**Fig. 5g**).

Expression changes and epiallelism at *EVD*-proximal loci

The all-or-nothing methylation process (**Fig. 4**) suggested that LTR-derived 24-nt siRNAs could silence in *trans* active *EVD* copies, which we tested in reciprocal crosses between wild-type, epi15 F_{11} and epi15 F_{14} plants. When introduced into wild-type plants, F_{14}-derived copies remained silenced in F_1 progeny, in which lower levels of LTR-derived 24-nt siRNAs correlated with the expected halved *EVD* copy number (**Fig. 6a**, lanes 3 and 7 versus 9, and **Supplementary Fig. 8a**). By contrast, F_{11}-derived copies remained transcriptionally active,

© 2013 Nature America, Inc. All rights reserved.

showing reduced *EVD* RNA levels that correlated with decreased 3′ *gag*–derived siRNAs, further underscoring the link between *EVD* transcription levels and RDR6-dependent siRNA production (**Fig. 6a**, lanes 2 and 4 versus 5). Notably, crossing F_{11} and F_{14} plants resulted in silencing of all F_{11}-derived *EVD* copies, as *EVD* RNA and 3′ *gag*–derived siRNAs of 21–22 nt were below detection in F_1 progeny, demonstrating potent *trans* silencing by LTR-derived 24-nt siRNAs (**Fig. 6a**, lanes 6 and 8).

A *Copia93* remnant in the Col-0 genome is a solo LTR that has >88% DNA sequence similarity to the *EVD* LTR and is embedded within the *RPP4* promoter (*AT4G16860*). Bisulfite sequencing showed that, as in Col-0, the solo-LTR *RPP4* DNA was unmethylated in both epi15 F_8 and epi15 F_{11} plants. However, coinciding with the induction of LTR-derived 24-nt siRNAs, it became increasingly methylated from the F_{14} generation onward (**Fig. 6b**,**c** and **Supplementary Fig. 9**), correlating with reduced *RPP4* expression, induced by treatments with the salicylate analog benzothiadizole (BTH; **Fig. 6d**). To investigate the impact of newly inserted, active *EVD* copies on gene expression, we focused on two homozygous insertions (I and B) mapped in the epi454 F_8 generation (**Fig. 4d**) to *AT1G68420* and downstream of *AT5G25550*, respectively. Both full-length *EVD* insertions markedly increased transcription at both loci in F_9 individuals, displaying active *EVD* transcription and no LTR methylation (**Figs. 4e** and **6e**–**h**). By contrast, in F_9 plants carrying LTR-methylated and silenced *EVD* copies, transcription at the disrupted *AT1G68420* locus and expression of the *EVD*-proximal *AT5G25550* locus were low, as also seen in F_{15} individuals, in which *EVD* was completely silenced (**Figs. 4e** and **6e**–**h**). Whereas *AT1G68420* illustrates gene disruption by *EVD*, the example of *AT5G25550* shows how *trans*-acting LTR-derived 24-nt siRNAs produced during the onset of *EVD* TGS could trigger functionally relevant epiallelism at an *EVD*-proximal locus existing before or formed during genome invasion.

DISCUSSION

Most studies of plant TEs have involved the transcriptional reactivation of highly proliferated and predominantly heterochromatic transposon remnants, whereas rare studies of intact, active elements have intrinsically precluded analyses of TE resilencing. Here we have resolved these conundrums through the unique use of epiRILs. By erasing the DNA methylation status in *met1* and *ddm1* mutant backgrounds, we awakened an endogenous single-copy functional retroelement and subsequently followed the regain of its epigenetic silencing in a nearly wild-type genome, effectively reconstructing the sequence of events underlying a retrotransposon burst (**Fig. 7** and **Supplementary Fig. 10**). This exercise unraveled notable features of the biology of *EVD* that undoubtedly underpin the success of its proliferation over multiple generations. First, confinement to L2 cells and meristematic exclusion of *EVD* probably explain the paucity of phenotypic aberrations associated with its host's vegetative growth and seed set. This preservation of host fitness and L2-specific gametophytic transmission patterns probably explain the extensive reservoir of genetic diversity constituted in successive generations of offspring. A second unexpected underpinning of the invasive success of *EVD* is its resistance to PTGS. The finding that the RDR6, DCL2-DCL4 and AGO1-AGO2 antiviral pathway constitutes the first layer of defense against *EVD* explains previous observations that TE-derived 21-nt siRNA production is a common response to transcriptional reactivation of TEs in *ddm1*-mutant or stressed plants[9–12,19]. The use by *EVD* of Gag-dependent protection rather than direct suppression of this PTGS response might be explained by the fact that *trans*-acting VSRs, as deployed by most viruses, concurrently interfere with microRNA functions, affecting plant development and fertility. This added deleterious effect would be unfavorable to *EVD*, which, unlike infectious plant viruses, coevolves with its host and must preserve its integrity.

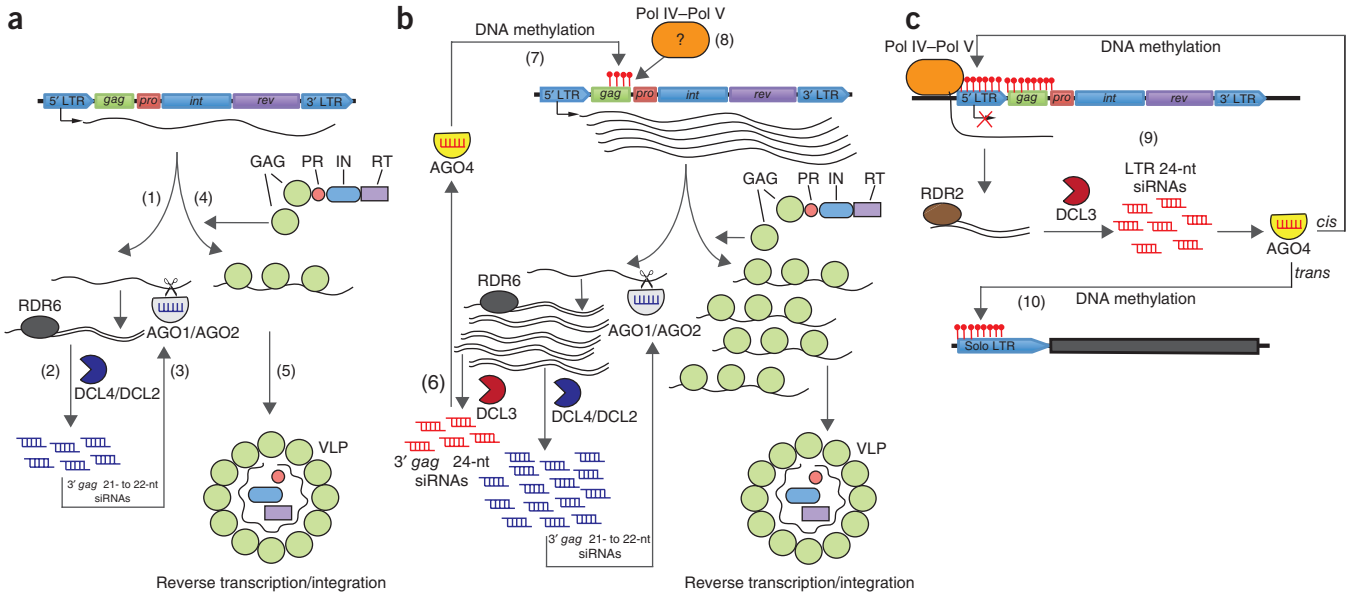


Figure 7 Rise and demise of *EVD*. (**a**) In early proliferation stages, a low *EVD* copy number accounts for moderate transcription and dsRNA synthesis by RDR6 (1). Processing of dsRNA by DCL4 or DCL2 produces 3′ *gag*–derived siRNAs of 21–22 nt in length (2), which are loaded onto AGO1-AGO2 (3). This complex can only partly degrade *EVD* RNA (3,4), owing to its protection by Gag as part of putative VLPs (5). (**b**) Because single seeds acquire distinct patterns and numbers of *EVD* insertion over generations, an increasing amount of progeny plants will attain the threshold of ~40 copies, beyond which saturating dsRNA levels allow processing by DCL3 of 3′ *gag*–derived 24-nt siRNAs loaded onto AGO4 (6). These siRNAs induce *de novo* methylation of 3′ *gag* DNA (7), which in turn initiates antisense transcription, possibly by Pol IV–Pol V, and coordinated spreading of methylation toward the LTR (8). (**c**) Eventual epigenetic silencing of *EVD* is achieved over a few subsequent generations, resulting in TGS actively maintained at the LTR by RdDM (9). The ensuing 24-nt LTR-derived siRNAs show potent *cis* and *trans* silencing properties (10).

© 2013 Nature America, Inc. All rights reserved.

Paradoxically, the resistance of *EVD* to PTGS constitutes the very foundation of its eventual demise, invariably initiated beyond a threshold of ~40 *EVD* copies per genome, regardless of the genetic backgrounds of the epiRILs or initial copy numbers. Our data suggest that this threshold underlies the limit of Pol II–dependent *EVD* transcript accumulation (itself determined by the number of active copies) beyond which RDR6-dependent dsRNA becomes accessible to the action of DCL3 in addition to those of DCL4 and DCL2; the ensuing AGO4-loaded 3′ *gag*–derived 24-nt siRNAs guide *de novo* methylation of 3′ *gag* DNA as a prelude to the eventual TGS of *EVD*. The key underlying notion that distinct DCL effectors may share a continuum of action and templates under circumstances of exceptionally high dsRNA accumulation fully concurs with and biologically explains numerous observations that abundant virus- or hairpin-derived dsRNAs normally diced into siRNAs of 21–22 nt in length effectively become a potent substrate for DCL3 in *dcl2/dcl4* double-mutant *A. thaliana*[25,28,35,36,38]. We rule out the possibility that 3′ *gag*–derived 24-nt siRNAs are products of the classical Pol IV–RDR2 pathway because (i) their abundance was strictly correlated with active, Pol II–dependent *EVD* transcription (abundance was high in the F_{11} generation and markedly reduced in the F_{14} generation upon TGS onset); (ii) these species shared a near-identical processing pattern with the Pol II– and RDR6-dependent 3′ *gag*–derived siRNAs of 21–22 nt in length; and (iii) introducing the *rdr6* mutation into *35S:EVD* transgenic plants strongly reduced the accumulation of these species without production of 24-nt species, as expected if a parallel RDR2-dependent pathway was operating.

The access of DCL3 to RDR6-dependent dsRNA, which itself is derived from Pol II transcripts, unveils a previously unexplored facet of RdDM, explaining how PTGS might acquire a DNA-level component, and is consistent with earlier work on transgene silencing[39–42] and more recent studies. Hence, a similar RDR6-AGO1-AGO2 pathway guides DNA methylation of intergenic regions and recently evolved pseudogenes: as with *EVD*, this methylation is also associated with 21-nt and 24-nt siRNA production and is sensitive to mutations in *AGO4* but not in *RDR2* (refs. 43,44). Furthermore, comprehensive methylome analyses of many *A. thaliana* silencing mutants[45] showed that, of all mutants not involved in canonical RdDM, *rdr6* and the related *rdr1* mutants showed the strongest hypomethylation affecting all cytosine contexts. RDR6-associated methylation was linked to Pol II–dependent transcription of active genes rather than Pol IV– and Pol V–dependent transcription of TE remnants and/or repeats. Moreover, as with *EVD*, it coincided with 21-nt and 24-nt siRNA production but with a clear predominance of the former species, contrasting with the near-exclusive 24-nt siRNA accumulation at canonical RdDM target loci.

De novo 3′ *gag* methylation eventually activated, via 3′-to-5′ spreading, RdDM at the 5′ LTR of *EVD*, leading to its final demise via TGS. However, unlike in most cases of RdDM affecting TE remnants in *cis*, the ensuing LTR-derived 24-nt siRNAs showed notable *trans* silencing abilities, as originally observed in transgene silencing studies[46–51]. Although the mechanism of 3′-to-5′ spreading involved in this process awaits further clarification, it is tempting to propose that initial 3′ *gag* methylation recruits antisense transcription driven by Pol V, known to guide AGO4 to active gene promoters via transcription of long noncoding RNAs[52,53]. Our work finally shows how *trans* silencing by 24-nt LTR-derived siRNAs can establish functional *de novo* epiallelism at *EVD*-proximal genes. Owing to its intrinsic reversibility and flexibility, this additional layer of epigenetic control may further optimize plant evolution and adaptation as a result of TE proliferation, in full agreement with Barbara McClintock's visionary concept of 'controlling elements' (ref. 1).

URLs. Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo/; Gateway vectors, https://gateway.psb.ugent.be/; Gypsy Database, http://www.gydb.org/; *in situ* hybridization protocol, http://www.its.caltech.edu/~plantlab/protocols/insitu.htm; Kismeth software, http://katahdin.mssm.edu/kismeth/revpage.pl; Fasteris, http://www.fasteris.com/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession code. Raw and processed data sets of sRNAs are deposited at NCBI's Gene Expression Omnibus (GEO) under accession GSE43412.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of the Voinnet laboratory for fruitful discussions and critical reading of the manuscript. C. Brosnan is specially thanked for his intellectual input and support all throughout this project. We are indebted to D. Meyer for help with *in situ* hybridization. We thank L. Navarro (Institut de Biologie de l'Ecole Nomrale Supérieure, IBENS) for providing the LTR:*GUS* transgenic lines. We thank D. Gilmer (Institut de Biologie Moléculaire des Plantes) for providing antiserum to GFP. This project was supported by a core grant from ETH-Z and a grant from the Swiss National Foundation Genetics of miRNA Action and Biogenesis (31003A_132907). A.M.-O. is supported by a PhD fellowship from the ETH-Z foundation.

AUTHOR CONTRIBUTIONS

O.V. and A.M.-O. conceived and designed the experiments, helped by V.C. on experiments involving *ddm1*-mutant epiRILs. A.M.-O., M.E. and A. Martin performed the experiments. O.V., V.C., A.M.-O. and M.E. analyzed the data. A. Marchais performed computer and statistical analyses. O.V. and A.M.-O. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216 (1956).
2. Ito, H. *et al.* An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**, 115–119 (2011).
3. Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J. & van der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**, 1527–1530 (2008).
4. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163 (2011).
5. Lisch, D. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* **60**, 43–66 (2009).
6. Kankel, M.W. *et al. Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122 (2003).
7. Matzke, M., Kanno, T., Daxinger, L., Huettel, B. & Matzke, A.J. RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* **21**, 367–376 (2009).
8. Zemach, A. *et al.* The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
9. Yu, A. *et al.* Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc. Natl. Acad. Sci. USA* **110**, 2389–2394 (2013).
10. Slotkin, R.K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461–472 (2009).
11. Teixeira, F.K. *et al.* A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**, 1600–1604 (2009).
12. Dowen, R.H. *et al.* Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. USA* **109**, E2183–E2191 (2012).
13. Pereira, V. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol.* **5**, R79 (2004).

© 2013 Nature America, Inc. All rights reserved.

14. Peterson-Burch, B.D., Nettleton, D. & Voytas, D.F. Genomic neighborhoods for *Arabidopsis* retrotransposons: a role for targeted integration in the distribution of the Metaviridae. *Genome Biol.* **5**, R78 (2004).
15. Pérez-Hormaeche, J. *et al.* Invasion of the *Arabidopsis* genome by the tobacco retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing. *Plant Physiol.* **147**, 1264–1278 (2008).
16. Liu, B. & Wendel, J.F. Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome* **43**, 874–880 (2000).
17. Tsukahara, S. *et al.* Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* **461**, 423–426 (2009).
18. Mirouze, M. *et al.* Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461**, 427–430 (2009).
19. Nuthikattu, S. *et al.* The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol.* **162**, 116–131 (2013).
20. Reinders, J. *et al.* Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* **23**, 939–950 (2009).
21. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**, e1000530 (2009).
22. Bowman, J.L., Baum, S.F., Eshed, Y., Putterill, J. & Alvarez, J. Molecular genetics of gynoecium development in *Arabidopsis*. *Curr. Top. Dev. Biol.* **45**, 155–205 (1999).
23. Nole-Wilson, S., Azhakanandam, S. & Franks, R.G. Polar auxin transport together with AINTEGUMENTA and REVOLUTA coordinate early *Arabidopsis* gynoecium development. *Dev. Biol.* **346**, 181–195 (2010).
24. Estrada-Luna, A.A. *et al.* Beyond promiscuity: from sexuality to apomixis in flowering plants. *In Vitro Cell Dev. Biol. Plant* **38**, 146–151 (2002).
25. Dunoyer, P. *et al.* An endogenous, systemic RNAi pathway in plants. *EMBO J.* **29**, 1699–1712 (2010).
26. Wang, X.B. *et al.* RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **107**, 484–489 (2010).
27. Voinnet, O. Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci.* **13**, 317–328 (2008).
28. Deleris, A. *et al.* Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* **313**, 68–71 (2006).
29. Wang, X.-B. *et al.* The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in *Arabidopsis thaliana*. *Plant Cell* **23**, 1625–1638 (2011).
30. Wu, Q., Wang, X. & Ding, S.W. Viral suppressors of RNA-based viral immunity: host targets. *Cell Host Microbe* **8**, 12–15 (2010).
31. Voinnet, O., Rivas, S., Mestre, P. & Baulcombe, D. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**, 949–956 (2003).
32. Sabot, F. & Schulman, A.H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**, 381–388 (2006).
33. Larsen, L.S.Z. *et al.* Ty3 capsid mutations reveal early and late functions of the amino-terminal domain. *J. Virol.* **81**, 6957–6972 (2007).
34. Llorens, C. *et al.* The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* **39**, D70–D74 (2011).
35. Blevins, T. *et al.* Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* **34**, 6233–6246 (2006).
36. Henderson, I.R. *et al.* Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–725 (2006).
37. Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A. & Voinnet, O. Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.* **39**, 848–856 (2007).
38. Gasciolli, V., Mallory, A.C., Bartel, D.P. & Vaucheret, H. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr. Biol.* **15**, 1494–1500 (2005).
39. Vaistij, F.E., Jones, L. & Baulcombe, D.C. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**, 857–867 (2002).
40. Jauvion, V., Rivard, M., Bouteiller, N., Elmayan, T. & Vaucheret, H. RDR2 partially antagonizes the production of RDR6-dependent siRNA in sense transgene-mediated PTGS. *PLoS ONE* **7**, e29785 (2012).
41. Fojtova, M. Epigenetic switch from posttranscriptional to transcriptional silencing is correlated with promoter hypermethylation. *Plant Physiol.* **133**, 1240–1250 (2003).
42. Fojtová, M. The *trans*-silencing capacity of invertedly repeated transgenes depends on their epigenetic state in tobacco. *Nucleic Acids Res.* **34**, 2280–2293 (2006).
43. Garcia, D. *et al.* Ago hook and RNA helicase motifs underpin dual roles for SDE3 in antiviral defense and silencing of nonconserved intergenic regions. *Mol. Cell* **48**, 109–120 (2012).
44. Pontier, D. *et al.* NERD, a plant-specific GW protein, defines an additional RNAi-dependent chromatin-based pathway in *Arabidopsis*. *Mol. Cell* **48**, 121–132 (2012).
45. Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V. & Jacobsen, S.E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
46. Kilby, N.J., Leyser, H.M. & Furner, I.J. Promoter methylation and progressive transgene inactivation in *Arabidopsis*. *Plant Mol. Biol.* **20**, 103–112 (1992).
47. Van Houdt, H., Bleys, A. & Depicker, A. RNA target sequences promote spreading of RNA silencing. *Plant Physiol.* **131**, 245–253 (2003).
48. Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765 (2005).
49. Vermeersch, L. *et al.* Transitive RNA silencing signals induce cytosine methylation of a transgenic but not an endogenous target. *Plant J.* **74**, 867–879 (2013).
50. You, W., Lorkovic, Z.J., Matzke, A.J.M. & Matzke, M. Interplay among RNA polymerases II, IV and V in RNA-directed DNA methylation at a low copy transgene locus in *Arabidopsis thaliana*. *Plant Mol. Biol.* **82**, 85–96 (2013).
51. Daxinger, L. *et al.* A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* **28**, 48–57 (2009).
52. Wierzbicki, A.T., Ream, T.S., Haag, J.R. & Pikaard, C.S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* **41**, 630–634 (2009).
53. Zheng, Q. *et al.* RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J.* **73**, 179–189 (2012).

© 2013 Nature America, Inc. All rights reserved.

ONLINE METHODS

Plant material. *A. thaliana* ecotype Columbia-0 (Col-0) was used in this study as wild type; *met1-3*, *dcl2-1, dcl4-2*, *rdr6-12* and *ddm1* mutants and the *met1*- or *ddm1*-derived epiRILs have been previously described[20,54–59]. Propagation of the epiRILs was performed by single-seed descent procedures.

Plasmid construction and transformation. *EVD* constructs were generated using single-site Gateway technology (Invitrogen); primer sequences are listed in **Supplementary Table 5**. The *gag-pol* ORF from the *EVD* chromosome 5 insertion was amplified from BAC F2K13 by PCR using Phusion HF DNA Polymerase (Thermo Scientific) and cloned into pJET1.2 (Thermo Scientific). Deletion of the *gag* sequence encoding the RBD was carried out by site-directed mutagenesis, following the QuickChange protocol developed by Stratagene. Sequences encoding HA and Flag tags were added by PCR. *EVD* and *EVD*$^{\Delta RBD}$ clones were recombined into pDONRP1P2. LR recombination was performed using the *35S* promoter–containing vector pB7WG2,0 (Gateway vectors; see URLs)[60], and the resulting construct was introduced into *Agrobacterium tumefaciens* GV3101. *A. thaliana* transformation was carried out by the floral dip method[61]. At least ten transgenic lines were analyzed. The LTR:*GUS-GFP* transgenic plant was a gift from L. Navarro. Transient expression assays were performed as previously described[31].

***In situ* hybridization.** Probe DNA for *EVD in situ* hybridization was cloned into pGEM-T Easy (Promega). Antisense *EVD* digoxigenin (DIG)-labeled riboprobe was synthesized *in vitro* from pGEM-*EVD*-ISH plasmid with T7 RNA polymerase (Promega) and DIG RNA Labeling Mix (Roche). Samples were fixed twice in FAA (3.2% formaldehyde, 5% acetic acid, 50% ethanol) by vacuum infiltration for 30 min before dehydration in ethanol and embedding in Paraplast Plus (McCornick Scientific). Sections of 10 µm were generated using a Leica microtome and were mounted on Polysine (Menzel-Glaser) slides. *In situ* hybridization was performed as described by J. Long (see URLs). Slides were hybridized with approximately 40-45 ng of DIG-labeled probe overnight at 52 °C. Inmunological detection of the DIG-labeled probes was performed using a DIG Nucleic Acid Detection kit (Roche).

Bisulfite sequencing–based DNA methylation analysis. Genomic DNA was extracted from *A. thaliana* inflorescences using the DNeasy Plant Mini kit (Qiagen). Bisulfite treatment was performed using the EpiTect Bisulfite kit (Qiagen). Bisulfite-treated DNA was then amplified by PCR. Primer design and PCR were performed following published recommendations[62]. PCR fragments were cloned into pGEM-T Easy, and individual colonies were sequenced. Sequences were analyzed with Kismeth software (see URLs)[63] to obtain the percentage of methylated sites for each sequence context. Wilson score intervals were used to find 95% confidence intervals for these percentages. Results shown were obtained from two independent experiments.

McrBC-qPCR–based DNA methylation analysis. McrBC digestion was performed on DNA from young leaves; qRT-PCR was performed on equal amounts of digested and undigested DNA. Results were expressed as percentage of loss of molecules after McrBC digestion. Error bars represent standard deviation from two independent experiments each involving duplicate PCRs.

qRT-PCR. *EVD* copy number quantification and expression analyses were conducted by qRT-PCR on a LightCycler480 II apparatus (Roche) using the SYBR Green qPCR kit (Eurogentech) and gene-specific primers (**Supplementary Table 5**). For *EVD* copy number quantification, genomic DNA was extracted from *A. thaliana* with the DNeasy Plant Mini kit. *EVD* genomic copies were subjected to absolute quantification and were normalized with the single-locus gene *ACTIN2* (*ACT2*, *AT3G18780*). For qRT-PCR, total RNA was extracted with the RNeasy Plant Mini kit. Total RNA (2 µg) was reverse transcribed into cDNA using the Maxima First-Strand cDNA Synthesis kit (Thermo Scientific) for relative quantification. Results were normalized using *ACT2* levels. For qRT-PCR analysis of genes adjacent to new *EVD* insertions (**Fig. 6**), cDNA was generated using only oligo(dT) to prime reverse transcription. Error bars represent standard deviation from three technical replicates of two biological replicates.

RNA blot analysis. Total RNA was extracted from *A. thaliana* and *N. benthamiana* with TRIzol reagent (Invitrogen). Unless otherwise indicted, ten plants of the different epiRIL generations were bulked for analysis, such that we always used the same number of plants per experiment per generation. Five to ten plants of the transgenic lines were used for RNA analysis. For RNA blot analysis of high-molecular-weight RNA, 10–15 µg of total RNA was resolved on denaturing 1.2% agarose gels with 2.2 M formaldehyde, capillary transferred to HyBond-NX membrane (GE Healthcare) and cross-linked by UV irradiation[64]. Ethidium bromide staining before transfer was used to confirm equal loading. For low-molecular-weight RNA blot analysis, low-molecular-weight RNA was isolated as described[65], and 5–10 µg was resolved on a denaturing 17.5% polyacrylamide-urea gel, transferred to membrane by electroblotting and chemically cross-linked[66]. Radiolabeled probes for the detection of *EVD* and *GFP* RNAs and siRNAs were made by random-priming reactions using the Prime-A-Gene kit (Promega) in the presence of [α-^{32}P]-dCTP (Hartmann Analytic). The templates used were PCR products (**Supplementary Table 5**). DNA oligonucleotides complementary to microRNAs, *trans*-acting siRNAs (tasiRNAs), *cis*-acting siRNAs or U6 small nuclear RNA (snRNA) were end labeled with [γ-^{32}P]-dATP using T4 PNK (Thermo Scientific).

Southern blot analysis. Unless otherwise indicted, ten plants of the different epiRIL generations were bulked for genomic DNA extraction using the CTAB method[67], and 1 µg of DNA was digested overnight with SspI (Thermo Scientific), separated by gel electrophoresis on a 1% agarose gel and transferred to HyBond-NX membrane and cross-linked by UV irradiation[64]. [^{32}P]-labeled DNA probes were generated by random priming using *EVD* PCR products as template (**Supplementary Table 5**). Detection and analysis of the *EVD* band pattern was performed using ImageQuant TL 7.0 software (GE Healthcare).

Protein analysis. Total proteins were extracted from 0.1–0.3 g of *A. thaliana* and *N. benthamiana* tissues ground in liquid nitrogen, homogenized with 0.6 ml of extraction buffer (0.7 M sucrose, 0.5 M Tris-HCl, pH 8, 5 mM EDTA, pH 8, 0.1 M NaCl, 2% 2-mercaptoethanol) containing 2 µM proteasome inhibitor MG-132 (Calbiochem, Millipore) and protease inhibitor cocktail (Complete, Roche) following the manufacturers' instructions and mixed for 5 min. An equal volume of water-saturated Tris-buffered phenol was added, and samples were mixed for 5 min. Phases were separated by 10-min centrifugation (12,000*g* at 4 °C). Proteins were precipitated from the phenol phase by the addition of 5 volumes of 0.1 M ammonium acetate in methanol and were incubated at −20 °C overnight. After a 15-min centrifugation step (5,000*g* at 4 °C), precipitate was washed with ammonium acetate in methanol and resuspended (3% SDS, 62.3 mM Tris-HCl, pH 8, 10% glycerol). Total protein (100–300 ng) was resolved on SDS-PAGE, transferred by electroblotting onto Inmobilon-P PVDF membrane (Millipore) and incubated with antibodies in PBS with 0.1% Tween-20 and 5% nonfat dried milk, following standard protein blotting procedures[64]. After incubation with HRP-conjugated secondary goat antibody to rabbit (Sigma), detection was performed with the ECL Western Blotting Detection kit (GE Healthcare). Equal loading was verified by Coomassie staining of the membrane after protein blotting. Affinity-purified antibodies were used at the indicated dilutions: AGO1 (1:8,000), AGO2 (1:4,000), AGO4 (1:2,000), DCL3 (1:1,000), HA (1:5,000) and Flag (1:7,000). Antiserum to DCL4 was used at a 1:5,000 dilution, and antiserum to GFP was used at a 1:15,000 dilution.

AGO protein immunoprecipitation. *A. thaliana* inflorescences were ground in liquid nitrogen and homogenized in 1 ml of IP buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 0.1% NP-40) containing 2 µM proteasome inhibitor MG-132 and protease inhibitor cocktail (Complete) following the manufacturers' instructions for 1 h at 4 °C. Cell debris was removed by centrifugation for 30 min (12,000*g* at 4 °C). Protein was quantified in extracts to normalize total protein inputs. Extracts were precleared by incubation for 30 min with 10 µl of Protein A–agarose beads (Sigma) at 4 °C (10 r.p.m. (revolutions per minute)), and beads were removed by centrifugation for 1 min (12,000*g* at 4 °C). Clarified lysates were incubated with antibodies for 2 h at 4 °C (10 r.p.m.) and for an additional 1 h with 10 µl of Protein A–agarose beads. Beads were collected by centrifugation (4,000*g* at 4 °C), and immunoprecipitates

© 2013 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2703

were washed three times (20 min each) with 1 ml of IP buffer. For RNA analysis, immune complexes were subjected to TRIzol extraction. Aliquots of input, flow-through (before washes) and immunoprecipitated fractions were subjected to protein blot analysis to assess the efficiency of inmunoprecipitation. For immunoprecipitation of AGO1, antibody was used at a 1:1,000 dilution, and, for immunoprecipitation of AGO2 and AGO4, antibodies were used at 1:500 dilutions.

Antibodies. Peptides used to raise rabbit polyclonal antibodies for AGO1, AGO2, AGO4, DCL3 and DCL4 were previously described[43,68–70]. Efficiency of affinity-purified antibodies from AGO1, AGO2, AGO4 and DCL3 antisera and antiserum to DCL4 were previously validated by comparing protein detection in wild-type plants and the corresponding mutant[43,71]. Antiserum to GFP was a gift of D. Gilmer. Affinity-purified HRP-coupled antibodies to HA and Flag (Sigma) were used for the detection of tagged *EVD*-encoded proteins.

BTH treatment. To induce *RPP4* expression, 0.5 mM BTH (Sigma) in water was sprayed on 2-week-old plants. Rosette leaves were harvested for RNA extraction 24 h after treatment[72,73].

Protein alignments. Retroviral and retrotransposon Gag protein sequences were retrieved from the Gypsy Database (see URLs)[34]. Sequence alignments were performed with Jalview software[74] by ClustalW Multiple-Sequence Alignment.

Deep sequencing. Total RNA (10 µg) was extracted from *A. thaliana* inflorescences with TRIzol reagent, processed into sequencing libraries using adapted Illumina protocols and sequenced at Fasteris (Switzerland) using the HiSeq 2000 sequencer. Reads were mapped against the TAIR9 version of the *A. thaliana* genome using Bowtie software[75] (options: -e 50 -a–best–strata -m 5000–nomaqround -y) together with the SAMtools[76] and BEDTools[77] suites. In-house R cran scripts were implemented to display the mapping results.

54. Vongs, A., Kakutani, T., Martienssen, R.A. & Richards, E.J. *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928 (1993).
55. Saze, H., Scheid, O.M. & Paszkowski, J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **34**, 65–69 (2003).
56. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
57. Xie, Z., Allen, E., Wilken, A. & Carrington, J.C. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **102**, 12984–12989 (2005).
58. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L. & Poethig, R.S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**, 2368–2379 (2004).
59. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**, e1000530 (2009).
60. Karimi, M., Inzé, D. & Depicker, A. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195 (2002).
61. Clough, S.J. & Bent, A.F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998.).
62. Henderson, I.R., Chan, S.R., Cao, X., Johnson, L. & Jacobsen, S.E. Accurate sodium bisulfite sequencing in plants. *Epigenetics* **5**, 47–49 (2010).
63. Gruntman, E. *et al.* Kismeth: analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* **9**, 371 (2008).
64. Sambrook, J. & Russell, D.W. *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
65. Lu, C., Meyers, B.C. & Green, P.J. Construction of small RNA cDNA libraries for deep sequencing. *Methods* **43**, 110–117 (2007).
66. Pall, G.S. & Hamilton, A.J. Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.* **3**, 1077–1084 (2008).
67. Richards, E., Reichardt, M. & Rogers, S. Preparation of genomic DNA from plant tissue. *Curr. Protoc. Mol. Biol.* **Chapter 2**, Unit 2.3 (2001).
68. Qi, Y., Denli, A.M. & Hannon, G.J. Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol. Cell* **19**, 421–428 (2005).
69. Nakazawa, Y., Hiraguri, A., Moriyama, H. & Fukuhara, T. The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 *in vivo* and functions in the *trans*-acting siRNA pathway. *Plant Mol. Biol.* **63**, 777–785 (2007).
70. Azevedo, J. *et al.* Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. *Genes Dev.* **24**, 904–915 (2010).
71. Schott, G. *et al.* Differential effects of viral silencing suppressors on siRNA and miRNA loading support the existence of two distinct cellular pools of ARGONAUTE1. *EMBO J.* **31**, 2553–2565 (2012).
72. Huang, X., Li, J., Bao, F., Zhang, X. & Yang, S. A gain-of-function mutation in the *Arabidopsis* disease resistance gene *RPP4* confers sensitivity to low temperature. *Plant Physiol.* **154**, 796–809 (2010).
73. Lawton, K.A. *et al.* Benzothiadiazole induces disease resistance in *Arabidopsis* by activation of the systemic acquired resistance signal transduction pathway. *Plant J.* **10**, 71–82 (1996).
74. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
75. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
76. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
77. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

© 2013 Nature America, Inc. All rights reserved.

Reconstructing *de novo* silencing of an active plant retrotransposon.

Arturo Marí-Ordóñez, Antonin Marchais, Mathilde Etcheverry, Antoine Martin, Vincent Colot and Olivier Voinnet.

SUPPLEMENTARY INFORMATION:

Supplementary Figure 1: EVD vs ATR expression and LTR methylation.

Supplementary Figure 2: Arabidopsis gynoecium and ovule development.

Supplementary Figure 3: EVD in situ hybridization in WT and two additional epiRILs.

Supplementary Figure 4: Suppression of silencing assay in N. benthamiana.

Supplementary Figure 5: GAG RNA binding domain conservation and EVD polyprotein product production in EVD constructs.

Supplementary Figure 6: Mapping of 21-22-nt siRNAs to EVD in epi15 F14 and EVD copy number in epi454 and 438 F8.

Supplementary Figure 7: Bisulfite sequencing dot-plots of EVD LTR in epi15 generations and EVD copy number and DNA methylation in epi439 generations.

Supplementary Figure 8: Correlation of LTR 24-nt siRNAs abundance with EVD copy number and 21-24-nt siRNAs mapping correlation on EVD.

Supplementary Figure 9: Bisulfite sequencing dot-plots of RPP4 Solo-LTR in epi15 generations.

Supplementary Figure 10: EVD life cycle in Arabidopsis.

Supplementary Table 1: Figure 1 and Supplementary Figure 1 bisulfite sequencing raw data.

Supplementary Table 2: Figure 4 and Supplementary Figure 7 bisulfite sequencing raw data.

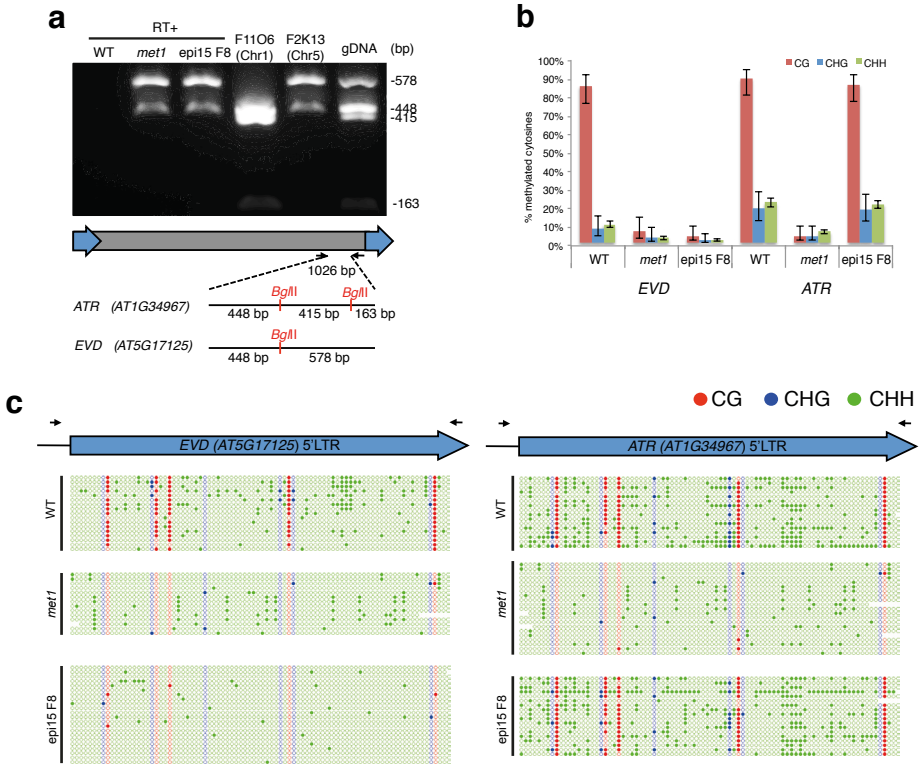
Supplementary Table 3: Figure 5 bisulfite sequencing raw data.

Supplementary Table 4: Figure 6 and Supplementary Figure 9 bisulfite sequencing raw data.

Supplementary Table 5: List of primers.

Nature Genetics: doi:10.1038/ng.2703

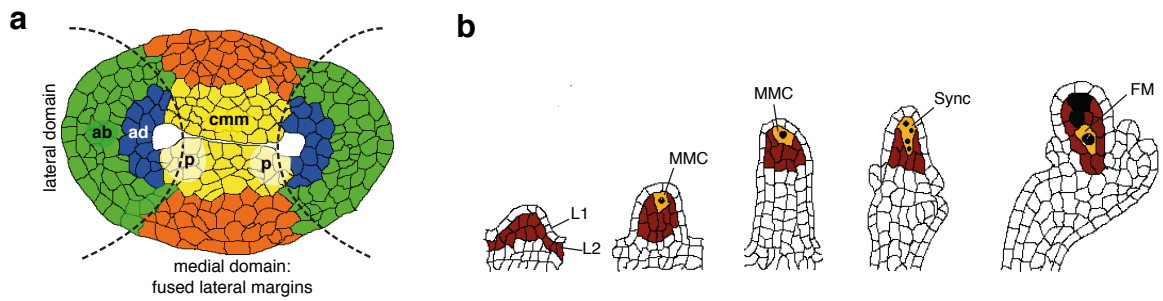
Supplementary Figure 1



Supplementary Figure 1: *EVD* vs *ATR* expression and LTR methylation.

(a) Restriction fragment length polymorphism (RFLP) analysis of RT-PCR products shown in Figure 1a. genomic DNA (gDNA), and PCR products obtained from F11O6 BAC (comprising chromosome 1 *ATR, AT1G34967*, copy) and F2K13 BAC (comprising chromosome 5 *EVD, AT5G17125*, copy), were digested with BglII to determine whether transcripts originated from *EVD* or *ATR*. **(b)** Bisulfite sequencing-based methylation analysis at the *EVD* and *ATR* 5'LTR in WT, *met1* and epi15 F8. The 5'LTR of *AT5G17125* (*EVD*) and *AT1G34967* (*ATR*) was specifically amplified using a primer inside the LTR and a specific primer outside of the LTR 5' end. Cytosine methylation levels are given at CG, CHG and CHH sites. Error bars represent the 95% confidence limits given by the Wilson score interval (see Supplementary Table 1 for details). **(c)** Dot-plots depiction of bisulfite sequencing data obtained for individual clones used to generate the results shown in Figure 1b, Supplementary Figure 1b and Supplementary Table 1. Filled circles indicate methylated cytosines and empty circles unmethylated cytosines.

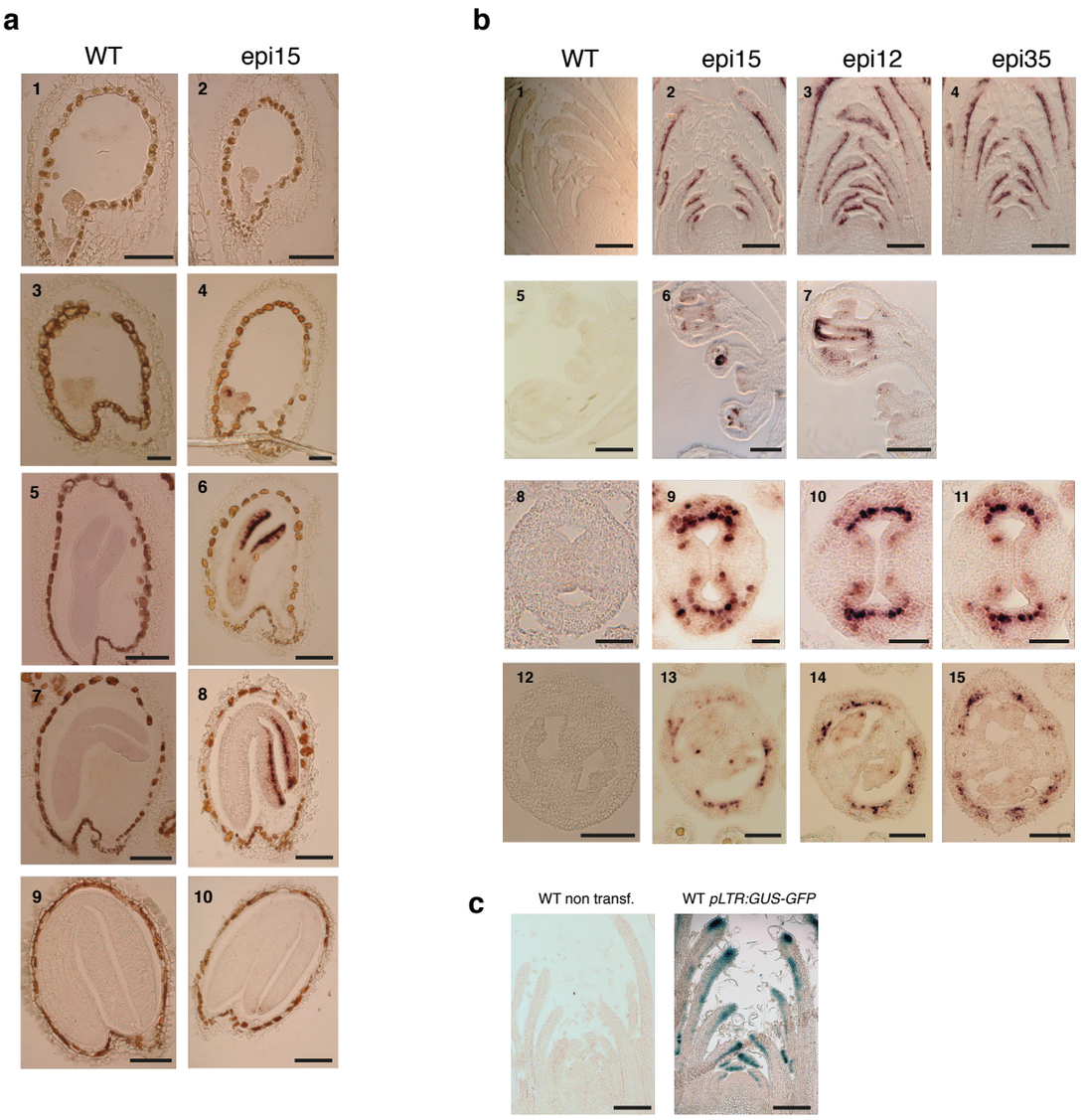
Supplementary Figure 2



Supplementary Figure 2: Arabidopsis gynoecium and ovule development.

(a) Scheme of a cross-section of the Arabidopsis gynoecium. The gynoecium is formed by the congenital fusion of two carpels (modified foliar organs that bear the ovules) through their lateral margins (cmm, yellow). Placental tissues (p, pale yellow), from which ovules arise (see b), differentiate from the cmm at the intersections of lateral and medial domains[21-23]. ab, Abaxial lateral domain (green); ad, adaxial lateral domain (blue); abaxial medial domain (orange). Adapted from ref. 23. **(b)** Ovule development in Arabidopsis. Ovule primordia emerge from the placentas, where a somatic cell from the sub-epidermal L2 cell layer (in red) differentiates into the megaspore mother cell (MMC) that undergoes meiosis, forming a syncytium (sync) of haploid nuclei, of which three degenerate, leaving a single haploid product: the female megaspore (FM). The FM then gives rise to the female gametophyte. Adapted from ref. 24.

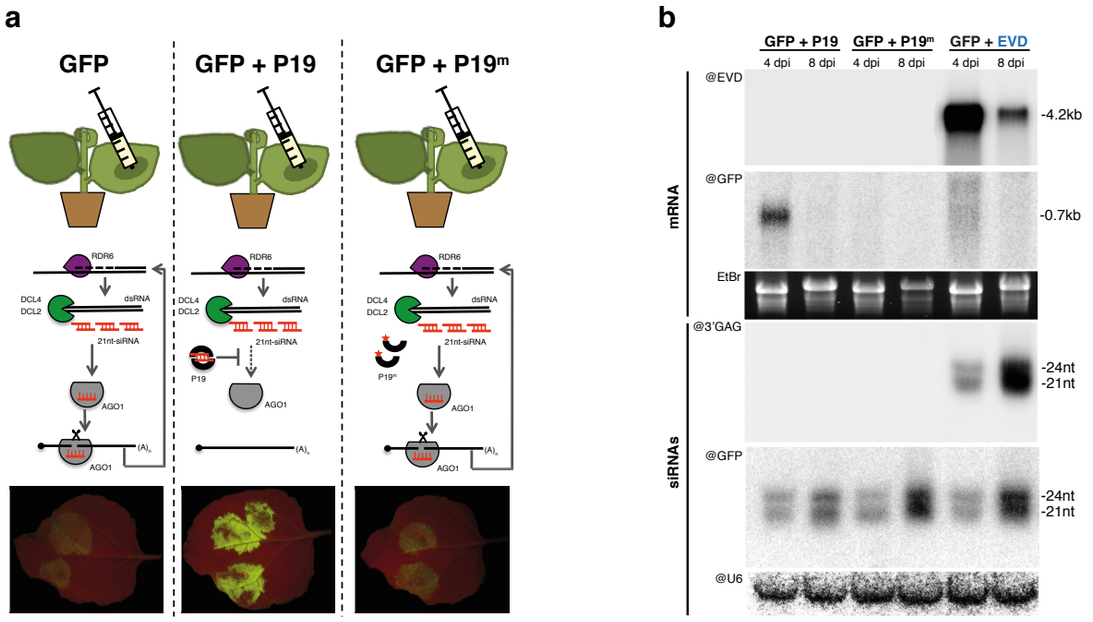
Supplementary Figure 3



Supplementary Figure 3: *EVD in situ* hybridization in WT and two additional epiRILs.

(a) *In-situ* hybridization study of *EVD* transcript accumulation, using an antisense probe, in developing embryos of WT and epi15 F10. Longitudinal sections of embryos at globular- (1-2), heart- (3-4), mid-torpedo- (5-6), late-torpedo- (7-8) and mature green-stage (9-10). **(b)** *In-situ* hybridization study of *EVD* transcript accumulation, using an antisense probe, in various tissues of WT, epi15 F10 and two other epiRILs expressing *EVD*: epi12 (F9) and epi35 (F9) previously described in ref.18. Longitudinal section of the SAM region from a young Arabidopsis rosette (1-4). Longitudinal sections of flower bud (5-7). Gynoecium transversal sections at floral developmental stage 8 (8-11). Longitudinal sections of ovule primordia at floral developmental stage 12 (12-15). **(c)** Longitudinal section of plant rosettes at the SAM region of non-transgenic WT and pLTR:GUS-GFP transgenic WT plants, following GUS-staining. Bars = 50 µm for **(a)** 1-4, **(b)** 12-15; 100 µm for **(a)** 5-10, **(b)** 1-7 and **(c)**; 10 µm for **(b)** 8-11.

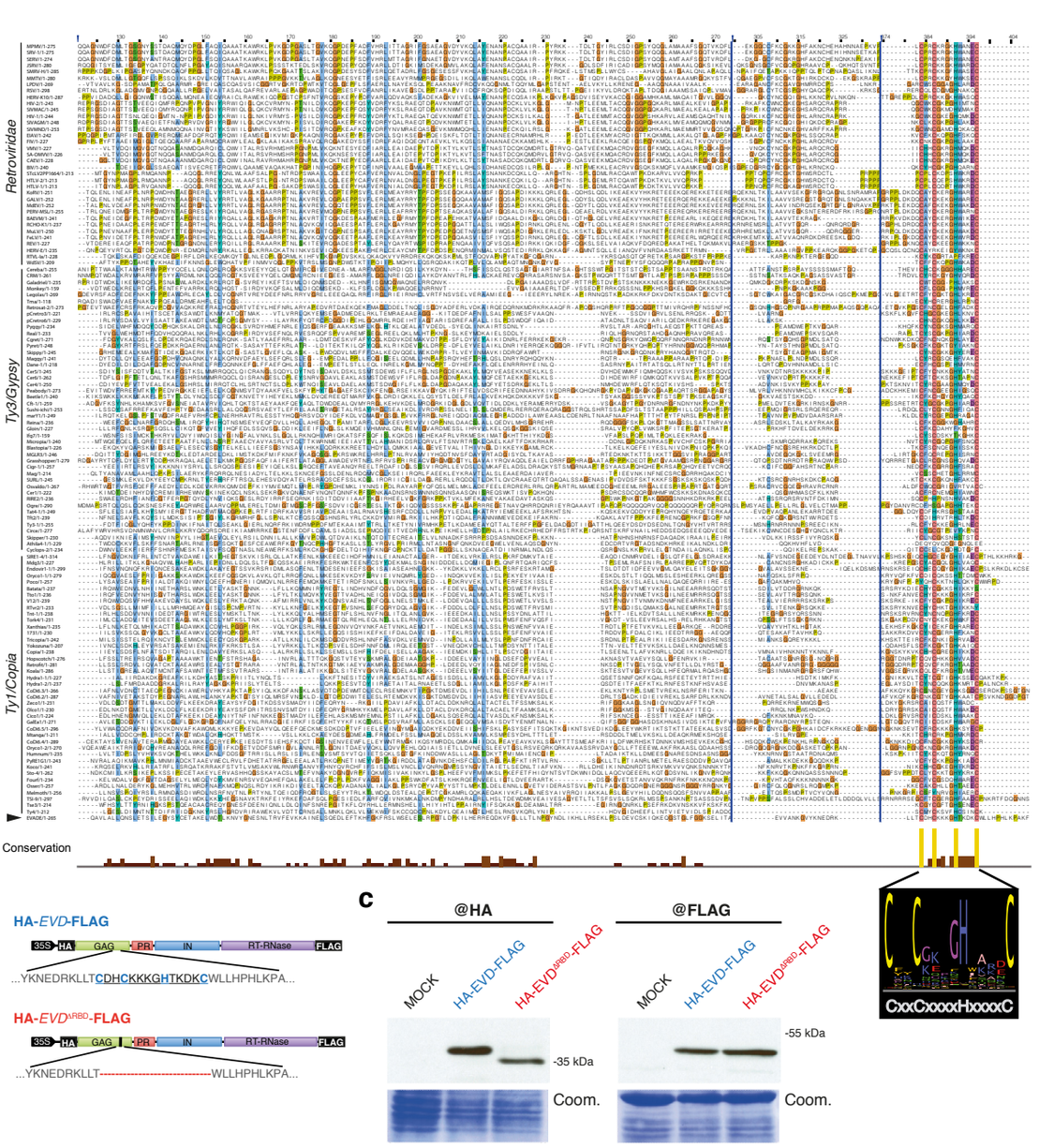
Supplementary Figure 4



Supplementary Figure 4: Suppression of silencing assay in *N. benthamiana*.

(a) Suppression of silencing assay by Agrobacterium-mediated transient expression in *N. benthamiana* leaves. *Agrobacterium tumefaciens*, infiltrated into plant leaves mediates transfer of transgenes into plant cells, allowing ectopic expression of the genes of interests, but it rapidly decreases 3-4 days post infiltration (dpi) due the onset of PTGS[30]. This feature of the system has been routinely used to characterize proteins with VSR activity by co-infiltrating the protein of interest with a GFP reporter transgene acting both as a trigger and a target of PTGS[30] . In the scheme, when infiltrated alone, GFP protein levels (visualized under UV illumination) are reduced due to the action of PTGS; however co-infiltration with the VSR P19, prevents the onset of PTGS in the infiltrated area allowing significantly higher levels of GFP accumulation. P19 acts as a head-to-tail homodimer that specifically binds 21-nt small RNA duplexes, preventing their loading into AGO. This property is abolished in the point mutant P19^{m}, which dimerizes but is unable to bind small RNAs. **(b)** Analysis of p35S:EVD transient expression in *N. benthamiana*. The potential suppression of silencing activity of EVD was tested by co-infiltration of EVD constructs (Figure 2g) with the p35S:GFP reporter construct. p35S:GFP co-infiltrated with p35S:P19 or its mutant version p35S:P19m were used as controls for positive and negative suppression activity, respectively. Northern analysis of High-molecular weight (mRNA) and low-molecular weight RNA (siRNA) was carried out on samples extracted from infiltrated leaves at 4 and 8 dpi. Ethidium bromide (EtBr) staining of ribosomal RNA and detection of U6 are shown as loading controls.

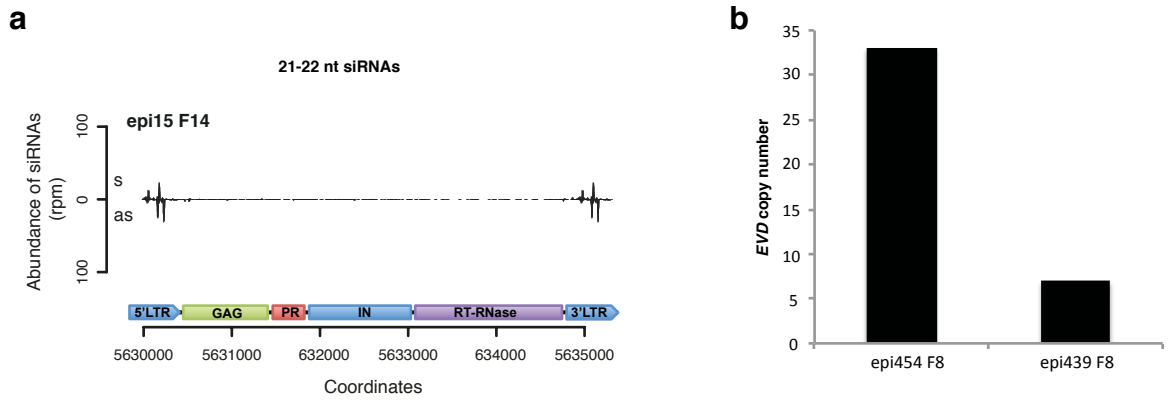
Supplementary Figure 5



Supplementary Figure 5: GAG RNA binding domain conservation and *EVD* polyprotein product production in *EVD* constructs.

(e) Alignment of GAG amino acid sequences from representative members of *Retroviridae* and *Pseudoviridae* (Ty3/Gypsy and Ty1/Copia). The bottom sequence (black arrow) corresponds to the *EVD*-encoded GAG protein. Coloured columns in the alignment indicate residues with conserved physico-chemical properties. The black box displays the consensus sequence logo of highly conserved residues corresponding to the RNA-binding motif ($Cx_2Cx_4Hx_4C$). **(c)** Scheme of p35S:HA-*EVD*-FLAG and p35S:HA-*EVD*$^{\Delta RBD}$-FLAG constructs, see methods for details. **(d)** Accumulation of the *EVD* polyprotein N-terminal and C-terminal products from p35S:HA-*EVD*-FLAG and p35S:HA-*EVD*$^{\Delta RBD}$-FLAG was assessed by Western analysis of total proteins extracted from *N. benthamiana*-infiltrated leaves at 4 dpi. Equal loading was verified by Coomassie staining of the membranes after blotting.

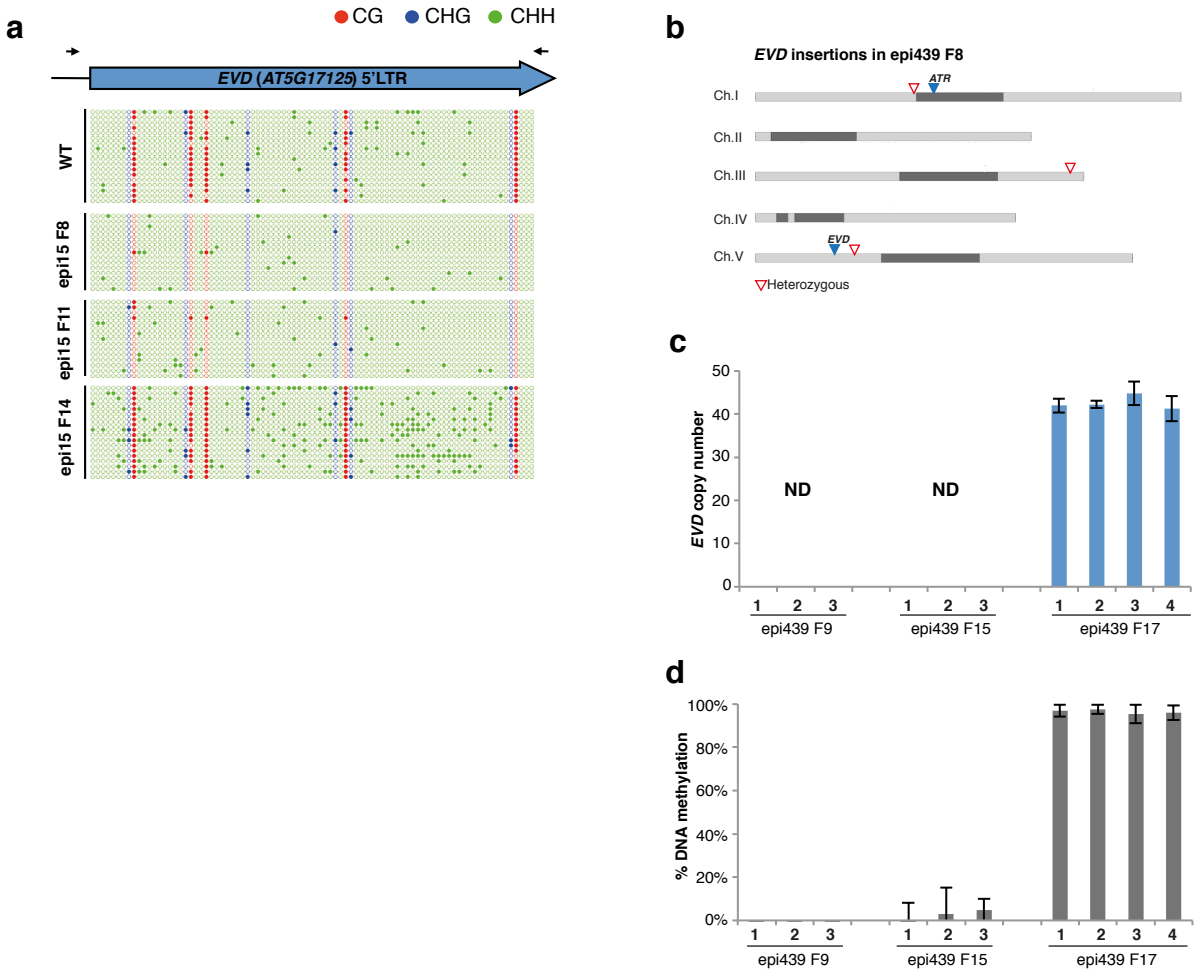
Supplementary Figure 6



Supplementary Figure 6: Mapping of 21-22-nt siRNAs to *EVD* in epi15 F14 and *EVD* copy number in epi454 and 438 F8.

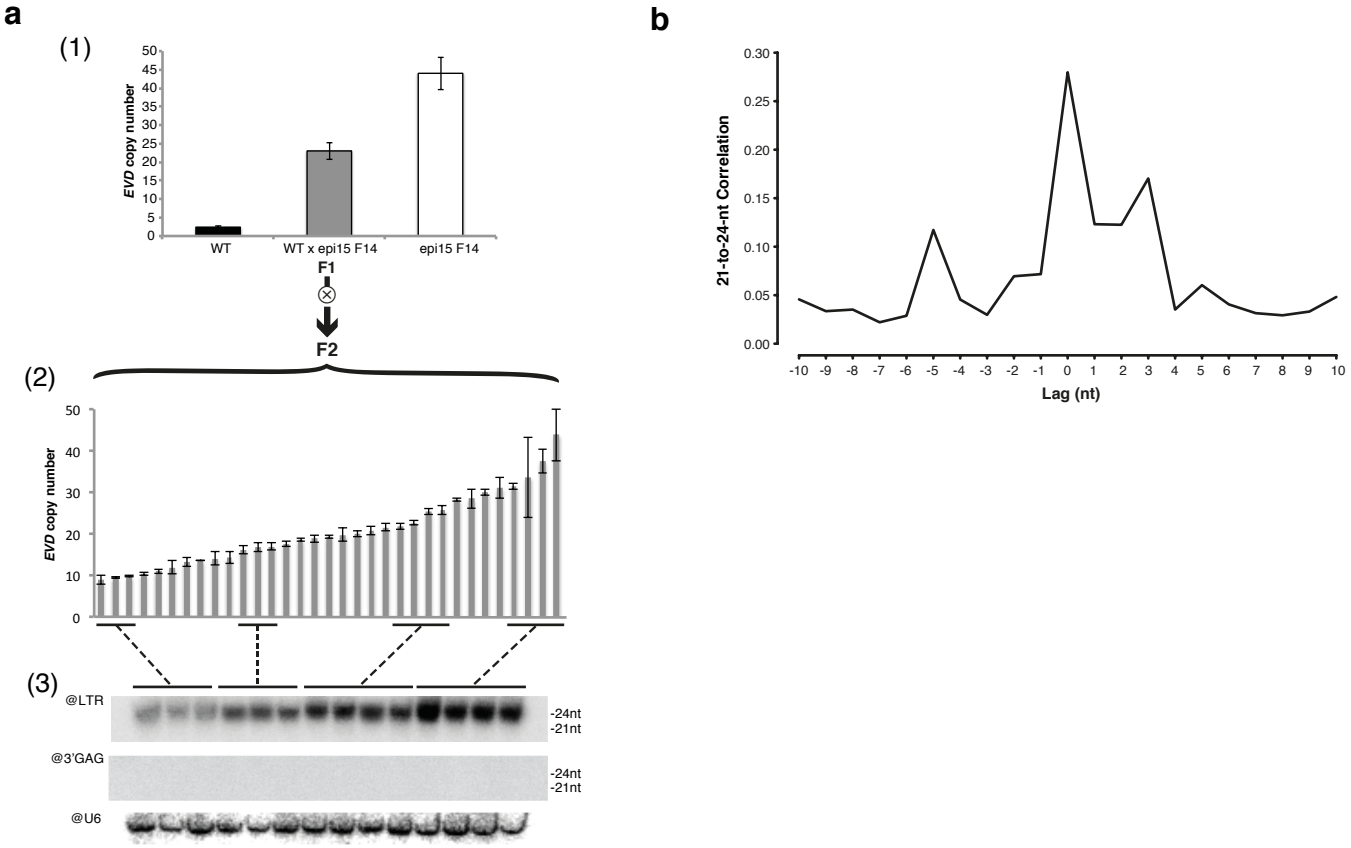
(a) Deep sequencing analysis of 21-22-nt siRNAs mapping to *EVD* in epi15 F14. **y** axis: number of normalized small RNA reads per million sequences (rpm); **x** axis: Arabidopsis chromosome 5 coordinates corresponding to the *EVD* locus found in Col-0 (TAIR 9). **s**, sense; **as**, antisense. **(b)** *EVD* copy number in epi454 and 439 F8 calculated as the total *EVD* insertions inferred from the number of heterozygous and homozygous insertions obtained by sequencing (Gilly et al, submitted) including the original *EVD* and *ATR* loci.

Supplementary Figure 7



Supplementary Figure 7: Bisulfite sequencing dot-plots of *EVD* LTR in epi15 generations and *EVD* copy number and DNA methylation in epi439 generations. **(a)** Dot-plots depiction of bisulfite sequencing data obtained for individual clones used to generate the results shown in Figure 4a and Supplementary Table 2. Filled circles indicate methylated cytosine and empty circles unmethylated cytosine. **(b)** Scheme of Arabidopsis chromosomes showing *EVD* insertions found in the *ddm1*-derived epi439 line at F8. The original *EVD* and *ATR* loci are marked with blue triangles, red triangles indicate new transposition events. The three new insertions were selected for further analysis. **(c)** *EVD* copy number measured by qPCR in individual plants of epi439 F17. *EVD* copy number was not determined (ND) in the F9 and F15 generations. Error bars represent standard deviation from two independent experiments involving duplicate PCRs each. **(d)** McrBC-qPCR-based methylation analysis at the LTRs of *EVD* new insertions shown in (b) in the epi439 F9, -F15 and -F17 individual plants used in (c). Error bars represent standard deviation from duplicate PCRs in c-d.

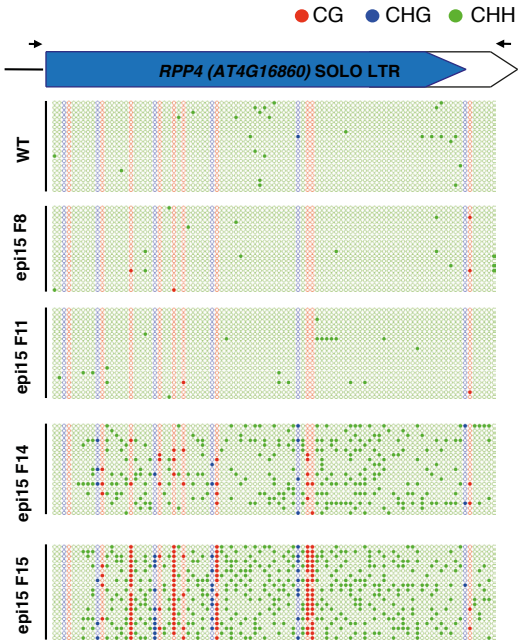
Supplementary Figure 8



Supplementary Figure 8: Correlation of LTR 24-nt siRNAs abundance with EVD copy number and 21-24-nt siRNAs mapping correlation on EVD.

(a) To explain the PTGS-to-TGS shift in EVD silencing, we considered the possible emergence of discrete EVD loci with peculiar features allowing production of 24-nt LTR siRNAs with potent trans-TGS activities. This hypothesis had two predictable and interrelated implications: (i) such loci would produce disproportionate amounts of 24nt siRNAs independently of EVD copy numbers and, (ii) these loci would be easily segregated away from an F14. We therefore tested these ideas here. (1) EVD copy number measured by qPCR in WT, epi15 F14 and the F1 progeny of the F14 X WT cross. (2) Same as in (1) but in distinct plants of the F2 progeny, grouped according to their respective copy numbers. Error bars represent standard deviation from two independent experiments involving triplicate PCRs each. (3) Northern analysis of EVD LTR 24-nt siRNA levels in F2 progeny plants containing similar numbers of EVD insertions as determined in (2). U6 detection is shown as loading control. The results clearly show that (i) all F2s, with no exception, produce 24nt siRNAs and (ii) 24nt-LTR siRNA levels strictly correlate with the copy number of segregating EVD copies, making the emergence of the hypothesized, specific EVD loci highly unlikely.
(b) Pearson correlation coefficient (Y-axis) between the 5'-end of all 24mers and all 21-mers mapping the 2,000 first nucleotides of EVD with the computation of virtual a lag from -10 to +10 nucleotides (X-axis) between the first nucleotide of all 21-mers and all 24-mers. This allows to assess if DCLs process the 24-mers and the 21-mers at the same position of a shared dsRNA substrate.

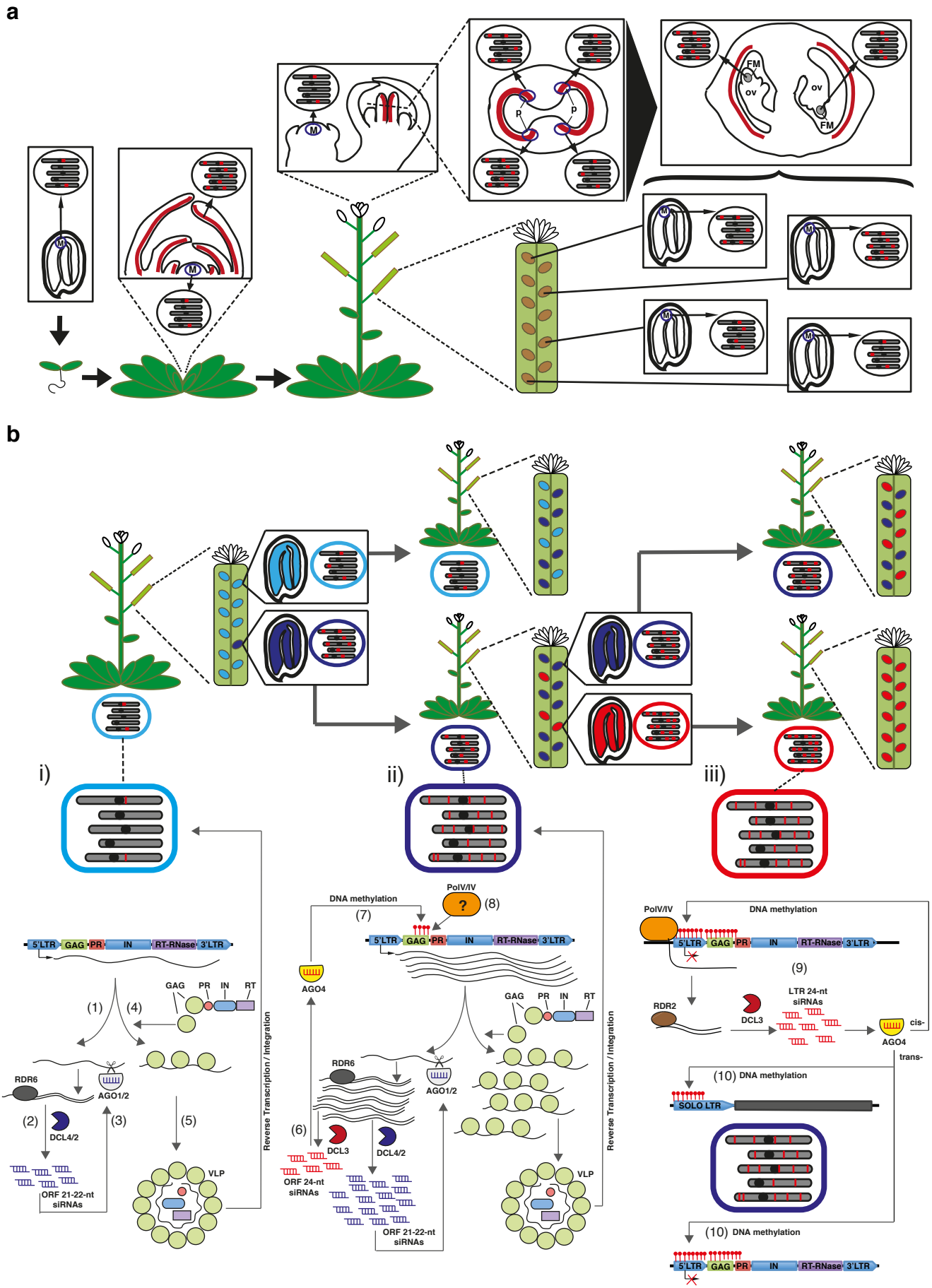
Supplementary Figure 9



Supplementary Figure 9: Bisulfite sequencing dot-plots of *RPP4* Solo-LTR in epi15 generations.

Dot-plots of bisulfite sequencing from individual clones for results shown in Figure 6b and Supplementary Table 4. Filled circles indicate methylated cytosine and empty circles unmethylated cytosine.

Supplementary Figure 10



Supplentary Figure 10: *EVD* life cycle in Arabidopsis.

(a) *EVD* expression pattern ensures host integrity through developnet and transmission of new insertions to the next generation. Once reactivated, adaxial L2-specific *EVD* expression (in red) restricts potential transposition to somatic tissue. Meristem (harbouring the stem cells from which aerial organs derive; M) exclussion will then allow the plant to develop without the accumulation of *EVD* new insertions. However, L2 expression and transposition in gynoecia (flower female reproductive organ), particularly at the placental regions (P) which give rise to the ovules (ov), from which a L2 cell differentiates into the haploid female megaspore (FM) , will allow transmission of *EVD* new insertions (see Supplementary Fig.2 for more details). Therefore, after fertilization, and embryo development, each seed will display different number and pattern of *EVD* insertions. **(b)** Rise and demise of *EVD*. In early proliferation stages (A), a low *EVD* copy number (light blue; i) accounts for moderate transcription and dsRNA synthesis by RDR6 (1). Processing of dsRNA by DCL4 and DCL2 produces 3'GAG 21-22-nt siRNAs (2) loaded into AGO1/2 (3) that can only partly degrade the EVD RNA (4) due to its protection by GAG as part of putative viral-like particles (VLPs; 5). Because single seeds acquire distinct patterns and numbers of *EVD* insertion over generations, an increasing amount of progeny plants will attain the ~40 copies threshold (dark blue; ii) beyond which saturating dsRNA levels allow processing, by DCL3, of 3'GAG 24-nt siRNAs loaded into AGO4 (6), these induce *de novo* methylation of the 3'GAG DNA (7), which, in turn, initiates antisense transcription, possibly by PolIV/PolV, and coordinated spreading of methylation towards the LTR (8). Eventual epigenetic silencing of *EVD* is achieved over a few subsequent generations (red, iii) resulting in TGS maintained at the LTR by RdDM (9). Ensuing 24-nt LTR siRNAs display potent *cis*- and *trans*-silencing properties (10).

Supplementary Table 1

		Total			% methylated			# Clones	# Exp	Wilson Score Interval (95% confidence level)					
		CG	CHG	CHH	CG	CHG	CHH			CG int-	CHG int-	CHH int-	CG int+	CHG int+	CHH int+
EVD	WT	85	102	1274	84,70%	7,84%	10,12%	17	2	75,57%	4,03%	8,58%	90,83%	14,71%	11,90%
	met1-3	80	96	1199	6,25%	3,12%	3,00%	14	2	2,70%	1,07%	2,17%	13,81%	8,78%	4,13%
	epi15	110	132	1647	3,66%	1,51%	1,94%	22	2	1,44%	0,41%	1,38%	9,01%	5,35%	2,73%
ATR	WT	80	96	1197	88,75%	18,75%	22,05%	16	2	79,98%	12,20%	19,79%	93,97%	27,70%	24,49%
	met1-3	103	125	1560	3,88%	4,00%	6,28%	21	2	1,52%	1,72%	5,18%	9,56%	9,02%	7,59%
	epi15	89	106	1345	85,39%	17,92%	21,04%	18	2	76,59%	11,78%	18,95%	91,26%	26,30%	23,30%

Supplementary Table 1: Figure 1 and Supplementary Figure 1 bisulfite sequencing raw data.

Raw data from the bisulfite sequencing analysis in Figure 1b and Supplentary Figure 1. The number of CG, CHG and CHH sites as well as the number of sequenced clones are indicated. Wilson score intervals are shown for a confidence level of 0.95.

Supplementary Table 2

	Total			% methylated			# Clones	# Exp	Wilson Score Interval (95% confidence level)					
	CG	CHG	CHH	CG	CHG	CHH			CG int -	CHG int -	CHH int -	CG int +	CHG int +	CHH int +
WT	90	108	1350	82,22%	10,28%	3,92%	18	2	73,05%	5,85%	3,01%	88,75%	17,44%	5,09%
epi15 F8	75	90	1125	2,66%	1,12%	2,76%	15	2	0,73%	0,20%	1,95%	9,20%	6,04%	3,89%
epi15 F11	75	90	1125	8,00%	3,33%	3,82%	15	2	3,72%	1,14%	2,85%	16,37%	9,34%	5,11%
epi15 F14	90	108	1350	93,33%	21,29%	13,85%	18	2	86,21%	14,63%	12,11%	96,91%	29,93%	15,80%

Supplementary Table 2: Figure 4 and Supplementary Figure 7 bisulfite sequencing raw data.

Raw data from the bisulfite sequencing analysis in Figure 4a and Supplementary Figure 7. The number of CG, CHG and CHH sites as well as the number of sequenced clones are indicated. Wilson score intervals are shown for a confidence level of 0.95.

Supplementary Table 3

		Total			% methylated			# Clones	# Exp	Wilson Score Interval					
		CG	CHG	CHH	CG	CHG	CHH			CG int-	CHG int-	CHH int-	CG int+	CHG int+	CHH int+
5'GAG	WT	176	80	581	94,88%	32,50%	8,60%	16	2	90,56%	23,24%	6,58%	97,28%	43,36%	11,16%
	epi15 F8	182	89	660	4,39%	3,37%	2,57%	18	2	2,24%	1,15%	1,61%	8,43%	9,45%	4,08%
	epi15 F11	194	90	663	26,28%	4,44%	2,26%	18	2	20,59%	1,74%	1,37%	32,89%	10,87%	3,70%
	epi15 F14	252	115	850	60,31%	33,91%	6,58%	23	2	54,16%	25,90%	5,10%	66,15%	42,96%	8,45%
3'GAG	WT	117	104	585	77,77%	2,88%	0,68%	13	2	69,42%	0,98%	0,26%	84,35%	8,13%	1,74%
	epi15 F8	162	144	810	5,55%	0,69%	0,86%	18	2	2,95%	0,12%	0,42%	10,21%	3,82%	1,77%
	epi15 F11	198	176	990	30,80%	11,36%	5,65%	22	2	24,79%	7,47%	4,38%	37,54%	16,90%	7,27%
	epi15 F14	180	159	899	77,33%	8,80%	3,67%	20	2	70,68%	5,31%	2,62%	82,84%	14,23%	5,11%

Supplementary Table 3: Figure 5 bisulfite sequencing raw data.

Raw data from the bisulfite sequencing analysis in Figure 5d-g. The number of CG, CHG and CHH sites as well as the number of sequenced clones are indicated. Wilson score intervals are shown for a confidence level of 0.95.

Supplementary Table 4

	Total			% methylated			# Clones	# Exp	Wilson Score Interval (95% confidence level)					
	CG	CHG	CHH	CG	CHG	CHH			CG int-	CHG int-	CHH int-	CG int+	CHG int+	CHH int+
WT	189	112	1451	0,00%	0,89%	1,72%	19	2	0,00%	0,16%	1,17%	1,99%	4,88%	2,53%
epi15 F8	180	108	1385	2,22%	0,00%	1,01%	18	2	0,87%	0,00%	0,60%	5,57%	3,43%	1,69%
epi15 F11	190	113	1460	1,05%	0,00%	1,43%	19	2	0,29%	0,00%	0,94%	3,75%	3,29%	2,18%
epi15 F14	189	114	1460	13,75%	10,52%	17,39%	19	2	9,56%	6,12%	15,53%	19,39%	17,49%	19,42%
epi15 F15	200	119	1533	51,11%	26,05%	25,11%	20	2	44,23%	19,00%	23,00%	57,95%	34,60%	27,34%

Supplementary Table 4: Figure 6 and Supplementary Figure 9 bisulfite sequencing raw data.

Raw data from the bisulfite sequencing analysis in Figure 1b and Supplementary Figure 9. The number of CG, CHG and CHH sites as well as the number of sequenced clones are indicated. Wilson score intervals are shown for a confidence level of 0.95.

Supplementary Table 5: List of primers.

(Non-converted bisulfite primers sequence is indicated in braquets)

EXPERIMENT	PRIMER	SEQUENCE (5'-3')
EVD RT-PCR/RFLP	EVD-RT F	GATCTTGGTGAACTCAAATACTT
	EVD-RT R	TACTCCCCCTCAAGCGCAAGAC
EVD copy# qPCR/ qRT-PCR	EVD-qRT F	GATAGAGGAGATAGAAGATCTACAACTGG
	EVD-qRT R	CTCTATACTCCGATTCTGCACTCGAACA
EVD ISH/Southern/ HMW-Northern probe	EVD-P F	TTTCATGGTATCAGAGCATAAGAT
	EVD-P R	CACCCGAGTCTATCACAATTGAAC
LTR Bisulfite PCR	EVD-LTR BS F	TTTATAARCATAAAAACATAATCTTATRCTCTAATACCATA (TTTATAAGCATAAAGACATGATCTTATGCTCTGATACCATG)
	EVD-LTR BS R	GGAAAAAAAYGAAGGAATTTTAGAATAYGATTTGGTAATTT (GGAAAAAAACGAAGGAACCCTAGAACACGATTTGGCAATTC)
	ATR-LTR BS R	TTTTAATGAGGTGATGAATTATGTGTTTTAAGTTTAGATTT (TTTTAATGAGGTGATGAATCATGTGTTCCAAGCTTAGACTC)
LTR probe	EVD-LTR F	TTGATCAAGACTCAAATAAGAAAGGCC
	EVD-LTR R	TATGCTCTGATACCATGAAAGAATAT
5'GAG probe	EVD-P F	TTTCATGGTATCAGAGCATAAGAT
	EVD-5G R	TTTTTAAGTGTATCCCACAACTCTT
3'GAG probe	EVD-3G F	TAAGTCAAGAAGACTTAGAGTTTA
	EVD-3G R	ACTTTGCTCCTCATGAGTTTCTT
PR-IN probe	EVD-PI F	GGTGGTTCAAAAGGAGAACTCTATGTTCT
	EVD-PI R	GAAACATCATTGACCGCGCCACCTCCAT
IN-RT probe	EVD-IR F	GAAGGATTGGGAGAACCTAAAAGATCTTAC
	EVD-IR R	CTTATACTTGATTGTGAATACCCACCTTGA
RT-RN probe	EVD-RR F	CAAAGACGGTATAGACTCTACCAAGAC
	EVD-RR R	CTCTATACTCCGATTCTGCACTCGAACA
RN probe	EVD-RN F	GATAGAGGAGATAGAAGATCTACAACTGG
	EVD-RN R	TACTCCCCCTCAAGCGCAAGAC
mGFP probe	mGFP-F	AGTAAAGGAGAAGAACTTTTCACT
	mGFP-R	TTCCGTCCTCCTTGAAATCGA
EVD WT GAG-POL ORF Cloning	EVD ORF attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGACTTCTCAAAAGATG
	EVD ORF attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAAGAGTGAGATAGATCC
EVD dRBD Site-directed Mut.	EVD dRBD F	AGGATAGAAAACTGTTGACTTGGTTGCTACATCCACACCT
	EVD dRBD R	AGGTGTGGATGTAGCAACCAAGTCAACAGTTTTCTATCCT
EVD HA-GAG-POL-FLAG Cloning	HA-EVD F	ATGTACCCATACGATGTTCCAGATTACGCTGGTGGAGGCGAGACTTCTCAAAAGATGA
	EVD-FLAG R	TCACTTGTCATCGTCATCCTTGTAATCGCCACCTCCAGAGTGAGATAGATCCACAA
	HA-EVD attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTACCCATACGATGTTCC
	EVD-FLAG attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTTGTCATCGTCATCCTTGT
epi454/439 *EVD* insertions zygosity	454 EVD-A F	CTACTGGCGTGGATGCTTTT
	454 EVD-A R	GTAAACAGAGTTCAGATTTAAACCAA
	454 EVD-B F	TTCTTCCAAAATTTGATTACTTGC
	454 EVD-B R	TCCGAAAATGGCATAAGAGG
	454 EVD-C F	TTGGAATTCGATTCAACTAGAGG
	454 EVD-C R	CTCAAATTTCATTGGAGACACATC
	454 EVD-D F	TGTGCGTAGCTTTTCCTTGA
	454 EVD-D R	ATTGAGGAACCAGCATCACC
	454 EVD-E F	GCCTTAATGAATGGCCCAATA
	454 EVD-E R	ACTCCACGGGACTACCTCAC
	454 EVD-F F	GACCCCAACCAATCAATCAC
	454 EVD-F R	GACGTTATGGTGGAATATGTGC
	454 EVD-G F	AAAATGTCACTTATCATTGCAAGC
	454 EVD-G R	TTTTAGTGGTGCGGGCTTTA
	454 EVD-H F	CCAACGTGGGTTAGATCCAT
	454 EVD-H R	ACGGCTATTGGAGTTTTTGA
	454 EVD-I F	CGGTGTTTGTAGTGGGAACC
	454 EVD-I R	GCAAAAACATTCGTACATTTCG
	454 EVD-J F	TCTCTTGTCTTGTGCGACTCA
	454 EVD-J R	CACACCATGGATAATGCTCTT
	439 EVD-ChI F	ATCAACAAATAGATATCTTTGCATTTT
	439 EVD-ChI R	TTCTTTTCGGATTCTGTGTGG
	439 EVD-ChIII F	ATGGTTCGATTGTGGAGGAA

	439 EVD-ChIII R	CAACATTCTAGGGTTTTACCTTGC
	439 EVD-ChV F	TCTCCCACCCAAAATCTCAG
	439 EVD-ChV R	GCGTGTTGTTTTGCAGTGAT
epi454/439 *EVD* insertion specific McRB-qPCR methylation analysis	EVD-LB R	ACTGGTCATTGGTTGACCTTG
	EVD-RB F	CCAAGACCAAGGAACCTCTTC
	454 EVD-A qF	ATGAAAATATTTGGCGTGTTCAT
	454 EVD-A qR	TCGAATCTGTCTTCTAGATTATGTGTG
	454 EVD-B qF	GCTCAGAGGCAGTTGTGAAAG
	454 EVD-B qR	ACTTTTCAAGCAACGTTGTGC
	454 EVD-C qF	CATCAGCTTGATGATCTACAGAGAG
	454 EVD-D qF	TTGTCCCGAGGATAGTCTTC
	454 EVD-D qR	ACCAGAGTATATAGAACATGATATTGAAGTG
	454 EVD-E qF	CGTTTTAAGACGAAGGGTGTG
	454 EVD-E qR	AGGACTGACGTCACACCCTA
	454 EVD-F qF	CCTTTCGAGTCGAGATCCAAT
	454 EVD-F qR	TCATATTCACATATGTTGGTGTTGG
	454 EVD-G qF	CTGTACATTTCAATATTTTATCCATCTTG
	454 EVD-G qR	AAATCACAACATATGTAAATGTATCAAGT
	454 EVD-H qF	GAAGACTTTATGCCGTTCATCA
	454 EVD-H qR	TTACATTCTTAGCGTTTCCTGGT
	454 EVD-I qF	TTCAGTTTCATTAATAAGAACCAGGTTAG
	454 EVD-J qF	AAAACCTGGTACCTTCTCTGTGA
	454 EVD-J qR	TCAAAAACATTCAGAAGGTCAGAG
	439 EVD-ChI qF	AAAGCAGTCGAGAATAAGAACCA
	439 EVD-ChI qR	AGATCTTCTGTTTTGATGCAAGTC
	439 EVD-ChIII qF	TGAGGAGTTGAATTCTGTGATGA
	439 EVD-ChIII qR	CTTGAGCCCCAAGGAAGTC
	439 EVD-ChV qF	CCAGTTTGACGAAATTAAAACAA
	439 EVD-ChV qR	CGTAACCAAACATAAGGAGAACC
5'GAG Bisulfite PCR	EVD-5GAG BS F	ATTTTAAGGGGGAAATTATTTAATTTGGTTAAGAATAAT (ACTTCAAGGGGGAAACTACCTAACTTGGTCAAGAACAAC)
	EVD-5GAG BS F	ACCTAAACTTACCAAAATATTTAATAAACTCTAAATCTTCTTA (ACCTGAACTTACCAAAGTGTTTAGTAAACTCTAAGTCTTCTTG)
3'GAG Bisulfite PCR	EVD-3GAG BS F	GTGTTTGAAGTGGAAGAAGGYGATTAAYGAATTAA (GTGTTTGAAGTGAAGAAGGCGATTAACGAACTAA)
	EVD-3GAG BS F	ATAACCCRACTTAACTTTRCTCCTCATAAATTTCTTAAA (ATGACCCGGCTTGACTTTGCTCCTCATGAGTTTCTTGAG)
RPP4 qRT-PCR	RPP4-qRT F	ACATGCGAGGACAAACCAGAGG
	RPP4-qRT R	TTCAATTCCGACGAAGTCACC
RPP4 SOLO LTR Bisulfite PCR	RPP4-SLTR BS F	GGGAGGYGTTTAGGTAATTTATAYGTATATTATATATTGATT (GGGAAGGCGTCCAGGTAATTCATACGTATATTATATATTGATC)
	RPP4-SLTR BS R	TTCCTCTACAAAAATATTTTAAAAACACAAATTTTATAATTATCAA (TTCCTCTACAAAGATATTTTGAAAACACAAATTTTGTGGTTATCAG)
AT1G68420 qRT-PCR	AT1G68420 qF	GTCTCACACTTCAAGGCATCTTGC
	AT1G68420 qR	TTCATAGCTTCCTCTACACCGGCA
AT5G25550 qRT-PCR	AT5G25550 qF	GTTACAAGCGTGGAAACGTGCCAT
	AT5G25550 qR	TGTAGTCGCACACACGTGATCCAA
qRT-PCR Controls	ACT2 qF	GCACCCTGTTCTTCTTACCG
	ACT2 qR	AACCCTCGTAGATTGGCACA
	AT5G13440 qF	ACAAGCCAATTTTTGCTGAGC
	AT5G13440 qR	ACAACAGTCCGAGTGTCATGGT
	AT2G36060 qF	TGAAGTCGTGAGACAGCGTTG
	AT2G36060 qR	GGGCTTCTCCATTGTTGGTC
	AT4G29130 qF	GGCGTTTTCTGATAGCGAAAA
	AT4G29130 qR	ATGGATCAGGCATTGGAGCT

Discussion complémentaire sur Mari-Ordonez et al. 2013 :

Bien que ces travaux aient permis d'établir les étapes clefs permettant l'établissement du contrôle épigénétique via la méthylation de l'ADN sur les copies nouvellement insérées d'*ATCOPIA93*, certaines zones d'ombre subsistent et mériteraient des analyses complémentaires.

Tout d'abord, si le ciblage de cette région par les siARN dépendant de RDR6 (21nt puis 24nt) est essentiel pour l'établissement du contrôle épigénétique, on ne sait pas à ce jour pourquoi cette région est spécifiquement ciblée. Les travaux sur le rétrotransposon à LTR COPIA Ty1 chez la levure ont permis de mettre en évidence un transcrit antisens initié dans cette région (Berretta et al. 2008, Matsuda and Garfinkel 2009). On peut également évoquer le ciblage potentiel de cette région par un microARN qui engendrerait la production de siARN secondaires dépendants de RDR6 comme cela est le cas dans la voie des trans-acting siARN.

D'autre part, le passage subit du RdDM de GAG vers les LTR reste encore à élucider. Comme proposé dans l'article, il pourrait être dû à un phénomène de « spreading », ici encore causé par un transcrit antisens initié dans GAG. Un tel transcrit serait synthétisé par l'une des polymérases du RdDM (POLIV ou POLV) qui ciblerait GAG une fois l'établissement de la méthylation initié dans cette zone. Des croisements entre les epiRIL et les mutants pour ces polymérases permettront de mettre en évidence leur implication dans l'établissement de la méthylation sur la région GAG et/ou dans le passage vers les LTR.

Enfin, comme décrit précédemment, chez la plante sauvage, *EVADE* est contrôlé quasi-exclusivement par MET1 et DDM1. Il semble cependant que son ciblage par le RdDM, nécessaire à l'établissement de la méthylation, perdure dans les epiRIL sur plusieurs générations après la mise en place de la méthylation sur les LTR (au moins 6 générations plus tard dans l'epiRIL 454). Il serait donc intéressant de tester si ce ciblage par le RdDM est désormais nécessaire (et suffisant ?) pour le maintien de la répression transcriptionnelle. Ceci pourra être testé en croisant les epiRIL avec des mutants affectés dans les différentes voies de la méthylation, notamment dans le RdDM. En outre, on peut se demander si ce ciblage par le RdDM va un jour être perdu au profit d'une régulation basée essentiellement

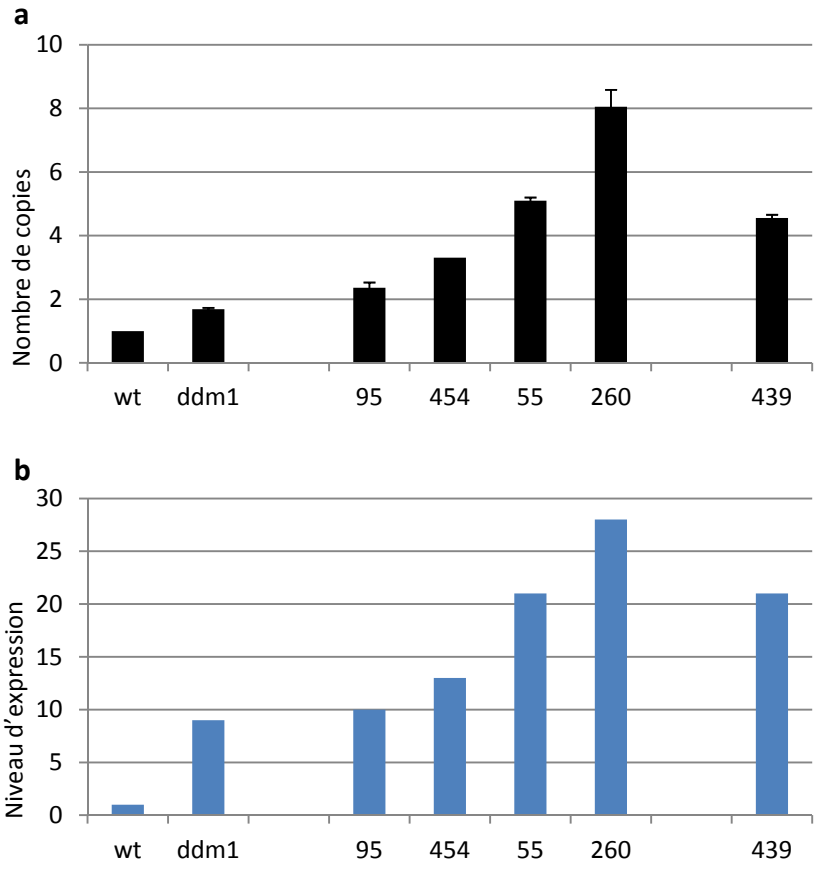


Figure 3.3 : Corrélation entre le nombre de copies d'*AT2TE42810* (*VANDAL21*) et le niveau d'expression chez *ddm1* et les epiRILs.
(**a**) Evaluation du nombre de copies par qPCR en utilisant des amorces spécifiques de la région interne d'*AT2TE42810*. (**b**) Représentation du niveau d'expression global d'*AT2TE20205* établi à partir des données d'ARNseq et normalisé en fonction de la taille des banques

sur l'action de MET1 et DDM1. Ceci pourra être exploré dans des générations très avancées des epiRIL.

3.1.2 Etude de la mise en place du contrôle épigénétique sur les transposons à ADN

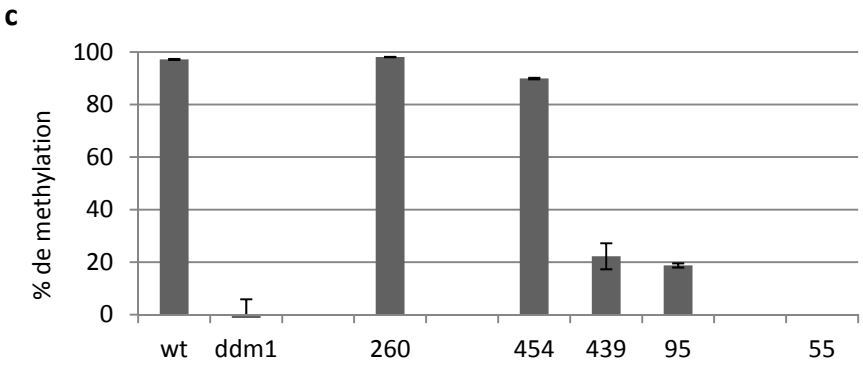
Les nombreux travaux réalisés sur le contrôle des transposons à ADN, notamment chez le maïs, ont mis en évidence des systèmes de régulation complexes et qui diffèrent selon les familles d'ET. Il apparait également qu'il existe souvent un découplage entre contrôle de l'expression et de la mobilisation. Le cas des transposons de type EN/SPM décrit dans la partie 2 en fournit un exemple frappant. Chez le maïs, la protéine TNPA, codée par l'élément, est nécessaire à la mobilisation et joue un rôle de régulateur positif car elle induit l'hypométhylation en *trans* et la réactivation transcriptionnelle de copies silencieuses appelées cryptiques. Cependant, l'expression élevée de TNPA peut également réprimer la transposition en se fixant sur les régions promotrices de l'ET si elles sont fortement hypométhylées, ce qui a pour effet de bloquer la transcription (Raina et al. 1998, Schläppi et al. 1994). Cet exemple met donc en évidence la relation complexe qui peut exister entre méthylation de l'ADN et mobilisation des transposons à ADN. De fait, il a été montré, toujours chez le maïs, que l'hypométhylation complète des séquences subterminales de l'élément *Ac* n'est pas requise pour la transposition et que la forme hémiméthylée produite juste après la réplication peur être reconnue par la transposase et mobilisée (Wang et al. 1996, Ros and Kunze 2001). Il a même été montré *in vitro* une affinité plus élevée de la transposase pour la forme hémiméthylée que pour celle complètement hypométhylée (Kunze and Starlinger 1989).

Néanmoins, chez Arabidopsis, la méthylation élevée des transposons à ADN observée chez le sauvage est quasi-systématiquement associée à leur répression transcriptionnelle et à l'absence de mobilisation. Je me suis donc intéressée à la mise en place du contrôle épigénétique sur les nouvelles copies d'*ATENSPM3* et *VANDAL21*, les deux transposons à ADN les plus mobiles dans les epiRIL.



b

	260	454	439	95	55
origine parentale du donneur	wt	wt	wt	wt	ddm1 (absente)
insertion(s) partagée(s)	0	2	0	0	0
insertion(s) unique(s)	0	0	4	8	11



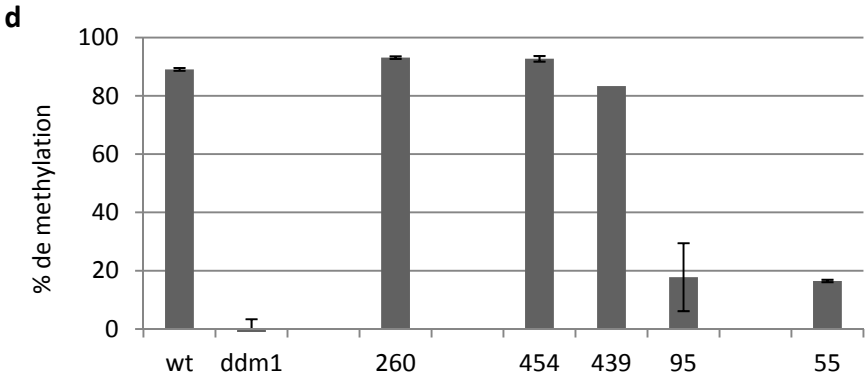


Figure 3.4 : Etude comparative des profils de méthylation et d'expression d'*AT2TE20205* (*ATENSPM3*) chez le mutant *ddm1* et les épiRIL.

(**a**) Représentation schématique du positionnement des amorces de qPCR utilisée pour mesurer le % de méthylation de l'extrémité 5' (1) et de la partie interne (2). (**b**) Table récapitulative de l'origine parentale du donneur et de la présence de nouvelles insertions uniques et/ou partagées pour les lignées étudiées. (**c**) % de méthylation de l'extrémité 5' du locus donneur mesuré par McrBC-qPCR. (**d**) % de méthylation moyen sur l'ensemble des copies au niveau de la partie interne d'*AT2TE20205* mesuré par McrBC-qPCR. Les barres d'erreur représentent l'écart type entre deux répliques techniques.

- *VANDAL21*

Certaines observations faites dans la partie 2 semblent indiquer que *VANDAL21* est actif dans toutes les epiRIL présentant de nouvelles insertions. En effet, les analyses faites sur la méthylation en 5' du locus donneur (*AT2TE42810*) ont montré soit l'absence de reméthylation lorsque ce dernier est hérité du parent *ddm1* soit son hypométhylation systématique lorsqu'il est hérité du parent sauvage et que de nouvelles insertions sont également présentes (fig. 2.17). Dans le premier cas, la copie issue de *ddm1* n'a pas reméthylé, dans le second, au moins une des nouvelles insertions doit être active pour causer la déméthylation en trans du locus donneur (car ce phénomène n'est pas stable en l'absence de copie active). On ne peut cependant pas conclure que toutes les copies sont actives au sein de chaque epiRIL. Afin de déterminer le niveau d'expression d'*AT2TE42810* dans les epiRIL, j'ai tiré avantage de données de transcriptome (ARNseq) de cinq epiRIL ainsi que de lignées sauvage et *ddm1* apparentées aux deux lignées parentales. Les cinq epiRIL présentent toutes des nouvelles insertions d'*AT2TE42810* mais des caractéristiques diverses concernant la dynamique et la fréquence d'insertion (fig. 3.3a et 2.12). L'expression d'*AT2TE42810* (copie donneuse et nouvelles insertions confondues) a pu être mise en évidence dans les cinq epiRIL (fig. 3.3b). En outre, on observe une corrélation quasi parfaite entre le nombre de copies d'*AT2TE42810* et la quantité de transcrits. Cette observation suggère une expression égale de toutes les copies (donneuse et nouvelles insertions) au sein de chaque epiRIL. Notons que le nombre relativement élevé de copies dans l'epiRIL 260 ne promeut pas l'établissement de la répression (contrairement à ce qui est observé pour *ATCOPIA93*). Des études sur les générations avancées des epiRIL présentant déjà un grand nombre d'insertions de *VANDAL21* en F9 devraient permettre de déterminer s'il existe un nombre limite de nouvelles insertions. A ce stade des analyses nous n'avons donc aucune indication de l'établissement de la répression transcriptionnelle sur les nouvelles insertions de *VANDAL21* dans les epiRIL.

- *ATENSPM3*

J'ai ici concentré mes analyses sur l'élément mobile pleine longueur potentiellement autonome *AT2TE20205*. Contrairement à ce qui vient d'être présenté pour *VANDAL21*, la

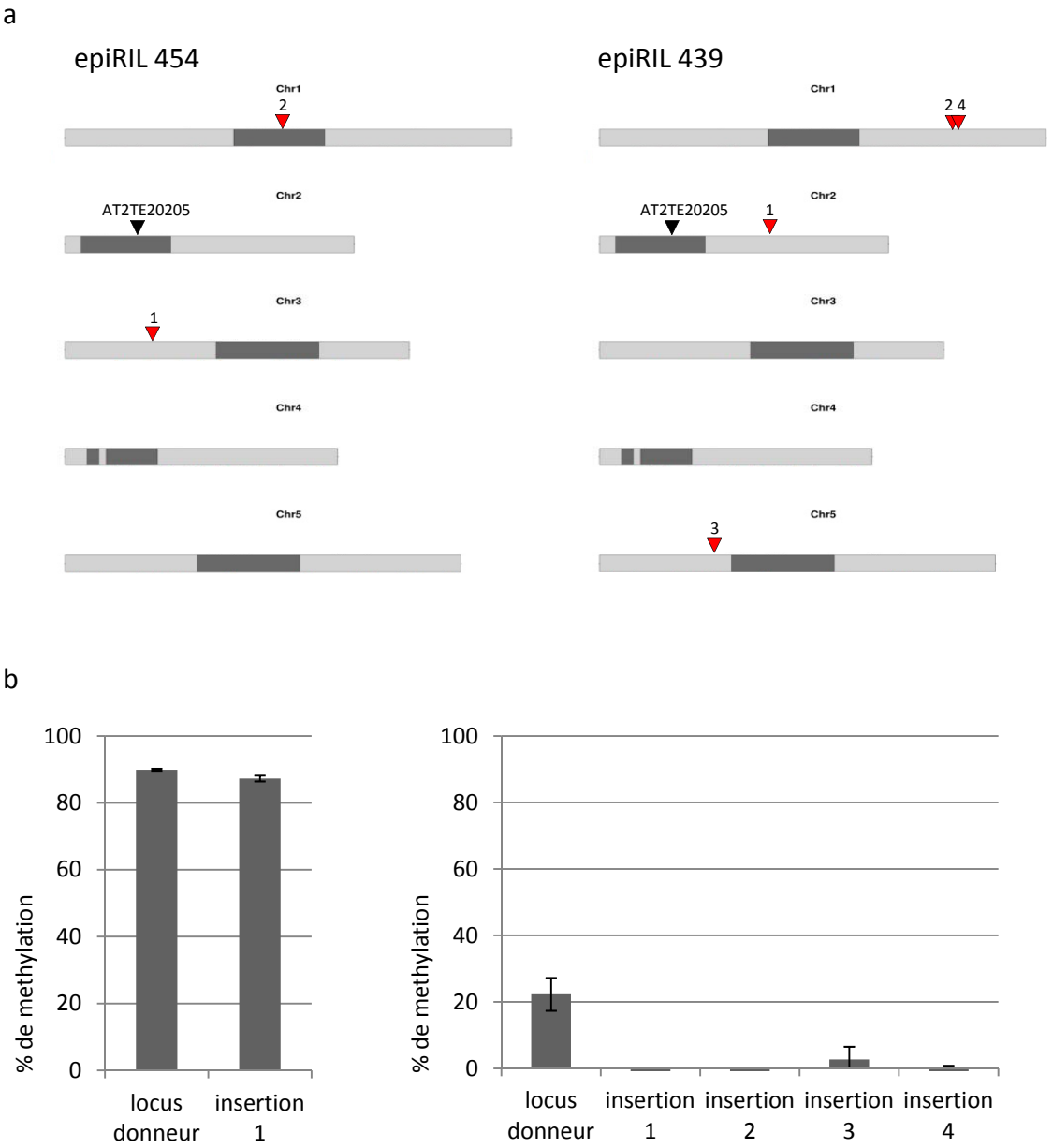


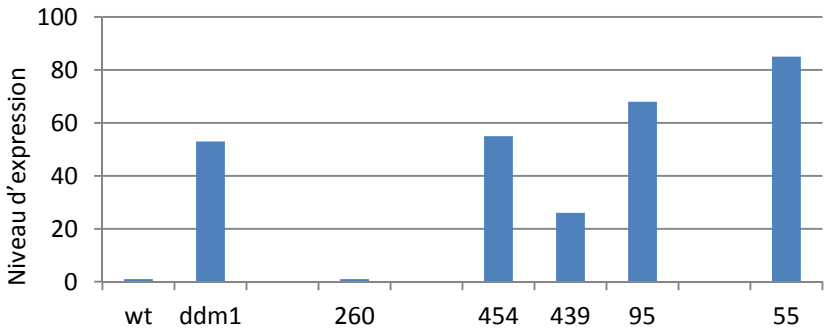
Figure 3.5 : Le % de méthylation au niveau de l'extrémité 5' des nouvelles insertions est identique à celui du locus donneur.

(**a**) Localisation du locus donneur (triangle noir) et des nouvelles insertions (triangles rouges) d'*AT2TE20205* (*ATENSPM3*) le long du génome des epiRIL 454 et 439. (**b**) % de méthylation par McrBC qPCR de l'extrémité 5' du locus donneur et des nouvelles insertions dans les epiRIL 454 et 439. Les barres d'erreur représentent l'écart type entre deux répliques techniques. (Remarque : la séquence nucléotidique de la région flanquant l'insertion 2 de l'epiRIL 454 ne m'a pas permis de dessiner d'amorces efficaces pour la qPCR).

présence de nouvelles copies d'*AT2TE20205* n'est pas systématiquement associée avec l'hypométhylation de la copie au locus donneur héritée du parent sauvage (fig. 2.16). Cette observation pourrait indiquer que, dans certaines lignées, le contrôle épigénétique a été établi sur les *ATENSPM3* (donneur et nouvelles insertions) car l'absence d'hypométhylation en trans peut signifier que les nouvelles insertions, normalement responsables de ce phénomène ne s'expriment plus dans ces lignées. Rappelons ici qu'il ne semblait exister aucun lien entre le nombre de nouvelles insertions, leur ancienneté (partagé vs unique) et l'absence d'hypométhylation en trans du locus donneur (cf. partie 2.2).

Afin de tester l'acquisition de la méthylation de l'ADN sur l'ensemble des copies d'*AT2TE20205,* j'ai estimé par McrBC-qPCR le niveau moyen de méthylation sur l'ensemble des copies en utilisant des amorces situées dans la partie interne de l'ET. Ainsi, un niveau de méthylation de 100% indique que toutes les copies (donneur et nouvelles insertions) sont méthylées dans cette région, et de 0% signifie qu'aucune copie n'est méthylée. Un niveau intermédiaire de méthylation peut refléter soit la présence de copies méthylées et de copies non méthylées soit une méthylation intermédiaire sur l'ensemble des copies. J'ai réalisé cette expérience sur 5 epiRIL présentant des caractéristiques différentes concernant l'origine parentale du locus donneur, la présence de nouvelles insertions partagées ou uniques, l'état de méthylation en 5' du locus donneur hérité du sauvage ainsi que sur la plante sauvage et le mutant *ddm1*. Comme attendu, la plante sauvage et l'epiRIL 260 (qui ne présente pas de nouvelles insertions et qui a hérité le locus donneur du parent sauvage) sont fortement méthylées dans cette région. Cette méthylation est complètement perdue dans le mutant *ddm1*. Les epiRIL 454 et 439 présentent un niveau élevé de méthylation alors que les epiRIL 95 et 55 montrent une absence totale de méthylation (fig. 3.4d). Il semble donc que nous ayons ici trois cas de figure : (i) absence de méthylation dans la région interne sur l'ensemble des copies et hypométhylation de la région 5' du donneur, (ii) méthylation dans la région interne sur l'ensemble des copies et hypométhylation du locus donneur, (iii) méthylation interne de l'ensemble des copies et absence d'hypométhylation du locus donneur. La méthylation de l'ensemble des copies y compris les nouvelles insertions dans la région interne n'empêcherait donc pas l'hypométhylation en trans du locus donneur (epiRIL 439). Pour les epiRIL 439 et 454 qui présentent toutes deux la méthylation interne sur l'ensemble des copies mais des profils de méthylation différents de la copie donneuse, j'ai cherché à

a



b

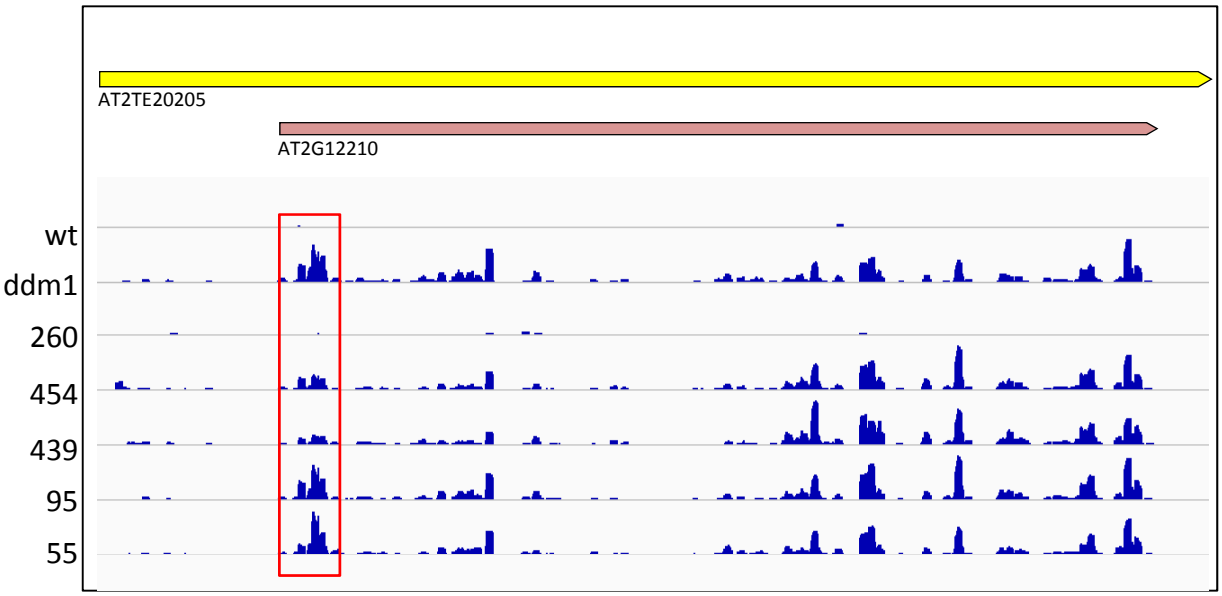


Figure 3.6 : Expression d'*AT2TE20205* (*ATENSPM3*) chez le sauvage, *ddm1* et les epiRILs.

(**a**) Niveau d'expression d'*AT2TE20205* établi à partir des données d'ARNseq et normalisé en fonction de la taille des banques (**b**) Représentation de la couverture des lectures le long d'*AT2TE20205* à partir de données d'ARN-seq. La région encadrée en rouge présente des différences de profil entre les lignées.

déterminer l'état de méthylation des nouvelles insertions une à une sur l'extrémité 5'. Il apparait que les nouvelles insertions se comportent de la même façon que la copie donneuse au niveau de l'extrémité 5' : méthylées dans l'epiRIL 454, hypométhylées dans l'epiRIL 439 (fig. 3.5). On peut alors supposer que la répression transcriptionnelle est due à l'état méthylé de l'extrémité 5', ce qui expliquerait que seule l'epiRIL 454 ne soit pas hypométhylée en trans. Afin de tester cette hypothèse, j'ai utilisé les données d'ARNseq disponibles pour ces lignées. Contrairement à ce qui pouvait être attendu, l'expression d'*AT2TE20205* est mise en évidence dans le mutant *ddm1* ainsi que dans toutes les epiRIL possédant des nouvelles insertions y compris l'epiRIL 454 qui présente pourtant des niveaux de méthylation comparables à la plante sauvage et à l'epiRIL 260 pour les régions étudiées (fig. 3.6a). Une méthylation dans la région 5' et dans la région interne (pourtant situé en 5' du gène *AT2G12210* codé par *AT2TE20205*) ne semble donc pas suffisante pour bloquer l'expression de l'élément.

Les études réalisées chez le maïs avaient mis en évidence une structure particulière du gène des éléments Spm et révélé notamment l'existence d'un épissage alternatif donnant lieu entre autres, à deux protéines aux fonctions distinctes (TNPA et la transposase TNPD) (Masson et al. 1989, Yuan and Wessler 2011). En m'appuyant sur les données de couverture obtenues par ARNseq, j'ai examiné le profil des transcrits correspondant à *AT2TE20205* dans les cinq epiRIL. Deux profils distincts sont observables pour les epiRIL présentant des nouvelles insertions et le mutant *ddm1*. Ils diffèrent au niveau de la partie 5' du gène *AT2G12210*, qui présente une diminution de couverture par rapport à la partie 3' du gène dans les epiRIL 454 et 439 en comparaison avec les epiRIL 95, 55 et le mutant *ddm1* (fig. 3.6b). Notons que les epiRIL 454 et 439 sont également celles qui présentaient une méthylation sur l'ensemble des copies dans cette même région (fig. 3.4), mais à ce stade des analyses il est impossible de déterminer s'il y a un lien de causalité entre ces deux observations ni même si elles sont liées car il faudrait étendre l'analyse sur plus de lignées.

Enfin, bien que le lien entre répression transcriptionelle et méthylation de l'ADN soit loin d'être évident pour les *ATENSPM3*, j'ai néanmoins cherché à déterminer la dynamique d'acquisition de la méthylation sur les nouvelles copies (et de réacquisition sur la copie donneuse), tant pour l'extrémité 5' que pour la région interne au cours des générations (fig. 3.7). Pour les deux epiRIL analysées, 454 et 439, on observe l'acquisition (ré-acquisition pour

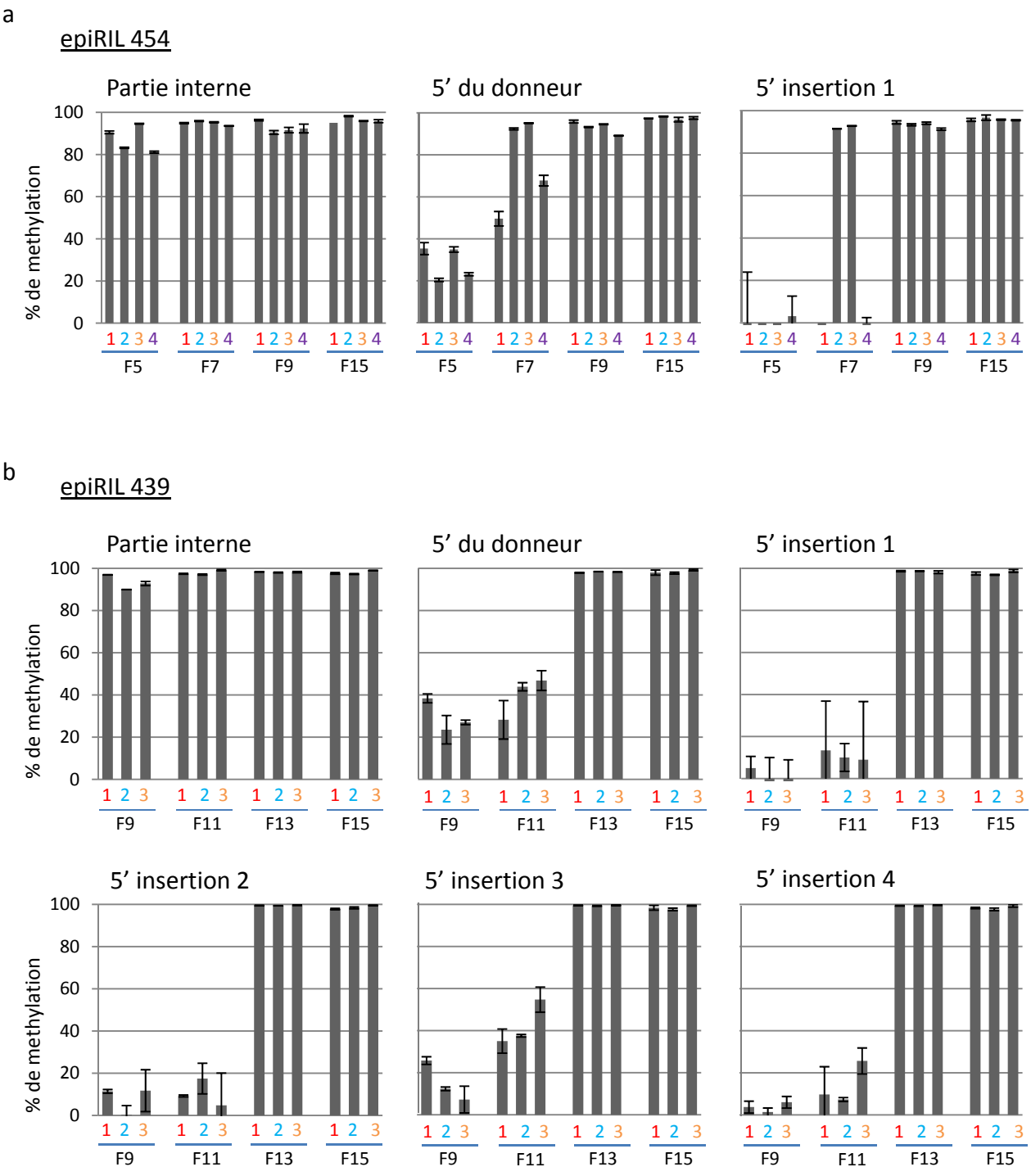


Figure 3.7 : Etude de l'établissement de la méthylation de l'ADN sur *AT2TE20205* (locus donneur et nouvelles insertions) au cours des générations dans les epiRIL 454 et 439 (mesures faites sur des plantes individuelles, 4 plantes par génération pour l'epiRIL 454 et 3 pour l'epiRIL 439). Les barres d'erreur représentent l'écart type entre deux répliques techniques.

le locus donneur) progressive de la méthylation de l'ADN au niveau de l'extrémité 5' des différentes copies au cours des générations. Concernant la partie interne en revanche, l'établissement de la méthylation n'a pu être établi. En effet, cette région apparait déjà complètement méthylée dans l'epiRIL 454 dès la F5 (fig. 3.7). On ne peut donc pas conclure à ce stade si cette méthylation interne est acquise à partir d'un état non méthylé ou si elle a toujours été présente dans ces epiRIL. On ne peut en effet pas exclure que les nouvelles insertions soient dues à la mobilisation du donneur hérité du parent sauvage qui, bien qu'hypométhylé en trans au niveau de l'extrémité 5', a pu conserver sa méthylation dans la partie interne. Cette hypothèse est pourtant peu vraisemblable car la mobilisation de la copie donneuse héritée du parent sauvage n'a jamais été mise en évidence (cf. partie 2). L'acquisition de la méthylation dans la partie interne pourra être testée dans les générations précoces de l'epiRIL 439 ainsi que dans les générations avancées des epiRIL 95 et 55.

3.1.3 Conclusions et perspectives

Les travaux menés sur *ATCOPIA93* ainsi que sur *ATCOPIA78, VANDAL21* et *ATENSPM3*, bien qu'encore préliminaires, ont déjà permis de mettre en évidence des modalités d'établissement du contrôle épigénétique contrastées pour les différentes familles d'ET.

Dans le cas d'*ATCOPIA78* la répression des nouvelles insertions est mise en place rapidement car la mobilisation est stoppée dès la F2, sans doute de façon concomitante à l'établissement de la méthylation de l'ADN. On ne peut cependant pas exclure qu'une étape de PTGS, bien plus brève que dans le cas d'*ATCOPIA93* précède l'établissement de la méthylation. En absence d'indication d'une production abondante de siARN de 21nt dans le mutant *ddm1*, une approche génétique permettrait d'élucider ce point. Dans le cas ou le ciblage par le RdDM serait immédiat, il serait intéressant de déterminer s'il est dû à l'action en trans des siARN de 24nt produits par les copies résidentes ou si les *ATCOPIA78* présentent une caractéristique intrinsèque leur permettant d'être reconnus par le RdDM immédiatement après insertion. Cette hypothèse pourrait être testée par transformation d'une accession, si elle existe, ne possédant pas *ATCOPIA78* dans son génome.

Pour *ATCOPIA93*, l'un des aspects les plus frappants de l'extinction est qu'elle dépend strictement du nombre de copies présentes dans le génome. Cet effet dose-dépendant est fréquent pour les ET puisqu'on le trouve pour l'élément Ac du maïs et l'élément I de la drosophile, pour ne citer que ces deux exemples. Cependant, les mécanismes impliqués sont très différents (Chaboissier et al. 1998, Heinlem et al. 1994, Scofield et al. 1993).

L'analyse d'*ATENSPM3* a permis de mettre en évidence une acquisition (réacquisition) de la méthylation de l'ADN sur l'extrémité de la copie donneuse et des nouvelles insertions dans deux epiRIL. Ces mêmes lignées présentent également un niveau élevé de méthylation dans une région interne d'*AT2TE20205*. A ce stade, il n'est pas possible de conclure si cette méthylation de la région interne est établie à partir d'un état hypométhylé, ou si elle résulte de la mobilisation de copies méthylées. On peut cependant faire le parallèle avec l'exemple d'*ATCOPIA93* où l'établissement de la méthylation dans la région interne GAG précède et est même requis pour l'établissement de la méthylation des extrémités de l'ET. Cependant, ici, il ne semble pas exister de relation entre établissement de la méthylation et nombre de copies. Enfin, il n'a pas été possible d'établir un lien entre établissement de la méthylation et répression transcriptionnelle. Ces études préliminaires mettent en évidence une relation complexe entre méthylation de l'ADN et régulation de l'expression des *ATENSPM3* chez Arabidopsis. Ces observations renvoient notamment aux nombreux travaux réalisés sur les éléments Spm du maïs qui avaient entre autre mis en évidence plusieurs régions critiques au sein de l'élément impliqués dans le contrôle de l'expression et de la transposition, ainsi que la production de quatre protéines distinctes par épissages alternatif (en plus de TNPA et TNPD précédemment citées, il existe TNPB et TNPC dont les fonctions sont encore inconnues) (Fedoroff 2013).

Pour finir, nos résultats montrent une absence de contrôle épigénétique des *VANDAL21* dans toutes les epiRIL analysées à la génération F9, et ce malgré des variations du nombre de nouvelles insertions présentes dans celles-ci. La propagation continue de certaines epiRIL devrait nous permettre d'identifier les conditions nécessaires à la mise en place de ce contrôle.

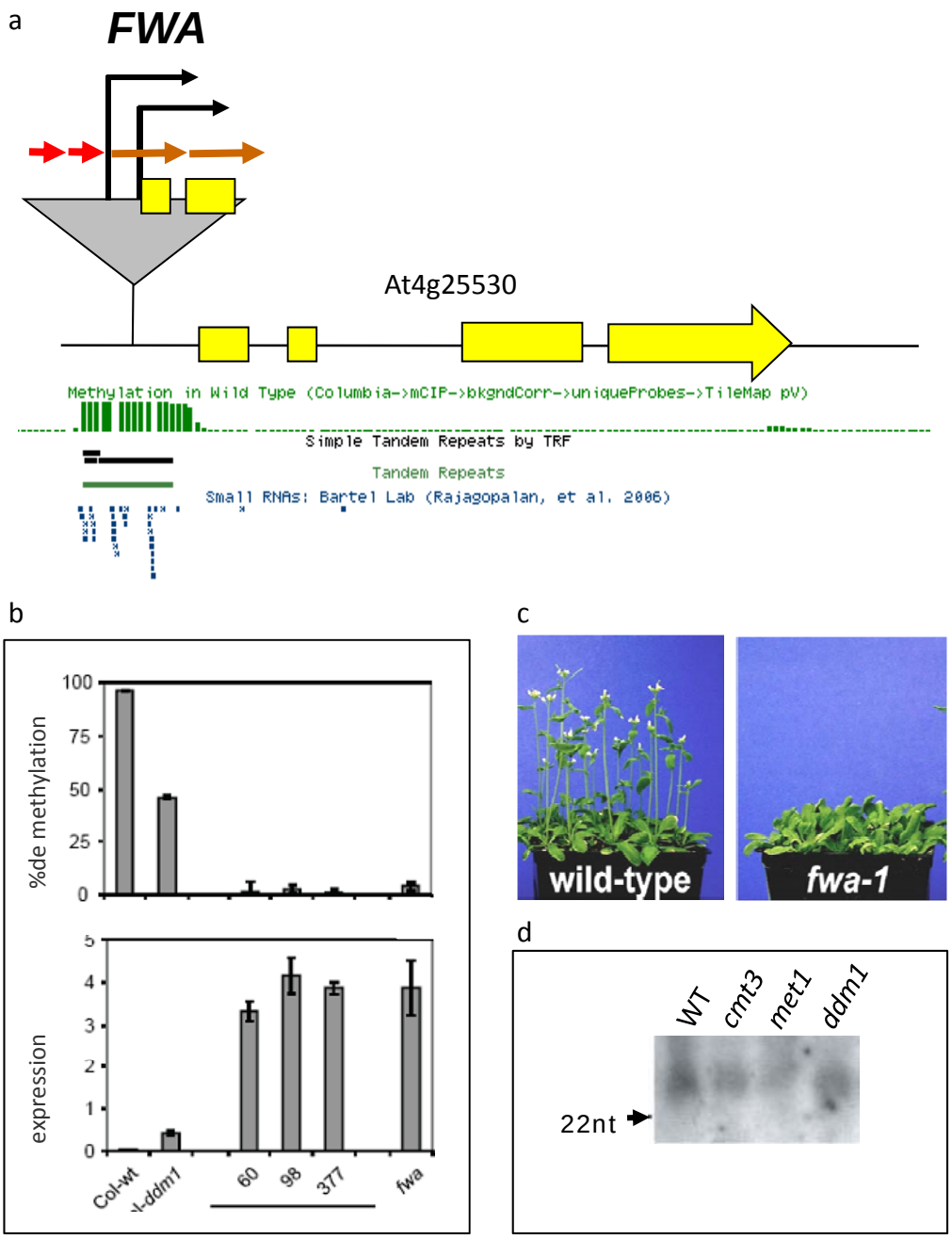


Figure 3.8 : caractéristiques et propriétés du locus *FWA*.
(**a**) représentation schématique du locus, les flèches rouges et ocres représentes les répétitions en tandem (Lippman et al., 2004). (**b**) Méthylation et expression de FWA chez le sauvage, *ddm1,* les epiRIL à floraison très tardive et l'épimutant fwa (Johannes et al, 2009). (**c**) phénotype de l'épimutant *fwa-1* (www.mcdb.ucla.edu/Research/Jacobsen/LabWebSite). (**d**) Northern blot mettant en évidence les siARN ciblant la région promotrice du gène *FWA* chez le sauvage et différents mutants (Lippman et al., 2004).

3.2 Identification de régulateurs du processus d'établissement/rétablissement de la méthylation de l'ADN par crible génétique

3.2.1 Description du crible

L'approche du crible génétique pour identifier des régulateurs du processus d'établissement/rétablissement de la méthylation de l'ADN découle d'observations faites sur la population d'epiRIL. Comme il a été présenté dans l'introduction générale, un sous-ensemble des séquences qui perdent leur méthylation en contexte *ddm1* sont capables de recouvrer celle-ci après ségrégation de la mutation *ddm1* grâce au ciblage par la machinerie du RdDM. Cette réversion, bien que robuste et fidèle, présente une pénétrance variable selon les séquences. C'est ainsi que la séquence répétée en tandem dérivée d'un ET (*SINE*) qui compose le promoteur et les deux premiers exons du gène *FWA* (fig. 3.8a) a échappé à la reméthylation dans trois des 125 epiRIL dans lesquelles elle a été héritée du parent *ddm1* (Johannes et al. 2009) (fig. 3.8b). Ces trois epiRIL présentent une expression ectopique du gène FWA et un important retard de floraison (fig. 3.8b).

Dans ces trois epiRIL, le locus *FWA* présente donc certaines similarités avec un ET nouvellement inséré. Il s'agit en effet d'une séquence qui possède une capacité intrinsèque à être méthylée au niveau des cytosines dans tous les contextes nucléotidiques mais qui est ici hypométhylée et exprimée. Compte-tenu du phénotype d'avancée de floraison que causerait le rétablissement de la méthylation de l'ADN sur la région promotrice de *FWA*, j'ai décidé de développer un crible génétique visant à identifier des mutants capables de forcer la reméthylation de l'épiallèle fortement hypométhylé. De fait, de tels mutants pourraient également favoriser l'établissement de la méthylation sur les ET nouvellement insérés. Ce crible génétique est totalement original, car il repose d'une part sur le suivi d'une séquence endogène et non pas de transgène, source d'effets complexes, et sur sa reméthylation/répression, plutôt que sur sa déméthylation/réactivation. S'il n'est pas exclu d'obtenir des mutations gain de fonction dans des régulateurs positifs de l'établissement de la méthylation, nous nous attendons plutôt ici à identifier des régulateurs négatifs de ce processus. Notre approche se base essentiellement sur l'idée qu'il existe une compétition entre répression (établissement de la méthylation) et expression autrement dit qu'un ou

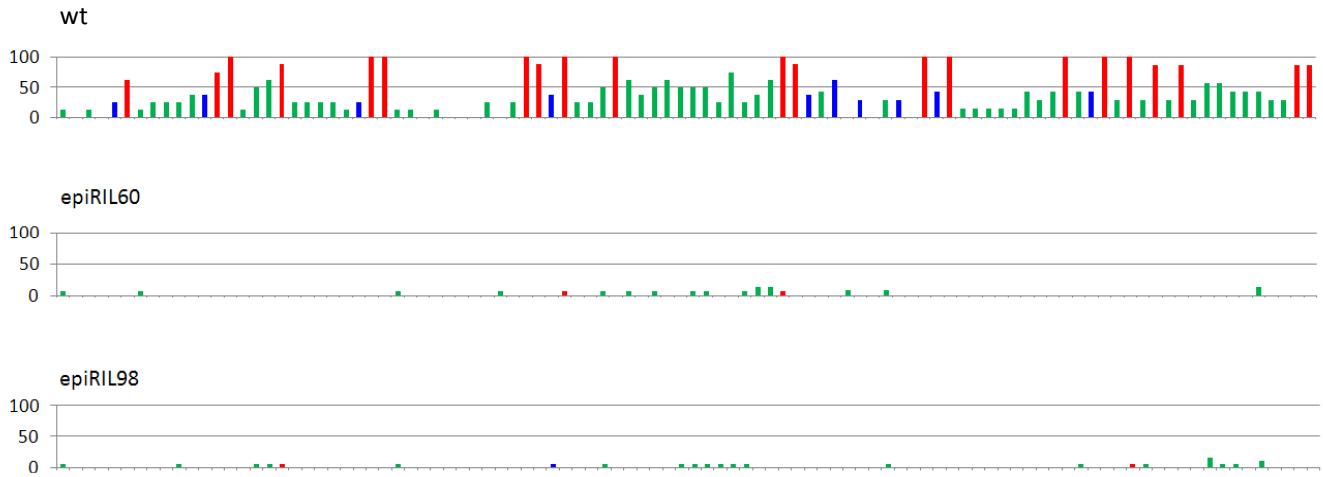


Figure 3.9 : Analyse de la méthylation de la séquence répétée en tandem qui compose le promoteur du gène *FWA* par séquençage bisulfite. Chaque barre représente une cytosine.

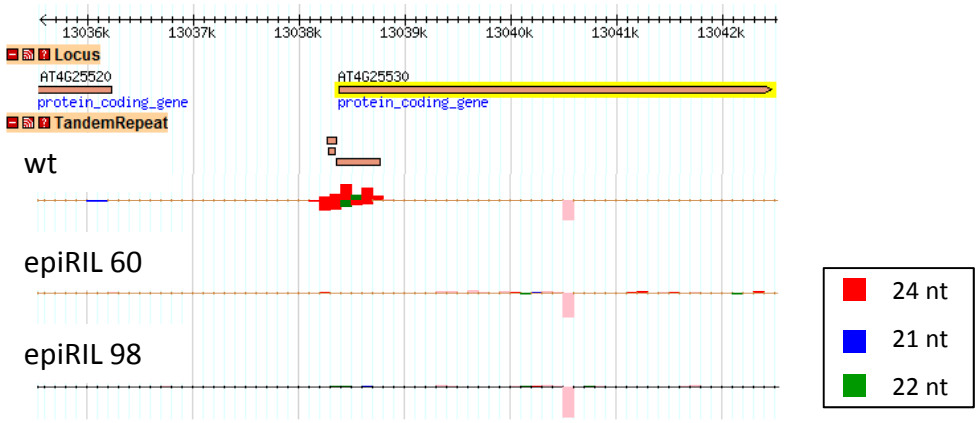


Figure 3.10 : Vue Gbrowse des données de séquençage profond des petit ARN au niveau du locus *FWA* pour le sauvage et les epiRIL 60 et 98.

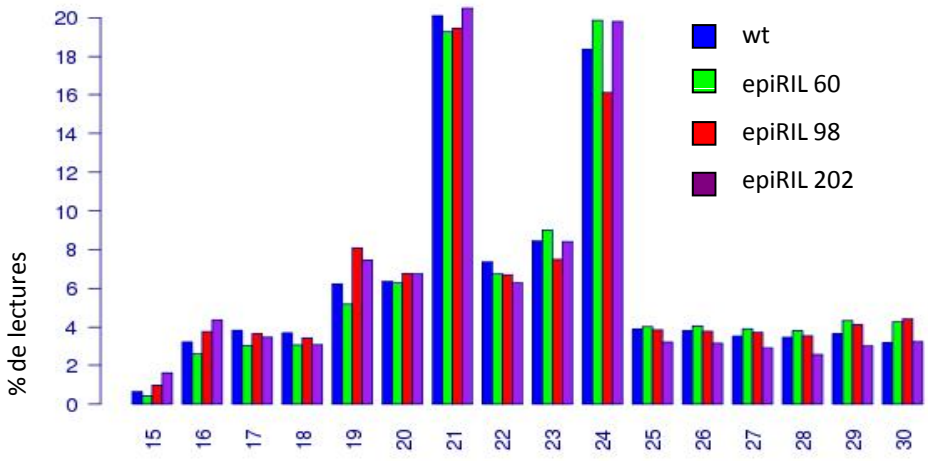


Figure 3.11 : Analyse des populations de siARN chez le sauvage et les épiRIL.

plusieurs mécanismes « empêchent » la reméthylation de *fwa* dans ces epiRIL. On peut notamment penser à une compétition entre les polymérases POLII et POLIV ou encore, entre les machineries responsables de la déposition de marques d'histone ayant des rôles antagonistes.

3.2.2 Caractérisation du locus *FWA* dans les epiRIL et analyses préliminaires

J'ai restreint mon travail à deux des trois lignées epiRIL *fwa* car l'epiRIL 377 présentait une fertilité amoindrie ce qui est peu compatible avec une approche de crible nécessitant la production d'un grand nombre de graines.

Les données préliminaires avaient mis en évidence une perte drastique de méthylation de l'ADN au niveau du locus *FWA* pour les epiRIL 60 et 98 par des expériences de McrBC-qPCR ainsi que la surexpression de *FWA* (fig. 3.8b, Johannes et al. 2009). J'ai dans un premier temps cherché à caractériser plus en détail les épiallèles *fwa* dans ces deux epiRIL 60 et 98. Pour ce faire, j'ai réalisé le séquençage bisulfite de la quasi-totalité de la séquence répétée en tandem. Cette séquence est bien fortement hypométhylée dans les deux epiRIL (fig 3.9).

Les données de séquençage profond de petits ARNs indiquent par ailleurs une perte complète des siARN de 24nt ciblant le locus *FWA* dans les deux lignées en accord avec le fait que ce locus soit stablement hypométhylé dans celles-ci (fig. 3.10).

J'ai ensuite vérifié que la stabilité de l'hypométhylation forte de la région promotrice de *FWA* et l'absence de siARN ciblant ce locus n'était pas due à un RdDM non fonctionnel dans ces lignées. En effet, des ET étant actifs dans ces lignées, il était envisageable, bien que peu probable, que leur mobilisation ait engendré la disruption d'un gène impliqué dans la voie du RdDM. Les données de séquençage profond des petits ARNs montrent que leur accumulation à l'échelle du génome n'est pas significativement différente entre ces lignées et la plante sauvage (p-value>0,05 avec un test kolmogoroff-smirnov) (fig. 3.11). Pour confirmer cette première indication d'un RdDM fonctionnel, j'ai également évalué par McrBC-qPCR le niveau de méthylation de deux séquences dont la méthylation est essentiellement sous contrôle du RdDM (fig 3.12a) ainsi que de sept séquences préalablement définies comme reméthylables (Teixeira et al. 2009) et toutes localisées dans

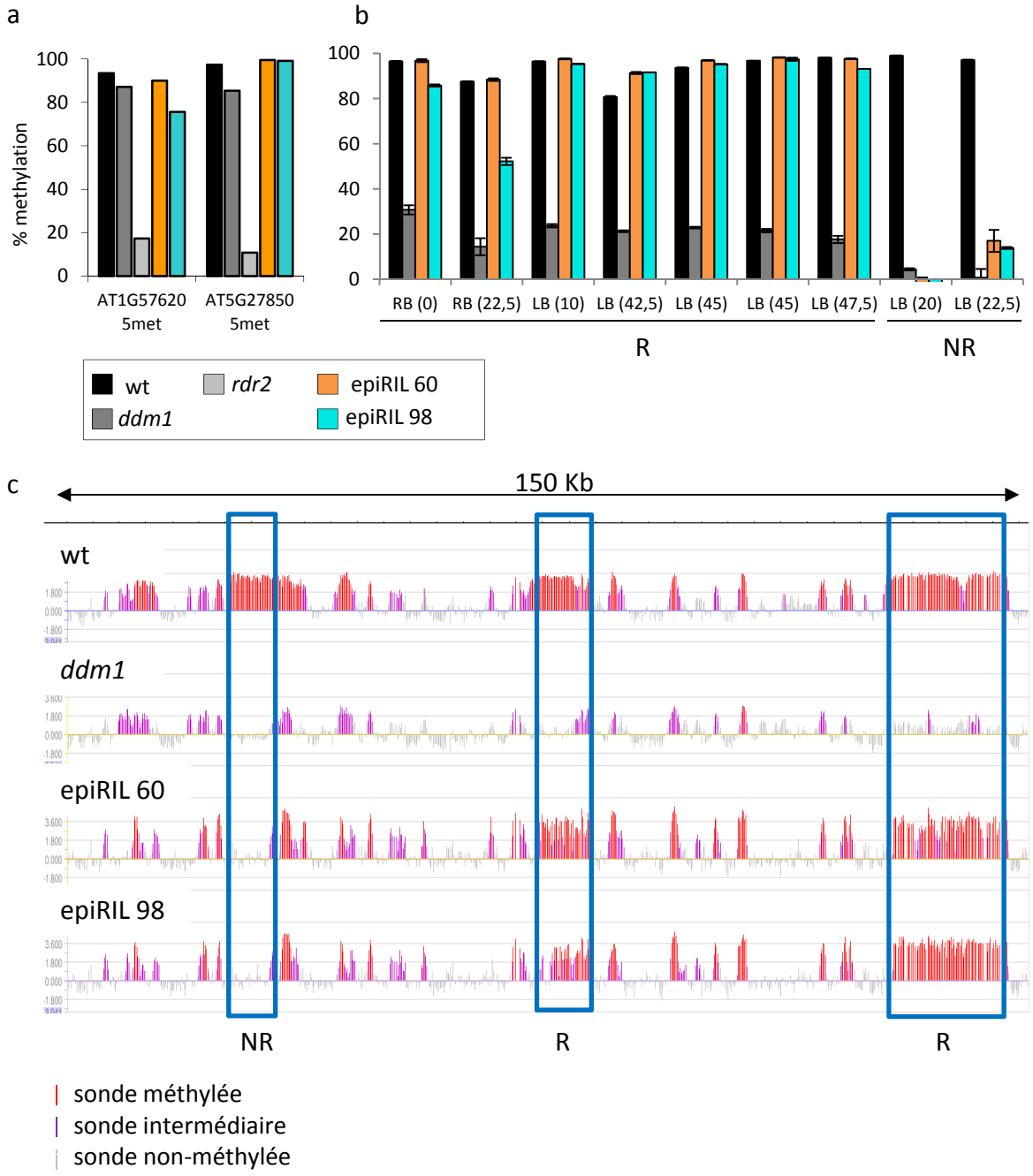


Figure 3.12 : Analyse de la méthylation de l'ADN dans les epiRIL 60 et 98.
(**a**) Mesure du % de méthylation (McrBC-qPCR) de deux séquences contrôlées majoritairement par le RdDM chez le sauvage, les mutant *ddm1* et *rdr2* et les epiRIL 60 et 98. (**b**) Mesure du % de méthylation de sept séquences reméthylables (R) et de deux séquences non-reméthylables (NR) chez le sauvage, *ddm1* et les epiRIL 60 et 98. Les barres d'erreur représentent l'écart type entre deux répliques techniques. (**c**) Exemple de profil de méthylation obtenu par MeDIP-chip pour le sauvage, *ddm1* et les epiRIL 60 et 98. Les cadres mettent en évidence une région non-reméthylable (NR) et deux régions reméthylables (R).

une même région héritée du parent *ddm1* (fig. 3.12b). Si le processus général de reméthylation est altéré dans les lignées 60 et 98, ces neuf séquences devraient être présentes à l'état hypométhylé, ce qui n'est pas observé, prouvant ainsi que la non-réversion observée au locus *FWA* ne s'étend pas à l'intégralité du génome. J'ai confirmé ce dernier résultat en montrant par MeDIP-chip que les lignées 60 et 98 présentent un profil de méthylation compatible avec un RdDM totalement fonctionnel (fig. 3.12c), même si, comme dans toutes les autres epiRIL analysées au laboratoire, un petit nombre de locus présentent un comportement similaire à celui de *FWA* (données non montrées).

3.2.3 Mutagénèse et phénotypage

Les graines issues des deux lignées epiRIL 60 et 98 ont été mutagénisées par traitement au méthanesulfonate d'éthyle (EMS) qui engendre l'alkylation des résidus guanines pour former des O^6-ethylguanines ne s'appariant plus avec les cytosines mais avec les thymines. La mutagénèse EMS induit donc principalement des substitutions C/G vers T/A.

J'ai semé ces graines en serre (INRA de Versailles), avec l'aide de l'ensemble de l'équipe, afin d'obtenir 6000 plantes M1 (3000 pour chaque epiRIL), chacune identifiée par un matricule indiquant sa position dans la serre (fig 3.13). Ce mode de numérotation a été mis au point afin de pouvoir tracer d'éventuelles contaminations. L'efficacité de la mutagénèse est révélée par la présence d'individus présentant des secteurs albinos (fig 3.14). Les plantes M1 n'ont pas été criblées pour l'avancement de la floraison car les mutations dominantes sont rares et doivent de toute façon se retrouver à la génération suivante. D'autre part, on ne s'attend pas à ce que l'établissement de la méthylation puisse se mettre en place sans le passage pas un cycle de reproduction.

Nous avons récolté les graines de chaque plante M1 et semé, dans un premier temps, 2000 lots de graines M2 (1000 issus de chacune des deux mutagénèses). Chaque lot est composé de 20 graines, ce qui permet théoriquement d'obtenir au moins une plante homozygote pour une mutation donnée avec une probabilité de 95%. En effet, les graines contiennent deux cellules progénitrices ce qui implique pour une même plante M1, deux évènements distincts de mutagénèse. Dès lors, une mutation donnée ségrégera dans la moitié seulement

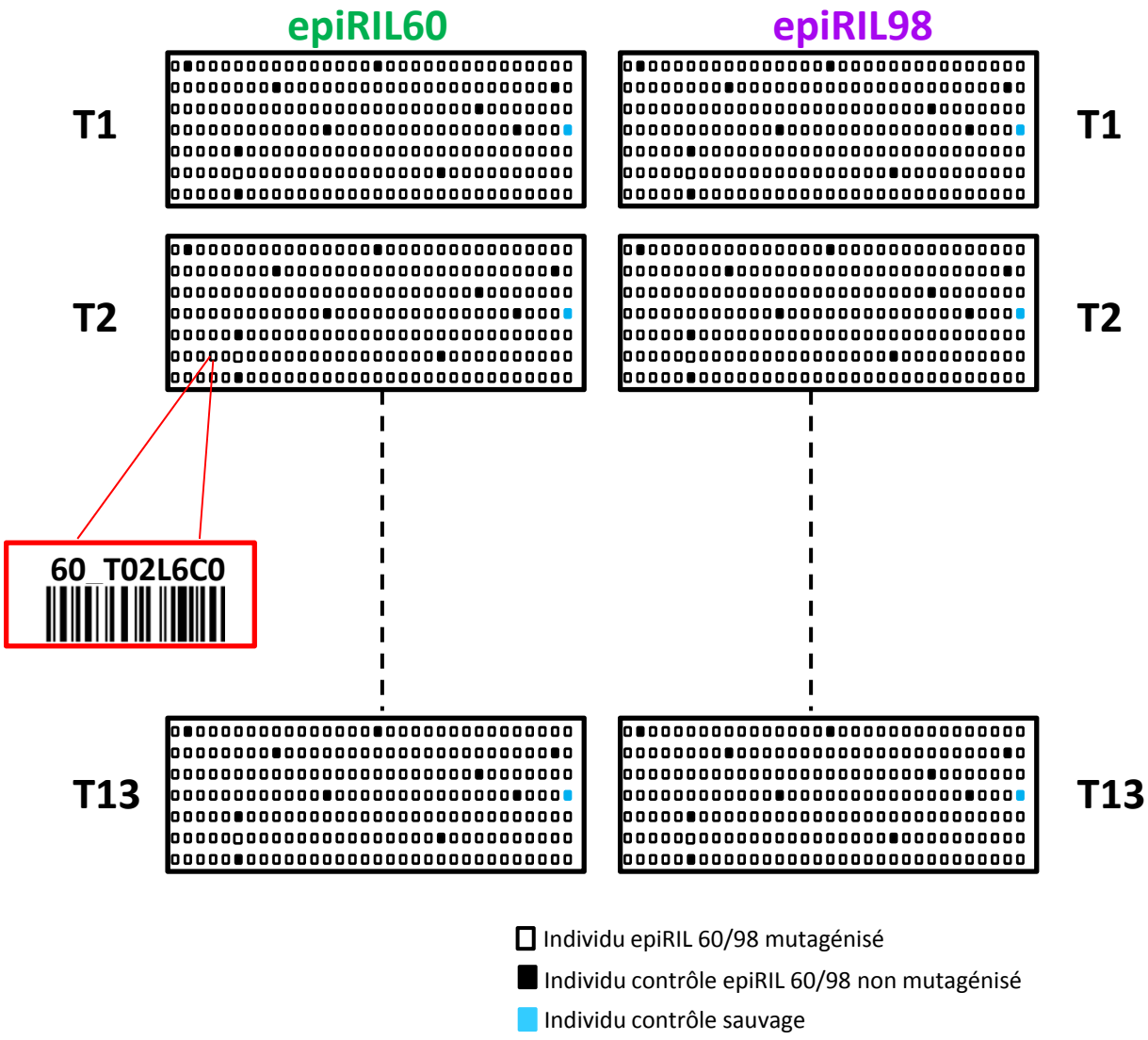


Figure 3.13 : Plan de la serre pour le semi des 6000 plantes M1.

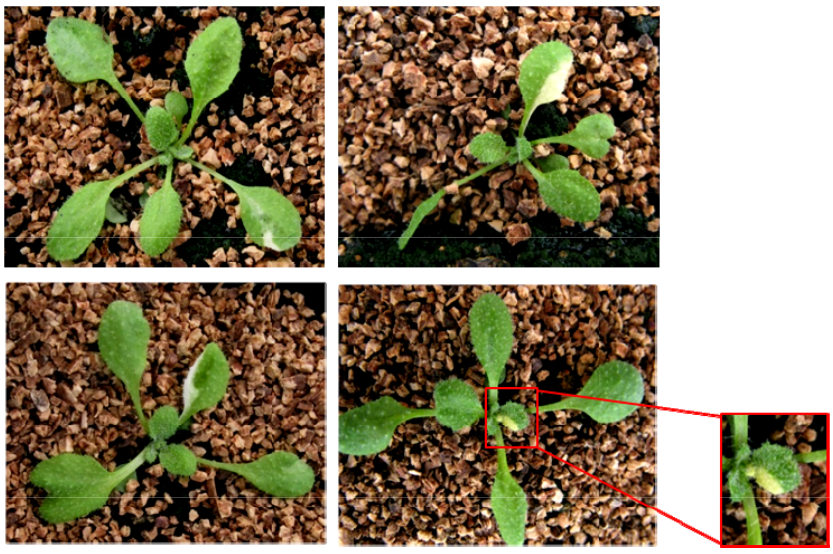


Figure 3.14 : Exemple d'individus présentant des secteurs albinos caractéristiques des plantes ayant subi une mutagénèse par traitement à l'EMS.

des graines issues de cette plante et sera donc présente à l'état homozygote dans seulement une graine sur huit (fig. 3.15).

Afin d'estimer le taux de germination, j'ai compté le nombre de graines germées sur un sous-ensemble de familles (210) réparties sur l'ensemble de la serre (fig. 3.16). Il est de 63% ce qui signifie qu'en moyenne 12,6 graines ont germé par famille. Ce taux est nettement inférieur à celui attendu même dans le cadre d'une mutagénèse EMS. Il est probablement dû aux conditions de chaleur très élevée pendant la culture des plantes M1 qui a pu engendrer un séchage accéléré empêchant la maturation complète des graines M2. La conséquence de cette germination réduite est que la probabilité d'obtenir au moins une plante homozygote pour une mutation donnée au sein d'une famille M2 chute à 80%. Par extension, en moyenne, il sera très peu probable d'obtenir plus de deux plantes homozygotes pour une mutation donnée au sein d'une même famille.

J'ai phénotypé les plantes M2 pour l'avancement de la floraison et ai ainsi sélectionné 52 candidats répartis en 3 catégories : plantes ayant fleuri (0) avant la lignée sauvage de référence, (1) en même temps que celle-ci et (2) après celle-ci mais significativement plus tôt que les epiRIL 60 et 98 (annexe 6). Comme attendu dans le cas de mutations récessives, j'ai identifié au maximum une à deux plantes présentant le phénotype recherché par famille. Il est important de noter qu'aucune des 500 plantes contrôles (c'est-à-dire des epiRIL 60 et 98 non mutagénisées) n'a présenté une réversion spontanée du phénotype de floraison tardive. Parallèlement, 500 autres familles M2 ont été également phénotypées par nos collaborateurs à l'écotron de Foljuif et 23 candidats ont été sélectionnés. Les graines M3 issues des candidats ont été récoltées pour chaque plante individuellement pour la suite des analyses.

Avant d'entamer les analyses moléculaires sur les candidats identifiés en serre, j'ai effectué un phénotypage dans des conditions contrôlées et homogènes (chambre de culture en « rotonde ») des plantes M3 issues de l'autofécondation de chacun des candidats (64 candidats ont ainsi été étudié, les autres étaient stériles ou sont morts avant de donner des graines). Au total 37 des candidats M2 testés présentent un avancement de la floraison dans leur descendance et ont été analysés plus avant.

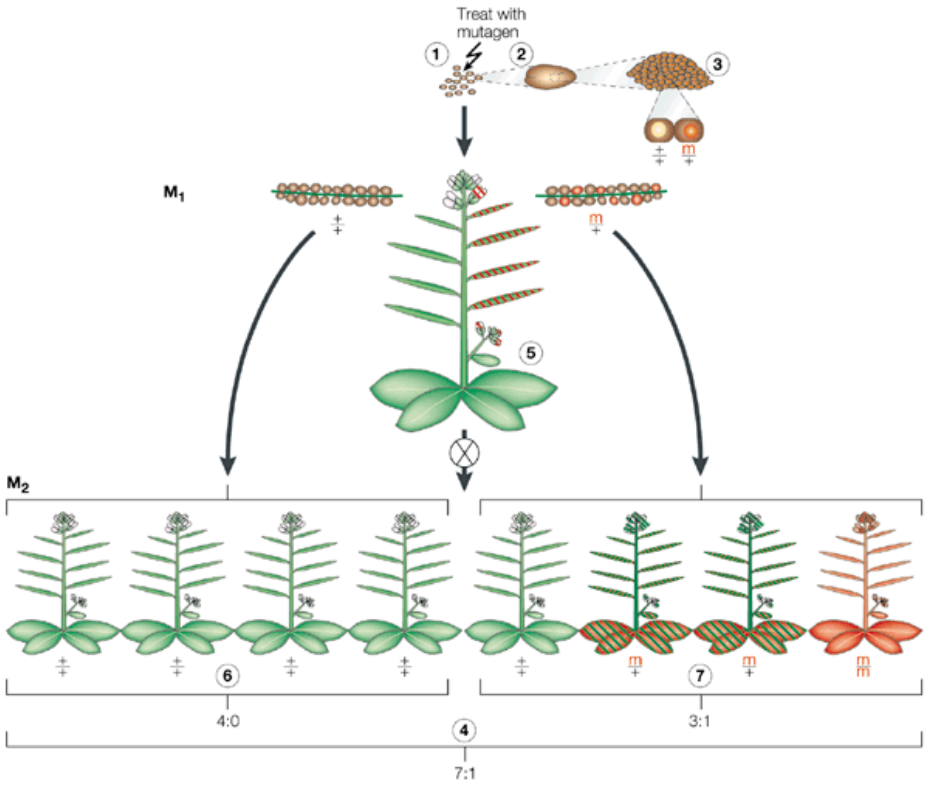


Figure 3.15 : Schéma explicatif illustrant la ségrégation des mutations suite à une mutagénèse par traitement à l'EMS. (Page et al., 2002)

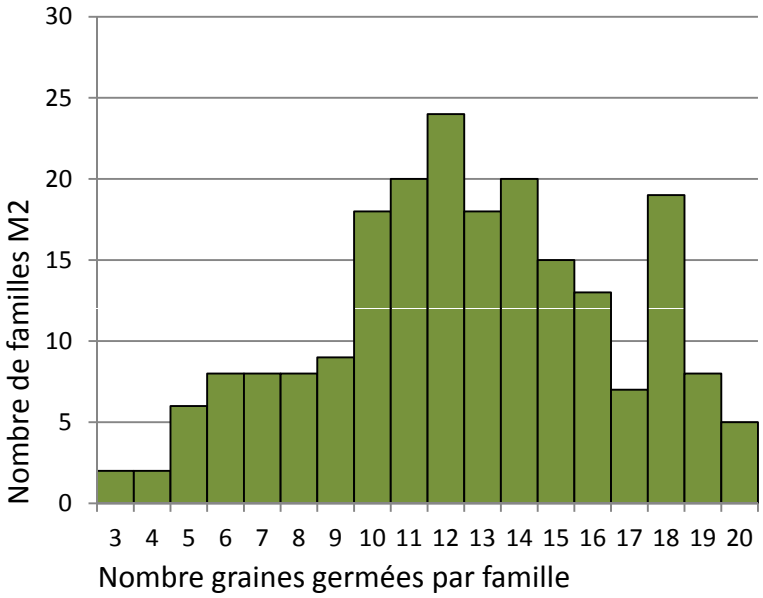


Figure 3.16 : Distribution du taux de germination des plantes mutagénisées.

3.2.4 Analyses complémentaires et résultats

Parmi ces 37 candidats, 10 présentent une expression très réduite de *FWA* (au moins 5x inférieure à celle des contrôles epiRIL 60 et 98) et sont donc potentiellement reméthylés (fig. 3.17). J'ai donc poursuivi leur analyse, en commençant par écarter par génotypage l'hypothèse qu'il puisse s'agir de plantes contaminantes d'origine sauvage. Pour ce faire, j'ai tiré avantage de la présence d'une nouvelle insertion d'*ATCOPIA93* présente à l'état homozygote dans les epiRIL 60 et 98 (fig. 3.18). Quatre des dix candidats se sont avérés être des contaminants sauvages.

Etant donné que l'on ne s'attend pas nécessairement à ce que la méthylation établie au niveau du locus *FWA* soit très élevée (la reméthylation/l'établissement de la méthylation sont des processus progressifs au cours des générations et les plantes analysées ne sont qu'à deux cycles d'autofécondation après mutagénèse) et qu'il n'est pas certain qu'elle soit en tout point semblable à celle de l'épiallèle sauvage, j'ai choisi d'analyser les six candidats restants par séquençage bisulfite de l'intégralité de la répétition en tandem. Les résultats ne montrent aucun gain de méthylation significatif en comparaison avec les epiRIL 60 et 98 non mutagénisées (fig. 3.19). La perte d'expression de *FWA* dans ces six candidats est donc indépendante du niveau de méthylation de la séquence répétée en tandem située dans son promoteur. Plusieurs causes peuvent ici être évoquées expliquer l'extinction de *FWA* dans ces candidats. La premières et la plus triviale est une mutation perte de fonction dans l'un des régulateurs positif d'*FWA* comme un facteur de transcription. La seconde est une mutation en cis dans la région promotrice de *FWA* empêchant la fixation d'un facteur de transcription, et la troisième est une mutation dans la région codante du gène causant un STOP prématuré pouvant être à l'origine de la dégradation du transcrit (RNA decay). Les deux dernières hypothèses sont aisément testables par séquençage du locus *FWA*.

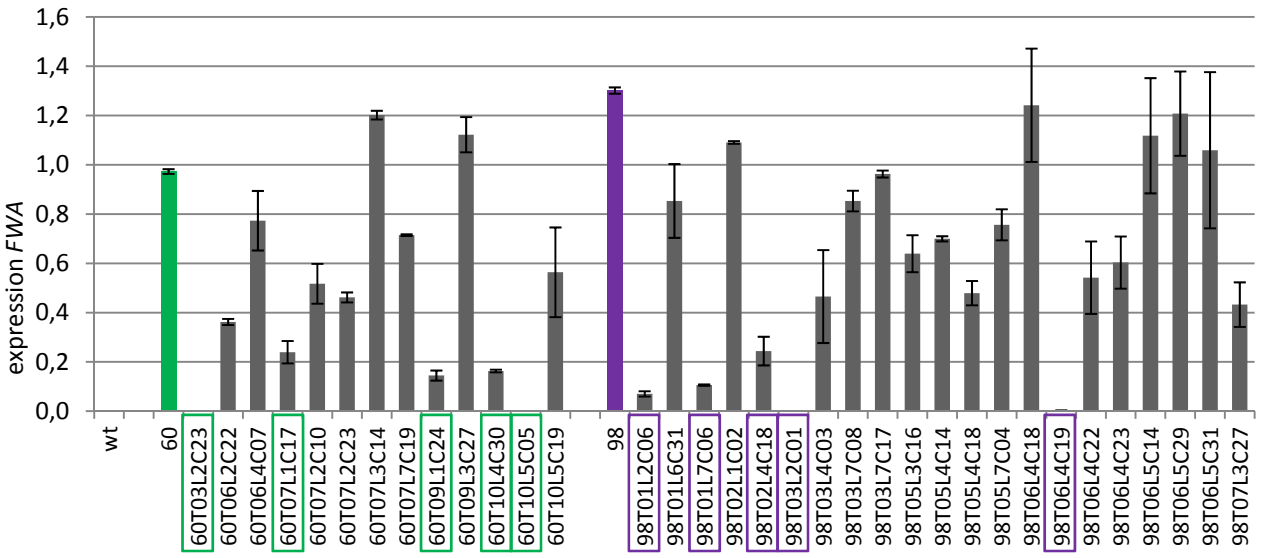


Figure 3.17 : Mesure de l'expression de *FWA* par qPCR exprimée en fonction du gène de l'actine chez les candidats isolés sur la base du critère phénotypique « avancement de la floraison ». Les candidats sélectionnés sont encadrés.

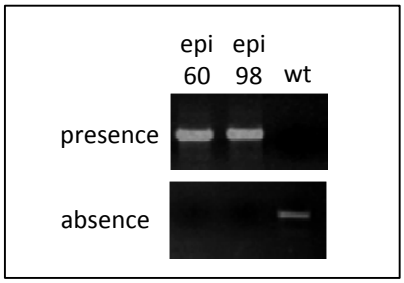


Figure 3.18 : Mise en évidence de la présence/absence par PCR d'une nouvelle insertion d'*ATCOPIA93* utilisée pour génotyper les candidats.

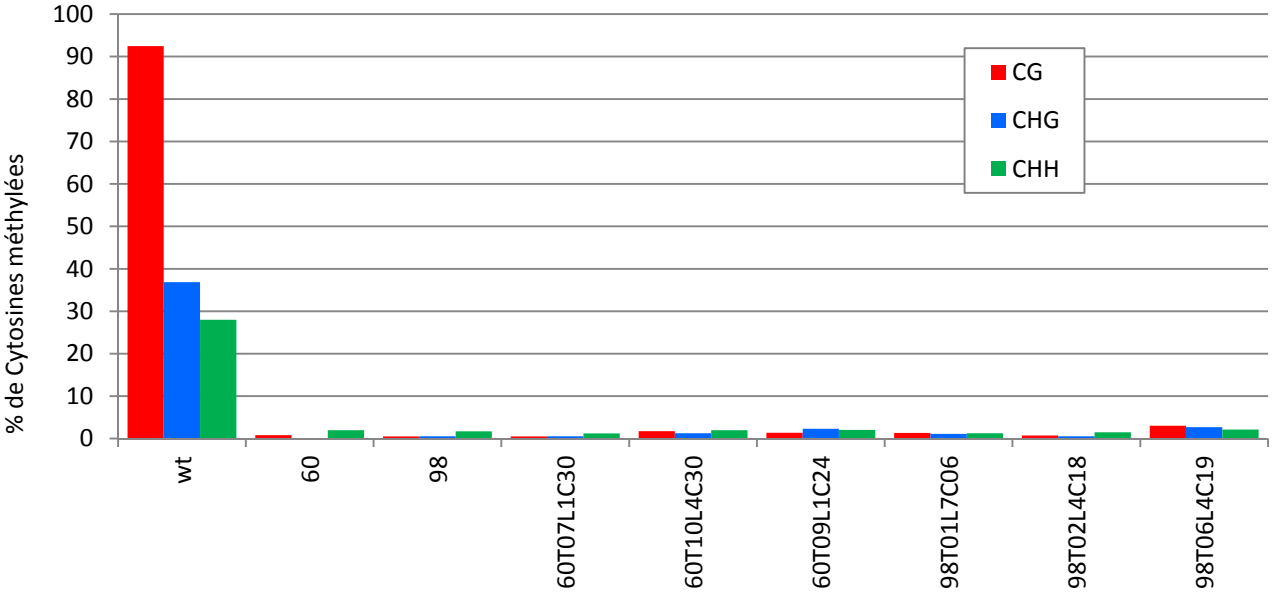


Figure 3.19 : Résultats du séquençage bisulfite de la séquence répétée en tandem de *FWA* chez les 6 candidats.

3.2.5 Discussion et perspectives

L'approche choisie n'a pas permis de mettre en évidence de révertants exhibant une reméthylation du locus *FWA*. L'une des raisons possibles est que le processus de reméthylation tel qu'il a été décrit, est progressif au cours des générations. Or j'ai effectué le criblage phénotypique sur les plantes M2, autrement dit, une seule génération seulement après mutagénèse et la première génération présentant une mutation donnée à l'état homozygote. On peut donc supposer que même si le processus de reméthylation est amorcé dès la M2 chez quelques individus, le niveau de méthylation atteint n'est pas suffisant pour éteindre le gène et ne se traduira donc pas par un avancement de la floraison. Il faudrait alors s'affranchir du criblage phénotypique et analyser directement la méthylation au niveau du locus *FWA* par une approche de crible moléculaire afin de pouvoir mettre en évidence une méthylation même faible du locus. Dans ce but, des prélèvements de tissu foliaire ont été effectués pour chacune des familles M2 et pourront être analysés. Nous avons également envisagé d'effectuer le crible sur une génération plus avancée après la mutagénèse afin que la reméthylation, si elle a lieu, soit plus prononcée. Nous avons ainsi récolté les graines M3 pour toutes les familles M2 et pourrons les propager sur une ou plusieurs générations.

L'une des hypothèses de départ était qu'il existait une compétition entre la machinerie de transcription (dépendante de la polymérase II) et la machinerie d'établissement de la méthylation au niveau du promoteur de *FWA*. Afin de tester cette hypothèse nous pourrons utiliser les candidats qui présentaient une forte chute d'expression de *FWA* afin de déterminer si l'arrêt de la transcription permet le rétablissement (progressif ?) de la méthylation dans les générations suivantes.

4
Influence de la méthylation des éléments transposables sur les gènes situés à proximité

4.1 Conséquences de la perte de la méthylation de l'ADN des ET sur l'expression des gènes situés à proximité

Comme décrit dans l'introduction, il semble qu'une perte importante de la méthylation de l'ADN au niveau des séquences répétées, telle celle induite par une mutation dans le gène *ddm1*, n'ait que peu de conséquences sur la régulation des gènes (Lippman et al. 2004, Vongs et al. 1993). Cette observation est renforcée par le fait que le mutant *ddm1* ne présente pas d'altérations phénotypiques majeures (du moins dans les premières générations d'autofécondation). Cependant, ces informations sont quelque peu contradictoires avec les résultats d'une étude montrant que les ET méthylés semblent avoir un effet négatif sur l'expression des gènes à proximité desquels ils sont localisés (Hollister et al. 2011).

Afin d'analyser plus avant les conséquences de la perte de la méthylation de l'ADN au niveau des séquences répétées, j'ai participé aux travaux initiés par le précédent étudiant en thèse, Felipe Teixeira, présentés dans le manuscrit en préparation ci-après. Ma contribution principale concerne les expériences d'ARN-chip utilisant les puces à ADN CATMA réalisées afin de comparer les transcriptomes des différents mutants. J'ai également généré le triple mutant *ddm1rdr2T-sdc* et ai participé à la validation des gènes candidats potentiellement contrôlés directement par la méthylation des séquences répétées situées à proximité.

A DNA methylation-based double lock system acting over repeat sequences ensures normal development in Arabidopsis

Felipe Karam Teixeira1,2,a,*, Mathilde Etcheverry1,*, Alexis Sarazin1, Ludivine Taconnat2, Anne-Valérie Gendrel 2b, Alexandra To1,2,c, Agnès Bulski1,2,d, Jean-Pierre Renou2,d, Vincent Colot1,2†

1Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Centre National de la Recherche Scientifique (CNRS) UMR8197 - Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, 46 rue d'Ulm, 75230 Paris cedex, France

2Unité de Recherche en Génomique Végétale, Evry Cedex, France

Present address: aNew York University School of Medicine, Skirball Institute, New York, NY 10016, USA; bInstitut Curie, Bâtiment Biologie du Développement, 75248 PARIS Cedex 05; cINRA, Institut Jean-Pierre Bourgin, F-78026 Versailles Cedex, France; dMuséum National d'Histoire Naturelle, CNRS UMR7205, 75005 Paris, France; dINRA Centre d'Angers, Beaucouzé Cedex, France

†Corresponding author: colot@biologie.ens.fr

*These authors contributed equally to this work

Running title: DNA methylation pathways and genome control

Abstract

DNA methylation is an epigenetic mechanism that is essential for silencing repeat elements in mammals and plants. However, the extent to which repeat-associated DNA methylation impacts the expression of neighboring genes is unclear. Here we show in Arabidopsis that severe and fully penetrant phenotypic alterations and strong misregulation of hundreds of genes are only observed when compromising simultaneously the RNA-directed DNA methylation (RdDM) and DDM1-dependent maintenance of DNA methylation pathways, which both target repeat sequences specifically. However, upregulation of repeat sequences is not detectably accentuated or more widespread under these conditions that when only DDM1 function is compromised. Further analysis indicates that the two DNA methylation pathways act together to maintain the normal expression of only a very small number of genes near repeats and that these genes tend to have pleiotropic effects. These findings demonstrate that most Arabidopsis genes are not detectably sensitive to the epigenetic status of neighboring repeats and that a double-lock system ensures that sensitive genes retain proper expression.

Keywords: Epigenetics/chromatin/RNA interference/transposable elements.

Introduction

The genetic dissection of DNA methylation in Arabidopsis has uncovered a complex interplay between DNA methyltransferases (MTases), DNA demethylases, histone-modifying or remodeling enzymes, RNA interference (RNAi) components, and RNA polymerases [1]. Among the actors identified so far, the chromatin remodeler DECREASE IN DNA METHYLATION1 (DDM1; [2]) stands alone as a master regulator of DNA methylation maintenance over repeat sequences [3]. Indeed, mutations in *DDM1* lead to a loss of DNA methylation over most repeat elements present in the genome but not over genes. Furthermore, *ddm1* mutants exhibit a massive accumulation of transcripts corresponding to

repeat sequences, but do not substantially alter the expression of neighboring genes, with few exceptions [2, 4]. Consistent with these observations, early generation *ddm1* mutants exhibit only mild phenotypic alterations. However, severe but sporadic phenotypic alterations are observed in advanced generations and result from the mobilization of transposable elements (TEs) into or near genes as well as rare late onset epigenetic alteration of gene expression [5-10]. In contrast, the RNA-dependent DNA methylation (RdDM) pathway, which targets a majority of methylated repeat elements but not genes [11-14] contributes little to the overall DNA methylation level or the silencing of its targets [12, 14-17].

We have previously shown that RdDM endows its targets with the ability to regain wild type (wt) DNA methylation following *ddm1*-induced loss of DNA methylation [14]. Moreover, we demonstrated that the RdDM pathway is active in *ddm1* plants, being able to promote low-level DNA methylation over its targets. These observations indicate that DDM1-dependent DNA methylation and RdDM act concomitantly over a subset of repeat elements. Here, we show that the simultaneous disruption of these two DNA methylation pathways leads to severe, fully penetrant phenotypic alterations as well as widespread changes in gene expression. However, detailed analysis indicates that most of these defects are caused by the misexpression of a few pleiotropic genes and that most genes are insensitive to the epigenetic status of neighboring repeat elements. Taken together, our findings reveal a limited yet critical direct function of RdDM and DDM1-dependent DNA methylation in preserving normal expression of genes in Arabidopsis.

Results and Discussion

Compromising together DDM1-dependent methylation and RdDM leads to severe and fully penetrant phenotypic alterations

To understand better the genetic interplay between RdDM and DDM1-dependent DNA methylation and their impact on the expression of genes located near repeat elements, we generated mutant plants affected in these two pathways. Double mutant plants were obtained at the expected Mendelian ratio of 1/16 in the F2 progeny of a cross between *ddm1* and a mutant line for the RNA DEPENDENT RNA POLYMERASE 2 gene, which is a key component of RdDM (supplementary Table S1). In contrast to these two parental lines however, first generation *ddm1rdr2* double mutant plants exhibited severe and fully or nearly fully penetrant phenotypic alterations. These alterations were characterized in the following generation (F3) and included wrinkled leaves, thin stems, small stature, late-flowering, and partial sterility (Fig 1). Further selfing led to even more severe defects as well as to the stochastic appearance of additional phenotypes. Similar phenotypic alterations including progressive degeneracy were also observed when the *ddm1* mutation was combined with a mutation in the genes NUCLEAR RNA POLYMERASE D2 (NRPD2) or DICER-LIKE 2 (DCL2) and DCL3, which encode other key components of RdDM.(supplementary Fig S1). Thus, compromising simultaneously DDM1-dependent DNA methylation and RdDM causes a robust suite of phenotypic defects. We therefore conclude that these two pathways act redundantly to enable normal plant development.

Most genes near methylated repeats are insensitive to the simultaneous loss of DDM1-dependent methylation and RdDM

We next performed comparative transcriptome analysis of wild type, *ddm1* and *ddm1rdr2* plants, using 10-days-old seedlings and a gene-based array (CATMA) containing ~24000 gene-specific tags from Arabidopsis [18]. In agreement with previous analyses, few genes were misregulated in *rdr2* when compared to wt (Fig 2A,B; 62 upregulated and 22 downregulated; [12, 17]). Up to three times more genes were misregulated in *ddm1* and

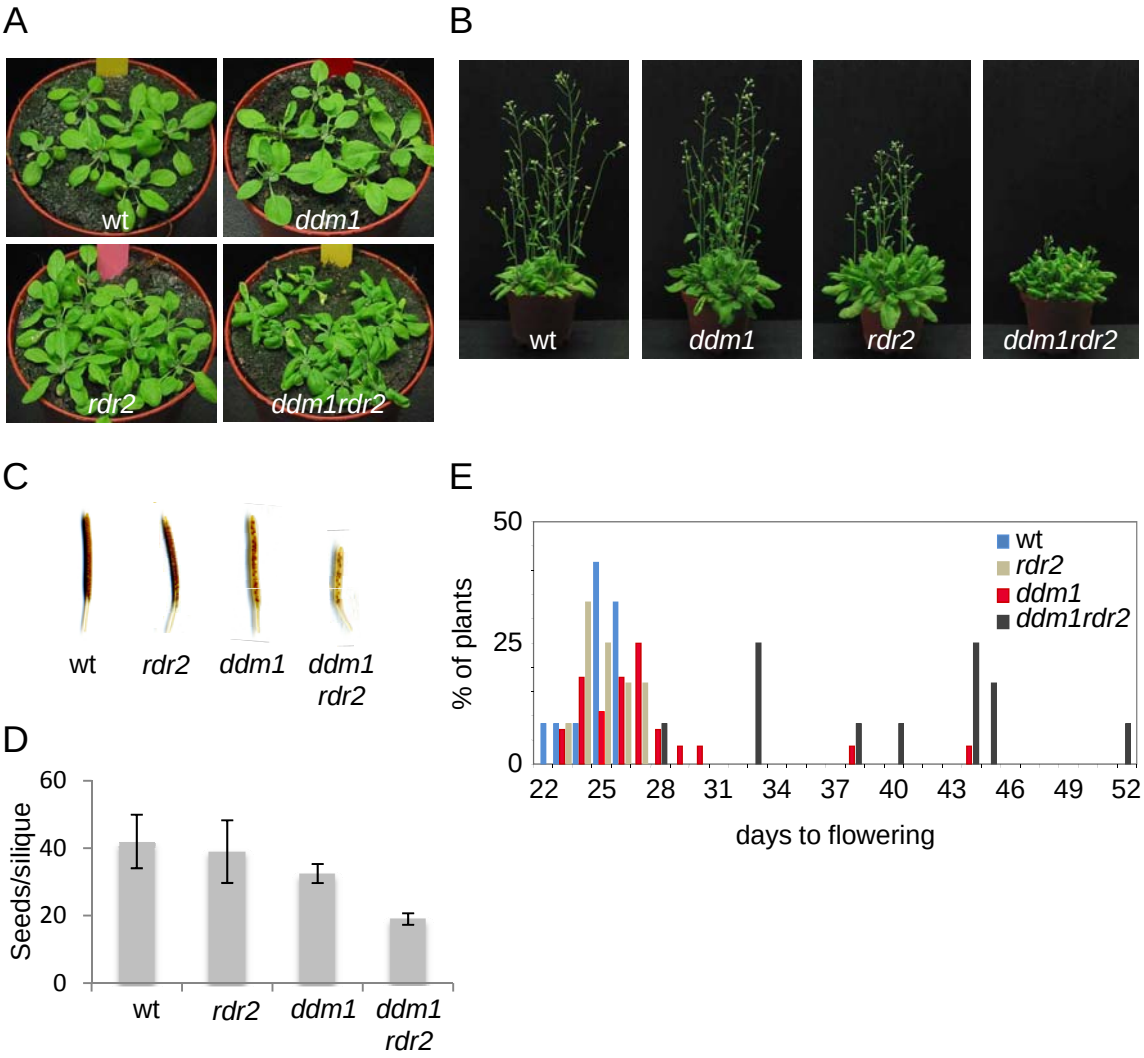


Figure 1. Combined loss of DDM1 and RNAi leads to fully penetrant phenotypic alterations. (A) wt, *ddm1*, *rdr2*, and *ddm1rdr2* plants of 23-days-old and (B) 40-days-old. (C) Siliques for each genotype, discolored using ethanol 70%. (D) Average number of seeds per silique based on 7-10 siliques. (E) Analysis of flowering time as measured by the number of days to flowering for 24 plants of each genotype.

ddm1rdr2 when these mutants were individually compared to wt, with *ddm1rdr2* exhibiting slightly more misregulated genes than *ddm1* (Fig 2A,B). Direct comparisons between single and double mutants were also carried out (supplementary Fig S2). Altogether, these analyses identified a set of ~360 genes that are misexpressed in plants compromised in one or both DNA methylation pathways. We therefore conclude that that these two pathways act redundantly to enable normal plant development.

Misregulated genes were then individually inspected for their proximity (within 1kb) to targets of DDM1-dependent methylation and RdDM, using a genome browser display of publically available single-nucleotide resolution methylome datasets [13, 19]. This revealed that less than one third (~100) of misregulated genes are close to such targets and thus likely to be directly affected by loss of DNA methylation over these sequences. As expected, this group included the three genes for which mis-expression in response to compromised DDM1-dependent methylation or RdDM has been extensively characterized, namely FLOWERING WAGENINGEN A (FWA, At4g25530; [5, 9]), SUPPRESSOR OF drm1drm2cmt3 (SDC, At2g17690; [20]), and At5g27850 [16]. Validation of the microarray results and direct examination of DNA methylation in matched samples was carried out using these three genes as well as nine randomly selected genes (supplementary Table S2, S3). With one exception (At5g22500; excluded from further analysis), RT-qPCR analysis validated the transcriptome data (supplementary Table S3). Changes in DNA methylation, as measured by McrBC-qPCR, were tightly correlation with changes in gene expression for all three previously characterized genes as well as six of the eight remaining genes that were examined (Fig 2C; supplementary Table S4). For four genes (SDC, At3g26520, FWA, and At4g30960), strongest misregulation was observed in *ddm1rdr2* and was always associated with accentuated loss of DNA methylation when compared to both ddm1 and *rdr2*. This indicates that at these genes, DDM1-dependent DNA methylation and RdDM act synergistically or additively rather than strictly redundantly in preserving normal expression. For the other direct target genes tested, expression and DNA methylation were already severely affected in either *ddm1* (At1g36240, At3g28950, and At5g26270) or *rdr2* (At1g57620 and At5g27850), with no further changes in the double mutant.

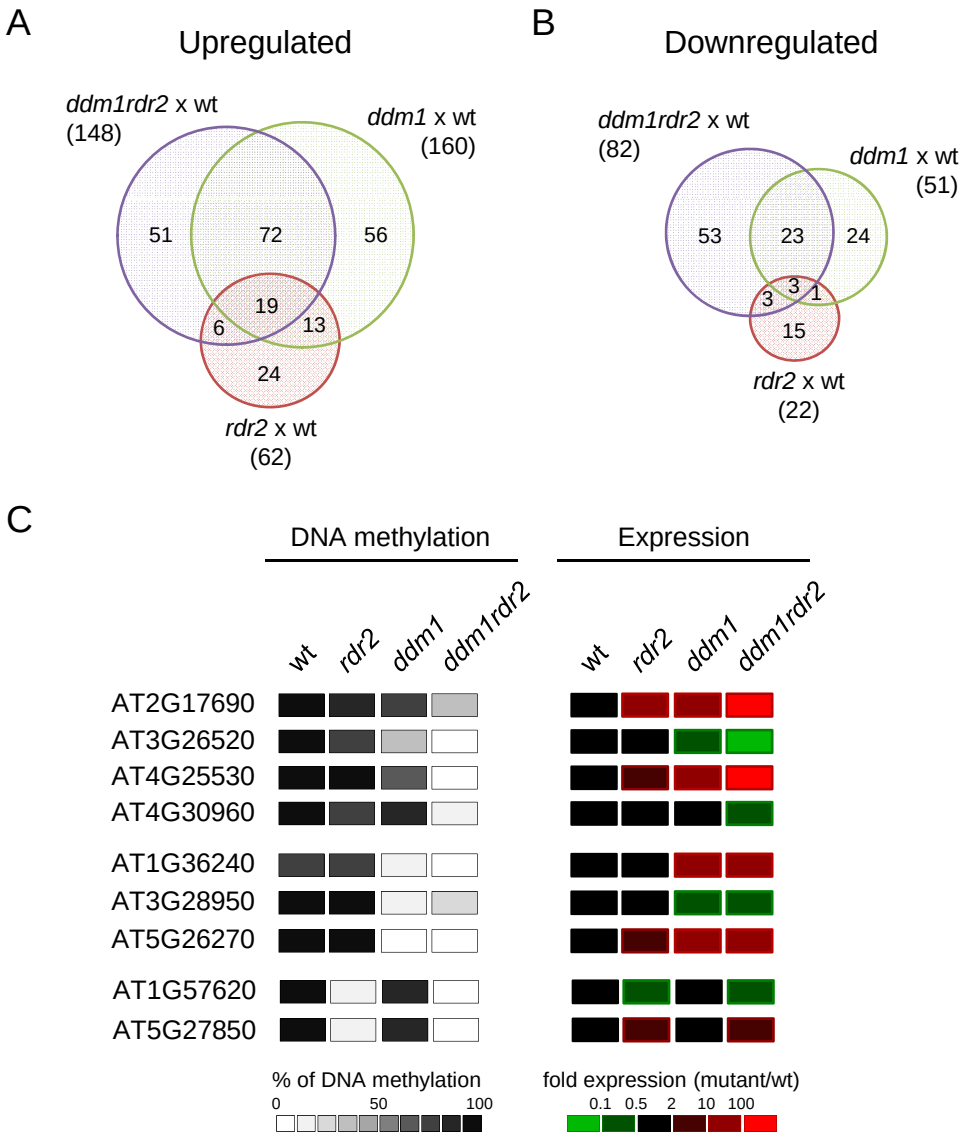


Figure 2. Genome-wide analysis allows the identification of a set of genes that are directly regulated by nearby DNA methylation. (A-B) Venn diagram representing the overlap of genes upregulated (A) and downregulated (B) in comparisons between *ddm1* and wt (green), *rdr2* and wt (red), and *ddm1rdr2* and wt (purple) as accessed by transcriptome analysis using gene-based CATMA arrays. All results are summarized in Supplemental Table S2. (C) RT-qPCR and McrBC-qPCR analyses conducted with a set of candidate genes. DNA methylation results are expressed as percentage of loss of molecules after McrBC digestion. Expression results were normalized to the average of three control genes, and are represented as fold-induction in comparison to wt levels (=1). Results are summarized in Supplemental Tables S4-S5.

Fully penetrant phenotypic alterations observed in *ddm1rdr2* are mainly caused by misregulation of the two genes FWA and SDC

It has been previously shown that ectopic overexpression of FWA and SDC is associated with late-flowering and wrinkled leaves phenotypes, respectively [5, 9, 20]. Given that loss of DNA methylation and overexpression of these two genes is most pronounced in *ddm1rdr2*, this prompted us to test the implication of FWA and SDC in the severe and fully penetrant phenotypic alterations specifically observed in *ddm1rdr2*. To this end, we crossed *ddm1rdr2* with *rdr2* and examined the DDM1/DDM1, *rdr2/rdr2* (F2) progeny (Fig 3A). These plants were recovered at the expected frequency and exhibited the late-flowering (L) and wrinkled leaf (W) phenotypes singly or in combination at frequencies compatible with Mendelian inheritance of independently segregating dominant mutations (3:1). We then characterized the DNA methylation status of the repeats associated with FWA and SDC in three (F2) DDM1/DDM1, *rdr2/rdr2* individuals with different phenotypes (Fig 3B). The wrinkled leaf, early-flowering 10H6(F2) plant was homozygous for the *sdc-dr* allele as well as for the fully methylated FWA epiallele. The 10H8(F2) plant was homozygous for the methylated SDC allele and for the *fwa-dr* allele, in agreement with its late-flowering and normal leaf phenotype. The late-flowering and wrinkled leaf 10F9(F2) plant was shown to be heterozygous at both loci, presenting average DNA methylation levels between those measured in *rdr2* and *ddm1rdr2*. These three (F2) individuals were subsequently selfed and 24 (F3) plants of each line were phenotyped (supplementary Fig S4). As expected, all 10H6 F3 progeny had wrinkled leaves and was early-flowering, while all 10H8 F3 progeny had was normal leaves and was late-flowering . As expected from the double heterozygosity of the 10F9 F2 parent, the late-flowering and wrinkled leaf phenotypes segregate independently in its F3 progeny and at the expected frequencies for two unlinked dominant mutations.

Lastly, we took advantage of a T-DNA insertion disrupting the SDC gene (*T-sdc*) to generate a triple *ddm1rdr2T-sdc* mutant (Fig 3C). Replacement of the wt SDC copy with the *T-sdc* allele sufficient to rescue the wrinkled leaf phenotype of *ddm1rdr2* plants. However, triple mutant plants still exhibited a late-flowering phenotype, consistent with the fact that this alteration is induced by ectopic expression of FWA. Altogether, our results indicate that *ddm1rdr2*-induced alterations of DNA methylation and gene expression in the imprinted FWA and SDC

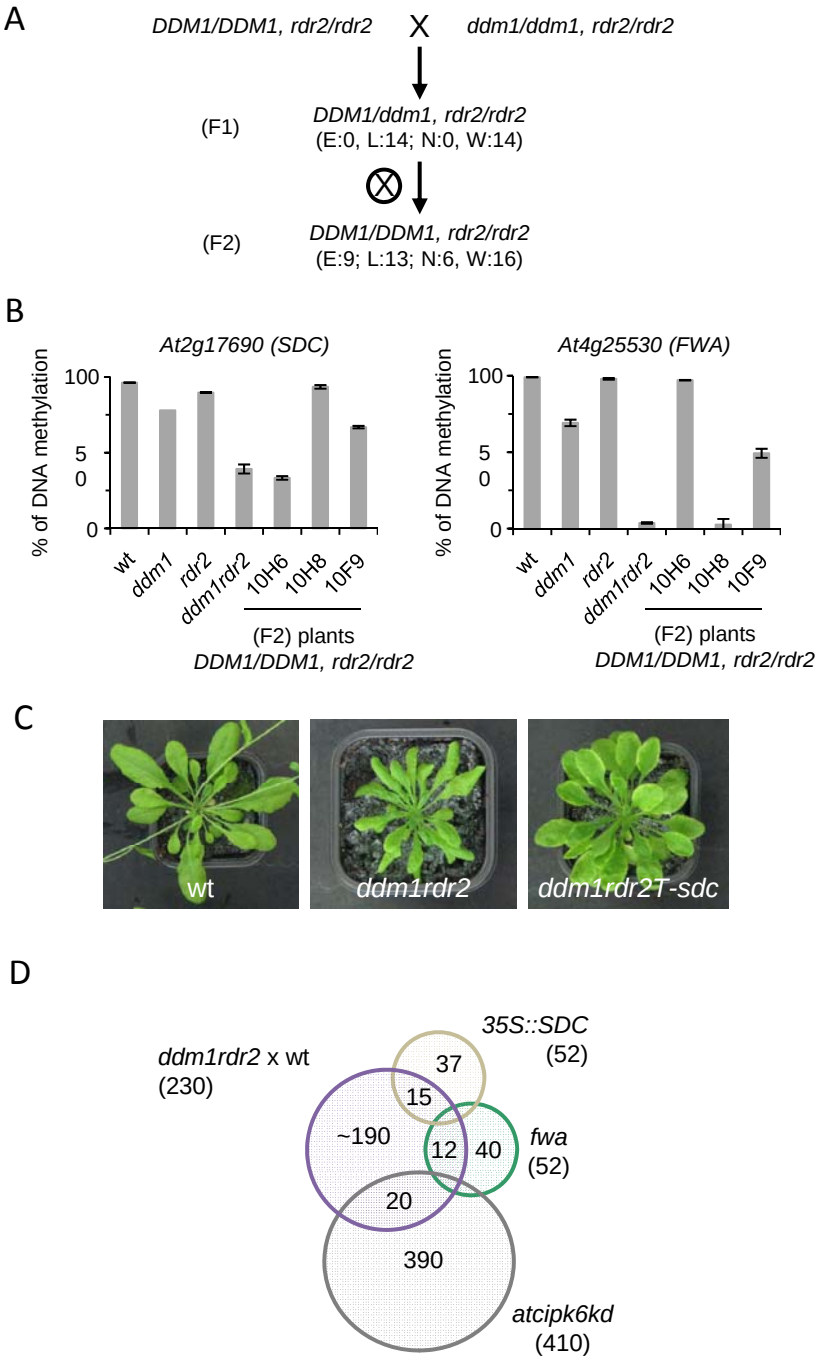


Figure 3. Phenotypic alterations and gene misregulation observed in *ddm1rdr2* directly results from misregulation of few primary targets. (A) Crossing scheme used to test the stability and segregation of late-flowering and wrinkled leaf phenotypes in a RNAi mutant background. (F1) individuals derived from *ddm1rdr2* x *rdr2* crosses were selfed and (F2) plants were genotyped. Phenotypic segregation for flowering time and leaf structure was characterized in *DDM1/DDM1*, *rdr2/rdr2* (F2) plants (E: early-flowering; L: late-flowering; N: normal leaf; W: wrinkled leaf). (B) DNA methylation over the promoter regions of *FWA* and *SDC* genes was accessed by McrBC-qPCR in wt, *ddm1*, *rdr2*, *ddm1rdr2*, and three (F2) *DDM1/DDM1*, *rdr2/rdr2* individuals: 10H6 (wrinkled leaf and early-flowering), 10H8 (normal leaf and late-flowering), and 10F9 (wrinkled leaf and late-flowering). (C) Wrinkled leaf phenotype observed in *ddm1rdr2* is rescued to wt in *ddm1rdr2T-sdc* triple mutant plants. (D) Venn diagram representing the overlap of genes misregulated in *ddm1rdr2* x *ddm1*, *35S::SDC*, *fwa*, and *atcpik6kd* (re-analyzed published data).

genes underlie much of the fully penetrant phenotypic alterations that are only observed when DDM1-dependent DNA methylation and RdDM are simultaneously compromised.

Few genes are misregulated directly as a consequence of loss of repeat-associated DNA methylation

We next tested if overexpression of FWA and SDC could also explain most of the changes in gene expression seen in *ddm1rdr2* mutant plants. To this end, we first compared the transcriptome of wt and *fwa* epimutant seedlings and identified 52 differentially expressed genes (Fig 3D). The impact of the ectopic expression of SDC was assessed using previously published data obtained with35S::SDC transgenic plants, which mimic the *sdc* epimutant [20]. Again, 52 genes showed altered expression as a result of overexpressing SDC. Additionally, we took advantage of publically available transcriptome data for another gene likely affected directly by loss of DNA methylation in *ddm1rdr2* mutant plants. This genes, called CBL-INTERACTING PROTEIN KINASE 6 (CIPK6, At4g30960) is downregulated and lose DNA methylation over its promoter region specifically in *ddm1rdr2* (Fig. 2). Disruption of the CIPK6 gene by a T-*DNA* insertion into the 3'UTR (atcipk6kd) leads to the misregulation of 410 genes [21].

When combined together, these three gene sets overlap with ~19% of the genes that are misexpressed in *ddm1rdr2* (Fig 3D).

In conclusion, our results indicate that despite most genes being insensitive to the epigenetic status of neighboring repeat sequences in Arabidopsis, DDM1-dependent DNA methylation and RdDM are critical for maintaining the normal expression of a few key genes with pleiotropic effects. However, this picture may be radically different in plant species that have repeat-rich genomes, such as maize, where phenotypic alterations and widespread gene misregulation are already observed in plants that are only defective in RdDM [22].

Methods

Plant material and growth conditions

Plants were from the A. thaliana Col-0 accession. *ddm1-2, met1-1, dcl2-1, dcl3-1, rdr2-1, nrpd1a-1,* and *nrpd1b-1* mutants have been previously described [2, 23-25]. *fwa* and 35S::SDC lines were kindly provided by Dr. Tetsuji Kakutani (NIG, Shizuoka, Japan) and Dr. Steve Jacobsen (UCLA, CA, USA), respectively [9, 20]. The SDC insertion (*T-sdc*) is Salk T-DNA_017593. Plants were grown under long-day conditions, either in vitro in liquid for 10- and 12-day-old seedlings [4], or in soil otherwise.

Transcriptome analysis

Total RNA was extracted using RNeasy Qiagen kit. Microarray analyses were conducted using the CATMA arrays [18]. Two independent biological replicates were produced for each comparison, and one technical replication with dye-swap was performed for each biological replicate. Labeling, hybridization to slides, and scanning were performed as described [27]. Normalization and statistical analysis were performed as described [28]. Published transcriptome for atcipk6kd was conducted on Affymetrix GeneChip Arabidopsis ATH1 genome array [21]. For our analysis, we only considered the differentially expressed genes that are present in the CATMA array.

Analysis of DNA methylation and transcription

DNA and RNA were extracted from seedlings using DNeasy and RNeasy Qiagen kits, respectively. Digestion with the McrBC enzyme (New England Biolabs) was performed on 500 ng of genomic DNA, as previously described [14]. Quantitative PCR (qPCR) was performed using a Roche LightCycler® 480 system machine and LightCycler® 480 SYBR Green I Master. Primers are listed in supplementary Table S5. Results were expressed as percentage of loss of molecules after McrBC digestion.

Reverse transcription (RT) was performed on 1 ug of total RNA using oligodT and Superscript II (Invitrogen). qPCR was performed as above and results were expressed as percentage of expression relative to the mean value obtained for three control genes (At2g36060; At4g29130; At5g13440; [14]). Primers are listed in supplementary Table S6.

References

1. Law, J.A. and S.E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet, 2010. 11(3): p. 204-20.
2. Vongs, A., et al., Arabidopsis-Thaliana DNA Methylation Mutants. Science, 1993. 260(5116): p. 1926-1928.
3. Teixeira, F.K. and V. Colot, Repeat elements and the Arabidopsis DNA methylation landscape. Heredity (Edinb), 2010. 105(1): p. 14-23.
4. Lippman, Z., et al., Role of transposable elements in heterochromatin and epigenetic control. Nature, 2004. 430(6998): p. 471-6.
5. Kakutani, T., Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana. Plant Journal, 1997. 12(6): p. 1447-1451.
6. Kakutani, T., et al., Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proceedings of the National Academy of Sciences of the United States of America, 1996. 93(22): p. 12406-12411.
7. Miura, A., et al., Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature, 2001. 411(6834): p. 212-4.
8. Saze, H. and T. Kakutani, Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. Embo Journal, 2007. 26(15): p. 3641-3652.
9. Soppe, W.J.J., et al., The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Molecular Cell, 2000. 6(4): p. 791-802.
10. Tsukahara, S., et al., Bursts of retrotransposition reproduced in Arabidopsis. Nature, 2009. 461(7262): p. 423-U125.
11. Ahmed, I., et al., Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res, 2011. 39(16): p. 6919-31.
12. Kasschau, K.D., et al., Genome-wide profiling and analysis of Arabidopsis siRNAs. Plos Biology, 2007. 5(3): p. 479-493.
13. Lister, R., et al., Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell, 2008. 133(3): p. 523-36.
14. Teixeira, F.K., et al., A Role for RNAi in the Selective Correction of DNA Methylation Defects. Science, 2009. 323(5921): p. 1600-1604.
15. Cao, X.F. and S.E. Jacobsen, Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Current Biology, 2002. 12(13): p. 1138-1144.
16. Huettel, B., et al., Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. Embo Journal, 2006. 25(12): p. 2828-2836.
17. Kurihara, Y., et al., Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. Biochem Biophys Res Commun, 2008. 376(3): p. 553-7.
18. Sclep, G., et al., CATMA, a comprehensive genome-scale resource for silencing and transcript profiling of Arabidopsis genes. BMC Bioinformatics, 2007. 8: p. 400.
19. Cokus, S.J., et al., Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature, 2008. 452(7184): p. 215-219.
20. Henderson, I.R. and S.E. Jacobsen, Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes & Development, 2008. 22(12): p. 1597-1606.
21. Tripathi, V., et al., CIPK6, a CBL-interacting protein kinase is required for development and salt tolerance in plants. Plant J, 2009. 58(5): p. 778-90.

22. Jia, Y., et al., Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genet, 2009. 5(11): p. e1000737.
23. Kankel, M.W., et al., Arabidopsis MET1 Cytosine Methyltransferase Mutants. Genetics, 2003. 163(3): p. 1109-22.
24. Pontier, D., et al., Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes & Development, 2005. 19(17): p. 2030-2040.
25. Xie, Z.X., et al., Genetic and functional diversification of small RNA pathways in plants. Plos Biology, 2004. 2(5): p. 642-652.
26. Turck, F., et al., Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. Plos Genetics, 2007. 3(6): p. 855-866.
27. Lurin, C., et al., Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell, 2004. 16(8): p. 2089-103.
28. Gagnot, S., et al., CATdb: a public access to Arabidopsis transcriptome data from the URGV-CATMA platform. Nucleic Acids Res, 2008. 36(Database issue): p. D986-90.

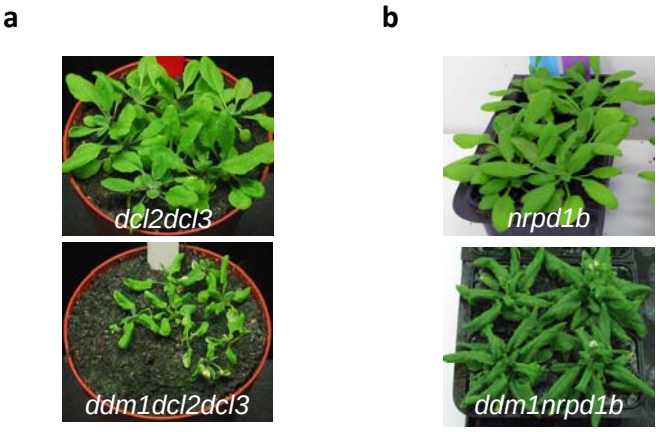
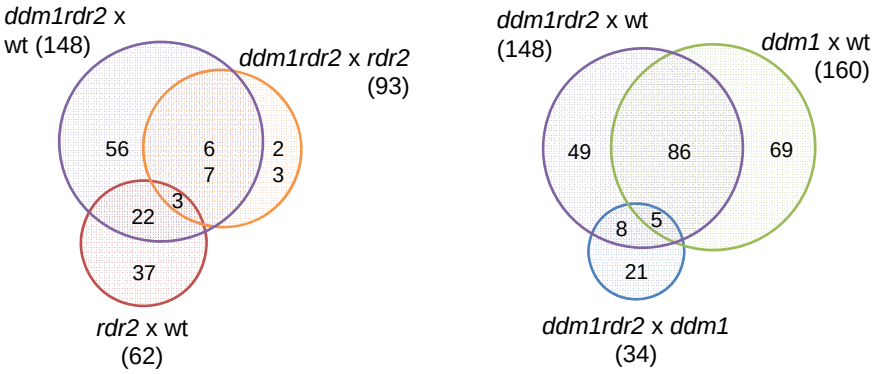


Figure S1. **Phenotypic alteration in *ddm1dcl2dcl3* and *ddm1nrpd1b* mutants. (a)** 23-days-old *dcl2dcl3* and *ddm1dcl2dcl3* plants. **(b)** 25-days-old *nrpd1b* and 40-days-old *ddm1nrpd1b* plants.

a



b

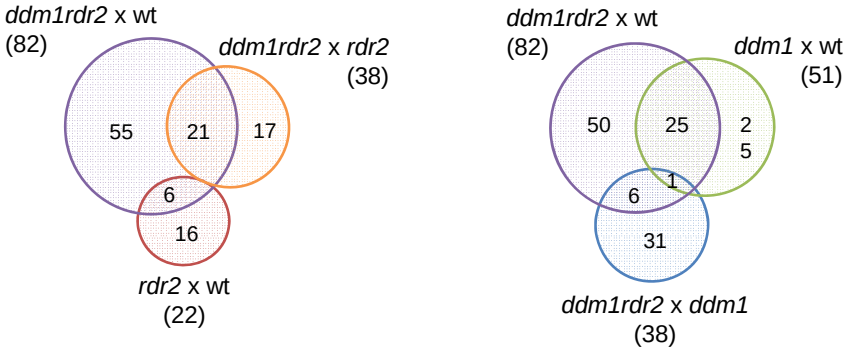
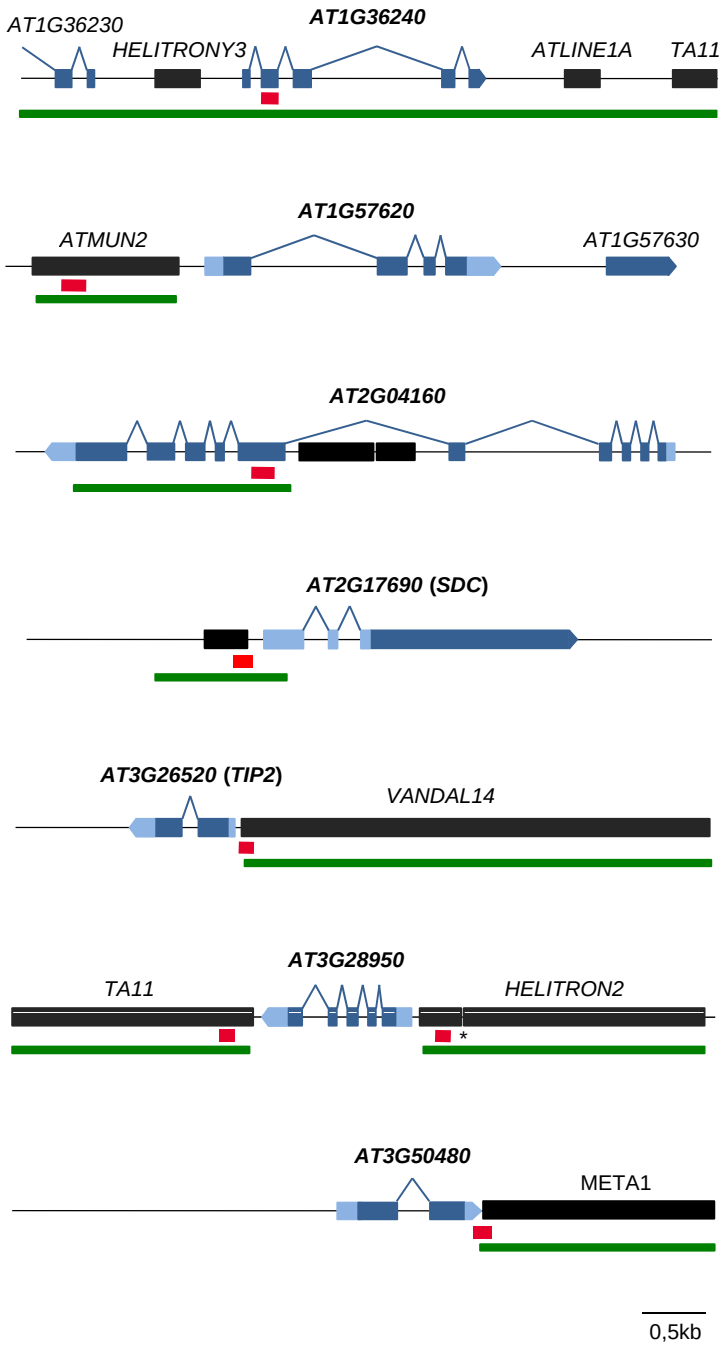


Figure S2. Additional CATMA analysis comparing single and double mutants.


AT1G36230
AT1G36240
HELITRONY3
ATLINE1A
TA11
AT1G57620
ATMUN2
AT1G57630
AT2G04160
AT2G17690 (SDC)
AT3G26520 (TIP2)
VANDAL14
AT3G28950
TA11
HELITRON2
*
AT3G50480
META1
0,5kb

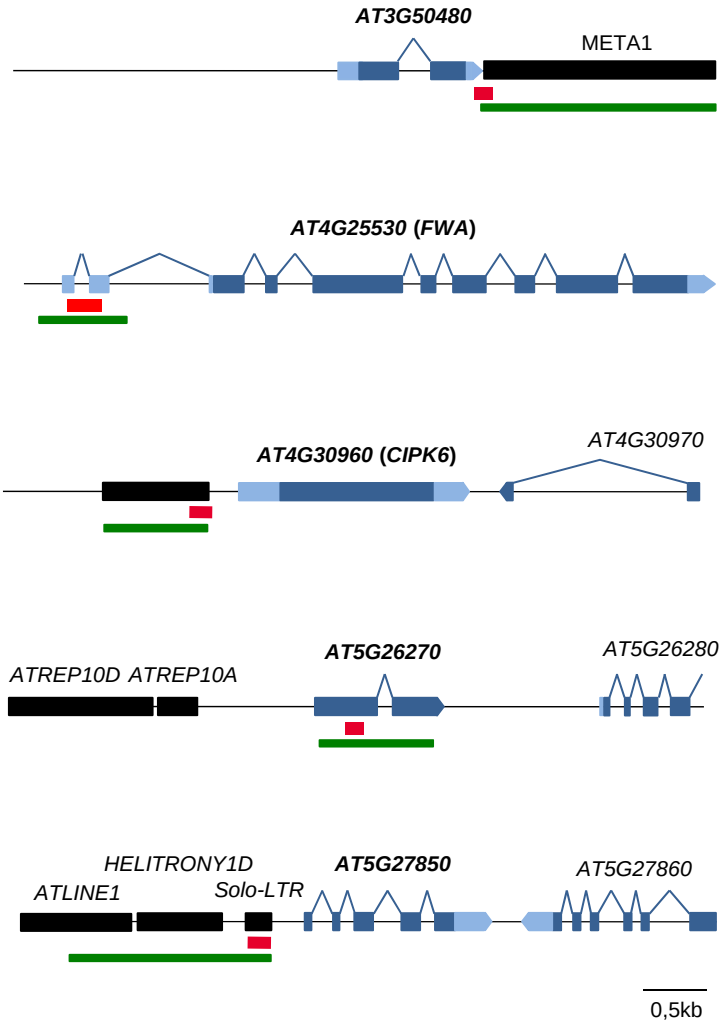


Figure S3. Genomic structure of eleven tested genes. Densely methylated sequences are indicated as green lines. Densely methylated sequences probed for DNA methylation by McrBC-qPCR are shown in red. Starts are indicate nearby probes for which changes in DNA methylation do not show correlation with changes in gene expression. Annotation: UTRs – light blue; coding sequences - dark blue; repeat elements – black.

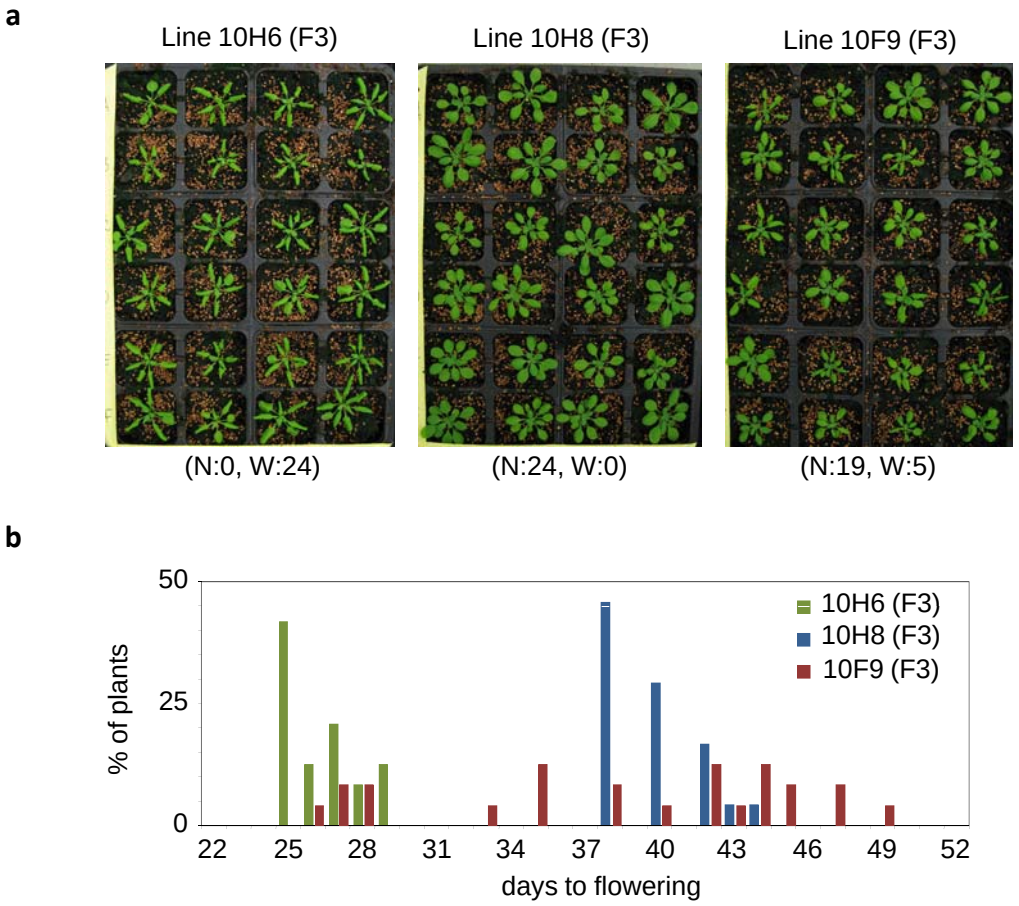


Figure S4. Segregation of leaf structure (A) and flowering time (B) phenotypes in 24 (F3) siblings originated through selfing of three (F2) *DDM1/DDM1*, *rdr2/rdr2* individuals: 10H6 (wrinkled leaf and early-flowering), 10H8 (normal leaf and late-flowering), and 10F9 (wrinkled leaf and late-flowering).

Table S1. Segregation of *DDM1* and *RDR2* alleles in F2 progeny

Genotype	**Observed**	**Expected**
DDM1/DDM1, rdr2/rdr2	9	8,75
DDM1/DDM1, RDR2/rdr2	14	17,5
DDM1/DDM1, RDR2/RDR2	12	8,75
DDM1/ddm1, rdr2/rdr2	18	17,5
DDM1/ddm1, RDR2/rdr2	36	35
DDM1/ddm1, RDR2/RDR2	19	17,5
ddm1/ddm1, rdr2/rdr2	9	8,75
ddm1/ddm1, RDR2/rdr2	18	17,5
ddm1/ddm1, RDR2/RDR2	5	8,75
Total	140	140

Table S2. List of the 12 candidate genes tested.

Gene ID	**Annotation (TAIR)**	**Position of methylated region probed**
AT1G36240	60S ribosomal protein L30	2nd exon
AT1G57620	Emp24/gp25L/p24 family protein	Upstream and Downstream
AT2G04160	Auxin-induced in root culture 3	6th exon
AT2G17690	F-box domain containing protein (SDC)	Upstream
AT3G26520	Gamma tonoplast intrinsic protein 2 (TIP2).	Upstream
AT3G28950	Avirulence induced gene (AIG) protein	Upstream and Downstream
AT3G50480	Homolog of RPW8 4 (HR4)	Downstream
AT4G25530	Homeodomain-containing transcription factor (FWA)	5' and 1st exon
AT4G30960	CBL-INTERACTING PROTEIN KINASE 6 (CIPK6)	Upstream
AT5G22500	Acyl CoA reductase/male-sterility protein (FAR1)	8th intron
AT5G26270	Unknown protein	1st exon
AT5G27850	60S ribosomal protein L18	Upstream

Table S3. Raw RT-qPCR data. Average percentage (a minimum of two measurements) of expression in relation to controls and standard deviation are presented.

Name of primer pair	% of control genes				STDEV			
	wt	*ddm1*	*rdr2*	*ddm1rdr2*	wt	*ddm1*	*rdr2*	*ddm1rdr2*
AT1G36240	1,27	50,18	1,56	50,00	0,17	0,74	0,05	0,25
AT1G57620	5,60	5,22	1,49	1,58	0,07	0,37	0,07	0,26
AT2G04160	2,83	1,51	1,43	1,14	0,14	0,09	0,03	0,03
AT2G17690	0,05	3,05	0,61	11,16	0,00	0,16	0,02	0,55
AT3G26520	981,05	241,62	716,23	40,10	69,66	6,51	24,57	2,75
AT3G28950	24,96	8,96	19,82	8,50	0,06	0,09	0,49	0,48
AT3G50480	13,68	35,23	18,69	53,42	0,47	0,52	0,18	2,09
AT4G25530	0,00	0,17	0,02	1,15	0,00	0,02	0,02	0,12
AT4G30960	3,51	2,56	2,49	1,25	0,01	0,13	0,04	0,13
AT5G22500	0,71	0,91	0,64	0,19	0,01	0,01	0,08	0,01
AT5G26270	1,02	25,53	3,27	23,90	0,10	0,75	0,11	0,47
AT5G27850	519,64	585,65	1322,71	1861,58	1,27	5,74	45,37	95,76

Table S4. Raw McrBC-qPCR data. Average percentage (a minimum of two measurements) of molecules lost through McrBC digestion and standard deviation are presented.

Name of primer pair	% of methylation				STDEV			
	wt	*ddm1*	*rdr2*	*ddm1rdr2*	wt	*ddm1*	*rdr2*	*ddm1rdr2*
AT1G36240	78,83	16,19	72,16	7,93	0,62	2,05	0,96	2,12
AT1G57620_5met	93,40	87,12	17,33	8,21	1,85	1,26	2,84	5,84
AT2G04160	99,24	7,07	99,08	20,17	0,14	10,00	0,09	0,39
AT2G17690(2)	96,27	77,93	89,74	39,25	0,02	0,00	0,05	2,98
AT3G26520_5met	98,01	39,33	75,00	2,71	0,06	0,98	0,25	3,01
AT3G28950_3met	99,54	18,49	99,19	24,21	0,09	0,40	0,04	3,55
AT3G28950_5met	98,01	85,58	62,24	41,96	0,08	0,92	0,56	0,85
AT3G50480_3met	87,24	62,37	13,54	7,63	0	0,74	0,85	3,17
AT4G25530(FWATR)	99,04	69,19	97,94	3,74	0,01	2,11	0,55	0,47
AT4G30960_5met(2)	94,88	89,88	78,79	17,34	1,38	0,25	0,00	2,03
AT5G26270	99,02	5,59	95,56	3,04	0,05	6,93	0,24	3,33
AT5G27850_5met	97,33	85,31	10,79	5,84	0,07	5,28	7,53	3,64

Table S5. List of primer pairs used for McrBC-qPCR analysis, sequence annotation and the position in relation to gene.

Name of primer pair	Chr.	Position (TAIR7)	Forward	Reverse	Annotation	Position in relation to gene
AT1G36240	1	13616208..13616307	AGTCCCATGAAGGAATCAACA	TTGGAACTGCGAAGAGATTTG	-	2nd exon
AT1G57620_5met	1	21345158..21345327	TAAGCTTGGGAAAAATGTCG	TTGGCCTGGTCAACGAAT	ATMUN2 (DNA/MuDR)	5'
AT2G04160	2	1403259..1403436	TGAGTGTGATCGCTGCCTAC	GGGGCTCCAAGAAGGATAAC	-	6th exon
AT2G17690(2)	2	7689165..7689268	CGTCAGCCCTAACACAAAACA	GCGGCGTGAGTCTCTAAGAT	Single tandem repeats	5'
AT3G26520_5met	3	9725039..9725275	GCCGAGTGTTTTCTTCTTCTTG	ACAAGTCTGCTGTCGGGAAT	VANDAL14	5'
AT3G28950_3met	3	10976761..10977039	CCCAACGCTCTATGGCTAAA	AGGATGAGCCGTTTACGATG	TA11	3'
AT3G28950_5met	3	10978833..10979036	GTCGGTATCCGTGCTACACA	CCATGCTACAAACCGAGGAA	HELITRONY2	5'
AT3G50480_3met	3	18745119..18745283	CAAAGAAATGAAGGCCCAGAT	GATGATGGGCTTAGATGATGG	META1 (LTR/Copia)	3'
AT4G25530(FWATR)	4	13038394..13038584	GCCATTGGTCCAAGTGCTAT	CGGTGCTCGTATGAATGTTG	Single tandem repeats	5' and 1st exon
AT4G30960_5met(2)	4	15066625..15066785	CCGCTGGTTTACCAGTCATAA	TAGGGCCTGACGGATTAAAA	Dispersed repeat	5'
AT5G26270	5	9206718..9206926	GCTGTGTGCAGTCCCTTCTA	GCCATGGAACGAGTAGGAGA	-	1st exon
AT5G27850_5met	5	9872531..9872826	TGCTTTCTCTTTCTCTTCTCTTTTTC	AAACCGGATAAGTATGGATGTCA	HELITRONY1D (RC/Helitron)	5'
META1_5met	4	7816672..7816958	GAGGTGATTAGCTAGCACTTTCAA	GGAAATGAGTGTGTTATTCCATTG	META1 (LTR/Copia)	-

Table S6. List of primer pairs used for RT-qPCR analysis and gene annotation

Name of primer pair	Chr.	Position (TAIR7)	Forward Primer	Reverse Primer	Annotation
AT1G36240	1	13616208..13616307	AGTCCCATGAAGGAATCAACA	TTGGAACTGCGAAGAGATTTG	60S ribosomal protein L30
AT1G57620	1	21347625..21347787	CACCTTATGGCACCGTGTTA	GCGATTCCAGTTTTCCAGTC	Emp24/gp25L/p24 family protein
AT2G04160	2	1403259..1403436	TGAGTGTGATCGCTGCCTAC	GGGGCTCCAAGAAGGATAAC	Auxin-induced in root culture 3
AT2G17690	2	7690614..7690902	ACTAACGGTTCTTGCGGAGA	CTATCGTTTTGGGAGGGTCA	F-box domain containing protein (SDC)
AT3G26520	3	9724077..9724274	TGACCCCAAGAACGGTAGTC	GCGAGTCCACCACCAATAAG	Gamma tonoplast intrinsic protein 2 (TIP2).
AT3G28950	3	10977537..10977707	CGGAGAAATTCGTGGAGACT	CAAAACCCACTCTTTACTTACTCAC	Avirulence induced gene (AIG) protein
AT3G50480	3	18744185..18744361	CCGGCTATCAGATCGTCTTC	AGCCGTCGTAAATGACTTGG	Homolog of RPW8 4 (HR4)
AT4G25530	4	13038394..13038584	GCCATTGGTCCAAGTGCTAT	CGGTGCTCGTATGAATGTTG	Homeodomain-containing transcription factor (FWA)
AT4G30960	4	15067667..15067829	CACGAAGTCATGGCGAGTAA	CACCACGGCTATGACAGAAA	CBL-INTERACTING PROTEIN KINASE 6 (CIPK6)
AT5G22500	5	7472038..7472190	AGCGGTGAACAACCTGGAT	TTGCGAGATCTCTTCATCAGAA	Acyl CoA reductase/male-sterility protein (FAR1)
AT5G26270	5	9206718..9206926	GCTGTGTGCAGTCCCTTCTA	GCCATGGAACGAGTAGGAGA	Unknown protein
AT5G27850	5	9874172..9874332	TGAAGCAGTGAAGCATTTCG	ACCAAGCAAAACGAAACCAG	60S ribosomal protein L18
Controls					
At2g36060	2	15149917..15150067	TGAAGTCGTGAGACAGCGTTG	GGGCTTCTCCATTGTTGGTC	Ubiquitin E2 variant 1C
At4g29130	4	14352280..14352408	GGCGTTTTCTGATAGCGAAAA	ATGGATCAGGCATTGGAGCT	Hexokinase (HXK1)
At5g13440	5	4308303..4308423	ACAAGCCAATTTTTGCTGAGC	ACAACAGTCCGAGTGTCATGGT	Ubiquinol-cytochrome C reductase iron-sulfur subunit

4.2 Influence des nouvelles insertions d'ET sur l'expression des gènes situés à proximité avant et après établissement du contrôle épigénétique

Les analyses présentées dans le manuscrit ci-dessus indiquent que relativement peu de gènes sont directement influencés par la méthylation des ET (et autres séquences répétées) localisés à proximité chez Arabidopsis. Quoiqu'il en soit, ces analyses ne permettent de voir que ce qui a passé le filtre de la sélection naturelle et ne nous renseigne pas sur l'influence qu'un ET tout juste inséré peut avoir sur l'expression des gènes situés à proximité. J'ai donc cherché à déterminer les conséquences que les ET nouvellement insérés pouvaient avoir sur les gènes situés à proximité dans les epiRIL avant et après acquisition du contrôle épigénétique. Je me suis focalisée sur les nouvelles insertions d'*EVADE* (*ATCOPIA93*), le seul ET pour lequel l'établissement de la méthylation de l'ADN associé à la répression transcriptionnelle a pu être clairement décrit.

Je me suis intéressée à un premier jeu de quatre gènes qui présentent des nouvelles insertions d'*EVADE* à l'état homozygote en F9 dans l'épiRIL 454 : *AT1G01010*, *AT1G68410*, *AT5G2550* et *AT5G25560* (fig. 4.1). J'ai testé l'expression de ces gènes par RT-qPCR dans des plantes individuelles de l'épiRIL 454 présentant les copies d'*EVADE* dans un état méthylé ou non. Parmi les quatre gènes étudiés, *AT5G2550* qui présente une nouvelle insertion d'*EVADE* à ~300 pb de son extrémité 3', montre une variation significative d'expression corrélée avec le changement d'état de méthylation de la copie d'*EVADE* située à proximité (fig. 4.2). Afin de déterminer si les transcrits détectés ici sont produits à partir du promoteur du gène ou s'il s'agit de transcrits « read through » initiés dans l'ET, il faudra réaliser des expériences de 5' Race. Les trois autres gènes ne sont pas affectés par l'insertion d'*ATCOPIA93* située à proximité (données non montrées). Dans Mari-Ordonez et al., 2013, les données concernant le gène *AT1G68420* sont également présentées (figure 6 de l'article). Néanmoins, étant donné qu'*EVADE* est localisé dans la séquence codante du gène (fig. 4.1), ce dernier est vraisemblablement invalidé.

Cette analyse sera étendue à plus d'insertions et poursuivie de façon systématique sur les autres epiRIL présentant des nouvelles insertions d'*ATCOPIA93*. Les domaines d'expression des candidats pourront être comparés avec celui d'*ATCOPIA93* qui est restreint à la L2.

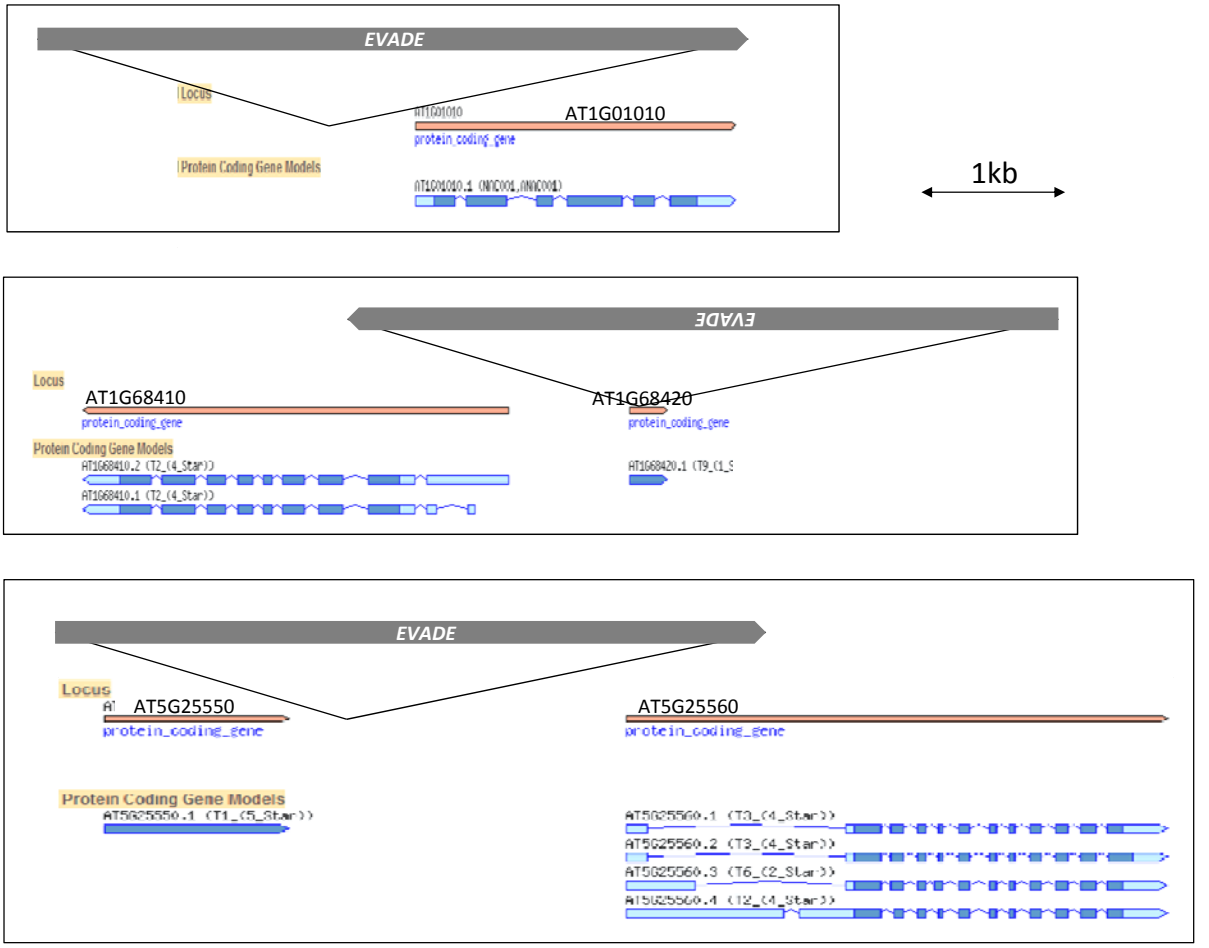


Figure 4.1 : Positionnement de trois nouvelles insertions d'*EVADE* (*ATCOPIA93*) dans l'epiRIL 454 sur une vue Gbrowse.

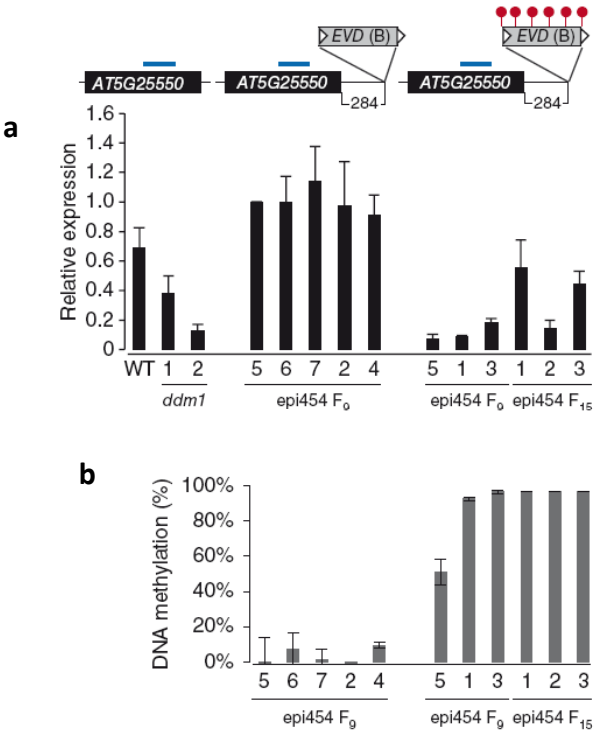


Figure 4.2 : Conséquences de l'insertion et de la méthylation d'*EVADE* sur l'expression du gène *AT5G25550*. (**a**) Niveau d'expression (RT-qPCR) du gène *AT5G2550* chez le sauvage, *ddm1* et des plantes individuelles de l'épiRIL 454 présentant une insertion d'*EVADE* en 3'. Les traits bleus représentent la région amplifiée par qPCR. (**b**) Méthylation (McrBC-qPCR) de l'insertion d'*EVADE* localisée en 3' d'*AT5G25550* dans des plantes individuelles de l'épiRIL 454.

5
Conclusions générales et perspectives

5.1 Conclusions générales

Les travaux rapportés dans cette thèse portent sur le rôle joué par la méthylation de l'ADN dans le contrôle de la mobilisation des ET chez la plante de référence *Arabidopsis thaliana*. En m'appuyant essentiellement sur l'étude d'une population d'epiRIL, j'ai pu confirmer que seul un petit nombre d'ET est mobilisé suite à la perte de la méthylation de l'ADN (partie 2), et ce alors même que plusieurs centaines d'ET sont réactivés transcriptionnellement dans ces conditions. D'autre part, mes travaux indiquent que l'établissement du contrôle épigénétique sur les ET nouvellement insérés diffère d'une famille d'ET à l'autre, même lorsque celles-ci appartiennent à la même superfamille (Partie 3). C'est dans ce cadre que j'ai contribué à la caractérisation approfondie des modalités de mise en place de la répression épigénétique sur *EVADE* (partie 3 et Mari-Ordonez et al. 2013), l'élément le plus mobile dans les epiRIL.

J'ai également montré que la distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis (partie 2) diffère radicalement de celle des copies résidentes, qui sont fortement regroupées dans les régions péricentromériques. Ce regroupement résulterait donc plus de l'élimination des insertions situées dans les bras chromosomiques que d'un ciblage préférentiel des ET dans les régions péricentromériques. Cette conclusion est en accord avec des analyses récentes qui montrent que le génome d'*Arabidopsis thaliana* subit une réduction de taille rapide depuis la divergence avec *Arabidopsis lyrata*, causée principalement par un grand nombre de délétions de petite taille qui affecterait majoritairement les ET (Hu et al. 2011).

Enfin, je me suis intéressée à l'influence directe des variations de méthylation des ET sur la régulation des gènes situés à proximité. En m'appuyant sur les nouvelles insertions d'*EVADE* dans les epiRIL, j'ai montré que l'expression d'un gène pouvait être affectée par l'état de méthylation d'une nouvelle insertion d'ET située à proximité (partie 4). J'ai également contribué à montrer, en collaboration avec un précédent étudiant en thèse au laboratoire, Felipe Teixeira, que seul un petit nombre de gènes est directement affecté par la perte de méthylation des séquences répétées situées à proximité chez Arabidopsis. Néanmoins, le maintien de la méthylation de l'ADN sur quelques locus est assuré de manière redondante

par l'action de DDM1 et du RdDM, et la dérégulation de ces gènes suite à une perte de méthylation conduit à des altérations phénotypiques sévères (partie 4).

5.2 Perspectives

5.2.1 Contrôle de la mobilisation des ET chez *Arabidopsis thaliana*

Comme nous l'avons vu dans la partie 1, il existe un décalage important entre le nombre d'ET réactivés transcriptionnellement et le nombre d'éléments mobiles dans *ddm1* et les epiRIL. Nous avons évoqué plusieurs hypothèses pour expliquer ces résultats, dont notamment une spécificité cellulaire ou tissulaire du contrôle des ET telle qu'aucun événement de transposition ne puisse être transmis à la génération suivante. Rappelons en effet que notre schéma expérimental ne permet d'identifier que ce type d'événement et nullement la mobilisation somatique des ET, qui reste totalement inexplorée chez Arabidopsis.

L'absence d'expression de certains ET spécifiquement dans les tissus qui participent à la génération suivante peut résulter de mécanismes de répression particulièrement efficaces dans ces tissus. Cette hypothèse est appuyée par les nombreuses indications, présentées dans l'introduction, d'un contrôle spécifique des ET au moment de la reproduction (de la gamétogénèse jusqu'au début du développement de la graine) faisant potentiellement intervenir les différentes voies de l'interférence ARN. Si tel est le cas, nous devrions observer la mobilisation d'un plus grand nombre d'éléments (de façon transmissible) dans les mutants affectant simultanément la fonction DDM1 et les voies de l'ARNi (RdDM et PTGS). A ce sujet, nous pouvons déjà noter que le mutant double *ddm1rdr2* présente une fertilité très réduite par rapport aux deux mutants simples (partie 4). D'un autre côté, le mutant double *ddm1rdr6* est quant à lui quasiment stérile (données du laboratoire non publiées). Ces observations pourraient s'apparenter aux phénomènes de catastrophe méiotique et de dysgénésie des hybrides observés respectivement chez la souris et la drosophile dans des conditions causant la réactivation des ET (Bourc'his and Bestor 2004, Castro and Carareto 2004, Bucheton 1990). Il est donc envisageable que la transposition dans ces mutants soit autrement plus importante, concerne plus d'ET, et puisse être (du moins partiellement) à l'origine de cette stérilité. Cependant, si le RdDM et/ou le PTGS permettent de réprimer l'activité des ET dans certains tissus, il restera à déterminer pourquoi, dans *ddm1*, certains ET peuvent échapper à ce contrôle. L'étude d'*EVADE* fournit une première explication,

puisque dans ce cas nous avons vu que la protéine GAG « protège » l'ARN messager d'*EVADE* de sa dégradation par le PTGS (partie 3 et Mari-Ordonez et al. 2013). Il en va probablement de même pour les autre rétroéléments à LTR.

Quoiqu'il en soit, s'il existe bien un contrôle renforcé des ET à certains stades de développement alors que ces mêmes ET sont exprimés (voire mobiles) dans d'autres tissus dans le mutant *ddm1,* on peut se demander pourquoi ce contrôle ne s'étend pas à toute la plante. Outre le fait que la transposition somatique puisse n'avoir que peu de conséquences sur la fitness de l'individu (et que la mise en place d'un contrôle renforcé a potentiellement un coût), ne peut-on pas également supposer que la réactivation (transcriptionnelle notamment) des ET dans certains tissus confère un avantage quelconque à la plante ? Nous savons que la réactivation transcriptionnelle de certains ET s'accompagne de la production de siARN (Teixeira et al. 2009, Slotkin et al. 2009) et des travaux récents ont mis en évidence la mobilité des siARN dans la plante (Melnyk et al. 2011, Molnar et al. 2010, Brosnan and Voinnet 2011). Si l'activation des ET dans les epiRIL a été induite artificiellement par l'utilisation du mutant *ddm1*, nous savons que des conditions « naturelles » comme des stress peuvent également induire la réactivation des ET (Pecinka et al. 2010, Tittel-Elmer et al. 2010, Yu et al. 2013). On peut alors supposer que la production de petits ARN mobiles à partir des ET puisse être impliquée dans la réponse au stress.

5.2.2 Prolifération et élimination des ET chez *Arabidopsis thaliana*

L'étude de la distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis a révélé un profil drastiquement différent de celui des copies résidentes. Cela suggère donc que la suraccumulation des ET dans les régions péricentromériques chez Arabidopsis relève plus de l'élimination des ET insérés dans les bras chromosomiques que d'un ciblage préférentiel. Cette élimination peut être le fruit de la sélection naturelle, autrement dit la contre-sélection des insertions ayant des effets délétères. Ce peut notamment être le cas des insertions ayant lieu dans des gènes ou à proximité. De fait, près de 70% des nouvelles insertions d'*ATCOPIA93*, *ATENSPM3* et *VANDAL21* sont trouvées dans ou à proximité des gènes. Si au laboratoire il est possible de maintenir de telles plantes, dans la nature en revanche, il est probable qu'elles soient rapidement contre-sélectionnées. Des expériences

de compétition entre epiRIL dans des conditions de culture variées permettraient d'estimer l'impact des nouvelles insertions d'ET sur la fitness des lignées concernées.

L'autre force évolutive permettant l'élimination des ET est la recombinaison homologue qui, chez Arabidopsis, a lieu quasi essentiellement le long des bras chromosomiques. Or, nous savons que la recombinaison homologue inégale entre les deux LTR d'un rétroélément aboutit à l'élimination de l'ET pour ne laisser qu'une seule LTR (solo-LTR). La perte de l'ET peut alors se fixer soit car elle est positivement sélectionnée soit par dérive. Les epiRIL fournissent un matériel de choix pour tester la formation de solo-LTR à partir des nouvelles insertions et le rôle éventuel de la méthylation de l'ADN dans cette formation. De fait, chez certaines espèces, la méthylation de l'ADN est un puissant inhibiteur de la recombinaison homologue (Colot et al. 1996 Maloisel and Rossignol 1998). Cela ne semble cependant pas être le cas chez Arabidopsis car le taux de recombinaison mesuré dans les régions péricentromériques des epiRIL n'est pas plus élevé que dans des croisements classiques entre différentes accessions (Colome-Tatche et al. 2012). Néanmoins, il est possible que localement, la méthylation de l'ADN puisse affecter la recombinaison.

Pour finir, il reste à déterminer dans quelle mesure les résultats obtenus à partir des epiRIL reflètent l'activité récente des ET dans le génome d'Arabidopsis, notamment concernant l'identité des ET mobilisés. En effet, les perturbations induites par la mutation *ddm1* diffèrent de celles observées en réponse à des variations environnementales décrites à ce jour, qui semblent être de moindre intensité et transitoires (Pecinka et al. 2010, Tittel-Elmer et al. 2010, Yu et al. 2013). Le séquençage du génome de 1001 accessions d'Arabidopsis devrait permettre de tester rapidement si oui on non les ET mobiles dans les épiRIL sont également les plus mobiles dans la nature (Cao et al. 2011).

6
Matériel et Méthodes

6.1 Matériel biologique et conditions de culture

Tous les travaux présentés dans cette thèse ont été effectués sur l'accession Columbia (col) *d'Arabidopsis thaliana*.

L'allèle mutant pour le gène DDM1 est *ddm1-2* (Vongs et al. 1993) et l'allèle mutant pour le gène RDR2 est *rdr2-1* (Xie et al. 2004).

Le mutant *ddm1* utilisé pour toutes les expériences a subit quatre cycles d'autofécondation à partir de l'établissement de la mutation *ddm1-2* à l'état homozygote. Le double mutant *ddm1rdr2* a été obtenu comme décrit dans le manuscrit Teixeira et al., par croisement du mutant *ddm1-2* avec le mutant *rdr2-1.* Les plantes utilisées ont subit deux générations d'autofécondation après établissement des deux mutations à l'état homozygote.

La population d'epiRIL a été obtenue à partir d'un croisement entre le mutant *ddm1-2* et un individu sauvage comme décrit dans Johannes et al. 2009.

A l'exception des deux cas mentionnés ci-après, toutes les expériences ont été réalisées sur des plantules entières cultivées dans du milieu liquide (1/2MS, 7% de sucrose) en conditions de jour long (16 heures de lumière à 22°C et 8 heures de nuit à 19°C).

Les analyses de la méthylation de l'ADN et d'expression dans les générations précoces et avancées des epiRIL 260, 454 et 439 (partie 3.1) ont été réalisées sur des feuilles de rosette de plantes individuelles cultivées en terre en conditions de jour court (8h de lumière, 16h de nuit).

Les analyses des candidats du crible et des contrôles (plante sauvage et epiRIL 60 et 98) (partie 3.2) ont été réalisées sur des feuilles de rosette de plantes cultivées en terre en conditions de jour long.

6.2 Identification des évènements d'insertion ayant eu lieu dans les epiRIL

Les ADN d'epiRIL utilisés pour le séquençage sont issus de lots de plantes F9, les insertions détectées correspondent donc à celles présentent dans la plante mère unique F8 pour

chaque lignée. L'ADN du mutant *ddm1* a été extrait à partir d'un lot de plantes cousines de la plante *ddm1* à l'origine des epiRIL.

Le séquençage et l'analyse des 53 lignées epiRIL ainsi que des deux sauvages et du mutant *ddm1* et l'identification des nouvelles insertions sont décrits dans le manuscrit Gilly et al. présenté dans la partie 2.

Les amorces utilisées pour la validation des évènements d'insertions sont listées dans la table supplémentaire 12 du manuscrit Gilly et al.

Les insertions considérées pour les analyses (annexe 2) sont les listes de sortie de TE-tracker auxquelles ont été enlevés les évènements suivants :

- les évènements impliquant des régions très répétées du génome qui ne sont pas présentes sur le génome de référence comme les répétitions centromériques ou dans les répétitions d'ADN ribosomique (ADNr). De fait le programme a identifié des évènements d'insertion dans ces régions et qui sont quasiment toujours du même type : des rétroéléments LTR GYPSY de la superfamille *ATHILA* dans les répétitions centromériques et de la superfamille ATGP dans l'ADNr. Le fait que de tels types d'évènements aient été identifiés chez les deux sauvages que nous avons séquencé suggère qu'il s'agit de faux positifs. De fait, il a été montré que les centromères d'Arabidopsis sont constitués de répétitions satellites de 180 pb entrecoupées par des rétroéléments LTR GYPSY mais ces régions ne sont pas présentes dans le génome de référence.

- les évènements faisant intervenir des donneurs qui ne recouvrent pas ou quasiment pas (<10%) des ET annotés. Une analyse de ces régions a montré qu'il s'agit dans la plupart des cas de répétitions centromériques imbriquées les unes dans les autres ou insérées dans des *ATHILA*. Pour les mêmes raisons que celles évoquées plus haut, il s'agit probablement de faux positifs. Certains des donneurs non annotés qui ne semblent pas correspondre à des répétitions centromériques ont été inspectés pour la présence de structures caractéristiques d'ET (TIR et LTR) mais aucune n'a pu être mise en évidence.

- les évènements identifiés également chez les sauvages (qui pour la plupart appartenaient aux deux catégories précédentes).

6.3 Analyses de la méthylation de l'ADN

- MeDIP-chip

Les méthylomes de wt, *ddm1*, *rdr2*, *ddm1rdr2* (partie 2.2) ainsi que des epiRIL 60 et 98 (partie 3.2) ont été réalisés par MeDIP-chip. Dans cette méthode, l'ADN simple brin méthylé est immunoprécipité à l'aide d'un anticorps spécifique puis hybridé, en comparaison avec l'ADN total, sur une puce à ADN (Nimblegen). Cette puce contient 711320 sondes homothermales définies dans des fenêtres de 165nt tout le long de l'ensemble du génome d'*Arabidopsis thaliana.* Le protocole expérimental est décrit dans Cortijo et al., 2012.

- McrBC-qPCR

Les analyses du pourcentage de méthylation des nouvelles insertions d'ET et des copies résidentes (partie 2.2 et partie 3.1), des séquences reméthylables et non-reméthylables dans les epiRIL 60 et 98 (partie 3.2) ainsi que des séquences méthylées situées à proximité des gènes (partie 4.2, manuscrit Teixeira et al.) ont été réalisées par McrBC-qPR. L'ADN génomique est extrait à partir de matériel congelé avec le DNeasy Mini Kit (Qiagen). La digestion par l'enzyme McrBC (New England Biolabs), qui coupe spécifiquement l'ADN méthylé, est réalisée sur 200 ng d'ADN génomique et 200 ng sont utilisés pour le contrôle non digéré. La qPCR est réalisée parallèlement sur 2 ng d'ADN des échantillons digérés et non digérés en utilisant la machine Roche LightCycler® 480 et le LightCycler® 480 SYBR Green I Master. Le pourcentage de méthylation est obtenu par le calcul suivant :

%méthylation = 1-[2^(Cp ADN non digéré - Cp ADN digéré)]*100

Deux séquences contrôles sont utilisées pour évaluer l'efficacité de la digestion (contrôle positif) et vérifier que la quantité d'ADN des échantillons digéré et non digéré est équivalente (contrôle négatif).

Amorces utilisées :

Utilisation/nom	Amorce 1 (5'-3')	Amorce 2 (5'-3')
5' du locus donneur *AT2TE42810* (*VANDAL21*)	TTTTTGGTTTCAAAATGTTTTCTACA	TCTTAGTTCCTCCTCACAGAACG
5' du locus donneur *AT2TE20205* (*ATENSPM3*)	GAAGCATTGGAGCTGGTTTC	TAAGTGTTGGCGCTGAAGTG
région interne d'*AT2TE20205* (*ATENSPM3*)	GCAAAGTTAGCAGATGCTGAAATAC	CATATCCCATATCAAAGCTCCTTAG
5' de l'insertion 1 d'*AT2TE20205* dans l'éiRIL 454	GAAGGAAAAGAACATGGAAAACC	TAAGTGTTGGCGCTGAAGTG
5' de l'insertion 1 d'*AT2TE20205* dans l'éiRIL 439	CGCCACAACAATGATGCTT	TAAGTGTTGGCGCTGAAGTG
5' de l'insertion 2 d'*AT2TE20205* dans l'éiRIL 439	CTAAAAATTTAAGACCAGTTTTATTATAGGTATGT	TAAGTGTTGGCGCTGAAGTG
5' de l'insertion 3 d'*AT2TE20205* dans l'éiRIL 439	CAAAGGACCGTGTCATAAGAAAG	TAAGTGTTGGCGCTGAAGTG
5' de l'insertion 4 d'*AT2TE20205* dans l'éiRIL 439	AAGACCATCTACGTTCATAAAGAGTTT	TAAGTGTTGGCGCTGAAGTG
5' du locus donneur *AT3TE89830* (*ATCOPIA78*)	GCGAAAACATCTGTGCTTGTAG	TGTTTAAGGAGGTGGAAGATG
5' du locus donneur *AT3TE92525* (*ATCOPIA78*)	CCTATCAATGGTATAAAGTGTTGTATCTT	TGTTTAAGGAGGTGGAAGATG
5' du locus donneur *AT5TE15240* (*ATCOPIA78*)	ACACAACTAAAGAAAAACATTAGAGGAA	TGTTTAAGGAGGTGGAAGATG
5' de l'insertion 1 d'*ATCOPIA78*	AGGGTTTAACTGGAATAGAATTTGG	TGTTTAAGGAGGTGGAAGATG
5' de l'insertion 2 d'*ATCOPIA78*	TCTCTTGAGTCTAATAACGAGTAAGAACA	TGTTTAAGGAGGTGGAAGATG
5' de l'insertion 3 d'*ATCOPIA78*	TTTCTTCTTCTCCGATTCTCTCC	TGTTTAAGGAGGTGGAAGATG
5' de l'insertion 4 d'*ATCOPIA78*	TCCCAGTAAATCTCCGGCTAC	TGTTTAAGGAGGTGGAAGATG
AT1G57620_5met	TAAGCTTGGGAAAAATGTCG	TTGGCCTGGTCAACGAAT
AT5G27850_5met	TGCTTTCTCTTTCTCTTCTCTTTTTC	AAACCGGATAAGTATGGATGTCA
RB(0)	CGACGAGAGGCTGCATCTAA	CACAGAACAAGGAACAAACGAG
RB(22,5)	ACCCAGGTGGTCGAGTGATA	CAAGCCGAGTACCACCATCT
LB(10)	TCCCTCGCTGGAGGTATATG	AGTCCGCCAACTGATGATGT
LB(42,5)	AATCAGGCGGCTTTCAGTT	TGACATCATTGTTGATCCTCAG
LB(45)	TTCTTAGGATTTGCTCCGTCA	GCCTTTGAACTTTGTGAATGG
LB(47,5)	GCGGTGAACGTGTCAAATAC	GGCAATGGCTTACTGCTAGG
LB(20)	GGCGCTTATCTCCTGTTCTG	ATTTTGGGAAATCGGGAAAC
LB(22,5)	GGTTGTAACCAGCGAATCGT	TGGAGGGTCTGCGTAGATTT
AT5G13440 : contrôle négatif	ACAAGCCAATTTTTGCTGAGC	ACAACAGTCCGAGTGTCATGGT
AT5g36220_3met : contrôle positif	CCGAACACTTCACCAGATCA	CAGACCCGGGTAACTTTTGA

Les amorces utilisées pour l'analyse de la méthylation des nouvelles insertions d'*ATCOPIA93* sont listées dans Mari-ordonnez et al, 2013 présenté dans la partie 3 et celles pour tester la méthylation des séquences méthylées à proximité des gènes dans le matériel et méthode du manuscrit Teixeira et al. présenté dans la partie 4.

Toutes les amorces ont été dessinées avec le Primer3 (http://frodo.wi.mit.edu/primer3/) en utilisant les paramètres suivants : %GC des amorces compris entre 40% et 60%, longueur maximale du produit PCR de 300 bp et Tm =60°C.

- séquençage bisulfite (BS-seq)

L'analyse de la méthylation du locus *FWA* chez le sauvage, les epiRIL 60 et 98 ainsi que les candidats du crible a été réalisée par BS-seq (partie 3.2). Le traitement de l'ADN au bisulfite engendre la conversion des cytosines non méthylées en uracile. Cette méthode permet d'obtenir des informations précises sur la position des cytosines méthylées au sein d'une séquence. L'ADN génomique a été extrait avec le DNeasy Mini Kit (Qiagen) et 300ng ont été traité au bisulfite en utilisant le EZ DNA Methylation-Gold™ Kit (ZYMO RESEARCH) en suivant les étapes du protocole alternatif 2 proposé par le fabricant. L'amplification par PCR de la séquence répétée de FWA ainsi que d'une séquence contrôle non méthylée ont été réalisées avec les amorces suivantes :

séquence	Amorce 1 (5'-3')	Amorce 2 (5'-3')
FWA-BS	GGGTTTAGTGTTTAYTTGTTTAAGG	CAACAAAAATCTRATTRTCARTATCCTAC
AT5G15440-BS	GGTAAGAATTGAAAGATGAGAGAT	TAATTATCATRTTARAACACCRTCATT

Les produits PCR ont été purifiés avec le QIAquick PCR purification kit (Qiagen) puis clonés dans un vecteur PGEM-T (Promega). Une douzaine de clones a été séquencée pour chaque lignée. Les séquences ont étés alignées avec le programme MEGA et l'analyse de la méthylation a été faite avec CyMATE (http://www.cymate.org/).

6.4 Analyses d'expression

- ARN-chip

L'analyse différentielle d'expression des ET dans le mutant *ddm1* par rapport au sauvage (partie 2.2) a été réalisée à partir de données issues d'une expérience d'ARN-chip publiée dans Colome-Tatche et al. 2012

Les analyses différentielles d'expression présentées dans le manuscrit Teixeira et al. (partie 4.2) ont également été réalisées par ARN-chip mais en utilisant des puces CATMA. Les ARN ont été extraits avec le RNeasy Mini Kit (Qiagen) et les étapes d'amplification, marquage et hybridation ont été réalisés comme décrit dans Lurin et al. 2004.

- ARN-seq

Les ARN totaux ont été extraits avec le RNeasy Mini Kit (Qiagen) et leur qualité a été validée en utilisant le RNA 6000 Nano Chip (Agilent). Les ADNc ont été synthétisés en utilisant un un oligodT et Les banques ont été réalisées et séquencées en utilisant la technologie GA IIX (Illumina). Suite au test de qualité, environs 70 000 000 de lectures de 51 nucléotides par banque ont été alignées sur le génome de référence d'*Arabidopsis thaliana* (TAIR10) en utilisant Bowtie2. L'abondance en ARNm pour chaque gene/ET (comptage) a été estimée comme la moyenne de position alignées sur la région transcrite par rapport au nombre total de positions couvertes. Pour la comparaison entre plusieurs banques, le nombre de lectures a été normalisé par rapport à la taille de chaque banque en utilisant l'approche DESeq (http://www.ncbi.nlm.nih.gov/pubmed/20979621). L'expression différentielle correspond au log du ratio du comptage pour le mutant (ou l'epiRIL) par rapport au comptage pour le sauvage.

- RT-qPCR

Les ARN ont été extraits avec le RNeasy Mini Kit (Qiagen). La RT a été réalisée sur 1ug d'ARN en utilisant un oligodT et la Superscript II (Invitrogen). La qPCR a été réalisée comme décrit plus haut. L'expression des gènes testés est exprimée relativement par rapport à l'expression de trois gènes contrôles (*At2g36060*; *At4g29130*; *At5g13440*) ou du gène de l'actine (quand c'est précisé).

Les amorces utilisées pour les expériences de RT-qPCR sont listées dans Mari-Ordonez et al. 2013 (partie 3.1) et dans le manuscrit Teixeira et al. (partie 4.2)

6.5 Analyse des siARN

L'extraction, la purification sur gel et le clonage des petits ARN ont étés réalisés comme décrit précédemment (Pfeffer et al. 2005) à partir de 200 µg d'ARN totaux. Les banques ont été séquencées en utilisant la technologie Hi-seq 2000 (Illumina) pour le sauvage et le mutant *ddm1* et GA IIX (Illumina) pour le double mutant *ddm1rdr2.* L'alignement sur le génome d'*Arabidopsis thaliana* a été effectué en utilisant le logiciel MUMmer (Kurtz et al. 2004).

6.6 Etude de la présence/excision des transposons à ADN

- Les méthodes et amorces PCR utilisées pour tester la présence et l'absence d'*AT2TE42810* au site donneur sont détaillés dans Fu et al. 2013 (annexe 3)

- La présence d'*AT2TE20205* au locus donneur a été testée par PCR en utilisant les amorces suivantes : 5' GAAGCATTGGAGCTGGTTTC 3' et 5' TAAGTGTTGGCGCTGAAGTG 3'

Etant donné que qu'*AT2TE20205* est localisé dans une séquence répétée au locus donneur, son excision ne fait que reconstruire une séquence qui existe déjà ailleurs dans le génome (à l'exception de l'empreinte laissée par l'excision). Il n'est donc pas possible de tester positivement son absence par PCR en utilisant des amorces situées dans les régions flanquantes.

- La présence/absence de l'insertion partagée d'*AT2TE20205* (Chr4 : 1648255..16482638) a été testée par PCR en utilisant les amorces suivantes : 5' CTTGCTCCAACCTCGTTAGC 3' et 5' GCTGTATTCACGTGGACGAA 3'

6.6 Mutagénèse EMS

Les graines ont été incubées dans une solution de 0,3% EMS (SIGMA) pendant huit heures et rincées trois fois à l'eau.

6.7 Génotypage

Le génotypage de la mutation *ddm1-2* est effectué par amplification de la région du gène *DDM1* contenant la mutation à l'aide des amorces :

dCAPS-Col-ddm1_F 5'-ACG AAG CAA CCA AGG AAG AA -3'

dCAPS_Col-ddm1_R 5'-GAG CCA TGG GTT TGT GAA ACG TA -3'

Le produit de la PCR est ensuite digéré par l'enzyme de restriction RsaI (Invitrogen) de 2H à 37°C et analysé par migration sur gel 2,5% d'agarose.

Le génotypage des candidats du crible pour l'élimination des contaminants sauvages a été réalisé sur la base de la présence d'une nouvelle insertion d'*ATCOPIA93* dans les lignées 60 et 98 testée par PCR en utilisant les amorces :

5TE20395i1LB_F 5-atgctgcaacctctgaacct-3'
5TE20395_LB_R 5'-caagcacaaacggactgatg-3'
5TE20395i1RB_R 5'-tctcatgtgctcttggttgg-3'

Présence de l'insertion =5TE20395i1LB_F + 5TE20395_LB_R

Absence de l'insertion = 5TE20395i1LB_F + 5TE20395i1RB_R

Références

AGI (2000). "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*." Nature **408**(6814): 796-815.

Ahmed, I., A. Sarazin, C. Bowler, V. Colot and H. Quesneville (2011). "Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis." Nucleic acids research **39**(16): 6919-6931.

Aye, M., S. L. Dildine, J. A. Claypool, S. Jourdain and S. B. Sandmeyer (2001). "A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration." Mol Cell Biol **21**(22): 7839-7851.

Barkan, A. and R. A. Martienssen (1991). "Inactivation of maize transposon *Mu* suppresses a mutant phenotype by activating an outward-reading promoter near the end of *Mu1*." Proc Natl Acad Sci U S A **88**(8): 3502-3506.

Baulcombe, D. C. (1996). "RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants." Plant Mol Biol **32**(1-2): 79-88.

Bennetzen, J. L., K. Schrick, P. S. Springer, W. E. Brown and P. SanMiguel (1994). "Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA." Genome **37**(4): 565-576.

Bernatavichute, Y. V., X. Zhang, S. Cokus, M. Pellegrini and S. E. Jacobsen (2008). "Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in <italic>Arabidopsis thaliana</italic>." PLoS ONE **3**(9): e3156.

Berretta, J., M. Pinskaya and A. Morillon (2008). "A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae." Genes Dev **22**(5): 615-626.

Bostick, M., J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan and S. E. Jacobsen (2007). "UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells." Science **317**(5845): 1760-1764.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.

Brabbs, T. R., Z. He, K. Hogg, A. Kamenski, Y. Li, K. H. Paszkiewicz, K. A. Moore, P. O'Toole, I. A. Graham and L. Jones (2013). "The stochastic silencing phenotype of Arabidopsis morc6mutants reveals a role in efficient RNA-directed DNAmethylation." The Plant Journal: n/a-n/a.

Brosnan, C. A. and O. Voinnet (2011). "Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications." Current Opinion in Plant Biology **14**(5): 580-587.

Bucheton, A. (1990). "I transposable elements and I-R hybrid dysgenesis in Drosophila." Trends Genet **6**(1): 16-21.

Buchmann, R. C., S. Asad, J. N. Wolf, G. Mohannath and D. M. Bisaro (2009). "Geminivirus AL2 and L2 Proteins Suppress Transcriptional Gene Silencing and Cause Genome-Wide Reductions in Cytosine Methylation." Journal of Virology **83**(10): 5005-5013.

Buisine, N., H. Quesneville and V. Colot (2008). "Improved detection and annotation of transposable elements in sequenced genomes using multiple reference sequence sets." Genomics **91**(5): 467-475.

Bureau, T. E. and S. R. Wessler (1994). "Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants." Plant Cell **6**(6): 907-916.

Calarco, J. P., F. Borges, M. T. Donoghue, F. Van Ex, P. E. Jullien, T. Lopes, R. Gardner, F. Berger, J. A. Feijo, J. D. Becker and R. A. Martienssen (2012). "Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA." Cell **151**(1): 194-205.

Cao, J., K. Schneeberger, S. Ossowski, T. Gunther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, X. Wang, F. Ott, J. Muller, C. Alonso-Blanco, K. Borgwardt, K. J. Schmid and D. Weigel (2011). "Whole-genome sequencing of multiple Arabidopsis thaliana populations." Nat Genet **43**(10): 956-963.

Cao, X., W. Aufsatz, D. Zilberman, M. F. Mette, M. S. Huang, M. Matzke and S. E. Jacobsen (2003). "Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation." Curr Biol **13**(24): 2212-2217.

Cao, X. and S. E. Jacobsen (2002). "Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes." Proc Natl Acad Sci U S A **99 Suppl 4**: 16491-16498.

Cao, X. and S. E. Jacobsen (2002). "Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing." Curr Biol **12**(13): 1138-1144.

Cao, X., N. M. Springer, M. G. Muszynski, R. L. Phillips, S. Kaeppler and S. E. Jacobsen (2000). "Conserved plant genes with similarity to mammalian de novo DNA methyltransferases." Proc Natl Acad Sci U S A **97**(9): 4979-4984.

Castro, J. P. and C. M. Carareto (2004). "Drosophila melanogaster P transposable elements: mechanisms of transposition and regulation." Genetica **121**(2): 107-118.

Chaboissier, M.-C., A. Bucheton and D. J. Finnegan (1998). "Copy number control of a transposable element, the I factor, a LINE-like element in Drosophila." Proceedings of the National Academy of Sciences **95**(20): 11781-11785.

Chan, S. W., D. Zilberman, Z. Xie, L. K. Johansen, J. C. Carrington and S. E. Jacobsen (2004). "RNA silencing genes control de novo DNA methylation." Science **303**(5662): 1336.

Choi, Y., M. Gehring, L. Johnson, M. Hannon, J. J. Harada, R. B. Goldberg, S. E. Jacobsen and R. L. Fischer (2002). "DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis." Cell **110**(1): 33-42.

Cokus, S. J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini and S. E. Jacobsen (2008). "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." Nature **452**(7184): 215-219.

Colome-Tatche, M., S. Cortijo, R. Wardenaar, L. Morgado, B. Lahouze, A. Sarazin, M. Etcheverry, A. Martin, S. Feng, E. Duvernois-Berthet, K. Labadie, P. Wincker, S. E. Jacobsen, R. C. Jansen, V. Colot and F. Johannes (2012). "Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation." Proceedings of the National Academy of Sciences of the United States of America **109**(40): 16240-16245.

Colot, V., L. Maloisel and J. L. Rossignol (1996). "Interchromosomal transfer of epigenetic states in *Ascobolus*: transfer of DNA methylation is mechanistically related to homologous recombination." Cell **86**(6): 855-864.

Congdon, L. M., S. I. Houston, C. S. Veerappan, T. M. Spektor and J. C. Rice (2010). "PR-Set7-mediated monomethylation of histone H4 lysine 20 at specific genomic regions induces transcriptional repression." Journal of Cellular Biochemistry **110**(3): 609-619.

Cost, G. J., Q. Feng, A. Jacquier and J. D. Boeke (2002). "Human L1 element target-primed reverse transcription in vitro." EMBO J **21**(21): 5899-5910.

Deleris, A., M. V. C. Greenberg, I. Ausin, R. W. Y. Law, G. Moissiard, D. Schubert and S. E. Jacobsen (2010). "Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation." EMBO Rep **11**(12): 950-955.

Deleris, A., H. Stroud, Y. Bernatavichute, E. Johnson, G. Klein, D. Schubert and S. E. Jacobsen (2012). "Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in <italic>Arabidopsis thaliana</italic>." PLoS Genet **8**(11): e1003062.

Deragon, J.-M. and X. Zhang (2006). "Short Interspersed Elements (SINEs) in Plants: Origin, Classification, and Use as Phylogenetic Markers." Systematic Biology **55**(6): 949-956.

Devos, K., J. Beales, Y. Ogihara and A. Doust (2005). "Comparative Sequence Analysis of the Phytochrome C Gene and its Upstream Region in Allohexaploid Wheat Reveals New Data on the Evolution of its Three Constituent Genomes." Plant Molecular Biology **58**(5): 625-641.

Dooner, H. K. and A. Belachew (1989). "Transposition Pattern of the Maize Element Ac from the Bz-M2(ac) Allele." Genetics **122**(2): 447-457.

Durán-Figueroa, N. and J.-P. Vielle-Calzada (2010). "ARGONAUTE9-dependent silencing of transposable elements in pericentromeric regions of Arabidopsis." Plant Signaling & Behavior **5**(11): 1476-1479.

Earley, K., R. J. Lawrence, O. Pontes, R. Reuther, A. J. Enciso, M. Silva, N. Neves, M. Gross, W. Viegas and C. S. Pikaard (2006). "Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance." Genes Dev **20**(10): 1283-1293.

Ebbs, M. L. and J. Bender (2006). "Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase." Plant Cell **18**(5): 1166-1176.

Elmayan, T., S. Balzergue, F. Beon, V. Bourdon, J. Daubremet, Y. Guenet, P. Mourrain, J. C. Palauqui, S. Vernhettes, T. Vialle, K. Wostrikoff and H. Vaucheret (1998). "Arabidopsis mutants impaired in cosuppression." Plant Cell **10**(10): 1747-1758.

Eun, C., Z. J. Lorkovic, U. Naumann, Q. Long, E. R. Havecker, S. A. Simon, B. C. Meyers, A. J. M. Matzke and M. Matzke (2011). "AGO6 Functions in RNA-Mediated Transcriptional Gene Silencing in Shoot and Root Meristems in <italic>Arabidopsis thaliana</italic>." PLoS ONE **6**(10): e25730.

Fedoroff, N. (1989). "The heritable activation of *cryptic Suppressor-mutator* elements by an active element." Genetics **121**: 591-608.

Fedoroff, N. V. (2013). The Discovery of Transposition. Plant Transposons and Genome Dynamics in Evolution, Wiley-Blackwell**:** 3-13.

Fedoroff, N. V. (2013). McClintock and Epigenetics. Plant Transposons and Genome Dynamics in Evolution, Wiley-Blackwell**:** 61-70.

Feschotte, C., N. Jiang and S. Wessler (2002). "Plant transposable elements: where genetics meets genomics." Nat Rev Genet **3**: 329 - 341.

Fojtova, M., H. Van Houdt, A. Depicker and A. Kovarik (2003). "Epigenetic switch from posttranscriptional to transcriptional silencing is correlated with promoter hypermethylation." Plant Physiol **133**(3): 1240-1250.

Fu, Y., A. Kawabe, M. Etcheverry, T. Ito, A. Toyoda, A. Fujiyama, V. Colot, Y. Tarutani and T. Kakutani (2013). "Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor." EMBO J **advance online publication**.

Fuchs, J., D. Demidov, A. Houben and I. Schubert (2006). "Chromosomal histone modification patterns – from conservation to diversity." Trends in Plant Science **11**(4): 199-208.

Gao, X., Y. Hou, H. Ebina, H. L. Levin and D. F. Voytas (2008). "Chromodomains direct integration of retrotransposons to heterochromatin." Genome Res **18**(3): 359-369.

Gehring, M., K. L. Bubb and S. Henikoff (2009). "Extensive demethylation of repetitive elements during seed development underlies gene imprinting." Science **324**(5933): 1447-1451.

Gehring, M., W. Reik and S. Henikoff (2009). "DNA demethylation by DNA repair." Trends in genetics : TIG **25**(2): 82-90.

Giraut, L., M. Falque, J. Drouaud, L. Pereira, O. C. Martin and C. Mézard (2011). "Genome-Wide Crossover Distribution in <italic>Arabidopsis thaliana</italic> Meiosis Reveals Sex-Specific Patterns along Chromosomes." PLoS Genet **7**(11): e1002354.

Goll, M. G. and T. H. Bestor (2005). "EUKARYOTIC CYTOSINE METHYLTRANSFERASES." Annual Review of Biochemistry **74**(1): 481-514.

Goodrich, D. W. and P. H. Duesberg (1990). "Retroviral recombination during reverse transcription." Proceedings of the National Academy of Sciences **87**(6): 2052-2056.

Grandbastien, M. A., C. Audeon, E. Bonnivard, J. M. Casacuberta, B. Chalhoub, A. P. Costa, Q. H. Le, D. Melayah, M. Petit, C. Poncet, S. M. Tam, M. A. Van Sluys and C. Mhiri (2005). "Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae." Cytogenet Genome Res **110**(1-4): 229-241.

Grandbastien, M. A., C. Audeon, J. M. Casacuberta, P. Grappin, H. Lucas, C. Moreau and S. Pouteau (1994). "Functional analysis of the tobacco Tnt1 retrotransposon." Genetica **93**(1-3): 181-189.

Greenberg, M. V. C., I. Ausin, S. W. L. Chan, S. J. Cokus, J. T. Cuperus, S. Feng, J. A. Law, C. Chu, M. Pellegrini, J. C. Carrington and S. E. Jacobsen (2011). "Identification of genes required for de novo DNA methylation in Arabidopsis." Epigenetics **6**(3): 344-354.

Greenblatt, I. M. (1984). "A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element, modulator, in maize." Genetics **108**(2): 471-485.

Grimanelli, D. and F. Roudier (2013). "Epigenetics and development in plants: green light to convergent innovations." Curr Top Dev Biol **104**: 189-222.

Habu, Y. (2010). "Epigenetic silencing of endogenous repetitive sequences by MORPHEUS' MOLECULE1 in Arabidopsis thaliana." Epigenetics **5**(7): 562-565.

Havecker, E., X. Gao and D. Voytas (2004). "The diversity of LTR retrotransposons." Genome Biology **5**(6): 225.

Havecker, E. R., L. M. Wallbridge, T. J. Hardcastle, M. S. Bush, K. A. Kelly, R. M. Dunn, F. Schwach, J. H. Doonan and D. C. Baulcombe (2010). "The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci." Plant Cell **22**(2): 321-334.

He, X.-J., Y.-F. Hsu, S. Zhu, A. T. Wierzbicki, O. Pontes, C. S. Pikaard, H.-L. Liu, C.-S. Wang, H. Jin and J.-K. Zhu (2009). "An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4- and RNA-Binding Protein." Cell **137**(3): 498-508.

Heinlem, M., T. Brattig and R. Kunze (1994). "In vivo aggregation of maize Activator (Ac) transposase in nuclei of maize endosperm and Petunia protoplasts." The Plant Journal **5**(5): 705-714.

Henderson, I. R. and S. E. Jacobsen (2007). "Epigenetic inheritance in plants." Nature **447**(7143): 418-424.

Henderson, I. R. and S. E. Jacobsen (2008). "Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading." Genes Dev **22**(12): 1597-1606.

Henikoff, S. and L. Comai (1998). "A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis." Genetics **149**(1): 307-318.

Hirochika, H., H. Okamoto and T. Kakutani (2000). "Silencing of retrotransposons in Arabidopsis and reactivation by the *ddm1* mutation." Plant Cell **12**(3): 357-369.

Hollister, J. D. and B. S. Gaut (2007). "Population and Evolutionary Dynamics of Helitron Transposable Elements in Arabidopsis thaliana." Molecular Biology and Evolution **24**(11): 2515-2524.

Hollister, J. D. and B. S. Gaut (2009). "Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression." Genome Research **19**(8): 1419-1428.

Hollister, J. D., L. M. Smith, Y.-L. Guo, F. Ott, D. Weigel and B. S. Gaut (2011). "Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata." Proceedings of the National Academy of Sciences **108**(6): 2322-2327.

Hsieh, T. F., C. A. Ibarra, P. Silva, A. Zemach, L. Eshed-Williams, R. L. Fischer and D. Zilberman (2009). "Genome-wide demethylation of Arabidopsis endosperm." Science **324**(5933): 1451-1454.

Hu, T. T., P. Pattyn, E. G. Bakker, J. Cao, J.-F. Cheng, R. M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haberer, J. D. Hollister, S. Ossowski, R. P. Ottilar, A. A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J. C. Carrington, B. S. Gaut, J. Schmutz, K. F. X. Mayer, Y. Van de Peer, I. V. Grigoriev, M. Nordborg, D. Weigel and Y.-L. Guo (2011). "The Arabidopsis lyrata genome sequence and the basis of rapid genome size change." Nat Genet **43**(5): 476-481.

Ibarra, C. A., X. Feng, V. K. Schoft, T. F. Hsieh, R. Uzawa, J. A. Rodrigues, A. Zemach, N. Chumak, A. Machlicova, T. Nishimura, D. Rojas, R. L. Fischer, H. Tamaru and D. Zilberman (2012). "Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes." Science **337**(6100): 1360-1364.

Inagaki, S., A. Miura-Kamio, Y. Nakamura, F. Lu, X. Cui, X. Cao, H. Kimura, H. Saze and T. Kakutani (2010). "Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome." EMBO J **29**(20): 3496-3506.

Ito, H., H. Gaubert, E. Bucher, M. Mirouze, I. Vaillant and J. Paszkowski (2011). "An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress." Nature **472**(7341): 115-119.

Ivashuta, S., M. Naumkina, M. Gau, K. Uchiyama, S. Isobe, Y. Mizukami and Y. Shimamoto (2002). "Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (Medicago sativa)." The Plant Journal **31**(5): 615-627.

Jackson, J. P., A. M. Lindroth, X. Cao and S. E. Jacobsen (2002). "Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase." Nature **416**(6880): 556-560.

Jacob, Y., S. Feng, C. A. LeBlanc, Y. V. Bernatavichute, H. Stroud, S. Cokus, L. M. Johnson, M. Pellegrini, S. E. Jacobsen and S. D. Michaels (2009). "ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing." Nat Struct Mol Biol **16**(7): 763-768.

Jacob, Y., H. Stroud, C. LeBlanc, S. Feng, L. Zhuo, E. Caro, C. Hassel, C. Gutierrez, S. D. Michaels and S. E. Jacobsen (2010). "Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases." Nature **466**(7309): 987-991.

Jauvion, V., M. Rivard, N. Bouteiller, T. Elmayan and H. Vaucheret (2012). "RDR2 Partially Antagonizes the Production of RDR6-Dependent siRNA in Sense Transgene-Mediated PTGS." PLoS ONE **7**(1): e29785.

Jeddeloh, J. A., T. L. Stokes and E. J. Richards (1999). "Maintenance of genomic methylation requires a SWI2/SNF2-like protein." Nat Genet **22**(1): 94-97.

Johannes, F., E. Porcher, F. K. Teixeira, V. Saliba-Colombani, M. Simon, N. Agier, A. Bulski, J. Albuisson, F. Heredia, P. Audigier, D. Bouchez, C. Dillmann, P. Guerche, F. Hospital and V. Colot (2009). "Assessing the impact of transgenerational epigenetic variation on complex traits." PLoS Genet **5**(6): e1000530.

Johnson, L. M., M. Bostick, X. Zhang, E. Kraft, I. Henderson, J. Callis and S. E. Jacobsen (2007). "The SRA methyl-cytosine-binding domain links DNA and histone methylation." Curr Biol **17**(4): 379-384.

Jordan, I. K. and J. F. McDonald (1998). "Evidence for the role of recombination in the regulatory evolution of Saccharomyces cerevisiae Ty elements." J Mol Evol **47**(1): 14-20.

Jullien, P. E., A. Mosquna, M. Ingouff, T. Sakata, N. Ohad and F. Berger (2008). "Retinoblastoma and Its Binding Partner MSI1 Control Imprinting in <italic>Arabidopsis</italic>." PLoS Biol **6**(8): e194.

Jullien, P. E., D. Susaki, R. Yelagandula, T. Higashiyama and F. Berger (2012). "DNA Methylation Dynamics during Sexual Reproduction in Arabidopsis thaliana." Current biology : CB **22**(19): 1825-1830.

Jurka, J. (2000). "Repbase update: a database and an electronic journal of repetitive elements." Trends Genet **16**(9): 418-420.

Kakutani, T. (1997). "Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in *Arabidopsis thaliana*." Plant J **12**(6): 1447-1451.

Kakutani, T., J. A. Jeddeloh, S. K. Flowers, K. Munakata and E. J. Richards (1996). "Developmental abnormalities and epimutations associated with DNA hypomethylation mutations." Proc Natl Acad Sci U S A **93**(22): 12406-12411.

Kakutani, T., J. A. Jeddeloh and E. J. Richards (1995). "Characterization of an *Arabidopsis thaliana* DNA hypomethylation mutant." Nucleic Acids Res **23**(1): 130-137.

Kakutani, T., K. Munakata, E. J. Richards and H. Hirochika (1999). "Meiotically and mitotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of *Arabidopsis thaliana*." Genetics **151**(2): 831-838.

Kapitonov, V. and J. Jurka (1999). "Molecular paleontology of transposable elements from Arabidopsis thaliana." Genetica **107**: 27 - 37.

Kapitonov, V. V. and J. Jurka (2001). "Rolling-circle transposons in eukaryotes." Proceedings of the National Academy of Sciences **98**(15): 8714-8719.

Kapitonov, V. V. and J. Jurka (2006). "Self-synthesizing DNA transposons in eukaryotes." Proceedings of the National Academy of Sciences of the United States of America **103**(12): 4540-4545.

Kapitonov, V. V. and J. Jurka (2007). "Helitrons on a roll: eukaryotic rolling-circle transposons." Trends in Genetics **23**(10): 521-529.

Kashkush, K. and V. Khasdan (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**(4): 1975-1985.

Kazazian, H. H. (2004). "Mobile Elements: Drivers of Genome Evolution." Science **303**(5664): 1626-1632.

Kinoshita, T., R. Yadegari, J. J. Harada, R. B. Goldberg and R. L. Fischer (1999). "Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm." Plant Cell **11**(10): 1945-1952.

Kunze, R., U. Behrens, U. Courage-Franzkowiak, S. Feldmar, S. Kühn and R. Lütticke (1993). "Dominant transposition-deficient mutants of maize Activator (Ac) transposase." Proceedings of the National Academy of Sciences **90**(15): 7094-7098.

Kunze, R. and P. Starlinger (1989). "The putative transposase of transposable element Ac from Zea mays L. interacts with subterminal sequences of Ac." EMBO J **8**(11): 3177-3185.

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu and S. L. Salzberg (2004). "Versatile and open software for comparing large genomes." Genome Biol **5**(2): R12.

Lahmy, S., N. Bies-Etheve and T. Lagrange (2010). "Plant-specific multisubunit RNA polymerase in gene silencing." Epigenetics **5**(1): 4-8.

Lang-Mladek, C., O. Popova, K. Kiok, M. Berlinger, B. Rakic, W. Aufsatz, C. Jonak, M.-T. Hauser and C. Luschnig (2010). "Transgenerational Inheritance and Resetting of Stress-Induced Loss of Epigenetic Gene Silencing in Arabidopsis." Molecular Plant **3**(3): 594-602.

Laski, F. A., D. C. Rio and G. M. Rubin (1986). "Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing." Cell **44**(1): 7-19.

Law, J. A., J. Du, C. J. Hale, S. Feng, K. Krajewski, A. M. S. Palanca, B. D. Strahl, D. J. Patel and S. E. Jacobsen (2013). "Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1." Nature **498**(7454): 385-389.

Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nat Rev Genet **11**(3): 204-220.

Law, J. A., A. A. Vashisht, J. A. Wohlschlegel and S. E. Jacobsen (2011). "SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV." PLoS Genet **7**(7): e1002195.

Lenoir, A., L. Lavie, J.-L. Prieto, C. Goubely, J.-C. Cote, T. Pélissier and J.-M. Deragon (2001). "The Evolutionary Origin and Genomic Organization of SINEs in Arabidopsis thaliana." Molecular Biology and Evolution **18**(12): 2315-2322.

Lenoir, A., T. Pélissier, C. Bousquet-Antonelli and J. M. Deragon (2005). "Comparative evolution history of SINEs in Arabidopsis thaliana and Brassica oleracea: evidence for a high rate of SINE loss." Cytogenetic and Genome Research **110**(1-4): 441-447.

Lesage, P. and A. L. Todeschini (2005). "Happy together: the life and times of Ty retrotransposons and their hosts." Cytogenet Genome Res **110**(1-4): 70-90.

Levin, H. L. and J. V. Moran (2011). "Dynamic interactions between transposable elements and their hosts." Nat Rev Genet **12**(9): 615-627.

Lindroth, A. M., X. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum, S. Henikoff and S. E. Jacobsen (2001). "Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation." Science **292**(5524): 2077-2080.

Lindroth, A. M., D. Shultis, Z. Jasencakova, J. Fuchs, L. Johnson, D. Schubert, D. Patnaik, S. Pradhan, J. Goodrich, I. Schubert, T. Jenuwein, S. Khorasanizadeh and S. E. Jacobsen (2004). "Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3." Embo J **23**(21): 4146-4155.

Lippman, Z., A. V. Gendrel, M. Black, M. W. Vaughn, N. Dedhia, W. R. McCombie, K. Lavine, V. Mittal, B. May, K. D. Kasschau, J. C. Carrington, R. W. Doerge, V. Colot and R. Martienssen (2004). "Role of transposable elements in heterochromatin and epigenetic control." Nature **430**(6998): 471-476.

Lippman, Z., B. May, C. Yordan, T. Singer and R. Martienssen (2003). "Distinct Mechanisms Determine Transposon Inheritance and Methylation via Small Interfering RNA and Histone Modification." PLoS Biol **1**(3): E67.

Lister, R., R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar and J. R. Ecker (2008). "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." Cell **133**(3): 523-536.

Liu, S., C.-T. Yeh, T. Ji, K. Ying, H. Wu, H. M. Tang, Y. Fu, D. Nettleton and P. S. Schnable (2009). "<italic>Mu</italic> Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome." PLoS Genet **5**(11): e1000733.

Liu, X., C. W. Yu, J. Duan, M. Luo, K. Wang, G. Tian, Y. Cui and K. Wu (2012). "HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in Arabidopsis." Plant Physiol **158**(1): 119-129.

Lu, F., X. Cui, S. Zhang, C. Liu and X. Cao (2010). "JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis." Cell Res **20**(3): 387-390.

Lurin, C., C. Andres, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyere, M. Caboche, C. Debast, J. Gualberto, B. Hoffmann, A. Lecharny, M. Le Ret, M. L. Martin-Magniette, H. Mireau, N. Peeters, J. P. Renou, B. Szurek, L. Taconnat and I. Small (2004). "Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis." Plant Cell **16**(8): 2089-2103.

Malagnac, F., L. Bartee and J. Bender (2002). "An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation." Embo J **21**(24): 6842-6852.

Maloisel, L. and J. L. Rossignol (1998). "Suppression of crossing-over by DNA methylation in *Ascobolus*." Genes Dev **12**(9): 1381-1389.

Manning, K., M. Tor, M. Poole, Y. Hong, A. J. Thompson, G. J. King, J. J. Giovannoni and G. B. Seymour (2006). "A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening." Nat Genet **38**(8): 948-952.

Mari-Ordonez, A., A. Marchais, M. Etcheverry, A. Martin, V. Colot and O. Voinnet (2013). "Reconstructing de novo silencing of an active plant retrotransposon." Nat Genet **advance online publication**.

Martienssen, R., A. Barkan, W. C. Taylor and M. Freeling (1990). "Somatically heritable switches in the DNA modification of *Mu* transposable elements monitored with a suppressible mutant in maize." Genes Dev **4**(3): 331-343.

Martin, A., C. Troadec, A. Boualem, M. Rajab, R. Fernandez, H. Morin, M. Pitrat, C. Dogimont and A. Bendahmane (2009). "A transposon-induced epigenetic change leads to sex determination in melon." Nature **461**(7267): 1135-1138.

Masson, P., J. A. Banks and N. Fedoroff (1991). "Structure and function of the maize Spm transposable element." Biochimie **73**(1): 5-8.

Masson, P., G. Rutherford, J. A. Banks and N. Fedoroff (1989). "Essential large transcripts of the maize Spm transposable element are generated by alternative splicing." Cell **58**(4): 755-765.

Matsuda, E. and D. J. Garfinkel (2009). "Posttranslational interference of Ty1 retrotransposition by antisense RNAs." Proceedings of the National Academy of Sciences **106**(37): 15657-15662.

Matsunaga, W., A. Kobayashi, A. Kato and H. Ito (2012). "The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in Arabidopsis thaliana." Plant Cell Physiol **53**(5): 824-833.

Melnyk, C. W., A. Molnar, A. Bassett and D. C. Baulcombe (2011). "Mobile 24 nt small RNAs direct transcriptional gene silencing in the root meristems of Arabidopsis thaliana." Curr Biol **21**(19): 1678-1683.

Mhiri, C., J. B. Morel, S. Vernhettes, J. M. Casacuberta, H. Lucas and M. A. Grandbastien (1997). "The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress." Plant Mol Biol **33**(2): 257-266.

Mirouze, M., J. Reinders, E. Bucher, T. Nishimura, K. Schneeberger, S. Ossowski, J. Cao, D. Weigel, J. Paszkowski and O. Mathieu (2009). "Selective epigenetic control of retrotransposition in Arabidopsis." Nature **461**(7262): 427-430.

Misra, S. and D. C. Rio (1990). "Cytotype control of Drosophila P element transposition: The 66 kd protein is a repressor of transposase activity." Cell **62**(2): 269-284.

Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada and T. Kakutani (2001). "Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*." Nature **411**(6834): 212-214.

Miura, K., M. Agetsuma, H. Kitano, A. Yoshimura, M. Matsuoka, S. E. Jacobsen and M. Ashikari (2009). "A metastable DWARF1 epigenetic mutant affecting plant stature in rice." Proc Natl Acad Sci U S A **106**(27): 11218-11223.

Moissiard, G., S. J. Cokus, J. Cary, S. Feng, A. C. Billi, H. Stroud, D. Husmann, Y. Zhan, B. R. Lajoie, R. P. McCord, C. J. Hale, W. Feng, S. D. Michaels, A. R. Frand, M. Pellegrini, J. Dekker, J. K. Kim and S. E. Jacobsen (2012). "MORC family ATPases required for heterochromatin condensation and gene silencing." Science **336**(6087): 1448-1451.

Molnar, A., C. W. Melnyk, A. Bassett, T. J. Hardcastle, R. Dunn and D. C. Baulcombe (2010). "Small Silencing RNAs in Plants Are Mobile and Direct Epigenetic Modification in Recipient Cells." Science **328**(5980): 872-875.

Mosher, R. A., C. W. Melnyk, K. A. Kelly, R. M. Dunn, D. J. Studholme and D. C. Baulcombe (2009). "Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis." Nature **460**(7252): 283-286.

Mourrain, P., C. Beclin, T. Elmayan, F. Feuerbach, C. Godon, J. B. Morel, D. Jouette, A. M. Lacombe, S. Nikic, N. Picault, K. Remoue, M. Sanial, T. A. Vo and H. Vaucheret (2000). "Arabidopsis *SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance." Cell **101**(5): 533-542.

Naito, K., F. Zhang, T. Tsukiyama, H. Saito, C. N. Hancock, A. O. Richardson, Y. Okumoto, T. Tanisaka and S. R. Wessler (2009). "Unexpected consequences of a sudden and massive transposon amplification on rice gene expression." Nature **461**(7267): 1130-1134.

Naumann, K., A. Fischer, I. Hofmann, V. Krauss, S. Phalke, K. Irmler, G. Hause, A. C. Aurich, R. Dorn, T. Jenuwein and G. Reuter (2005). "Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis." Embo J **24**(7): 1418-1429.

Numa, H., J.-M. Kim, A. Matsui, Y. Kurihara, T. Morosawa, J. Ishida, Y. Mochizuki, H. Kimura, K. Shinozaki, T. Toyoda, M. Seki, M. Yoshikawa and Y. Habu (2010). "Transduction of RNA-directed DNA methylation signals to repressive histone marks in Arabidopsis thaliana." EMBO J **29**(2): 352-362.

Ogiwara, I., M. Miya, K. Ohshima and N. Okada (1999). "Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs." Molecular Biology and Evolution **16**(9): 1238-1250.

Olmedo-Monfil, V., N. Duran-Figueroa, M. Arteaga-Vazquez, E. Demesa-Arevalo, D. Autran, D. Grimanelli, R. K. Slotkin, R. A. Martienssen and J.-P. Vielle-Calzada (2010). "Control of female gamete formation by a small RNA pathway in Arabidopsis." Nature **464**(7288): 628-632.

Paterson, A. H., J. E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A. K. Bharti, J. Chapman, F. A. Feltus, U. Gowik, I. V. Grigoriev, E. Lyons, C. A. Maher, M. Martis, A. Narechania, R. P. Otillar, B. W. Penning, A. A. Salamov, Y. Wang, L. Zhang, N. C. Carpita, M. Freeling, A. R. Gingle, C. T. Hash, B. Keller, P. Klein, S. Kresovich, M. C. McCann, R. Ming, D. G. Peterson, R. Mehboob ur, D. Ware, P. Westhoff, K. F. X. Mayer, J. Messing and D. S. Rokhsar (2009). "The Sorghum bicolor genome and the diversification of grasses." Nature **457**(7229): 551-556.

Pavet, V., C. Quintero, N. M. Cecchini, A. L. Rosa and M. E. Alvarez (2006). "Arabidopsis displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by pseudomonas syringae." Mol Plant Microbe Interact **19**(6): 577-587.

Pecinka, A., H. Q. Dinh, T. Baubec, M. Rosa, N. Lettner and O. Mittelsten Scheid (2010). "Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis." The Plant cell **22**(9): 3118-3129.

Penterman, J., R. Uzawa and R. L. Fischer (2007). "Genetic interactions between DNA demethylation and methylation in Arabidopsis." Plant Physiol **145**(4): 1549-1557.

Peterson-Burch, B., D. Nettleton and D. Voytas (2004). "Genomic neighborhoods for Arabidopsis retrotransposons: a role for targeted integration in the distribution of the Metaviridae." Genome Biology **5**(10): R78.

Pfeffer, S., M. Lagos-Quintana and T. Tuschl (2005). Cloning of Small RNA Molecules. Current Protocols in Molecular Biology, John Wiley & Sons, Inc.

Pouteau, S., M.-A. Grandbastien and M. Boccara (1994). "Microbial elicitors of plant defence responses activate transcription of a retrotransposon." The Plant Journal **5**(4): 535-542.

Pritham, E. J., T. Putliwala and C. Feschotte (2007). "Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses." Gene **390**(1–2): 3-17.

Raina, R., M. Schlappi and N. Fedoroff (1998). "Epigenetic mechanisms in the regulation of the maize Suppressor-mutator transposon." Novartis Found Symp **214**: 133-140.

Reinders, J., B. B. Wulff, M. Mirouze, A. Mari-Ordonez, M. Dapp, W. Rozhon, E. Bucher, G. Theiler and J. Paszkowski (2009). "Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes." Genes Dev **23**(8): 939-950.

Ros, F. and R. Kunze (2001). "Regulation of Activator/Dissociation Transposition by Replication and DNA Methylation." Genetics **157**(4): 1723-1733.

Roudier, F., I. Ahmed, C. Berard, A. Sarazin, T. Mary-Huard, S. Cortijo, D. Bouyer, E. Caillieux, E. Duvernois-Berthet, L. Al-Shikhley, L. Giraut, B. Despres, S. Drevensek, F. Barneche, S. Derozier, V. Brunaud, S. Aubourg, A. Schnittger, C. Bowler, M. L. Martin-Magniette, S. Robin, M. Caboche and V. Colot (2011). "Integrative epigenomic mapping defines four main chromatin states in Arabidopsis." The EMBO journal **30**(10): 1928-1938.

Sabot, F. and A. H. Schulman (2006). "Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome." Heredity **97**(6): 381-388.

Saze, H. and T. Kakutani (2007). "Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1." EMBO J **26**(15): 3641-3652.

Saze, H. and T. Kakutani (2011). "Differentiation of epigenetic modifications between transposons and genes." Current Opinion in Plant Biology **14**(1): 81-87.

Scheid, O. M., A. V. Probst, K. Afsar and J. Paszkowski (2002). "Two regulatory levels of transcriptional gene silencing in Arabidopsis." Proceedings of the National Academy of Sciences **99**(21): 13659-13662.

Schläppi, M., R. Raina and N. Fedoroff (1994). "Epigenetic regulation of the maize Spm transposable element: Novel activation of a methylated promoter by TnpA." Cell **77**(3): 427-437.

Schlappi, M., D. Smith and N. Fedoroff (1993). "*TnpA* trans-activates methylated maize *Suppressor-mutator* transposable elements in transgenic tobacco." Genetics **133**(4): 1009-1021.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T. A. Graves, P. Minx, A. D. Reily, L. Courtney, S. S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S. M. Rock, E. Belter, F. Du, K. Kim, R. M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S. M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M. J. Levy, L. McMahan, P. Van Buren, M. W. Vaughn, K. Ying, C.-T. Yeh, S. J. Emrich, Y. Jia, A. Kalyanaraman, A.-P. Hsia, W. B. Barbazuk, R. S. Baucom, T. P. Brutnell, N. C. Carpita, C. Chaparro, J.-M. Chia, J.-M. Deragon, J. C. Estill, Y. Fu, J. A. Jeddeloh, Y. Han, H. Lee, P. Li, D. R. Lisch, S. Liu, Z. Liu, D. H. Nagel, M. C. McCann, P. SanMiguel, A. M. Myers, D. Nettleton, J. Nguyen, B. W. Penning, L. Ponnala, K. L. Schneider, D. C. Schwartz, A. Sharma, C. Soderlund, N. M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T. K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J. L. Bennetzen, R. K. Dawe, J. Jiang, N. Jiang, G. G. Presting, S. R. Wessler, S. Aluru, R. A. Martienssen, S. W. Clifton, W. R. McCombie, R. A. Wing and R. K. Wilson (2009). "The B73 Maize Genome: Complexity, Diversity, and Dynamics." Science **326**(5956): 1112-1115.

Scofield, S. R., J. J. English and J. D. G. Jones (1993). "High level expression of the Activator transposase gene inhibits the excision of Dissociation in tobacco cotyledons." Cell **75**(3): 507-517.

Searle, I. R., O. Pontes, C. W. Melnyk, L. M. Smith and D. C. Baulcombe (2010). "JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in Arabidopsis." Genes & Development **24**(10): 986-991.

Selinger, D. A. and V. L. Chandler (2001). "*B-Bolivia*, an allele of the maize *b1* gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream." Plant Physiol **125**(3): 1363-1379.

Sharif, J., M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano and H. Koseki (2007). "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." Nature **450**(7171): 908-912.

Singer, T., C. Yordan and R. A. Martienssen (2001). "Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation (DDM1)*." Genes Dev **15**(5): 591-602.

Singh, M., S. Goel, R. B. Meeley, C. Dantec, H. Parrinello, C. Michaud, O. Leblanc and D. Grimanelli (2011). "Production of viable gametes without meiosis in maize deficient for an ARGONAUTE protein." Plant Cell **23**(2): 443-458.

Slotkin, R. K., M. Freeling and D. Lisch (2003). "Mu killer causes the heritable inactivation of the Mutator family of transposable elements in Zea mays." Genetics **165**(2): 781-797.

Slotkin, R. K., M. Freeling and D. Lisch (2005). "Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication." Nat Genet **37**(6): 641-644.

Slotkin, R. K., M. Vaughn, F. Borges, M. Tanurdzic, J. D. Becker, J. A. Feijo and R. A. Martienssen (2009). "Epigenetic reprogramming and small RNA silencing of transposable elements in pollen." Cell **136**(3): 461-472.

Soppe, W. J., S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef and A. J. Peeters (2000). "The late flowering phenotype of *fwa* mutants is caused by gain-of- function epigenetic alleles of a homeodomain gene." Mol Cell **6**(4): 791-802.

Soppe, W. J., Z. Jasencakova, A. Houben, T. Kakutani, A. Meister, M. S. Huang, S. E. Jacobsen, I. Schubert and P. F. Fransz (2002). "DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis." Embo J **21**(23): 6549-6559.

Spillane, C., C. Baroux, J. M. Escobar-Restrepo, D. R. Page, S. Laoueille and U. Grossniklaus (2004). "Transposons and tandem repeats are not involved in the control of genomic imprinting at the MEDEA locus in Arabidopsis." Cold Spring Harb Symp Quant Biol **69**: 465-475.

Stroud, H., Maxim V. C. Greenberg, S. Feng, Yana V. Bernatavichute and Steven E. Jacobsen (2013). "Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome." Cell **152**(1–2): 352-364.

Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-476.

Tariq, M., Y. Habu and J. Paszkowski (2002). "Depletion of MOM1 in non-dividing cells of Arabidopsis plants releases transcriptional gene silencing." EMBO Rep **3**(10): 951-955.

Teixeira, F. K. and V. Colot (2010). "Repeat elements and the Arabidopsis DNA methylation landscape." Heredity **105**(1): 14-23.

Teixeira, F. K., F. Heredia, A. Sarazin, F. Roudier, M. Boccara, C. Ciaudo, C. Cruaud, J. Poulain, M. Berdasco, M. F. Fraga, O. Voinnet, P. Wincker, M. Esteller and V. Colot (2009). "A Role for RNAi in the Selective Correction of DNA Methylation Defects." Science **323**: 1600-1604.

Tessadori, F., R. K. Schulkes, R. v. Driel and P. Fransz (2007). "Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis." The Plant Journal **50**(5): 848-857.

Tessadori, F., M. van Zanten, P. Pavlova, R. Clifton, F. Pontvianne, L. B. Snoek, F. F. Millenaar, R. K. Schulkes, R. van Driel, L. A. C. J. Voesenek, C. Spillane, C. S. Pikaard, P. Fransz and A. J. M. Peeters (2009). "<italic>PHYTOCHROME B</italic> and <italic>HISTONE DEACETYLASE 6</italic> Control Light-Induced Chromatin Compaction in <italic>Arabidopsis thaliana</italic>." PLoS Genet **5**(9): e1000638.

Tittel-Elmer, M., E. Bucher, L. Broger, O. Mathieu, J. Paszkowski and I. Vaillant (2010). "Stress-Induced Activation of Heterochromatic Transcription." PLoS Genet **6**(10): e1001175.

To, T. K., J.-M. Kim, A. Matsui, Y. Kurihara, T. Morosawa, J. Ishida, M. Tanaka, T. Endo, T. Kakutani, T. Toyoda, H. Kimura, S. Yokoyama, K. Shinozaki and M. Seki (2011). "<italic>Arabidopsis</italic> HDA6 Regulates Locus-Directed Heterochromatin Silencing in Cooperation with MET1." PLoS Genet **7**(4): e1002055.

Tran, R. K., D. Zilberman, C. de Bustos, R. F. Ditt, J. G. Henikoff, A. M. Lindroth, J. Delrow, T. Boyle, S. Kwong, T. D. Bryson, S. E. Jacobsen and S. Henikoff (2005). "Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis." Genome Biol **6**(11): R90.

Tsukahara, S., A. Kawabe, A. Kobayashi, T. Ito, T. Aizu, T. Shin-i, A. Toyoda, A. Fujiyama, Y. Tarutani and T. Kakutani (2012). "Centromere-targeted de novo integrations of an LTR retrotransposon of Arabidopsis lyrata." Genes Dev **26**(7): 705-713.

Tsukahara, S., A. Kobayashi, A. Kawabe, O. Mathieu, A. Miura and T. Kakutani (2009). "Bursts of retrotransposition reproduced in Arabidopsis." Nature **461**(7262): 423-426.

Vaillant, I., I. Schubert, S. Tourmente and O. Mathieu (2006). "MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis." EMBO Rep **7**(12): 1273-1278.

van Zanten, M., M. A. Koini, R. Geyer, Y. Liu, V. Brambilla, D. Bartels, M. Koornneef, P. Fransz and W. J. Soppe (2011). "Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation." Proc Natl Acad Sci U S A **108**(50): 20219-20224.

van Zanten, M., F. Tessadori, L. Bossen, A. J. Peeters and P. Fransz (2010). "Large-scale chromatin de-compaction induced by low light is not accompanied by nucleosomal displacement." Plant Signal Behav **5**(12): 1677-1678.

Vaucheret, H. (2008). "Plant ARGONAUTES." Trends in Plant Science **13**(7): 350-358.

Vaughn, M. W., M. Tanurd Ic, Z. Lippman, H. Jiang, R. Carrasquillo, P. D. Rabinowicz, N. Dedhia, W. R. McCombie, N. Agier, A. Bulski, V. Colot, R. W. Doerge and R. A. Martienssen (2007). "Epigenetic Natural Variation in Arabidopsis thaliana." PLoS Biol **5**(7): e174.

Vieira, C., M. Fablet, E. Lerat, M. Boulesteix, R. Rebollo, N. Burlet, A. Akkouche, B. Hubert, H. Mortada and C. Biémont (2012). "A comparative analysis of the amounts and dynamics of transposable elements in natural populations of Drosophila melanogaster and Drosophila simulans." Journal of Environmental Radioactivity **113**(0): 83-86.

Vongs, A., T. Kakutani, R. A. Martienssen and E. J. Richards (1993). "*Arabidopsis thaliana* DNA methylation mutants." Science **260**(5116): 1926-1928.

Wang, L., M. Heinlein and R. Kunze (1996). "Methylation pattern of Activator transposase binding sites in maize endosperm." Plant Cell **8**(4): 747-758.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." Nat Rev Genet **8**(12): 973-982.

Wicker, T., N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf and B. Keller (2001). "Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution." The Plant Journal **26**(3): 307-316.

Wierzbicki, A. T., J. R. Haag and C. S. Pikaard (2008). "Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes." Cell **135**(4): 635-648.

Wierzbicki, A. T., T. S. Ream, J. R. Haag and C. S. Pikaard (2009). "RNA polymerase V transcription guides ARGONAUTE4 to chromatin." Nat Genet **41**(5): 630-634.

Wolff, P., I. Weinhofer, J. Seguin, P. Roszak, C. Beisel, M. T. A. Donoghue, C. Spillane, M. Nordborg, M. Rehmsmeier and C. Köhler (2011). "High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm." PLoS Genet **7**(6): e1002126.

Woo, H. R., T. A. Dittmer and E. J. Richards (2008). "Three SRA-Domain Methylcytosine-Binding Proteins Cooperate to Maintain Global CpG Methylation and Epigenetic Silencing in Arabidopsis." PLoS Genet **4**(8): e1000156.

Woo, H. R., O. Pontes, C. S. Pikaard and E. J. Richards (2007). "VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization." Genes Dev **21**(3): 267-277.

Xie, W., X. Gai, Y. Zhu, D. C. Zappulla, R. Sternglanz and D. F. Voytas (2001). "Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p." Mol Cell Biol **21**(19): 6606-6614.

Xie, Z., L. K. Johansen, A. M. Gustafson, K. D. Kasschau, A. D. Lellis, D. Zilberman, S. E. Jacobsen and J. C. Carrington (2004). "Genetic and functional diversification of small RNA pathways in plants." PLoS Biol **2**(5): E104.

Yokthongwattana, C., E. Bucher, M. Caikovski, I. Vaillant, J. Nicolet, O. Mittelsten Scheid and J. Paszkowski (2010). "MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing." The EMBO journal **29**(2): 340-351.

Yu, A., G. Lepère, F. Jay, J. Wang, L. Bapaume, Y. Wang, A.-L. Abraham, J. Penterman, R. L. Fischer, O. Voinnet and L. Navarro (2013). "Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense." Proceedings of the National Academy of Sciences **110**(6): 2389-2394.

Yuan, Y.-W. and S. R. Wessler (2011). "The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies." Proceedings of the National Academy of Sciences **108**(19): 7884-7889.

Zemach, A., M. Y. Kim, P.-H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, Stacey L. Harmer and D. Zilberman (2013). "The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin." Cell **153**(1): 193-205.

Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. W. Chan, H. Chen, I. R. Henderson, P. Shinn, M. Pellegrini, S. E. Jacobsen and J. R. Ecker (2006). "Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis." Cell **126**(6): 1189-1201.

Zheng, X., O. Pontes, J. Zhu, D. Miki, F. Zhang, W.-X. Li, K. Iida, A. Kapoor, C. S. Pikaard and J.-K. Zhu (2008). "ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis." Nature **455**(7217): 1259-1262.

Zheng, X., J. Zhu, A. Kapoor and J.-K. Zhu (2007). "Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing." EMBO J **26**(6): 1691-1701.

Zhu, J. K. (2009). "Active DNA demethylation mediated by DNA glycosylases." Annu Rev Genet **43**: 143-166.

Zilberman, D., D. Coleman-Derr, T. Ballinger and S. Henikoff (2008). "Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks." Nature **456**(7218): 125-129.

Zilberman, D., M. Gehring, R. K. Tran, T. Ballinger and S. Henikoff (2007). "Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription." Nat Genet **39**(1): 61-69.

Annexes

Annexe 1 : Tables supplémentaires S4-S9 du manuscrit Gilly et al. (partie2).

Supplementary Table 4: TE-Tracker output for epiRIL 55

	Acceptor		Donor									
ID	position	size	chr	start	stop	size	orientation	overlap	read count	acceptor annotation (TAIR10)	donor annotation (Quesneville natural transposons)	score
1	chr1:10549..13848	3299	chr4	4358130	4360287	2157	sens	2081	176	AT1G01030	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
2	chr1:1392803..1393537	734	chr5	5630359	5635230	4871	sens	4346	74	AT1G04920	ATCOPIA93(AT5TE20395)	*
3	chr1:3990167..3990279	112	chr1	12753526	12760388	6862	inverse	6549	614	AT1G11810	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
3	chr1:3990167..3990279	112	chr5	5629997	5635305	5308	inverse	5217	594	AT1G11810	ATCOPIA93(AT5TE20395)	116
4	chr1:12169241..12169264	23	chr5	5630036	5635281	5245	sens	5215	520	AT1G33560	ATCOPIA93(AT5TE20395)	*
5	chr1:12169385..12169578	193	chr1	12753527	12760364	6837	sens	6824	564	AT1G33560	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
5	chr1:12169447..12169578	131	chr5	5630036	5635281	5245	sens	5215	526	AT1G33560	ATCOPIA93(AT5TE20395)	123
6	chr1:12755682..12756558	876	chr4	9644282	9651445	7163	sens	142	484	.	.	*
7	chr1:12759250..12759501	251	chr4	9648854	9654711	5857	sens	218	644	.	.	*
8	chr1:14235045..14239798	4753	chr3	13588788	13592540	3752	sens	3167	54	.	.	2
8	chr1:14239470..14239798	328	chr3	13587648	13592540	4892	sens	3167	146	.	.	2
9	chr1:14452468..14452848	380	chr1	14338005	14345208	7203	sens	1488	130	.	ATHILA0_I(AT1TE47120);ATHILA0_I(AT1TE47130);ATHILA	*
10	chr1:14850123..14851379	1256	chr3	13588788	13592571	3783	sens	3153	94	.	.	*
11	chr1:14871346..14872602	1256	chr3	13588788	13592571	3783	inverse	3384	98	.	.	*
12	chr1:15041329..15042211	882	chr4	3950866	3955393	4527	inverse	3346	74	.	.	1
12	chr1:15041352..15041935	583	chr4	4006775	4011402	4627	inverse	3657	66	.	.	6
13	chr1:15042578..15043051	473	chr4	3983491	3986874	3383	inverse	3219	72	.	.	0
13	chr1:15042582..15042904	322	chr4	3950866	3955393	4527	inverse	3346	72	.	.	1
13	chr1:15042658..15042748	90	chr4	4006775	4011693	4918	inverse	2397	72	.	.	16
14	chr1:15082317..15082807	490	chr5	11138467	11145077	6610	inverse	6242	124	AT1G40104	ATGP2(AT5TE40350)	*
15	chr1:15083228..15083409	181	chr2	3225761	3230908	5147	inverse	4127	112	AT1G40104	ATLINE1_2(AT2TE13950);ATGP2(AT2TE13955)	*
16	chr1:15083681..15084463	782	chr4	1668008	1669208	1200	sens	980	50	AT1G40104	ATGP2(AT4TE08440);ATENSPM2(AT4TE08435)	0
16	chr1:15083696..15084436	740	chr2	58735	63335	4600	inverse	3855	88	AT1G40104	ATGP2(AT2TE00110)	0
16	chr1:15083723..15084618	895	chr2	15176	20147	4971	inverse	903	66	AT1G40104	ATGP2(AT2TE00015)	0
16	chr1:15084048..15084048	0	chr4	1796365	1801189	4824	sens	1918	84	AT1G40104	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	0
16	chr1:15084075..15084449	374	chr3	15038364	15044362	5998	sens	3121	84	AT1G40104	ATGP2(AT3TE61305);ATGP2(AT3TE61300);ATGP2(AT3TE6	0
17	chr1:15085247..15085298	51	chr2	4236750	4243149	6399	sens	2239	152	AT1G40104	ATGP2(AT2TE17875)	*
18	chr1:15085340..15093551	8211	chr3	13589365	13592578	3213	sens	2991	64	AT1G40104	.	*
19	chr1:15181523..15191277	9754	chr4	3970762	3971651	889	sens	764	48	.	.	0
19	chr1:15183320..15191443	8123	chr1	15082095	15085662	3567	inverse	3261	142	.	.	2
19	chr1:15188958..15191290	2332	chr4	3998321	4001315	2994	sens	2597	80	.	.	0
19	chr1:15190795..15191443	648	chr3	13588829	13592466	3637	inverse	2979	128	.	.	0
20	chr1:15191607..15191787	180	chr4	4004893	4010384	5491	sens	1469	90	.	.	*
21	chr1:16510870..16511134	264	chr3	13587012	13592481	5469	inverse	3018	174	.	.	*
22	chr1:16511492..16511993	501	chr4	3950570	3955405	4835	inverse	4733	104	.	.	*
23	chr1:16517564..16519334	1770	chr2	3605682	3607791	2109	inverse	2098	58	.	.	2
23	chr1:16518754..16518976	222	chr4	3950517	3955916	5399	sens	4736	332	.	.	7
23	chr1:16518756..16519684	928	chr2	3616523	3621302	4779	sens	1574	74	.	.	3

23	chr1:16518760..16519684	924	chr3	13586908	13592318	5410	sens	874	168	.	.	7
23	chr1:16518777..16519684	907	chr1	15082294	15084544	2250	sens	1865	108	.	.	7
23	chr1:16518809..16518976	167	chr3	13588885	13592589	3704	inverse	3578	322	.	.	4
23	chr1:16518809..16519307	498	chr2	3622603	3627887	5284	sens	2567	190	.	.	30
23	chr1:16519330..16519334	4	chr4	3950855	3956020	5165	inverse	4053	200	.	.	7
24	chr1:16519891..16520565	674	chr3	13587006	13592589	5583	inverse	5465	418	.	.	9
24	chr1:16519982..16520557	575	chr2	3625626	3627887	2261	sens	1828	170	.	.	9
24	chr1:16520209..16520299	90	chr4	3950517	3955916	5399	sens	5305	328	.	.	9
24	chr1:16520513..16520779	266	chr4	3950529	3955919	5390	inverse	4309	114	.	.	9
25	chr1:19449701..19455727	6026	chr5	18311649	18311777	128	inverse	127	184	AT1G52220;AT1G52230	ATCOPIA68(AT5TE65910)	1
25	chr1:19450272..19455727	5455	chr3	1855904	1856012	108	sens	103	362	AT1G52220;AT1G52230	.	0
25	chr1:19450272..19455727	5455	chr4	3655374	3655982	608	inverse	601	386	AT1G52220;AT1G52230	.	2
25	chr1:19450272..19455727	5455	chr5	12710433	12710661	228	sens	223	384	AT1G52220;AT1G52230	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	2
25	chr1:19450272..19455727	5455	chr5	971390	971582	192	inverse	181	288	AT1G52220;AT1G52230	.	0
26	chr1:20281062..20281407	345	chr5	5630087	5635202	5115	inverse	4309	76	.	ATCOPIA93(AT5TE20395)	*
27	chr1:20679800..20679956	156	chr1	12753513	12759771	6258	sens	4867	734	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	2
27	chr1:20679800..20679956	156	chr5	5629977	5635293	5316	sens	5304	752	.	ATCOPIA93(AT5TE20395)	158
28	chr1:21486707..21486915	208	chr1	11304548	11309784	5236	inverse	2903	146	AT1G58050	ATENSPM3(AT1TE36570)	0
28	chr1:21486769..21486915	146	chr2	4385874	4394212	8338	inverse	2940	670	AT1G58050	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	6
28	chr1:21486789..21486913	124	chr2	4900800	4909264	8464	sens	2923	1300	AT1G58050	ATENSPM3(AT2TE20205)	683
29	chr1:21603943..21604210	267	chr2	4901064	4909286	8222	inverse	3081	460	.	ATENSPM3(AT2TE20205);ATENSPM7(AT2TE20210)	226
29	chr1:21604019..21604285	266	chr2	4385852	4393861	8009	sens	3098	234	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	4
30	chr1:25815758..25815886	128	chr1	12753526	12759803	6277	sens	5436	644	AT1G68740	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
30	chr1:25815758..25815886	128	chr5	5629983	5635302	5319	sens	5192	664	AT1G68740	ATCOPIA93(AT5TE20395)	117
31	chr1:27100692..27105017	4325	chr2	4391979	4394222	2243	inverse	2088	206	AT1G71990;AT1G72000	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0
31	chr1:27101182..27103667	2485	chr2	4900803	4903338	2535	sens	2522	716	AT1G71990;AT1G72000	ATENSPM3(AT2TE20205)	372
31	chr1:27101554..27103667	2113	chr1	12942913	12946922	4009	sens	4008	426	AT1G71990;AT1G72000	ATENSPM3(AT1TE42210)	0
32	chr1:28367370..28371506	4136	chr3	15285379	15286610	1231	sens	1202	292	AT1G75540;AT1G75550;A	ATENSPM3(AT3TE62130);HELITRON2(AT3TE62100)	*
33	chr2:1765..1882	117	chr5	13963334	13967691	4357	sens	908	220	AT2G01008	VANDAL21(AT5TE49755)	0
33	chr2:1797..1882	85	chr3	11544047	11549107	5060	sens	1563	218	AT2G01008	VANDAL21(AT3TE48070)	0
34	chr2:2815..3354	539	chr3	14201456	14201778	322	sens	321	258	.	.	0
34	chr2:2833..5048	2215	chr5	3253128	3253276	148	inverse	146	3052	AT2G01010	.	0
34	chr2:3101..3253	152	chr4	5581569	5586526	4957	inverse	510	180	.	ATGP2N(AT4TE23410)	0
34	chr2:4865..5744	879	chr2	3623116	3628066	4950	inverse	2960	102	AT2G01010	.	0
34	chr2:4955..5417	462	chr4	3950855	3955711	4856	inverse	4179	136	AT2G01010	.	0
34	chr2:4955..6401	1446	chr2	3605698	3607817	2119	inverse	2078	70	AT2G01010;AT2G01020	.	0
34	chr2:5204..5417	213	chr5	11729688	11736151	6463	sens	3818	86	AT2G01010	.	34
34	chr2:5607..6264	657	chr2	65115	67045	1930	sens	1756	328	AT2G01020	ATGP1(AT2TE00115);ATGP2N(AT2TE00120)	0
34	chr2:5960..7752	1792	chr1	20476248	20477917	1669	sens	1592	134	AT2G01021	ATGP2N(AT1TE67645)	0
34	chr2:6323..13311	6988	chr4	3950857	3954992	4135	sens	4059	218	AT2G01021;AT2G01023	.	34
34	chr2:6475..6978	503	chr3	15040270	15042826	2556	inverse	2156	78	AT2G01021	ATGP2(AT3TE61305);ATGP2(AT3TE61300);ATGP2(AT3TE6	0
34	chr2:6546..6978	432	chr4	1798086	1803182	5096	inverse	2665	118	AT2G01021	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	0
34	chr2:6880..6978	98	chr2	4238761	4242586	3825	inverse	3646	172	.	ATGP2(AT2TE17875)	3
34	chr2:6880..6978	98	chr3	11418947	11422515	3568	inverse	3482	82	.	ATGP2(AT3TE47595)	4

34	chr2:7378..8193	815	chr3	13588877	13592311	3434	sens	2991	160	.	.	0
34	chr2:7745..8222	477	chr4	5581576	5586545	4969	inverse	2560	342	.	ATGP2N(AT4TE23410)	1
34	chr2:10239..16406	6167	chr1	15081958	15085401	3443	inverse	3281	110	.	.	34
34	chr2:10239..17909	7670	chr3	13589305	13592512	3207	inverse	3008	208	.	.	0
34	chr2:10239..18914	8675	chr2	3605688	3607822	2134	sens	2102	102	.	.	5
34	chr2:17218..17971	753	chr5	11702361	11709089	6728	sens	3976	94	.	.	0
34	chr2:17233..18914	1681	chr1	15081958	15084547	2589	inverse	2546	86	.	.	0
34	chr2:17233..19531	2298	chr5	11730450	11730908	458	inverse	444	46	.	.	34
34	chr2:17235..17909	674	chr4	3950750	3955538	4788	sens	4322	356	.	.	0
34	chr2:17277..22896	5619	chr3	13588923	13592492	3569	inverse	3131	136	.	.	1
34	chr2:22616..24525	1909	chr2	3626906	3627874	968	sens	922	118	.	.	0
34	chr2:23196..24354	1158	chr4	3950880	3954416	3536	sens	2933	264	.	.	0
34	chr2:23272..24459	1187	chr3	13588928	13592501	3573	inverse	2998	222	.	.	0
34	chr2:23348..24675	1327	chr5	11723521	11724656	1135	sens	958	64	.	.	0
35	chr2:24867..24936	69	chr4	3950907	3955175	4268	sens	1328	74	.	.	*
36	chr2:25294..25803	509	chr4	3952148	3955175	3027	sens	2490	78	.	.	*
37	chr2:26686..27313	627	chr5	11729248	11736148	6900	inverse	1993	232	.	.	*
38	chr2:43201..44627	1426	chr3	13590277	13592515	2238	inverse	2026	98	.	.	*
39	chr2:45499..45808	309	chr3	13589326	13592500	3174	inverse	2938	64	.	.	*
40	chr2:48041..58735	10694	chr1	15084436	15084720	284	inverse	273	58	.	.	0
40	chr2:48279..48857	578	chr4	3950842	3955810	4968	sens	3447	302	.	.	40
40	chr2:48904..58906	10002	chr2	3605683	3607822	2139	sens	2079	178	.	.	0
40	chr2:49686..58735	9049	chr3	13588928	13592507	3579	inverse	3165	366	.	.	0
41	chr2:60196..63044	2848	chr3	13589315	13592494	3179	inverse	2978	80	.	.	0
41	chr2:62654..62850	196	chr4	3950893	3955173	4280	sens	2609	274	.	.	25
42	chr2:2563665..2564896	1231	chr5	11733513	11736162	2649	inverse	2636	72	.	.	2
42	chr2:2564006..2564496	490	chr4	3950893	3955811	4918	sens	2491	308	.	.	0
43	chr2:2566515..2566655	140	chr3	13590262	13592500	2238	inverse	2012	66	.	.	*
44	chr2:2568634..2568837	203	chr3	13704696	13710747	6051	inverse	1073	90	.	ATDNA12T3_2(AT3TE55730);ATDNA12T3_2(AT3TE55765	*
45	chr2:2569311..2569600	289	chr4	3950876	3955708	4832	sens	2341	324	.	.	*
46	chr2:2679685..2683591	3906	chr1	14592403	14592672	269	sens	253	88	.	ATENSPM5(AT1TE48035)	*
47	chr2:3215890..3216071	181	chr1	12753506	12760354	6848	sens	6771	760	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
47	chr2:3215890..3216071	181	chr5	5629997	5635284	5287	sens	5211	736	.	ATCOPIA93(AT5TE20395)	130
48	chr2:3226082..3227077	995	chr3	13589303	13592484	3181	inverse	3154	146	.	.	0
48	chr2:3226589..3227077	488	chr4	3950629	3955454	4825	sens	682	224	.	.	3
48	chr2:3226589..3227168	579	chr5	11730450	11736165	5715	inverse	4302	144	.	.	0
49	chr2:3228804..3229046	242	chr4	3952220	3953830	1610	sens	1530	74	.	.	*
50	chr2:3275646..3275965	319	chr1	12753516	12760280	6764	inverse	6742	896	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
50	chr2:3275722..3275965	243	chr5	5629977	5635296	5319	inverse	5295	896	.	ATCOPIA93(AT5TE20395)	108
51	chr2:3371620..3371975	355	chr5	5630013	5635274	5261	inverse	5153	530	AT2G07701	ATCOPIA93(AT5TE20395)	58
51	chr2:3371639..3371915	276	chr1	12753589	12760357	6768	inverse	6622	536	AT2G07701	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
52	chr2:3487246..3490086	2840	chr3	13168410	13169005	595	sens	513	50	AT2G07806	ATGP1(AT3TE53860);ATGP7(AT3TE53850)	0
52	chr2:3487687..3488570	883	chr2	6263309	6271061	7752	inverse	7728	278	AT2G07806	ATGP1(AT2TE25540)	0
52	chr2:3487687..3488570	883	chr5	10764246	10772472	8226	sens	7965	376	AT2G07806	ATGP1(AT5TE39170)	0

52	chr2:3487687..3490086	2399	chr4	1644119	1644986	867	inverse	801	100	AT2G07806	ATGP1(AT4TE08355)	0
52	chr2:3487809..3488570	761	chr3	15317834	15318946	1112	sens	921	130	AT2G07806	ATGP1(AT3TE62245)	0
52	chr2:3488149..3488570	421	chr2	4558707	4565948	7241	inverse	3471	330	AT2G07806	ATGP1(AT2TE18980)	0
52	chr2:3488149..3488941	792	chr3	12960624	12967673	7049	inverse	6840	254	AT2G07806	ATGP1(AT3TE53170)	0
52	chr2:3488347..3488452	105	chr2	1627944	1639627	11683	inverse	4039	1658	AT2G07806	ATGP1(AT2TE07550);ATLINE1A(AT2TE07540);ATGP1(AT2	538
53	chr2:3505788..3506083	295	chr5	5629993	5635311	5318	inverse	5071	794	.	ATCOPIA93(AT5TE20395)	95
53	chr2:3505788..3506153	365	chr1	12753523	12760357	6834	inverse	6789	828	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	1
54	chr2:3606602..3606616	14	chr3	11417239	11424200	6961	inverse	6952	104	.	ATGP2(AT3TE47595)	*
55	chr2:3616737..3623416	6679	chr1	15191443	15192881	1438	inverse	1246	80	AT2G07981;AT2G08986	ATHILA2(AT1TE49850)	1
55	chr2:3616758..3617705	947	chr4	1796367	1801515	5148	sens	3997	80	AT2G07981	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	0
55	chr2:3616774..3617479	705	chr3	11420006	11424230	4224	sens	2122	102	AT2G07981	ATGP2(AT3TE47595)	0
55	chr2:3618979..3624342	5363	chr2	1120	4775	3655	inverse	3618	82	AT2G08986	.	1
55	chr2:3619029..3619740	711	chr5	11139456	11144459	5003	inverse	1683	108	AT2G08986	ATGP2(AT5TE40350)	0
55	chr2:3619781..3619990	209	chr4	1796645	1801500	4855	sens	1926	92	AT2G08986	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	0
55	chr2:3620329..3621038	709	chr3	13590320	13592578	2258	sens	1997	70	AT2G08986	.	15
55	chr2:3622483..3623428	945	chr3	15720408	15722226	1818	sens	1736	56	AT2G08986	ATHILA2(AT3TE63540)	0
55	chr2:3622592..3623471	879	chr5	11139816	11144496	4680	inverse	975	132	AT2G08986	ATGP2(AT5TE40350)	0
55	chr2:3623368..3623391	23	chr4	3744383	3749824	5441	sens	4737	80	AT2G08986	ATHILA6B(AT4TE16570)	1
55	chr2:3623484..3624355	871	chr4	3949947	3955937	5990	inverse	2052	144	AT2G08986	.	24
55	chr2:3623607..3624355	748	chr3	11417468	11422973	5505	sens	3306	92	AT2G08986	ATGP2(AT3TE47595)	0
55	chr2:3623912..3624459	547	chr1	16514760	16523571	8811	sens	1296	204	AT2G08986	ATHILA5(AT1TE54480)	92
55	chr2:3623914..3624544	630	chr3	14196411	14200830	4419	inverse	3724	94	AT2G08986	.	6
56	chr2:3627215..3627221	6	chr5	11138451	11144459	6008	sens	5912	104	.	ATGP2(AT5TE40350)	*
57	chr2:3949156..3950145	989	chr3	13587645	13592581	4936	sens	3200	104	.	.	*
58	chr2:3954727..3955205	478	chr3	13587703	13592579	4876	sens	3180	248	.	.	*
59	chr2:3956274..3965491	9217	chr3	13588805	13592579	3774	sens	3179	236	.	.	3
59	chr2:3956350..3964963	8613	chr1	15082060	15085670	3610	sens	3298	162	.	.	1
60	chr2:4171056..4171205	149	chr1	12753507	12760316	6809	sens	6735	826	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
60	chr2:4171056..4171205	149	chr5	5629977	5635294	5317	sens	5232	818	.	ATCOPIA93(AT5TE20395)	139
61	chr2:4237436..4237887	451	chr4	3950888	3955810	4922	inverse	4363	366	.	.	15
61	chr2:4237513..4238359	846	chr5	11729401	11736163	6762	sens	4229	180	.	.	3
61	chr2:4237557..4238091	534	chr3	13588923	13592490	3567	sens	3161	168	.	.	0
61	chr2:4237557..4243077	5520	chr3	13588916	13592496	3580	sens	3181	100	.	.	0
61	chr2:4238568..4238730	162	chr3	13705293	13711739	6446	sens	5149	104	.	ATDNA12T3_2(AT3TE55730);ATDNA12T3_2(AT3TE55765	2
61	chr2:4242793..4243418	625	chr4	3950887	3955539	4652	inverse	2681	286	.	.	15
61	chr2:4242896..4243917	1021	chr3	14191681	14195167	3486	inverse	2800	50	.	.	0
61	chr2:4243035..4243435	400	chr5	11731583	11736152	4569	sens	1681	116	.	.	3
61	chr2:4243064..4243376	312	chr3	13589342	13592483	3141	sens	2988	218	.	.	0
62	chr2:4390247..4391281	1034	chr3	13588826	13592394	3568	inverse	3424	142	.	.	0
62	chr2:4390495..4390820	325	chr5	11733095	11736125	3030	inverse	2939	58	.	.	0
63	chr2:4698082..4698725	643	chr3	13588822	13592568	3746	sens	3510	104	.	.	*
64	chr2:5567146..5577685	10539	chr4	3984178	3986880	2702	sens	2386	74	.	.	0
64	chr2:5567849..5577927	10078	chr4	3965753	3969392	3639	sens	3085	62	.	.	8
64	chr2:5569409..5572667	3258	chr1	15082099	15085400	3301	inverse	3097	104	.	.	1

64	chr2:5569477..5577483	8006	chr3	13588806	13592321	3515	inverse	3461	178	.	.	1
64	chr2:5575082..5577335	2253	chr4	3999026	4001201	2175	sens	1760	60	.	.	0
65	chr2:8267663..8272348	4685	chr3	10181982	10182386	404	sens	383	108	AT2G19090	.	0
65	chr2:8272168..8272305	137	chr1	12753511	12760382	6871	inverse	6860	1646	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	1
65	chr2:8272168..8272305	137	chr5	5629981	5635296	5315	inverse	5302	1608	.	ATCOPIA93(AT5TE20395)	316
66	chr2:19475806..19475903	97	chr1	12753520	12760391	6871	sens	6627	824	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
66	chr2:19475806..19475903	97	chr5	5629983	5635304	5321	sens	5313	818	.	ATCOPIA93(AT5TE20395)	174
67	chr3:3351852..3356952	5100	chr4	4358136	4360281	2145	inverse	2050	234	AT3G10710;AT3G10720	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
68	chr3:3549888..3550083	195	chr2	4385870	4394233	8363	sens	4055	226	AT3G11325	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0
68	chr3:3549888..3550083	195	chr2	4900810	4909268	8458	inverse	4038	434	AT3G11325	ATENSPM3(AT2TE20205)	202
69	chr3:3573618..3573970	352	chr5	5630194	5635156	4962	sens	4256	72	AT3G11402	ATCOPIA93(AT5TE20395)	*
70	chr3:4589949..4590097	148	chr1	12753511	12760318	6807	sens	6802	954	AT3G13900	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
70	chr3:4589949..4590097	148	chr5	5629982	5635305	5323	sens	5273	914	AT3G13900	ATCOPIA93(AT5TE20395)	168
71	chr3:5285401..5285724	323	chr5	11686028	11697079	11051	sens	2084	282	.	ATHILA6A(AT5TE42135)	0
71	chr3:5285492..5285728	236	chr5	11833032	11842164	9132	inverse	245	236	.	ATHILA6A(AT5TE42470)	8
71	chr3:5285507..5285724	217	chr5	11935296	11945289	9993	inverse	2647	334	.	ATHILA6A(AT5TE42710)	0
71	chr3:5285507..5285728	221	chr4	3274877	3285033	10156	inverse	2737	270	.	ATHILA0_I(AT4TE15025);ATHILA3(AT4TE15005);ATHILA6A	8
71	chr3:5285514..5285728	214	chr5	11810300	11817812	7512	sens	286	226	.	ATHILA6A(AT5TE42440)	0
72	chr3:8036011..8036108	97	chr1	12753567	12760366	6799	sens	6704	788	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
72	chr3:8036011..8036108	97	chr5	5630001	5635297	5296	sens	5241	760	.	ATCOPIA93(AT5TE20395)	134
73	chr3:9019495..9019591	96	chr2	4900831	4909285	8454	inverse	2854	542	.	ATENSPM3(AT2TE20205);ATENSPM7(AT2TE20210)	277
73	chr3:9019495..9019622	127	chr2	4385853	4393305	7452	sens	2871	266	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	2
74	chr3:9043436..9043547	111	chr2	4385919	4393469	7550	sens	2863	666	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	1
74	chr3:9043436..9043773	337	chr1	11302768	11309787	7019	sens	2313	146	.	ATENSPM3(AT1TE36570)	0
74	chr3:9043438..9043531	93	chr2	4900808	4909219	8411	inverse	3572	1306	.	ATENSPM3(AT2TE20205)	651
75	chr3:9142495..9142632	137	chr2	4901288	4906335	5047	inverse	2257	68	.	ATENSPM3(AT2TE20205)	*
76	chr3:10647573..10648089	516	chr1	11304562	11309062	4500	inverse	2537	64	AT3G28415	ATENSPM3(AT1TE36570)	0
76	chr3:10647578..10647846	268	chr2	4385873	4393315	7442	inverse	2745	382	AT3G28415	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	3
76	chr3:10647627..10647765	138	chr2	4900823	4909265	8442	sens	2950	758	AT3G28415	ATENSPM3(AT2TE20205)	366
77	chr3:11417966..11418907	941	chr4	3950893	3955525	4632	inverse	4335	112	.	.	6
77	chr3:11417966..11418907	941	chr5	11727470	11736163	8693	sens	4738	130	.	.	0
78	chr3:11422988..11423479	491	chr4	3950900	3955247	4347	inverse	1652	210	.	.	*
79	chr3:12209082..12212216	3134	chr3	13588790	13592577	3787	sens	3764	150	.	.	*
80	chr3:12225739..12232985	7246	chr3	13588920	13592573	3653	sens	2978	98	.	.	*
81	chr3:12245484..12247328	1844	chr3	13588828	13592574	3746	sens	3100	188	.	.	*
82	chr3:13187023..13187656	633	chr3	13587630	13592575	4945	inverse	3510	276	.	.	13
82	chr3:13187023..13187656	633	chr3	13588797	13592575	3778	inverse	3510	250	.	.	13
83	chr3:13400059..13407564	7505	chr4	3655373	3655990	617	sens	614	444	.	.	0
83	chr3:13400059..13407564	7505	chr5	12710434	12710663	229	inverse	226	448	.	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	1
83	chr3:13400059..13407564	7505	chr5	971392	971580	188	sens	186	364	.	.	0
83	chr3:13400059..13407613	7554	chr1	15214383	15214507	124	inverse	112	206	.	ATENSPM5(AT1TE49955)	4
83	chr3:13400059..13407613	7554	chr5	18311649	18311776	127	sens	122	238	.	ATCOPIA68(AT5TE65910)	0
83	chr3:13401546..13407564	6018	chr1	7024972	7025327	355	inverse	346	388	.	ATENSPM6(AT1TE22710)	0
84	chr3:13588896..13589969	1073	chr4	3953424	3953624	200	inverse	192	60	.	.	59

84	chr3:13589398..13590454	1056	chr2	3606157	3607805	1648	inverse	1629	50	.	.	3
84	chr3:13589499..13590498	999	chr3	14195102	14200446	5344	inverse	2823	84	.	.	8
84	chr3:13589503..13590038	535	chr5	11138651	11143655	5004	inverse	1140	180	.	ATGP2(AT5TE40350)	0
84	chr3:13589503..13590285	782	chr2	17909	23146	5237	inverse	416	164	.	ATGP2(AT2TE00015)	2
84	chr3:13589504..13590293	789	chr2	58735	63422	4687	inverse	474	134	.	ATGP2(AT2TE00110)	4
84	chr3:13589520..13590276	756	chr2	2565124	2569911	4787	inverse	2035	116	.	ATGP2(AT2TE11540)	0
84	chr3:13589520..13590279	759	chr4	1667979	1669207	1228	sens	1184	114	.	ATGP2(AT4TE08440);ATENSPM2(AT4TE08435)	0
85	chr3:13591846..13592255	409	chr4	3869451	3876878	7427	inverse	602	114	.	ATHILA(AT4TE16945)	21
85	chr3:13592164..13592173	9	chr5	11138954	11144681	5727	inverse	2858	98	.	ATGP2(AT5TE40350)	0
86	chr3:13705824..13706280	456	chr4	1797053	1804088	7035	sens	2069	110	.	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	*
87	chr3:13707886..13707886	0	chr2	4239528	4244276	4748	sens	1664	96	.	ATGP2(AT2TE17875)	*
88	chr3:13708935..13709371	436	chr4	1796836	1801049	4213	sens	2788	84	.	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	*
89	chr3:13709701..13716525	6824	chr3	11420347	11422350	2003	sens	1604	94	.	ATGP2(AT3TE47595)	*
90	chr3:14193741..14193771	30	chr4	5581677	5586395	4718	inverse	1410	222	.	ATGP2N(AT4TE23410)	*
91	chr3:14193979..14195877	1898	chr5	3253128	3253274	146	inverse	137	94	AT3G41762	.	1761
91	chr3:14195688..14195849	161	chr3	11543813	11549107	5294	sens	2758	318	AT3G41762	VANDAL21(AT3TE48070)	0
91	chr3:14195732..14195849	117	chr5	13963063	13967691	4628	sens	3388	270	AT3G41762	VANDAL21(AT5TE49755)	0
91	chr3:14195764..14195849	85	chr2	10001490	10007272	5782	inverse	3420	860	AT3G41762	VANDAL21(AT2TE42810)	271
91	chr3:14195764..14195849	85	chr2	1507006	1513183	6177	sens	3559	456	AT3G41762	VANDAL21(AT2TE06955)	0
92	chr3:14196613..14199019	2406	chr5	3253128	3253276	148	inverse	145	3400	AT3G41768	.	1777
92	chr3:14196916..14198573	1657	chr2	7190932	7191520	588	sens	581	1018	AT3G41768	.	440
92	chr3:14197132..14198758	1626	chr2	2594	3503	909	sens	904	414	AT3G41768	.	469
92	chr3:14198647..14198670	23	chr1	15081974	15085676	3702	sens	3313	130	AT3G41768	.	4
92	chr3:14198926..14199833	907	chr2	3605698	3607810	2112	inverse	2064	50	AT3G41768;AT3G41979	.	0
92	chr3:14198926..14199833	907	chr4	3950534	3955197	4663	inverse	4535	162	AT3G41768;AT3G41979	.	2
92	chr3:14198926..14202025	3099	chr2	3625069	3627385	2316	inverse	2176	62	AT3G41768;AT3G41979	.	0
92	chr3:14199175..14200159	984	chr2	3618070	3624298	6228	sens	2124	98	AT3G41768;AT3G41979	.	8
92	chr3:14199857..14203657	3800	chr1	13972015	13973894	1879	inverse	1575	54	AT3G41979	ATGP2N(AT1TE45860);ATLINE2(AT1TE45855)	0
92	chr3:14199931..14200862	931	chr1	12566495	12568080	1585	inverse	1418	70	.	ATGP2N(AT1TE40940)	0
92	chr3:14199931..14200862	931	chr1	20475946	20477917	1971	sens	1826	140	.	ATGP2N(AT1TE67645)	0
92	chr3:14199931..14200862	931	chr3	12511607	12513553	1946	inverse	1778	142	.	ATGP2N(AT3TE51565)	0
92	chr3:14199989..14202716	2727	chr3	14026873	14029519	2646	inverse	2291	90	.	ATGP2N(AT3TE57550)	0
92	chr3:14200046..14200862	816	chr2	6509126	6510486	1360	inverse	1338	68	.	ATGP2N(AT2TE26610)	0
92	chr3:14200046..14200862	816	chr2	9191759	9193697	1938	sens	1638	106	.	ATGP2N(AT2TE38905)	0
92	chr3:14200046..14200862	816	chr4	2361556	2363087	1531	sens	1383	106	.	ATGP2N(AT4TE10985);TA11(AT4TE10980)	6
92	chr3:14200130..14200949	819	chr2	43566	47795	4229	sens	2878	166	.	ATGP2(AT2TE00075)	67
92	chr3:14200130..14200976	846	chr2	3227436	3230586	3150	sens	2818	102	.	ATLINE1_2(AT2TE13950);ATGP2(AT2TE13955)	0
92	chr3:14200360..14200949	589	chr2	2564667	2569726	5059	sens	2818	138	.	ATGP2(AT2TE11540)	0
92	chr3:14200446..14200949	503	chr2	4237899	4242972	5073	inverse	2837	200	.	ATGP2(AT2TE17875)	0
92	chr3:14200446..14200949	503	chr3	15040270	15042826	2556	inverse	2156	78	.	ATGP2(AT3TE61305);ATGP2(AT3TE61300);ATGP2(AT3TE6	0
92	chr3:14200517..14200949	432	chr4	1797019	1803182	6163	inverse	4612	150	.	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	1
92	chr3:14200775..14201055	280	chr2	14498	22978	8480	sens	2892	162	.	ATGP2(AT2TE00015)	0
92	chr3:14200851..14200949	98	chr3	11418962	11422705	3743	inverse	3432	116	.	ATGP2(AT3TE47595)	0
92	chr3:14200851..14200976	125	chr5	11139987	11144503	4516	sens	2916	80	.	ATGP2(AT5TE40350)	2

92	chr3:14201702..14202112	410	chr2	3354	3503	149	sens	149	164	.	.	295
92	chr3:14201721..14202114	393	chr4	5581576	5586545	4969	inverse	506	386	.	ATGP2N(AT4TE23410)	234
92	chr3:14201762..14202716	954	chr2	9191756	9193697	1941	sens	1664	88	.	ATGP2N(AT2TE38905)	0
92	chr3:14202108..14202164	56	chr4	3950828	3955151	4323	inverse	2543	58	.	.	2
92	chr3:14202906..14203657	751	chr1	20475917	20477819	1902	sens	1807	74	.	ATGP2N(AT1TE67645)	0
92	chr3:14203263..14203655	392	chr4	5581581	5586457	4876	inverse	1394	132	.	ATGP2N(AT4TE23410)	234
93	chr3:14457093..14457312	219	chr2	4389271	4390592	1321	inverse	1077	80	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0
93	chr3:14457168..14457296	128	chr2	4903490	4909015	5525	sens	2484	232	.	ATENSPM3(AT2TE20205)	22
94	chr3:15712363..15719864	7501	chr4	3970231	3974118	3887	inverse	3386	106	.	.	0
94	chr3:15712706..15714657	1951	chr1	15082078	15085821	3743	sens	3389	104	.	.	7
94	chr3:15712950..15713465	515	chr4	3950519	3955401	4882	inverse	4084	112	.	.	3
94	chr3:15713098..15719874	6776	chr4	3984299	3987343	3044	inverse	2582	72	.	.	0
94	chr3:15713101..15719917	6816	chr4	4010166	4011690	1524	inverse	1490	52	.	.	6
94	chr3:15713133..15719864	6731	chr4	3999030	4002336	3306	inverse	2995	78	.	.	0
94	chr3:15713146..15719894	6748	chr4	3965771	3968851	3080	inverse	3067	84	.	.	5
95	chr3:15720220..15720981	761	chr3	13588831	13592586	3755	sens	3727	220	.	.	1
95	chr3:15720641..15721345	704	chr4	3969393	3973804	4411	inverse	770	70	.	.	0
96	chr3:16618765..16619206	441	chr5	5630176	5635179	5003	sens	4493	78	AT3G45290	ATCOPIA93(AT5TE20395)	*
97	chr3:16767338..16767437	99	chr2	4900833	4909247	8414	inverse	3467	800	.	ATENSPM3(AT2TE20205)	401
97	chr3:16767338..16767548	210	chr2	4385891	4394205	8314	sens	2517	378	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0
98	chr3:17341732..17341847	115	chr2	5852427	5857334	4907	inverse	4877	1384	AT3G47090	ATCOPIA12(AT2TE23850);ATCOPIA13(AT2TE23855)	*
99	chr4:1801439..1803802	2363	chr3	13589327	13592519	3192	sens	3147	180	.	.	0
99	chr4:1801439..1803802	2363	chr3	13590283	13592519	2236	sens	2048	154	.	.	0
100	chr4:1803980..1804289	309	chr4	3950884	3955463	4579	inverse	3980	200	.	.	*
101	chr4:3737237..3737687	450	chr4	3974753	3979457	4704	inverse	2885	66	.	.	4
101	chr4:3737427..3745571	8144	chr3	13588806	13592481	3675	sens	2989	96	.	.	0
101	chr4:3737501..3745639	8138	chr1	15082095	15086025	3930	sens	3250	102	.	.	2
101	chr4:3737506..3744383	6877	chr2	3605713	3607837	2124	inverse	2070	64	.	.	0
101	chr4:3737536..3744383	6847	chr5	11707955	11710354	2399	inverse	2172	50	.	.	0
101	chr4:3737577..3738329	752	chr4	3950688	3955705	5017	inverse	4113	114	.	.	0
101	chr4:3737702..3747443	9741	chr4	3970305	3973804	3499	inverse	3120	104	.	.	0
102	chr4:3866554..3866804	250	chr4	3950549	3957370	6821	sens	5388	346	.	ATHILA2(AT4TE17240)	*
103	chr4:3867884..3867897	13	chr3	13588853	13592579	3726	inverse	3195	192	.	.	*
104	chr4:3868185..3868195	10	chr4	3950608	3955920	5312	sens	3703	380	.	.	*
105	chr4:3951894..3952153	259	chr2	42815	48355	5540	sens	386	134	.	ATGP2(AT2TE00075)	*
106	chr4:3953465..3954347	882	chr2	3225761	3226663	902	sens	770	84	.	ATLINE1_2(AT2TE13950);ATGP2(AT2TE13955)	0
106	chr4:3953751..3954331	580	chr2	18697	20522	1825	sens	1595	162	.	ATGP2(AT2TE00015)	0
106	chr4:3953780..3954331	551	chr5	11138468	11143784	5316	sens	1866	176	.	ATGP2(AT5TE40350)	0
107	chr4:3955172..3965753	10581	chr2	5565128	5567925	2797	sens	2282	102	.	ATHILA2(AT2TE22505)	1
107	chr4:3955831..3965744	9913	chr4	3747184	3750000	2816	inverse	2550	148	.	ATHILA6B(AT4TE16570)	0
108	chr4:3968323..3970317	1994	chr1	15179527	15181761	2234	sens	2032	100	.	ATHILA6B(AT1TE49845)	0
108	chr4:3968323..3975649	7326	chr1	15832527	15834456	1929	inverse	1896	48	.	ATHILA2(AT1TE52125);ATHILA2(AT1TE52105)	3
108	chr4:3968769..3972681	3912	chr3	15719862	15721956	2094	inverse	1843	114	.	ATHILA2(AT3TE63540)	0
108	chr4:3968770..3970323	1553	chr3	15711545	15713222	1677	inverse	1597	52	.	ATHILA2(AT3TE63540)	3

108	chr4:3968781..3970462	1681	chr2	5565128	5568010	2882	sens	2508	172	.	ATHILA2(AT2TE22505)	5
108	chr4:3969379..3970293	914	chr1	15191728	15192908	1180	sens	1151	92	.	ATHILA2(AT1TE49850)	0
108	chr4:3971559..3974981	3422	chr5	11460346	11462087	1741	sens	1658	54	.	ATHILA2(AT5TE41425)	0
108	chr4:3973006..3978849	5843	chr4	3747275	3749974	2699	inverse	2455	164	.	ATHILA6B(AT4TE16570)	12
108	chr4:3973037..3977971	4934	chr5	12040202	12043298	3096	inverse	2750	68	.	ATHILA6A(AT5TE42975)	0
108	chr4:3973228..3974965	1737	chr2	5577287	5578921	1634	sens	1555	94	.	ATHILA2(AT2TE22520)	0
108	chr4:3974052..3974964	912	chr2	5565128	5567922	2794	sens	2384	162	.	ATHILA2(AT2TE22505)	5
108	chr4:3974750..3984332	9582	chr1	15179952	15181889	1937	sens	1629	120	.	ATHILA6B(AT1TE49845)	0
108	chr4:3978868..3984476	5608	chr1	15191621	15192967	1346	sens	1227	98	.	ATHILA2(AT1TE49850)	0
109	chr4:3988883..3995918	7035	chr3	13588800	13592549	3749	inverse	3395	64	.	.	*
110	chr4:4002204..4006211	4007	chr3	15711486	15712743	1257	inverse	1125	66	.	ATHILA2(AT3TE63540)	9
110	chr4:4003110..4009433	6323	chr2	5565808	5567930	2122	sens	1816	148	.	ATHILA2(AT2TE22505)	2
110	chr4:4004076..4006222	2146	chr5	12040214	12043276	3062	inverse	2661	78	.	ATHILA6A(AT5TE42975)	0
111	chr4:4026715..4026961	246	chr4	3950519	3955401	4882	inverse	3497	342	.	.	*
112	chr4:4330878..4331700	822	chr3	13590014	13592575	2561	sens	2328	52	.	.	*
113	chr4:4780534..4780713	179	chr5	11766393	11767434	1041	inverse	1017	58	.	ATHILA5(AT5TE42330)	*
114	chr4:4781359..4781396	37	chr2	3965163	3966125	962	inverse	837	84	.	ATHILA2(AT2TE16860)	*
115	chr4:4967000..4967959	959	chr3	13588829	13592564	3735	inverse	3112	130	.	.	*
116	chr4:8561572..8561742	170	chr5	5629980	5635289	5309	sens	5223	836	AT4G14970	ATCOPIA93(AT5TE20395)	159
116	chr4:8561648..8561742	94	chr1	12753512	12760363	6851	sens	6794	838	AT4G14970	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
116	chr4:8561648..8561842	194	chr3	10181988	10182392	404	inverse	397	70	AT4G14970	.	0
117	chr4:9320021..9320640	619	chr1	7714742	7722470	7728	sens	7318	1262	AT4G16550	ATCOPIA78(AT1TE24850)	2
117	chr4:9320021..9323263	3242	chr1	3360542	3361308	766	inverse	754	150	AT4G16550	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
117	chr4:9320075..9320640	565	chr1	18013198	18018259	5061	sens	4829	1300	AT4G16550	ATCOPIA78(AT1TE59755)	0
117	chr4:9320075..9320640	565	chr1	21524998	21529469	4471	inverse	4014	1336	AT4G16550	ATCOPIA78(AT1TE71045)	9
117	chr4:9320075..9320640	565	chr1	3780801	3785719	4918	sens	4508	1384	AT4G16550	ATCOPIA78(AT1TE12295)	0
117	chr4:9320075..9320640	565	chr3	22059583	22064328	4745	sens	4347	1258	AT4G16550	ATCOPIA78(AT3TE89830)	0
117	chr4:9320075..9320640	565	chr3	22695568	22700052	4484	inverse	4354	1414	AT4G16550	ATCOPIA78(AT3TE92525)	0
117	chr4:9320075..9320640	565	chr5	4208085	4213049	4964	inverse	4754	1392	AT4G16550	ATCOPIA78(AT5TE15240)	0
117	chr4:9320075..9320683	608	chr3	13369251	13373860	4609	sens	4485	930	AT4G16550	ATCOPIA78(AT3TE54550)	0
118	chr4:9649959..9650180	221	chr1	12753515	12760162	6647	sens	2782	1174	AT4G17190	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
118	chr4:9650040..9650151	111	chr5	5629977	5635306	5329	sens	5292	1756	AT4G17190	ATCOPIA93(AT5TE20395)	388
118	chr4:9650040..9650164	124	chr3	10181982	10182391	409	inverse	401	146	AT4G17190	.	0
119	chr4:9650230..9651158	928	chr5	5629980	5635282	5302	sens	5284	1594	AT4G17200;AT4G17190	ATCOPIA93(AT5TE20395)	*
120	chr4:10739583..10739988	405	chr1	21524992	21529530	4538	sens	4527	1102	AT4G19740	ATCOPIA78(AT1TE71045)	2
120	chr4:10739583..10739988	405	chr1	7714742	7722145	7403	inverse	7345	928	AT4G19740	ATCOPIA78(AT1TE24850)	0
120	chr4:10739590..10739988	398	chr3	13369256	13373817	4561	inverse	4550	660	AT4G19740	ATCOPIA78(AT3TE54550)	0
120	chr4:10739659..10739988	329	chr1	18013166	18018038	4872	inverse	4642	1016	AT4G19740	ATCOPIA78(AT1TE59755)	0
120	chr4:10739659..10739988	329	chr1	3780769	3785628	4859	inverse	4565	1102	AT4G19740	ATCOPIA78(AT1TE12295)	1
120	chr4:10739659..10739988	329	chr1	7717360	7722221	4861	inverse	4721	874	AT4G19740	ATCOPIA78(AT1TE24850)	0
120	chr4:10739659..10739988	329	chr3	22059539	22064250	4711	inverse	4404	1180	AT4G19740	ATCOPIA78(AT3TE89830)	42
120	chr4:10739659..10739988	329	chr3	22695564	22700155	4591	sens	4568	1144	AT4G19740	ATCOPIA78(AT3TE92520);ATCOPIA78(AT3TE92525)	16
120	chr4:10739659..10739988	329	chr5	4208080	4213081	5001	sens	4632	1040	AT4G19740	ATCOPIA78(AT5TE15240)	0
120	chr4:10739659..10740111	452	chr1	3360538	3361293	755	sens	747	156	AT4G19740	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0

121	chr4:10898002..10907196	9194	chr1	6706699	6707874	1175	sens	1125	58	AT4G20170;AT4G20190	META1(AT1TE21710)	0
121	chr4:10898002..10907715	9713	chr3	17364524	17365005	481	inverse	439	48	AT4G20170;AT4G20190	META1(AT3TE70410)	0
121	chr4:10899004..10907196	8192	chr1	4407462	4408638	1176	inverse	1126	82	AT4G20170;AT4G20190	META1(AT1TE14315)	0
121	chr4:10899004..10907196	8192	chr1	6540900	6542074	1174	sens	1124	76	AT4G20170;AT4G20190	META1(AT1TE21210)	0
121	chr4:10899004..10907196	8192	chr2	911728	912882	1154	sens	1104	74	AT4G20170;AT4G20190	META1(AT2TE04085)	0
121	chr4:10899004..10907196	8192	chr3	14958447	14959621	1174	inverse	1124	70	AT4G20170;AT4G20190	META1(AT3TE61000)	0
121	chr4:10899004..10907196	8192	chr3	18736760	18737934	1174	sens	1124	58	AT4G20170;AT4G20190	META1(AT3TE76010)	0
121	chr4:10899004..10907196	8192	chr5	14411844	14412995	1151	sens	1088	64	AT4G20170;AT4G20190	META1(AT5TE51845)	0
121	chr4:10899004..10907196	8192	chr5	14471144	14472295	1151	sens	1088	70	AT4G20170;AT4G20190	META1(AT5TE52075)	0
121	chr4:10899004..10907540	8536	chr5	9383160	9384272	1112	inverse	1085	66	AT4G20170;AT4G20190	META1(AT5TE34040)	0
121	chr4:10899004..10907715	8711	chr2	1397558	1398067	509	sens	478	48	AT4G20170;AT4G20190	META1(AT2TE06405)	0
121	chr4:10899004..10907715	8711	chr3	15816336	15816825	489	sens	452	48	AT4G20170;AT4G20190	META1(AT3TE63950)	0
121	chr4:10899004..10907715	8711	chr3	16040488	16041002	514	inverse	492	58	AT4G20170;AT4G20190	META1(AT3TE64915)	0
121	chr4:10899004..10907715	8711	chr5	10402188	10403361	1173	inverse	1123	62	AT4G20170;AT4G20190	META1(AT5TE37975)	0
122	chr4:12413348..12416017	2669	chr4	4358132	4360293	2161	inverse	1988	234	AT4G23870;AT4G23880	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
123	chr4:13251935..13253229	1294	chr1	3360539	3361312	773	sens	756	196	AT4G26150	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
123	chr4:13252207..13252518	311	chr1	21526212	21529538	3326	sens	3280	880	.	ATCOPIA78(AT1TE71045)	0
123	chr4:13252207..13252518	311	chr3	22696783	22700153	3370	sens	3341	842	.	ATCOPIA78(AT3TE92525)	0
123	chr4:13252449..13252518	69	chr1	7714733	7722338	7605	inverse	7390	894	.	ATCOPIA78(AT1TE24850)	0
123	chr4:13252449..13252518	69	chr5	4208082	4213062	4980	sens	4590	992	.	ATCOPIA78(AT5TE15240)	0
123	chr4:13252525..13252965	440	chr3	22695565	22700077	4512	sens	4491	954	.	ATCOPIA78(AT3TE92525)	0
123	chr4:13252525..13252965	440	chr5	4208082	4213062	4980	sens	4590	990	.	ATCOPIA78(AT5TE15240)	0
123	chr4:13252525..13253191	666	chr1	21524994	21525465	471	sens	421	106	.	ATCOPIA78(AT1TE71045)	0
123	chr4:13252525..13253326	801	chr1	18013165	18016901	3736	inverse	3698	1058	AT4G26150	ATCOPIA78(AT1TE59755)	77
123	chr4:13252525..13253326	801	chr1	3780768	3784504	3736	inverse	3618	974	AT4G26150	ATCOPIA78(AT1TE12295)	0
123	chr4:13252525..13253326	801	chr3	13369252	13372792	3540	inverse	3159	544	AT4G26150	ATCOPIA78(AT3TE54550)	0
123	chr4:13252525..13253326	801	chr3	22059538	22064238	4700	inverse	3705	1030	AT4G26150	ATCOPIA78(AT3TE89830)	1
123	chr4:13252525..13253657	1132	chr1	7717359	7721331	3972	inverse	3488	754	AT4G26150	ATCOPIA78(AT1TE24850)	0
124	chr4:13668518..13673852	5334	chr4	4358131	4360262	2131	inverse	2111	276	AT4G27290;AT4G27300	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
125	chr4:17154860..17155612	752	chr5	5630033	5635134	5101	sens	4976	70	AT4G36260	ATCOPIA93(AT5TE20395)	*
126	chr4:17511985..17512767	782	chr2	1265875	1271634	5759	inverse	1988	342	AT4G37210	VANDAL21(AT2TE05755)	0
126	chr4:17512096..17512273	177	chr3	11541551	11549124	7573	sens	3345	534	AT4G37210	VANDAL21(AT3TE48070)	0
126	chr4:17512169..17512273	104	chr2	1505598	1513124	7526	sens	3403	690	AT4G37210	VANDAL21(AT2TE06955)	0
126	chr4:17512208..17512273	65	chr2	10001502	10009608	8106	inverse	3457	1406	AT4G37210	VANDAL21(AT2TE42810)	593
126	chr4:17512208..17512273	65	chr5	13960878	13967940	7062	sens	3517	500	AT4G37210	VANDAL21(AT5TE49755)	0
126	chr4:17512208..17512880	672	chr3	13072520	13078376	5856	sens	2550	264	AT4G37210	VANDAL21(AT3TE53580)	0
127	chr5:913859..914446	587	chr1	11304563	11309787	5224	inverse	194	92	AT5G03600	ATENSPM3(AT1TE36570)[2]	*
128	chr5:915339..915383	44	chr2	4386308	4392316	6008	inverse	380	302	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)[2]	*
129	chr5:2123289..2123790	501	chr2	4901371	4909242	7871	inverse	1390	110	AT5G06839	ATENSPM3(AT2TE20205)	*
130	chr5:2610008..2612736	2728	chr4	4358137	4360243	2106	inverse	2066	422	AT5G08130	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
131	chr5:5977351..5983454	6103	chr3	14082605	14084028	1423	inverse	1403	66	AT5G18070;AT5G18065	ATENSPM3(AT3TE57715)	0
131	chr5:5980263..5980983	720	chr4	4358131	4360292	2161	inverse	2144	516	AT5G18065	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	212
132	chr5:9624750..9624842	92	chr2	4900822	4909218	8396	inverse	3259	588	AT5G27300	ATENSPM3(AT2TE20205)	296
132	chr5:9624750..9624881	131	chr2	4385920	4393310	7390	sens	2892	254	AT5G27300	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0

133	chr5:10248983..10251449	2466	chr4	3950909	3953686	2777	sens	2559	48	.	.	*
134	chr5:11204896..11213628	8732	chr5	11732020	11736166	4146	inverse	3872	116	.	.	*
135	chr5:11322671..11322828	157	chr5	11702184	11707206	5022	inverse	3537	182	.	.	*
136	chr5:11705653..11706236	583	chr5	11320537	11325747	5210	inverse	509	180	.	ATHILA6A(AT5TE41070)	*
137	chr5:11710918..11711791	873	chr4	3950518	3956007	5489	sens	4920	148	.	.	*
138	chr5:11728297..11728522	225	chr5	12007455	12016779	9324	inverse	859	146	.	ATHILA2(AT5TE42900)	*
139	chr5:11728605..11729400	795	chr2	4386593	4391184	4591	inverse	274	64	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0
139	chr5:11728622..11732218	3596	chr4	3951870	3953768	1898	inverse	1742	108	.	.	10
139	chr5:11729649..11730394	745	chr5	11813559	11819841	6282	inverse	644	284	.	ATHILA6A(AT5TE42440)	0
139	chr5:11729701..11731582	1881	chr3	13590427	13592578	2151	sens	2144	120	.	.	19
139	chr5:11730409..11730833	424	chr2	3619059	3624005	4946	sens	2379	68	.	.	10
139	chr5:11730822..11731589	767	chr2	43092	47951	4859	inverse	2465	124	.	ATGP2(AT2TE00075)	7
139	chr5:11730829..11731583	754	chr2	13072	17309	4237	inverse	1455	74	.	ATGP2(AT2TE00015)	0
139	chr5:11731495..11732251	756	chr2	4386461	4391184	4723	inverse	151	150	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	0
140	chr5:11732388..11733096	708	chr5	12010367	12015464	5097	inverse	841	94	.	ATHILA2(AT5TE42900)	12
140	chr5:11732707..11743406	10699	chr3	13590542	13592578	2036	sens	1785	44	.	.	16
140	chr5:11733429..11734361	932	chr2	2564896	2569387	4491	inverse	3394	148	.	ATGP2(AT2TE11540)	0
140	chr5:11733791..11734524	733	chr2	4904048	4908562	4514	sens	411	94	.	ATENSPM3(AT2TE20205)	23
140	chr5:11733820..11734524	704	chr5	11320515	11325890	5375	inverse	1786	330	.	ATHILA6A(AT5TE41070)	9
141	chr5:11819011..11819805	794	chr5	11727063	11733126	6063	sens	2181	190	.	.	*
142	chr5:12015757..12016996	1239	chr3	13588832	13592362	3530	inverse	3365	180	.	.	*
143	chr5:12030534..12031755	1221	chr3	13588823	13592569	3746	sens	3042	68	.	.	*
144	chr5:12112562..12123139	10577	chr3	13588912	13592486	3574	sens	3104	98	.	.	*
145	chr5:12125316..12128718	3402	chr3	13588788	13592368	3580	sens	3412	76	.	.	*
146	chr5:12485252..12486832	1580	chr2	4904534	4906825	2291	inverse	1886	62	.	ATENSPM3(AT2TE20205)	*
147	chr5:12603975..12604164	189	chr1	12753508	12759809	6301	inverse	6246	768	AT5G33370	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
147	chr5:12603975..12604164	189	chr5	5629977	5634971	4994	inverse	4936	774	AT5G33370	ATCOPIA93(AT5TE20395)	165
148	chr5:12949683..12950233	550	chr2	10002100	10009589	7489	sens	2202	100	.	VANDAL21(AT2TE42810)	*
149	chr5:14550370..14550707	337	chr3	10181982	10182381	399	sens	351	56	.	.	0
149	chr5:14550492..14550576	84	chr1	12753508	12759815	6307	inverse	6158	732	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
149	chr5:14550492..14550576	84	chr5	5629986	5634975	4989	inverse	4928	756	.	ATCOPIA93(AT5TE20395)	160
150	chr5:19519544..19519893	349	chr3	10182160	10182383	223	inverse	194	52	AT5G48140	.	0
150	chr5:19519546..19519749	203	chr1	12753520	12759818	6298	sens	6271	834	AT5G48140	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
150	chr5:19519546..19519749	203	chr5	5629984	5634967	4983	sens	4913	878	AT5G48140	ATCOPIA93(AT5TE20395)	179
151	chr5:21317975..21319073	1098	chr3	18631568	18632043	475	inverse	416	72	AT5G52530	ATDNA2T9A(AT3TE75575)	0
151	chr5:21318303..21318712	409	chr1	19728526	19729287	761	sens	704	114	AT5G52530	ATDNA2T9B(AT1TE65210);HELITRON2(AT1TE65205)	0
151	chr5:21318303..21318712	409	chr1	7960491	7961253	762	inverse	705	124	AT5G52530	ATDNA2T9B(AT1TE25615)	0
151	chr5:21318303..21318712	409	chr2	11983186	11983910	724	inverse	668	96	AT5G52530	ATDNA2T9B(AT2TE52165)	0
151	chr5:21318303..21318712	409	chr2	12739525	12740237	712	inverse	574	90	AT5G52530	ATDNA2T9A(AT2TE55690)	0
151	chr5:21318303..21318712	409	chr4	10539021	10539606	585	inverse	551	108	AT5G52530	ATDNA2T9A(AT4TE48190)	0
151	chr5:21318303..21318712	409	chr4	5471632	5472393	761	inverse	704	104	AT5G52530	VANDALNX2(AT4TE22920)	0
151	chr5:21318659..21318712	53	chr4	11080496	11081994	1498	sens	1468	178	AT5G52530	ATDNA2T9A(AT4TE50860)	3
151	chr5:21318660..21318712	52	chr4	2639591	2640981	1390	sens	1360	156	AT5G52530	ATDNA2T9A(AT4TE12345)	18
151	chr5:21318735..21320053	1318	chr1	26375141	26375959	818	sens	778	70	AT5G52530	ATREP10D(AT1TE86380);ATDNA2T9A(AT1TE86375)	7

151	chr5:21318735..21320053	1318	chr4	11081537	11082070	533	sens	521	50	AT5G52530	ATDNA2T9A(AT4TE50860)	3
151	chr5:21318736..21320053	1317	chr2	12738695	12739240	545	inverse	533	60	AT5G52530	ATDNA2T9A(AT2TE55690)	6
152	chr5:23373982..23374373	391	chr1	12753862	12758435	4573	inverse	3039	64	AT5G57690	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)[1]	*
153	chr5:23864824..23869728	4904	chr4	4358136	4360231	2095	sens	1996	198	AT5G59120	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
154	chr5:24028627..24028773	146	chr2	4385871	4393462	7591	sens	2790	344	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	4
154	chr5:24028643..24028773	130	chr1	11302767	11309127	6360	sens	2643	90	.	ATENSPM3(AT1TE36570)	0
154	chr5:24028648..24028773	125	chr2	4900855	4909267	8412	inverse	3551	742	.	ATENSPM3(AT2TE20205)	351
155	chr5:24196449..24197993	1544	chr1	12755359	12759037	3678	sens	3046	52	AT5G60090	ATCOPIA93(AT1TE41580)[1]	*
156	chr5:25775322..25775612	290	chr3	13072505	13078287	5782	sens	3209	250	AT5G64460	VANDAL21(AT3TE53580)	0
156	chr5:25775336..25775756	420	chr2	1265085	1271650	6565	inverse	2460	336	AT5G64460	VANDAL21(AT2TE05755)	0
156	chr5:25775384..25775570	186	chr2	1505627	1513201	7574	sens	3664	780	AT5G64460	VANDAL21(AT2TE06955)	0
156	chr5:25775442..25775570	128	chr3	11541535	11549125	7590	sens	3150	596	AT5G64460	VANDAL21(AT3TE48070)	0
156	chr5:25775460..25775570	110	chr2	10001516	10009548	8032	inverse	3671	1502	AT5G64460	VANDAL21(AT2TE42810)	653
156	chr5:25775460..25775570	110	chr5	13960862	13968467	7605	sens	3731	564	AT5G64460	VANDAL21(AT5TE49755)	0

[1] For these COPIA93 events, we manually re-ran the scoring procedure on the two donors that appear everywhere else. Indeed, these insertions were covered by such a low number of reads that only one donor could be called by the program.

[1] For these ATENSPM3 events, we manually re-ran the scoring procedure on all other potential ATENSPM3 donors that appear elsewhere. Indeed, these insertions were so close to each other that only extreme parameter sets detected them, which tended to produce an insufficient number of donors.

Supplementary Table 5: TE-Tracker output for epiRIL 60

	Acceptor		Donor									
ID	position	size	chr	start	stop	size	orientation	overlap	read count	acceptor annotation (TAIR10)	donor annotation (Quesneville natural transposons)	score
1	chr1:6261344..6261460	116	chr2	1505601	1513010	7409	inverse	3522	876	AT1G18190	VANDAL21(AT2TE06955)	2
1	chr1:6261367..6261460	93	chr2	10001481	10009628	8147	sens	3453	1734	AT1G18190	VANDAL21(AT2TE42810)	574
2	chr1:12178270..12181765	3495	chr3	15283668	15286610	2942	inverse	2935	1002	AT1G33590;AT1G33600	ATENSPM3(AT3TE62120);ATENSPM3(AT3TE62130);HELIT	*
3	chr1:14308591..14311622	3031	chr5	14524137	14524768	631	sens	535	70	.	ARNOLDY2(AT5TE52230)	*
4	chr1:16518841..16522983	4142	chr3	13586990	13587784	794	inverse	768	88	.	.	*
5	chr1:19449230..19456172	6942	chr3	1855904	1856012	108	sens	106	442	AT1G52220;AT1G52230;A	.	0
5	chr1:19449230..19456172	6942	chr4	3655374	3655992	618	inverse	596	468	AT1G52220;AT1G52230;A	.	0
5	chr1:19449230..19456172	6942	chr5	12710437	12710658	221	sens	198	470	AT1G52220;AT1G52230;A	ATENSPM5(AT5TE45000)	0
6	chr1:25956040..25956309	269	chr2	4900837	4909258	8421	sens	4697	928	.	ATENSPM3(AT2TE20205)	*
7	chr2:2845..3353	508	chr3	14201455	14205097	3642	sens	3203	510	.	.	431
7	chr2:2886..5169	2283	chr5	3253128	3253276	148	inverse	147	4492	AT2G01010	.	1320
8	chr2:9542..19176	9634	chr3	13587002	13587784	782	sens	772	100	AT2G01023	.	*
9	chr2:2676203..2684484	8281	chr1	14592405	14592674	269	sens	252	108	.	ATENSPM5(AT1TE48035)	*
10	chr2:3956333..3964953	8620	chr3	13586906	13587786	880	inverse	784	94	.	.	*
11	chr2:10321726..10321904	178	chr2	4385861	4393239	7378	sens	4367	970	AT2G24260	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	9
11	chr2:10321820..10321904	84	chr2	4900801	4909201	8400	inverse	5037	1924	AT2G24260	ATENSPM3(AT2TE20205)	641
11	chr2:10321820..10323255	1435	chr1	12945155	12946181	1026	inverse	831	308	AT2G24260	ATENSPM3(AT1TE42210)	0
12	chr2:12660556..12660713	157	chr5	13960866	13967658	6792	sens	3578	766	.	VANDAL21(AT5TE49755)	0
12	chr2:12660607..12660736	129	chr2	1505601	1513200	7599	sens	2713	1110	.	VANDAL21(AT2TE06955)	1
12	chr2:12660614..12660803	189	chr2	1265602	1271646	6044	inverse	2513	658	.	VANDAL21(AT2TE05755)	0
12	chr2:12660632..12660713	81	chr2	10001497	10009624	8127	inverse	3727	2246	.	VANDAL21(AT2TE42810)	666
12	chr2:12660632..12660713	81	chr3	11541539	11549124	7585	sens	3752	918	.	VANDAL21(AT3TE48070)	0
13	chr2:13660418..13663919	3501	chr3	15283667	15286611	2944	sens	2930	1096	AT2G32150;AT2G32160	ATENSPM3(AT3TE62120);ATENSPM3(AT3TE62130);HELIT	*
14	chr2:15677093..15677345	252	chr1	12753516	12760392	6876	sens	6846	1074	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
14	chr2:15677093..15677345	252	chr5	5629994	5635309	5315	sens	5263	1088	.	ATCOPIA93(AT5TE20395)	180
15	chr2:17095781..17097372	1591	chr1	12944728	12946120	1392	sens	1350	266	AT2G40960	ATENSPM3(AT1TE42210)	0
15	chr2:17097260..17097372	112	chr2	4900803	4909259	8456	sens	4924	1136	AT2G40960	ATENSPM3(AT2TE20205)	372
16	chr2:19390375..19394443	4068	chr3	15283666	15286591	2925	inverse	2869	562	AT2G47240	ATENSPM3(AT3TE62120);ATENSPM3(AT3TE62130);HELIT	*
17	chr3:3464127..3467214	3087	chr4	4358131	4360296	2165	inverse	2154	938	AT3G11050;AT3G11060	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
18	chr3:8761967..8764866	2899	chr3	13371219	13373958	2739	inverse	2470	310	AT3G24220	ATCOPIA78(AT3TE54550)	0
18	chr3:8764108..8764606	498	chr1	3360539	3361313	774	sens	750	188	.	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
18	chr3:8764387..8764584	197	chr1	18013168	18018238	5070	inverse	5044	1054	.	ATCOPIA78(AT1TE59755)	1
18	chr3:8764387..8764584	197	chr1	3780762	3785613	4851	inverse	4816	1246	.	ATCOPIA78(AT1TE12295)	0
18	chr3:8764387..8764606	219	chr3	22059532	22064309	4777	inverse	4706	1260	.	ATCOPIA78(AT3TE89830)	29
18	chr3:8764481..8764584	103	chr1	7714736	7722460	7724	inverse	7588	956	.	ATCOPIA78(AT1TE24850)	0
18	chr3:8764557..8764584	27	chr1	21524994	21529603	4609	sens	4501	1498	.	ATCOPIA78(AT1TE71045)	0
18	chr3:8764557..8764584	27	chr3	22695570	22700517	4947	sens	4744	1592	.	ATCOPIA78(AT3TE92525)	0
18	chr3:8764557..8764584	27	chr5	4208082	4213088	5006	sens	4967	1624	.	ATCOPIA78(AT5TE15240)	110

18	chr3:8764557..8764725	168	chr4	9320305	9320550	245	inverse	241	60	.	ATCOPIA78(AT4TE42010)	0
19	chr3:13399005..13407467	8462	chr5	971393	971583	190	sens	179	304	.	.	0
19	chr3:13399042..13407467	8425	chr4	3655372	3655993	621	sens	585	378	.	.	2
19	chr3:13399042..13407467	8425	chr5	12710434	12710660	226	inverse	222	378	.	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	0
19	chr3:13403058..13407467	4409	chr1	7024972	7025322	350	inverse	290	330	.	ATENSPM6(AT1TE22710)	0
20	chr3:13590651..13591215	564	chr4	3953384	3953665	281	inverse	228	126	.	.	*
21	chr3:14194760..14199594	4834	chr2	9902	10224	322	sens	306	176	AT3G41761;AT3G41762;A	.	241
21	chr3:14196645..14199140	2495	chr5	3253128	3253276	148	inverse	145	4896	AT3G41768	.	2132
21	chr3:14196919..14201208	4289	chr2	7190932	7191520	588	sens	529	1220	AT3G41768;AT3G41979	.	403
21	chr3:14197118..14201455	4337	chr2	2594	3503	909	sens	868	532	AT3G41768;AT3G41979	.	365
22	chr3:14201682..14203672	1990	chr2	9888	10137	249	sens	209	146	.	.	*
23	chr3:15773104..15778796	5692	chr4	4358130	4360295	2165	sens	2109	954	AT3G43960;AT3G43980;A	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
24	chr3:16946117..16951602	5485	chr4	4358128	4360296	2168	sens	2137	498	AT3G46130;AT3G46140;A	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
25	chr3:18274281..18277593	3312	chr3	15283667	15286610	2943	inverse	2933	1012	AT3G49290;AT3G49300	ATENSPM3(AT3TE62120);ATENSPM3(AT3TE62130);HELIT	*
26	chr3:22615358..22615562	204	chr3	22695570	22700517	4947	inverse	4911	2008	AT3G61113	ATCOPIA78(AT3TE92525)	*
27	chr4:1151189..1152101	912	chr3	13075454	13076263	809	sens	734	130	AT4G02620;AT4G02630	VANDAL21(AT3TE53580)	0
27	chr4:1151429..1152101	672	chr2	1266226	1271573	5347	inverse	1624	564	AT4G02630	VANDAL21(AT2TE05755)	0
27	chr4:1151507..1151621	114	chr2	1505628	1513198	7570	sens	3317	1130	AT4G02630	VANDAL21(AT2TE06955)	0
27	chr4:1151507..1151621	114	chr3	11541536	11549122	7586	sens	3273	868	AT4G02630	VANDAL21(AT3TE48070)	0
27	chr4:1151507..1151621	114	chr5	13960863	13968149	7286	sens	2999	770	AT4G02630	VANDAL21(AT5TE49755)	0
27	chr4:1151516..1151621	105	chr2	10001482	10009649	8167	inverse	3312	2116	AT4G02630	VANDAL21(AT2TE42810)	636
28	chr4:3737409..3745115	7706	chr3	13587628	13587782	154	inverse	129	56	.	.	*
29	chr4:3995606..3999331	3725	chr3	15719702	15720170	468	sens	463	62	.	ATHILA2(AT3TE63540)	*
30	chr4:4781471..4781747	276	chr1	16511493	16511826	333	inverse	293	90	.	ATHILA5(AT1TE54480)	*
31	chr4:5859205..5859488	283	chr2	10001510	10006094	4584	inverse	1807	486	AT4G09200	VANDAL21(AT2TE42810)	*
32	chr4:5876408..5876685	277	chr2	10001505	10009634	8129	inverse	2917	1076	AT4G09250	VANDAL21(AT2TE42810)	*
33	chr4:9320049..9322399	2350	chr1	3360538	3361317	779	inverse	755	230	AT4G16550	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
33	chr4:9320088..9320618	530	chr5	4208081	4213015	4934	inverse	4884	1644	AT4G16550	ATCOPIA78(AT5TE15240)	0
33	chr4:9320133..9320618	485	chr3	13369244	13373872	4628	sens	4555	1016	AT4G16550	ATCOPIA78(AT3TE54550)	1
33	chr4:9320164..9320669	505	chr1	7714745	7722365	7620	sens	4237	1404	AT4G16550	ATCOPIA78(AT1TE24850)	1
33	chr4:9320309..9320618	309	chr1	21524995	21529536	4541	inverse	4469	1642	AT4G16550	ATCOPIA78(AT1TE71045)	16
33	chr4:9320385..9320618	233	chr1	18013201	18018118	4917	sens	2087	862	AT4G16550	ATCOPIA78(AT1TE59755)	0
33	chr4:9320385..9320618	233	chr1	3780804	3785355	4551	sens	408	752	AT4G16550	ATCOPIA78(AT1TE12295)	0
33	chr4:9320385..9320618	233	chr3	22059574	22063868	4294	sens	4109	1060	AT4G16550	ATCOPIA78(AT3TE89830)	0
33	chr4:9320385..9320618	233	chr3	22695569	22700128	4559	inverse	4308	1652	AT4G16550	ATCOPIA78(AT3TE92525)	0
34	chr4:9650036..9650129	93	chr3	10181981	10182388	407	inverse	379	156	AT4G17190	.	0
34	chr4:9650038..9650112	74	chr1	12753511	12760293	6782	sens	6373	2508	AT4G17190	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	2
34	chr4:9650038..9650112	74	chr5	5629979	5635286	5307	sens	5291	2544	AT4G17190	ATCOPIA93(AT5TE20395)	412
35	chr4:10898129..10907323	9194	chr3	14958428	14958946	518	inverse	507	76	AT4G20170;AT4G20190	META1(AT3TE61000)	0
35	chr4:10898166..10906937	8771	chr1	4407443	4408639	1196	inverse	1112	128	AT4G20170;AT4G20190	META1(AT1TE14315)	0
35	chr4:10898166..10907323	9157	chr1	6541575	6542093	518	sens	507	82	AT4G20170;AT4G20190	META1(AT1TE21210)	0
35	chr4:10898166..10907323	9157	chr2	1397567	1398086	519	sens	508	84	AT4G20170;AT4G20190	META1(AT2TE06405)	0
35	chr4:10898166..10907323	9157	chr3	15816351	15816819	468	sens	451	60	AT4G20170;AT4G20190	META1(AT3TE63950)	0
35	chr4:10898166..10907323	9157	chr3	16040469	16040987	518	inverse	507	88	AT4G20170;AT4G20190	META1(AT3TE64915)	0

35	chr4:10898166..10907323	9157	chr3	17364505	17365002	497	inverse	489	82	AT4G20170;AT4G20190	META1(AT3TE70410)	0
35	chr4:10898166..10907323	9157	chr3	18737435	18737953	518	sens	507	68	AT4G20170;AT4G20190	META1(AT3TE76010)	0
35	chr4:10898166..10907690	9524	chr2	912403	912901	498	sens	487	66	AT4G20170;AT4G20190	META1(AT2TE04085)	0
35	chr4:10898166..10907690	9524	chr5	10402169	10402687	518	inverse	507	84	AT4G20170;AT4G20190	META1(AT5TE37975)	0
35	chr4:10898166..10907690	9524	chr5	14412516	14412994	478	sens	463	62	AT4G20170;AT4G20190	META1(AT5TE51845)	0
35	chr4:10898166..10907690	9524	chr5	14471816	14472294	478	sens	463	62	AT4G20170;AT4G20190	META1(AT5TE52075)	0
35	chr4:10898166..10907690	9524	chr5	9383141	9383660	519	inverse	508	86	AT4G20170;AT4G20190	META1(AT5TE34040)	0
35	chr4:10898166..10907818	9652	chr1	6707449	6707893	444	sens	423	64	AT4G20170;AT4G20190	META1(AT1TE21710)	0
36	chr5:11707257..11708113	856	chr4	3953439	3953826	387	sens	325	44	.	.	*
37	chr5:16740222..16745361	5139	chr4	4358148	4360294	2146	sens	1984	402	AT5G41820;AT5G41830	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
38	chr5:17050129..17053285	3156	chr4	4358190	4360292	2102	sens	2077	380	.	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
39	chr5:21318051..21318874	823	chr1	7960489	7961242	753	inverse	739	166	AT5G52530	ATDNA2T9B(AT1TE25615)	4
39	chr5:21318051..21318874	823	chr2	11983184	11983905	721	inverse	718	126	AT5G52530	ATDNA2T9B(AT2TE52165)	0
39	chr5:21318051..21318874	823	chr3	18631299	18631992	693	inverse	692	170	AT5G52530	ATDNA2T9A(AT3TE75575)	0
39	chr5:21318051..21318874	823	chr4	10538842	10539536	694	inverse	693	170	AT5G52530	ATDNA2T9A(AT4TE48190)	0
39	chr5:21318051..21318874	823	chr4	5471630	5472399	769	inverse	768	144	AT5G52530	VANDALNX2(AT4TE22920)	4
39	chr5:21318051..21318929	878	chr1	19728520	19729289	769	sens	768	158	AT5G52530	ATDNA2T9B(AT1TE65210);HELITRON2(AT1TE65205)	0
39	chr5:21318310..21318874	564	chr2	12738712	12740167	1455	inverse	1267	172	AT5G52530	ATDNA2T9A(AT2TE55690)	3
39	chr5:21318840..21318874	34	chr4	2639585	2640990	1405	sens	1405	218	AT5G52530	ATDNA2T9A(AT4TE12345)	16
39	chr5:21318845..21318874	29	chr4	11080490	11081983	1493	sens	1469	224	AT5G52530	ATDNA2T9A(AT4TE50860)	1

Supplementary Table 6: TE-Tracker output for epiRIL 439

	Acceptor		Donor									
ID	position	size	chr	start	stop	size	orientation	overlap	read count	acceptor annotation (TAIR10)	donor annotation (Quesneville natural transposons)	score
1	chr1:6621510..6622359	849	chr1	3780767	3785258	4478	inverse	4491	1706	AT1G19180	ATCOPIA78(AT1TE12295)	0
1	chr1:6621629..6622359	730	chr1	7717524	7722256	4432	inverse	4732	620	AT1G19180	ATCOPIA78(AT1TE24850)	0
1	chr1:6622143..6622359	216	chr5	4208084	4213083	4978	sens	4999	1812	AT1G19180	ATCOPIA78(AT5TE15240)	0
1	chr1:6622211..6622359	148	chr1	21524996	21529545	4390	sens	4549	1776	AT1G19180	ATCOPIA78(AT1TE71045)	0
1	chr1:6622211..6622359	148	chr3	22059537	22064329	4754	inverse	4792	1508	AT1G19180	ATCOPIA78(AT3TE89830)	0
1	chr1:6622219..6622359	140	chr3	22695567	22700520	4932	sens	4953	1868	AT1G19180	ATCOPIA78(AT3TE92525)	0
1	chr1:6622219..6622410	191	chr1	18013167	18018387	5145	inverse	5220	1718	AT1G19180	ATCOPIA78(AT1TE59755)	134
1	chr1:6622219..6622513	294	chr3	13369246	13373955	4691	inverse	4709	778	AT1G19180	ATCOPIA78(AT3TE54550)	0
2	chr1:6622443..6622524	81	chr1	3780909	3785144	3371	inverse	4235	1430	AT1G19180	ATCOPIA78(AT1TE12295)	*
2	chr1:11207269..11216121	8852	chr3	10181982	10182380	382	sens	398	58	AT1G31319;AT1G31320;A	.	1
2	chr1:11211710..11211794	84	chr1	12753522	12760389	6831	inverse	6867	1036	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
3	chr1:11211722..11211794	72	chr5	5629978	5635310	5307	inverse	5332	1020	.	ATCOPIA93(AT5TE20395)	172
3	chr1:19450482..19455687	5205	chr3	1855904	1856012	107	sens	108	724	AT1G52220;AT1G52230	.	0
3	chr1:19450482..19455687	5205	chr4	3655376	3655991	613	inverse	615	824	AT1G52220;AT1G52230	.	1
3	chr1:19450482..19455687	5205	chr5	12710433	12710660	222	sens	227	814	AT1G52220;AT1G52230	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	1
3	chr1:19450482..19455687	5205	chr5	18311648	18311777	124	inverse	129	418	AT1G52220;AT1G52230	ATCOPIA68(AT5TE65910)	4
4	chr1:19450482..19455687	5205	chr5	971392	971584	187	inverse	192	636	AT1G52220;AT1G52230	.	0
4	chr1:24091108..24091194	86	chr2	4385853	4393465	2078	inverse	7612	956	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	17
5	chr1:24091117..24091194	77	chr2	4900803	4909285	2209	sens	8482	1896	.	ATENSPM3(AT2TE20205);ATENSPM7(AT2TE20210)	621
5	chr1:24334468..24334564	96	chr2	4385854	4393659	2025	sens	7805	974	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	14
6	chr1:24334475..24334558	83	chr2	4900805	4909284	2147	inverse	8479	1970	.	ATENSPM3(AT2TE20205);ATENSPM7(AT2TE20210)	637
6	chr1:28631078..28631250	172	chr2	10001501	10009630	4812	inverse	8129	1778	AT1G76320	VANDAL21(AT2TE42810)	577
7	chr1:28631154..28631280	126	chr2	1505616	1510807	2808	sens	5191	634	AT1G76320	VANDAL21(AT2TE06955)	0
8	chr2:2880..3288	408	chr5	3253117	3253276	146	inverse	159	5206	.	.	*
9	chr2:3350..3385	35	chr3	14199585	14205119	106	sens	5534	714	.	.	*
9	chr2:5378..13175	7797	chr3	13586990	13587786	793	sens	796	98	AT2G01010;AT2G01020;A	.	0
9	chr2:6461..9320	2859	chr5	3253128	3253275	142	inverse	147	164	AT2G01021	.	2049
10	chr2:9019..13095	4076	chr3	13589280	13592515	3131	inverse	3235	386	AT2G01023	.	18
11	chr2:22685..23066	381	chr1	15081963	15082400	418	inverse	437	84	.	.	*
11	chr2:49192..58749	9557	chr1	15081963	15082400	423	inverse	437	88	.	.	0
12	chr2:49574..58692	9118	chr3	13590275	13592521	2096	inverse	2246	278	.	.	0
12	chr2:79850..80054	204	chr2	10001483	10009651	2649	inverse	8168	1786	AT2G01090	VANDAL21(AT2TE42810)	578
13	chr2:79892..80054	162	chr2	1505597	1513159	2555	sens	7562	898	AT2G01090	VANDAL21(AT2TE06955)	0
14	chr2:2676251..2684592	8341	chr1	14592402	14592667	240	sens	265	98	.	ATENSPM5(AT1TE48035)	*
15	chr2:3250447..3250633	186	chr2	3359041	3368453	211	inverse	9412	1424	.	.	*
16	chr2:3954731..3955112	381	chr3	13588796	13592574	3637	sens	3778	468	.	.	*
17	chr2:3956348..3964907	8559	chr3	13588796	13592574	3541	sens	3778	510	.	.	*
18	chr2:7190966..7191402	436	chr3	14199585	14204974	106	sens	5389	980	AT2G16586	.	*

18	chr2:10506730..10513564	6834	chr3	14083717	14083984	248	inverse	267	46	AT2G24693;AT2G24696;A	.	0
19	chr2:10508113..10512570	4457	chr4	4358130	4360294	2115	inverse	2164	854	AT2G24696	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	399
19	chr2:11505151..11505860	709	chr1	12942845	12947004	198	sens	4159	584	.	ATENSPM3(AT1TE42210)	58
19	chr2:11505759..11505852	93	chr2	4385864	4393387	2014	inverse	7523	992	.	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	13
20	chr2:11505759..11505852	93	chr2	4900801	4909274	2237	sens	8473	1960	.	ATENSPM3(AT2TE20205)	656
20	chr3:8760306..8764533	4227	chr4	9320299	9320549	220	inverse	250	80	AT3G24220	ATCOPIA78(AT4TE42010)	0
20	chr3:8762528..8764535	2007	chr3	13371219	13373932	2201	inverse	2713	382	AT3G24220	ATCOPIA78(AT3TE54550)	0
20	chr3:8764207..8764533	326	chr1	18013165	18018309	5054	inverse	5144	1208	.	ATCOPIA78(AT1TE59755)	1
20	chr3:8764264..8764535	271	chr1	7717814	7722471	4368	inverse	4657	1000	.	ATCOPIA78(AT1TE24850)	0
20	chr3:8764283..8764533	250	chr1	7714733	7722202	7455	inverse	7469	1114	.	ATCOPIA78(AT1TE24850)	0
20	chr3:8764283..8764598	315	chr1	3360541	3361202	659	sens	661	172	.	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
20	chr3:8764355..8764535	180	chr5	4208084	4213082	4916	sens	4998	1756	.	ATCOPIA78(AT5TE15240)	83
20	chr3:8764409..8764434	25	chr1	3780764	3785495	4719	inverse	4731	1414	.	ATCOPIA78(AT1TE12295)	0
20	chr3:8764431..8764434	3	chr1	21524991	21529604	4606	sens	4613	1612	.	ATCOPIA78(AT1TE71045)	0
20	chr3:8764431..8764434	3	chr3	22059538	22064329	4788	inverse	4791	1388	.	ATCOPIA78(AT3TE89830)	36
21	chr3:8764431..8764434	3	chr3	22695567	22700519	4931	sens	4952	1730	.	ATCOPIA78(AT3TE92525)	0
21	chr3:13399161..13406892	7731	chr1	7024972	7025326	351	inverse	354	492	.	ATENSPM6(AT1TE22710)	0
21	chr3:13399161..13406892	7731	chr4	3655372	3655988	610	sens	616	550	.	.	1
21	chr3:13399161..13406892	7731	chr5	12710437	12710662	221	inverse	225	554	.	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	0
21	chr3:13399161..13406892	7731	chr5	18311647	18311775	125	sens	128	314	.	ATCOPIA68(AT5TE65910)	0
21	chr3:13399161..13407071	7910	chr5	971392	971582	186	sens	190	436	.	.	0
22	chr3:13403059..13407140	4081	chr1	15214383	15214509	124	inverse	126	242	.	ATENSPM5(AT1TE49955)	2
23	chr3:13590654..13591070	416	chr2	3607057	3607369	261	inverse	312	58	.	.	*
23	chr3:14194018..14195444	1426	chr5	3253130	3253276	135	inverse	146	202	.	.	2904
23	chr3:14194221..14203653	9432	chr2	9895	10166	234	sens	271	226	AT3G41761;AT3G41762;A	.	499
23	chr3:14196845..14198867	2022	chr5	3253128	3253276	146	inverse	148	5468	AT3G41768	.	2904
23	chr3:14196913..14199585	2672	chr2	7190932	7191521	584	sens	589	1418	AT3G41768	.	484
23	chr3:14197041..14199585	2544	chr2	2594	3428	831	sens	834	1004	AT3G41768	.	657
24	chr3:14199646..14200871	1225	chr2	3626906	3627061	125	sens	155	44	AT3G41979	.	*
24	chr3:21970792..21970868	76	chr1	12753508	12760283	5762	sens	6775	1242	AT3G59440	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
25	chr3:21970792..21970868	76	chr5	5629978	5635281	5250	sens	5303	1258	AT3G59440	ATCOPIA93(AT5TE20395)	185
26	chr3:22615329..22615551	222	chr3	22695569	22700475	4815	inverse	4906	1868	AT3G61113	ATCOPIA78(AT3TE92525)	*
26	chr4:3951195..3953518	2323	chr3	13590512	13590813	275	inverse	301	58	.	.	14
27	chr4:3952695..3953397	702	chr3	13587636	13587795	130	sens	159	60	.	.	1
28	chr4:3957395..3965682	8287	chr2	3605770	3607825	1993	sens	2055	248	.	.	*
28	chr4:3973014..3982766	9752	chr1	15192391	15192772	330	sens	381	64	.	ATHILA2(AT1TE49850)	0
28	chr4:3973030..3982086	9056	chr1	14429589	14429903	259	sens	314	46	.	ATHILA2(AT1TE47545);ATHILA4C(AT1TE47540)	0
28	chr4:3974767..3983490	8723	chr2	5578359	5578923	470	sens	564	68	.	ATHILA2(AT2TE22520)	0
29	chr4:3982030..3983631	1601	chr3	12581658	12581930	225	inverse	272	52	.	ATHILA2(AT3TE51810)	0
30	chr4:3993295..3993368	73	chr1	15700289	15700541	236	inverse	252	82	.	ATGP1(AT1TE51700);ATHILA(AT1TE51705)	*
31	chr4:3995819..3998899	3080	chr3	15719640	15720158	429	sens	518	64	.	ATHILA2(AT3TE63540)	*
31	chr4:4002294..4006156	3862	chr1	15191611	15192893	1192	sens	1282	160	.	ATHILA2(AT1TE49850)	0
32	chr4:4004021..4006240	2219	chr1	14450684	14450942	230	sens	258	56	.	ATHILA2(AT1TE47570)	0
32	chr4:9320313..9320913	600	chr1	7717840	7722102	674	sens	4262	648	AT4G16550	ATCOPIA78(AT1TE24850)	0

32	chr4:9320395..9320651	256	chr1	18013204	18018260	4940	sens	5056	1564	AT4G16550	ATCOPIA78(AT1TE59755)	0
32	chr4:9320395..9320651	256	chr1	3780807	3785669	4538	sens	4862	1658	AT4G16550	ATCOPIA78(AT1TE12295)	0
32	chr4:9320395..9320651	256	chr1	7714748	7722341	7348	sens	7593	1458	AT4G16550	ATCOPIA78(AT1TE24850)	0
32	chr4:9320395..9320651	256	chr3	13369244	13373893	4542	sens	4649	1136	AT4G16550	ATCOPIA78(AT3TE54550)	0
32	chr4:9320395..9320651	256	chr3	22059577	22064316	4623	sens	4739	1520	AT4G16550	ATCOPIA78(AT3TE89830)	1
32	chr4:9320395..9320651	256	chr3	22695565	22700050	4153	inverse	4485	1616	AT4G16550	ATCOPIA78(AT3TE92525)	0
32	chr4:9320395..9320651	256	chr5	4208082	4212992	4864	inverse	4910	1610	AT4G16550	ATCOPIA78(AT5TE15240)	0
32	chr4:9320395..9320663	268	chr1	21524997	21529547	4535	inverse	4550	1622	AT4G16550	ATCOPIA78(AT1TE71045)	14
33	chr4:9320395..9322180	1785	chr1	3360539	3361311	754	inverse	772	228	AT4G16550	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
33	chr4:10739462..10739998	536	chr1	3360538	3361315	745	sens	777	268	AT4G19740	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
33	chr4:10739611..10740238	627	chr1	7717826	7722130	4297	inverse	4304	1284	AT4G19740	ATCOPIA78(AT1TE24850)	0
33	chr4:10739688..10739875	187	chr1	21524993	21529526	4517	sens	4533	1664	AT4G19740	ATCOPIA78(AT1TE71045)	0
33	chr4:10739764..10739875	111	chr1	18013169	18017742	4565	inverse	4573	1516	AT4G19740	ATCOPIA78(AT1TE59755)	0
33	chr4:10739764..10739875	111	chr1	3780772	3785502	4569	inverse	4730	1708	AT4G19740	ATCOPIA78(AT1TE12295)	1
33	chr4:10739764..10739875	111	chr3	22059542	22063886	4332	inverse	4344	1722	AT4G19740	ATCOPIA78(AT3TE89830)	40
33	chr4:10739764..10739875	111	chr3	22695564	22700474	4569	sens	4910	1756	AT4G19740	ATCOPIA78(AT3TE92520);ATCOPIA78(AT3TE92525)	26
33	chr4:10739764..10739916	152	chr1	7714739	7722144	7378	inverse	7405	1472	AT4G19740	ATCOPIA78(AT1TE24850)	0
33	chr4:10739764..10739916	152	chr3	13369263	13373850	4478	inverse	4587	1008	AT4G19740	ATCOPIA78(AT3TE54550)	0
34	chr4:10739764..10739916	152	chr5	4208083	4213080	4950	sens	4997	1582	AT4G19740	ATCOPIA78(AT5TE15240)	0
34	chr4:10898431..10907479	9048	chr5	14412500	14412990	480	sens	490	92	AT4G20170;AT4G20190	META1(AT5TE51845)	0
34	chr4:10898431..10907580	9149	chr5	14471800	14472290	480	sens	490	92	AT4G20170;AT4G20190	META1(AT5TE52075)	0
34	chr4:10898555..10907479	8924	chr1	6707358	6707896	512	sens	538	68	AT4G20170;AT4G20190	META1(AT1TE21710)	0
34	chr4:10898555..10907479	8924	chr5	9383138	9383676	515	inverse	538	104	AT4G20170;AT4G20190	META1(AT5TE34040)	0
34	chr4:10898555..10907580	9025	chr5	10402166	10402703	485	inverse	537	100	AT4G20170;AT4G20190	META1(AT5TE37975)	0
34	chr4:10900819..10907580	6761	chr3	16040466	16041003	438	inverse	537	112	AT4G20190	META1(AT3TE64915)	0
34	chr4:10900819..10907580	6761	chr3	17364502	17364999	406	inverse	497	80	AT4G20190	META1(AT3TE70410)	0
34	chr4:10900895..10904094	3199	chr1	4407440	4408624	1164	inverse	1184	156	.	META1(AT1TE14315)	0
34	chr4:10900895..10904094	3199	chr1	6540914	6542096	1162	sens	1182	150	.	META1(AT1TE21210)	0
34	chr4:10900895..10907479	6584	chr2	1397556	1398089	500	sens	533	102	AT4G20190	META1(AT2TE06405)	0
34	chr4:10900895..10907479	6584	chr3	14958425	14958962	514	inverse	537	92	AT4G20190	META1(AT3TE61000)	0
34	chr4:10900895..10907479	6584	chr3	15816335	15816825	480	sens	490	100	AT4G20190	META1(AT3TE63950)	0
34	chr4:10900895..10907580	6685	chr2	912387	912904	494	sens	517	90	AT4G20190	META1(AT2TE04085)	0
35	chr4:10900895..10907580	6685	chr3	18737432	18737956	499	sens	524	74	AT4G20190	META1(AT3TE76010)	0
36	chr5:969696..971395	1699	chr4	4358131	4360292	2137	sens	2161	584	AT5G03705;AT5G03710	ATENSPM3(AT4TE18510);ATLANTYS1(AT4TE18500);ATEN	*
36	chr5:6591834..6592026	192	chr1	12753507	12758416	4826	sens	4909	810	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
37	chr5:6591834..6592026	192	chr5	5629983	5634958	4911	sens	4975	856	.	ATCOPIA93(AT5TE20395)	127
37	chr5:7815368..7815495	127	chr2	4385859	4393661	2061	sens	7802	968	AT5G23212	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	2
38	chr5:7815404..7815488	84	chr2	4900819	4909279	2463	inverse	8460	1874	AT5G23212	ATENSPM3(AT2TE20205)	681
39	chr5:11730792..11733078	2286	chr4	3953439	3953724	241	inverse	285	88	.	.	*
40	chr5:11733747..11734631	884	chr4	3953388	3953767	377	inverse	379	84	.	.	*
40	chr5:21318238..21318829	591	chr2	11983205	11983903	682	inverse	698	118	AT5G52530	ATDNA2T9B(AT2TE52165)	0
40	chr5:21318351..21318829	478	chr1	19728523	19729268	729	sens	745	146	AT5G52530	ATDNA2T9B(AT1TE65210);HELITRON2(AT1TE65205)	1
40	chr5:21318351..21318829	478	chr1	7960510	7961256	730	inverse	746	162	AT5G52530	ATDNA2T9B(AT1TE25615)	2
40	chr5:21318351..21318829	478	chr2	12739544	12740243	683	inverse	699	132	AT5G52530	ATDNA2T9A(AT2TE55690)	0

40	chr5:21318351..21318829	478	chr4	10538863	10539609	730	inverse	746	160	AT5G52530	ATDNA2T9A(AT4TE48190)	0
40	chr5:21318351..21318969	618	chr4	5471651	5472395	658	inverse	744	126	AT5G52530	VANDALNX2(AT4TE22920)	1
40	chr5:21318681..21318829	148	chr3	18630715	18632069	1210	inverse	1354	190	AT5G52530	ATDNA2T9A(AT3TE75575)	1
40	chr5:21318819..21318829	10	chr4	2639584	2640989	1385	sens	1405	226	AT5G52530	ATDNA2T9A(AT4TE12345)	13
41	chr5:21318822..21318829	7	chr4	11080493	11081989	1443	sens	1496	232	AT5G52530	ATDNA2T9A(AT4TE50860)	5
41	chr5:21318898..21319536	638	chr1	26375193	26375954	699	sens	761	92	AT5G52530	ATDNA2T9A(AT1TE86375)	3
42	chr5:21318898..21319536	638	chr2	12738700	12739330	577	inverse	630	88	AT5G52530	ATDNA2T9A(AT2TE55690)	5

Supplementary Table 7: TE-Tracker output for epiRIL 454

	Acceptor		Donor									
ID	position	size	chr	start	stop	size	orientation	overlap	read count	acceptor annotation (TAIR10)	donor annotation (Quesneville natural transposons)	score
1	chr1:3039..3249	210	chr1	12753520	12760374	6854	sens	6841	898	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
1	chr1:3115..3249	134	chr5	5629979	5635291	5312	sens	5273	794	.	ATCOPIA93(AT5TE20395)	137
2	chr1:8775539..8776909	1370	chr5	14175236	14176776	1540	sens	1450	48	AT1G24822;AT1G24880	VANDAL5(AT5TE50620);VANDAL5A(AT5TE50615);VANDA	*
3	chr1:11191545..11191795	250	chr1	12753512	12760369	6857	inverse	6825	530	AT1G31290	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
3	chr1:11191545..11191795	250	chr5	5629978	5635286	5308	inverse	5217	520	AT1G31290	ATCOPIA93(AT5TE20395)	109
4	chr1:15040959..15041648	689	chr2	3605696	3607809	2113	inverse	2051	64	.	.	*
5	chr1:15042139..15043682	1543	chr2	3626173	3627899	1726	inverse	1720	48	.	.	*
6	chr1:15083368..15083679	311	chr2	4904200	4908816	4616	sens	806	168	AT1G40104	ATENSPM3(AT2TE20205)	*
7	chr1:15190796..15191671	875	chr3	13588839	13592548	3709	inverse	3628	142	.	.	*
8	chr1:16511530..16519208	7678	chr2	3625634	3627398	1764	sens	1709	48	.	.	12
8	chr1:16518908..16519291	383	chr3	13586990	13592588	5598	inverse	5431	208	.	.	17
8	chr1:16518913..16519291	378	chr4	3950517	3956013	5496	sens	5310	164	.	.	18
9	chr1:16519739..16519838	99	chr4	3950650	3955912	5262	sens	5190	158	.	.	*
10	chr1:16520274..16520543	269	chr2	3605742	3607833	2091	sens	2025	62	.	.	*
11	chr1:16520649..16521032	383	chr4	3950605	3955930	5325	inverse	5279	194	.	.	28
11	chr1:16520935..16521032	97	chr3	13587015	13592426	5411	sens	2952	150	.	.	26
12	chr1:19449535..19456944	7409	chr3	1855904	1856012	108	sens	106	358	AT1G52220;AT1G52230;A	.	0
12	chr1:19449535..19456944	7409	chr4	3655376	3655993	617	inverse	609	406	AT1G52220;AT1G52230;A	.	0
12	chr1:19449535..19456944	7409	chr5	12710435	12710660	225	sens	222	402	AT1G52220;AT1G52230;A	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	0
12	chr1:19449535..19456944	7409	chr5	18311649	18311775	126	inverse	122	206	AT1G52220;AT1G52230;A	ATCOPIA68(AT5TE65910)	7
12	chr1:19449535..19456944	7409	chr5	971392	971583	191	inverse	187	324	AT1G52220;AT1G52230;A	.	0
13	chr1:25653718..25654122	404	chr1	12753506	12759792	6286	inverse	6166	996	AT1G68420	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
13	chr1:25653718..25654122	404	chr5	5629983	5635303	5320	inverse	5300	1088	AT1G68420	ATCOPIA93(AT5TE20395)	205
14	chr1:29616178..29616578	400	chr1	12753608	12759721	6113	inverse	6043	398	AT1G78760	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
14	chr1:29616254..29616578	324	chr5	5630010	5635273	5263	inverse	5173	440	AT1G78760	ATCOPIA93(AT5TE20395)	77
15	chr2:1967..3746	1779	chr3	13588834	13592484	3650	inverse	3402	124	AT2G01008;AT2G01010	.	34
15	chr2:2752..3262	510	chr4	3951866	3955719	3853	sens	3504	128	AT2G01008	.	19
15	chr2:2831..3356	525	chr3	14199590	14204425	4835	sens	321	246	.	.	1446
15	chr2:2889..3115	226	chr5	3253128	3253276	148	inverse	147	3328	.	.	1731
16	chr2:5080..6875	1795	chr5	11730437	11732721	2284	sens	1961	72	AT2G01010;AT2G01020;A	.	13
16	chr2:5316..8209	2893	chr1	15081963	15083298	1335	sens	1186	72	AT2G01010;AT2G01020;A	.	4
16	chr2:5716..14529	8813	chr4	3950852	3954016	3164	sens	2909	158	AT2G01020;AT2G01021;A	.	30
16	chr2:6695..6820	125	chr3	13588954	13592519	3565	sens	3465	372	.	.	35
16	chr2:6862..7002	140	chr2	4238704	4242968	4264	inverse	3675	152	.	ATGP2(AT2TE17875)	*
16	chr2:7213..8042	829	chr3	13588842	13592186	3344	sens	3287	136	.	.	34
16	chr2:7835..18482	10647	chr3	13586989	13587786	797	sens	647	66	AT2G01023	.	1
16	chr2:8781..8795	14	chr3	13586107	13592422	6315	inverse	4757	350	.	ATHILA6A(AT3TE55255)	36
16	chr2:16710..17258	548	chr4	3950841	3955773	4932	sens	3501	270	.	.	44

16	chr2:16729..17758	1029	chr3	13588923	13592518	3595	inverse	3135	146	.	.	1
16	chr2:17244..18152	908	chr1	15081964	15085412	3448	inverse	3436	104	.	.	1
16	chr2:9638..13714	4076	chr5	11732040	11733865	1825	inverse	1608	62	AT2G01023	.	1
16	chr2:9933..17758	7825	chr2	3605679	3607838	2159	sens	2098	212	.	.	0
17	chr2:23027..23564	537	chr3	13586996	13592503	5507	inverse	3186	266	.	.	0
17	chr2:23342..24326	984	chr4	3950656	3955792	5136	sens	4857	338	.	.	48
18	chr2:25365..25665	300	chr4	3950897	3955181	4284	sens	3813	114	.	.	*
19	chr2:44680..45308	628	chr3	13588923	13592507	3584	inverse	3178	106	.	.	*
20	chr2:48247..48628	381	chr4	3950841	3955792	4951	sens	3393	308	.	.	*
21	chr2:48642..59042	10400	chr1	15081964	15082410	446	inverse	433	64	.	.	0
21	chr2:49656..58816	9160	chr3	13588915	13592495	3580	inverse	3182	260	.	.	0
22	chr2:60978..61202	224	chr4	3952186	3953815	1629	sens	1550	58	.	.	*
23	chr2:2563525..2563683	158	chr3	13588789	13592518	3729	inverse	3158	144	.	.	0
23	chr2:2563525..2564031	506	chr4	3950841	3955723	4882	sens	1707	226	.	.	13
24	chr2:3226598..3227144	546	chr4	3950990	3955516	4526	sens	3153	160	.	.	12
24	chr2:3226598..3227212	614	chr3	13589304	13592518	3214	inverse	3175	130	.	.	0
25	chr2:3486635..3490634	3999	chr5	10764368	10765373	1005	sens	806	54	AT2G07806	ATGP1(AT5TE39170)	0
25	chr2:3487212..3490611	3399	chr1	16360287	16360997	710	inverse	600	60	AT2G07806	ATGP1(AT1TE53915)	0
25	chr2:3488109..3489072	963	chr1	13883175	13890604	7429	sens	3419	222	AT2G07806	ATGP1(AT1TE45560)	0
25	chr2:3488185..3489052	867	chr3	12613334	12620849	7515	inverse	3928	274	AT2G07806	ATGP1(AT3TE51930)	1
25	chr2:3488185..3489052	867	chr5	15491130	15498563	7433	inverse	7418	206	AT2G07806	ATGP1(AT5TE56030)	0
25	chr2:3488356..3488488	132	chr2	1627851	1639885	12034	inverse	4071	1080	AT2G07806	ATGP1(AT2TE07550);ATLINE1A(AT2TE07540);ATGP1(AT2T	394
26	chr2:3620217..3620909	692	chr3	14199248	14204186	4938	inverse	1923	142	AT2G08986	.	*
27	chr2:3622656..3623325	669	chr1	15191369	15192880	1511	inverse	1243	48	AT2G08986	ATHILA2(AT1TE49850)	*
28	chr2:3623430..3626333	2903	chr3	14191794	14194122	2328	inverse	2001	62	AT2G08986	.	*
29	chr2:3626361..3626819	458	chr3	14194935	14200309	5374	sens	3841	158	AT2G08986	.	*
30	chr2:3954597..3964908	10311	chr1	15082080	15085827	3747	sens	3293	116	.	.	6
30	chr2:3956350..3964908	8558	chr3	13588803	13592548	3745	sens	3006	238	.	.	21
31	chr2:4237782..4238103	321	chr3	13589296	13592498	3202	sens	3169	244	.	.	0
31	chr2:4237901..4238041	140	chr4	3950606	3955540	4934	inverse	4583	410	.	.	14
32	chr2:4240495..4240855	360	chr3	13590444	13592502	2058	sens	2034	94	.	.	*
33	chr2:4243087..4243606	519	chr4	3950841	3955647	4806	inverse	4099	280	.	.	*
34	chr2:5566891..5577751	10860	chr3	13588887	13592431	3544	inverse	3362	96	.	.	*
35	chr2:11496506..11496664	158	chr1	12753512	12760305	6793	sens	5035	522	AT2G26940	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
35	chr2:11496506..11496664	158	chr5	5629976	5635276	5300	sens	5273	566	AT2G26940	ATCOPIA93(AT5TE20395)	94
36	chr3:807190..807542	352	chr1	12753574	12760366	6792	inverse	6338	436	AT3G03380	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
36	chr3:807190..807542	352	chr5	5629989	5635283	5294	inverse	5103	424	AT3G03380	ATCOPIA93(AT5TE20395)	75
37	chr3:1737581..1737875	294	chr1	12753538	12758598	5060	sens	5024	570	AT3G05830	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
37	chr3:1737581..1737875	294	chr5	5629978	5635293	5315	sens	5277	690	AT3G05830	ATCOPIA93(AT5TE20395)	123
38	chr3:3083438..3083594	156	chr1	12753537	12760312	6775	sens	6329	1064	AT3G10010	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
38	chr3:3083438..3083594	156	chr5	5629986	5635307	5321	sens	5295	1048	AT3G10010	ATCOPIA93(AT5TE20395)	214
39	chr3:4492935..4493027	92	chr1	12753547	12760380	6833	inverse	6169	534	AT3G13700;AT3G13710	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
39	chr3:4492935..4493027	92	chr5	5629985	5635297	5312	inverse	5187	532	AT3G13700;AT3G13710	ATCOPIA93(AT5TE20395)	83
40	chr3:5595062..5595173	111	chr2	4385856	4393657	7801	inverse	3042	516	AT3G16460	ATHILA3(AT2TE18410);ATENSPM3(AT2TE18415)	10

40	chr3:5595062..5595173	111	chr2	4900805	4909282	8477	sens	3123	960	AT3G16460	ATENSPM3(AT2TE20205)	425
41	chr3:8763540..8764993	1453	chr1	3360550	3361311	761	sens	750	126	.	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
41	chr3:8763824..8764993	1169	chr1	18013366	18017636	4270	inverse	4130	498	.	ATCOPIA78(AT1TE59755)	1
41	chr3:8763824..8765272	1448	chr1	7717818	7722179	4361	inverse	3577	428	.	ATCOPIA78(AT1TE24850)	1
41	chr3:8763824..8765272	1448	chr3	13369323	13373734	4411	inverse	3987	170	.	ATCOPIA78(AT3TE54550)	0
41	chr3:8764278..8764999	721	chr1	7714742	7722179	7437	inverse	4658	540	.	ATCOPIA78(AT1TE24850)	1
41	chr3:8764369..8764724	355	chr3	22059537	22064320	4783	inverse	4762	720	.	ATCOPIA78(AT3TE89830)	24
41	chr3:8764369..8764724	355	chr5	4208093	4213081	4988	sens	4976	860	.	ATCOPIA78(AT5TE15240)	77
41	chr3:8764445..8764724	279	chr1	21525005	21529604	4599	sens	4592	780	.	ATCOPIA78(AT1TE71045)	0
41	chr3:8764445..8764724	279	chr1	3780767	3785342	4575	inverse	4230	648	.	ATCOPIA78(AT1TE12295)	0
41	chr3:8764445..8764724	279	chr3	22695576	22700520	4944	sens	4907	832	.	ATCOPIA78(AT3TE92525)	0
42	chr3:11417981..11418855	874	chr4	3950891	3955341	4450	inverse	3367	230	.	.	*
43	chr3:11421153..11421366	213	chr4	3952137	3953845	1708	inverse	1524	56	.	.	*
44	chr3:11422921..11423727	806	chr4	3950874	3955756	4882	inverse	4015	184	.	.	*
45	chr3:13398954..13405984	7030	chr1	15214383	15214508	125	inverse	107	132	.	ATENSPM5(AT1TE49955)	0
45	chr3:13398954..13405984	7030	chr1	7024969	7025329	360	inverse	348	266	.	ATENSPM6(AT1TE22710)	0
45	chr3:13398954..13407550	8596	chr4	3655383	3655987	604	sens	580	306	.	.	1
45	chr3:13398954..13407550	8596	chr5	12710434	12710661	227	inverse	225	304	.	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	0
45	chr3:13398954..13407550	8596	chr5	971392	971583	191	sens	179	222	.	.	0
45	chr3:13398954..13407724	8770	chr5	18311649	18311777	128	sens	124	178	.	ATCOPIA68(AT5TE65910)	0
46	chr3:13589515..13590430	915	chr2	5277	5857	580	inverse	489	48	.	.	*
47	chr3:14194602..14201455	6853	chr2	2594	2987	393	sens	365	160	AT3G41761;AT3G41762;A	.	290
47	chr3:14194873..14199584	4711	chr2	9898	10139	241	sens	234	220	AT3G41761;AT3G41762;A	.	633
47	chr3:14196613..14199418	2805	chr5	3253128	3253276	148	inverse	145	3616	AT3G41768	.	2676
47	chr3:14197427..14199590	2163	chr2	7190932	7191521	589	sens	564	724	AT3G41768	.	517
47	chr3:14197891..14202606	4715	chr5	11733614	11735494	1880	inverse	1539	52	AT3G41768;AT3G41979	.	20
47	chr3:14197973..14199248	1275	chr2	3605680	3607826	2146	sens	2116	80	AT3G41768	.	5
47	chr3:14199178..14200977	1799	chr4	1799065	1801105	2040	inverse	1742	54	AT3G41768;AT3G41979	ATGP2N(AT4TE08985);ATGP2(AT4TE08980)	0
47	chr3:14199287..14202013	2726	chr2	3615941	3617076	1135	sens	1019	52	AT3G41768;AT3G41979	.	6
47	chr3:14199911..14202069	2158	chr1	15082117	15083174	1057	sens	1047	104	AT3G41979	.	3
47	chr3:14200064..14200828	764	chr4	3950522	3955946	5424	inverse	5327	336	.	.	56
47	chr3:14200771..14200973	202	chr2	43309	48466	5157	sens	3179	146	.	ATGP2(AT2TE00075)	78
47	chr3:14200849..14200973	124	chr2	4237804	4242968	5164	inverse	3706	166	.	ATGP2(AT2TE17875)	0
47	chr3:14200849..14201149	300	chr2	2564473	2569116	4643	sens	3511	128	.	ATGP2(AT2TE11540)	2
47	chr3:14201399..14202198	799	chr5	11729691	11735234	5543	sens	1203	152	.	.	48
47	chr3:14201673..14202630	957	chr5	11701923	11705761	3838	sens	3504	138	.	.	10
48	chr3:14203058..14203629	571	chr4	3950583	3955912	5329	sens	4542	402	.	.	*
49	chr3:15717150..15718059	909	chr4	3737307	3737629	322	sens	296	44	.	ATHILA2(AT4TE16565)	*
50	chr3:15718819..15721035	2216	chr1	15082074	15085838	3764	sens	3706	158	.	.	*
51	chr3:16580047..16580428	381	chr1	12753605	12758593	4988	sens	4300	450	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
51	chr3:16580047..16580428	381	chr5	5629982	5635269	5287	sens	5258	556	.	ATCOPIA93(AT5TE20395)	104
52	chr3:16894460..16894609	149	chr1	12753508	12760367	6859	inverse	6337	610	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
52	chr3:16894460..16894609	149	chr5	5629981	5635267	5286	inverse	5186	624	.	ATCOPIA93(AT5TE20395)	107
53	chr3:21980027..21980291	264	chr1	12753515	12760390	6875	inverse	6286	434	AT3G59470	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0

53	chr3:21980027..21980291	264	chr5	5629983	5635265	5282	inverse	5210	468	AT3G59470	ATCOPIA93(AT5TE20395)	105
54	chr3:22228074..22228854	780	chr1	12753560	12758621	5061	sens	3359	290	AT3G60160	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
54	chr3:22228074..22228854	780	chr5	5630052	5635073	5021	sens	3625	300	AT3G60160	ATCOPIA93(AT5TE20395)	25
55	chr3:22556611..22556699	88	chr1	12753506	12760391	6885	inverse	6347	1092	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
55	chr3:22556611..22556699	88	chr5	5629975	5635308	5333	inverse	5307	1116	.	ATCOPIA93(AT5TE20395)	193
56	chr4:1151216..1151639	423	chr2	1265611	1271523	5912	inverse	2107	238	AT4G02620;AT4G02630	VANDAL21(AT2TE05755)	0
56	chr4:1151429..1151604	175	chr3	11541649	11548547	6898	sens	2981	370	.	VANDAL21(AT3TE48070)	0
56	chr4:1151438..1151604	166	chr2	1505606	1513059	7453	sens	3467	534	.	VANDAL21(AT2TE06955)	1
56	chr4:1151438..1151604	166	chr5	13960986	13967602	6616	sens	3458	370	.	VANDAL21(AT5TE49755)	1
56	chr4:1151438..1151660	222	chr3	13072510	13078362	5852	sens	3119	218	AT4G02630	VANDAL21(AT3TE53580)	1
56	chr4:1151475..1151604	129	chr2	10001512	10009495	7983	inverse	3398	1002	.	VANDAL21(AT2TE42810)	373
57	chr4:2199494..2202716	3222	chr5	9640968	9644562	3594	sens	3541	352	.	ATMU10;ARNOLD2(AT5TE34985);Unassigned(AT5TE3498	*
58	chr4:3651904..3652219	315	chr1	12753512	12760311	6799	inverse	5686	560	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
58	chr4:3651904..3652219	315	chr5	5629980	5635304	5324	inverse	5279	558	.	ATCOPIA93(AT5TE20395)	99
59	chr4:3737408..3745200	7792	chr3	13588865	13592445	3580	sens	3449	128	.	.	*
60	chr4:3867417..3867840	423	chr4	3950528	3955848	5320	sens	4586	356	.	.	*
61	chr4:3868160..3868353	193	chr4	3950528	3955911	5383	sens	4481	358	.	.	*
62	chr4:3868392..3868444	52	chr4	3950538	3955421	4883	inverse	4627	192	.	.	*
63	chr4:3953519..3954249	730	chr2	3245	4143	898	inverse	844	50	.	.	10
63	chr4:3953525..3954361	836	chr3	14201588	14202828	1240	inverse	1013	60	.	.	31
64	chr4:3955848..3965683	9835	chr2	5565162	5567876	2714	sens	2212	72	.	ATHILA2(AT2TE22505)	*
65	chr4:3968795..3971002	2207	chr3	15711574	15712734	1160	inverse	1126	50	.	ATHILA2(AT3TE63540)	8
65	chr4:3969400..3970346	946	chr3	12223057	12225121	2064	inverse	1984	60	.	ATHILA2(AT3TE50670)	0
65	chr4:3969405..3970563	1158	chr1	15191314	15192926	1612	sens	1295	62	.	ATHILA2(AT1TE49850)	3
65	chr4:3969405..3976044	6639	chr3	15719977	15722611	2634	inverse	2247	82	.	ATHILA2(AT3TE63540)	0
65	chr4:3973640..3982321	8681	chr1	15180244	15181766	1522	sens	1274	78	.	ATHILA6B(AT1TE49845)	0
65	chr4:3981012..3982333	1321	chr3	15719807	15722198	2391	inverse	1929	104	.	ATHILA2(AT3TE63540)	0
66	chr4:3995608..3996433	825	chr3	13588800	13592558	3758	inverse	3193	102	.	.	33
66	chr4:3996043..3999305	3262	chr2	5566569	5568070	1501	inverse	1308	70	.	ATHILA2(AT2TE22505)	0
67	chr4:4003331..4006179	2848	chr1	15191760	15192945	1185	sens	949	58	.	ATHILA2(AT1TE49850)	8
67	chr4:4004168..4006078	1910	chr2	5565106	5567964	2858	sens	2759	154	.	ATHILA2(AT2TE22505)	8
68	chr4:4008366..4008847	481	chr2	5577916	5578958	1042	sens	849	46	.	ATHILA2(AT2TE22520)	*
69	chr4:4781311..4781800	489	chr3	15300682	15301156	474	inverse	412	56	.	ATHILA5(AT3TE62215)	3
69	chr4:4781460..4781491	31	chr5	11765487	11767433	1946	inverse	1900	76	.	ATHILA5(AT5TE42330);ATHILA5(AT5TE42320);ATHILA2(AT	15
69	chr4:4781460..4781812	352	chr1	16511326	16512215	889	inverse	790	100	.	ATHILA5(AT1TE54480)	11
69	chr4:4781460..4781812	352	chr2	3965170	3965730	560	inverse	535	66	.	ATHILA2(AT2TE16860)	4
70	chr4:5006317..5006790	473	chr5	5629981	5635277	5296	inverse	5255	554	.	ATCOPIA93(AT5TE20395)	105
70	chr4:5006393..5006790	397	chr1	12753535	12760360	6825	inverse	6272	554	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	1
71	chr4:7002644..7003242	598	chr1	12755000	12759346	4346	sens	3958	726	AT4G11560	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	1
71	chr4:7003182..7003242	60	chr3	10181994	10182392	398	inverse	393	116	AT4G11560	.	0
72	chr4:7003313..7003463	150	chr1	12753520	12760300	6780	sens	3691	882	AT4G11560	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	1
72	chr4:7003389..7003463	74	chr5	5629977	5635296	5319	sens	5290	1184	AT4G11560	ATCOPIA93(AT5TE20395)	240
73	chr4:7083314..7083461	147	chr1	12753511	12760376	6865	inverse	6113	548	AT4G11770	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
73	chr4:7083314..7083461	147	chr5	5629981	5635296	5315	inverse	5222	564	AT4G11770	ATCOPIA93(AT5TE20395)	109

74	chr4:8861548..8863563	2015	chr2	1267302	1269181	1879	inverse	1683	72	AT4G15510;AT4G15520	VANDAL21(AT2TE05755)	0
74	chr4:8862614..8862828	214	chr2	1505627	1513028	7401	sens	3168	452	AT4G15510;AT4G15520	VANDAL21(AT2TE06955)	0
74	chr4:8862635..8862828	193	chr5	13960862	13967693	6831	sens	3357	366	AT4G15510;AT4G15520	VANDAL21(AT5TE49755)	0
74	chr4:8862690..8862919	229	chr3	11541535	11548722	7187	sens	3194	350	AT4G15520	VANDAL21(AT3TE48070)	0
74	chr4:8862690..8863216	526	chr3	13072527	13078313	5786	sens	3076	192	AT4G15520	VANDAL21(AT3TE53580)	0
74	chr4:8862711..8862828	117	chr2	10001494	10009607	8113	inverse	3435	892	AT4G15520	VANDAL21(AT2TE42810)	326
75	chr4:9319462..9320717	1255	chr1	7717820	7722145	4325	sens	4264	690	AT4G16550	ATCOPIA78(AT1TE24850)	0
75	chr4:9319927..9320769	842	chr1	3360537	3361315	778	inverse	761	106	AT4G16550	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
75	chr4:9319942..9320709	767	chr1	21524992	21529547	4555	inverse	4422	822	AT4G16550	ATCOPIA78(AT1TE71045)	9
75	chr4:9319942..9320709	767	chr5	4208080	4212634	4554	inverse	3086	506	AT4G16550	ATCOPIA78(AT5TE15240)	0
75	chr4:9319942..9320717	775	chr3	13369245	13373785	4540	sens	4241	542	AT4G16550	ATCOPIA78(AT3TE54550)	0
75	chr4:9319954..9320709	755	chr1	18013199	18017627	4428	sens	4389	780	AT4G16550	ATCOPIA78(AT1TE59755)	0
75	chr4:9319954..9320709	755	chr1	3780802	3785272	4470	sens	4415	820	AT4G16550	ATCOPIA78(AT1TE12295)	0
75	chr4:9319954..9320709	755	chr3	22059572	22064257	4685	sens	4351	750	AT4G16550	ATCOPIA78(AT3TE89830)	1
75	chr4:9319954..9320709	755	chr3	22695572	22700368	4796	inverse	4516	874	AT4G16550	ATCOPIA78(AT3TE92525)	0
76	chr4:10738831..10740218	1387	chr1	3360538	3361241	703	sens	680	94	AT4G19740	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
76	chr4:10738873..10740970	2097	chr1	7717808	7722120	4312	inverse	3724	536	AT4G19740	ATCOPIA78(AT1TE24850)	0
76	chr4:10739264..10739932	668	chr1	7714737	7722120	7383	inverse	6799	632	AT4G19740	ATCOPIA78(AT1TE24850)	0
76	chr4:10739264..10740218	954	chr3	13369266	13373111	3845	inverse	3833	442	AT4G19740	ATCOPIA78(AT3TE54550)	0
76	chr4:10739334..10739851	517	chr1	18013164	18017944	4780	inverse	4523	720	AT4G19740	ATCOPIA78(AT1TE59755)	0
76	chr4:10739334..10739851	517	chr1	21524993	21529513	4520	sens	4506	728	AT4G19740	ATCOPIA78(AT1TE71045)	0
76	chr4:10739334..10739851	517	chr1	3780767	3785547	4780	inverse	4501	778	AT4G19740	ATCOPIA78(AT1TE12295)	1
76	chr4:10739334..10739851	517	chr3	22059537	22064209	4672	inverse	4625	824	AT4G19740	ATCOPIA78(AT3TE89830)	32
76	chr4:10739334..10739851	517	chr3	22695564	22700436	4872	sens	4502	794	AT4G19740	ATCOPIA78(AT3TE92520);ATCOPIA78(AT3TE92525)	6
76	chr4:10739334..10739932	598	chr5	4208081	4212869	4788	sens	4732	700	AT4G19740	ATCOPIA78(AT5TE15240)	0
77	chr4:13251176..13253849	2673	chr1	7717807	7721253	3446	inverse	2977	496	AT4G26150	ATCOPIA78(AT1TE24850)	1
77	chr4:13251345..13253496	2151	chr1	3360540	3361313	773	sens	747	138	AT4G26150	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
77	chr4:13252121..13252644	523	chr1	7714734	7721253	6519	inverse	6052	618	.	ATCOPIA78(AT1TE24850)	1
77	chr4:13252121..13252644	523	chr3	22695566	22700441	4875	sens	4481	764	.	ATCOPIA78(AT3TE92525)	0
77	chr4:13252197..13252644	447	chr1	18013164	18017782	4618	inverse	3734	782	.	ATCOPIA78(AT1TE59755)	36
77	chr4:13252197..13252644	447	chr1	21525006	21529547	4541	sens	4461	704	.	ATCOPIA78(AT1TE71045)	0
77	chr4:13252197..13252644	447	chr1	3780764	3784426	3662	inverse	3519	752	.	ATCOPIA78(AT1TE12295)	0
77	chr4:13252197..13252644	447	chr3	13369246	13373807	4561	inverse	3184	384	.	ATCOPIA78(AT3TE54550)	0
77	chr4:13252197..13252644	447	chr3	22059534	22064085	4551	inverse	3734	754	.	ATCOPIA78(AT3TE89830)	0
77	chr4:13252197..13252644	447	chr5	4208083	4213080	4997	sens	4757	700	.	ATCOPIA78(AT5TE15240)	0
78	chr4:16682304..16682584	280	chr5	5629982	5635298	5316	sens	5228	552	AT4G35050	ATCOPIA93(AT5TE20395)	107
78	chr4:16682558..16682584	26	chr1	12753522	12760126	6604	sens	6169	562	AT4G35050	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
79	chr5:729472..729675	203	chr1	12753539	12760310	6771	sens	5584	466	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
79	chr5:729472..729675	203	chr5	5630049	5635303	5254	sens	5081	484	.	ATCOPIA93(AT5TE20395)	99
80	chr5:8895666..8895869	203	chr1	12753506	12758415	4909	sens	4892	984	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	0
80	chr5:8895666..8895869	203	chr3	10182141	10182387	246	inverse	220	50	.	.	0
80	chr5:8895666..8895869	203	chr5	5629980	5634870	4890	sens	4865	1054	.	ATCOPIA93(AT5TE20395)	178
81	chr5:11323169..11323303	134	chr5	11704673	11709983	5310	inverse	3121	152	.	.	*
82	chr5:11705564..11706344	780	chr2	6453	10462	4009	inverse	3852	106	.	.	*

83	chr5:11709049..11711963	2914	chr3	5279762	5284057	4295	sens	3738	130	.	VANDAL17(AT3TE22200)	*
84	chr5:11729860..11730400	540	chr5	11813541	11819114	5573	inverse	192	208	.	ATHILA6A(AT5TE42440)	*
85	chr5:11730830..11731586	756	chr2	2563003	2568115	5112	inverse	2433	158	.	ATGP2(AT2TE11540)	0
85	chr5:11730830..11731588	758	chr2	4239208	4244204	4996	sens	2433	136	.	ATGP2(AT2TE17875)	2
86	chr5:13332941..13333190	249	chr1	12753517	12758403	4886	inverse	1917	344	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
86	chr5:13333017..13333469	452	chr5	5629978	5635101	5123	inverse	4956	470	.	ATCOPIA93(AT5TE20395)	96
87	chr5:21317601..21318977	1376	chr1	7960494	7961244	750	inverse	625	56	AT5G52530	ATDNA2T9B(AT1TE25615)	1
87	chr5:21317601..21318977	1376	chr2	11983189	11983906	717	inverse	624	46	AT5G52530	ATDNA2T9B(AT2TE52165)	5
87	chr5:21318604..21318977	373	chr4	11080505	11082070	1565	sens	1378	106	AT5G52530	ATDNA2T9A(AT4TE50860)	0
87	chr5:21318680..21318977	297	chr2	12738695	12740200	1505	inverse	1428	114	AT5G52530	ATDNA2T9A(AT2TE55690)	7
87	chr5:21318680..21318977	297	chr3	18629777	18631977	2200	inverse	2084	114	AT5G52530	ATDNA2T9C(AT3TE75580);ATDNA2T9A(AT3TE75575)	0
87	chr5:21318680..21318977	297	chr4	10537172	10539521	2349	inverse	2314	106	AT5G52530	ATDNA2T9A(AT4TE48190);VANDAL21(AT4TE48180);ATDN	0
87	chr5:21318680..21318977	297	chr4	2639600	2641060	1460	sens	1335	96	AT5G52530	ATDNA2T9A(AT4TE12345)	6
87	chr5:21318680..21319624	944	chr1	26375147	26375959	812	sens	765	70	AT5G52530	ATREP10D(AT1TE86380);ATDNA2T9A(AT1TE86375)	13
88	chr5:22921846..22922268	422	chr1	12753520	12759729	6209	inverse	6163	318	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
88	chr5:22921922..22922268	346	chr5	5630004	5634903	4899	inverse	4826	322	.	ATCOPIA93(AT5TE20395)	62
89	chr5:23913607..23913992	385	chr3	10182145	10182385	240	inverse	215	48	.	.	0
89	chr5:23913821..23913992	171	chr5	5629982	5634892	4910	sens	4871	790	.	ATCOPIA93(AT5TE20395)	145
89	chr5:23913897..23913992	95	chr1	12753527	12758394	4867	sens	4758	726	.	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41575)	0
90	chr5:25834575..25834843	268	chr1	12753510	12758977	5467	sens	1709	184	AT5G64630	ATCOPIA93(AT1TE41580);ATCOPIA93(AT1TE41570);ATCO	1
90	chr5:25834575..25834843	268	chr5	5630011	5634975	4964	sens	4869	274	AT5G64630	ATCOPIA93(AT5TE20395)	52

Supplementary Table 8: TE-Tracker output for wt1

	Acceptor		Donor									
ID	position	size	chr	start	stop	size	orientation	overlap	read count	acceptor annotation (TAIR10)	donor annotation (Quesneville natural transposons)	score
1	chr1:8775594..8776081	487	chr5	14175222	14176744	1522	sens	1483	64	AT1G24822;AT1G24880	VANDAL5(AT5TE50620);VANDAL5A(AT5TE50615);VANDA	*
2	chr1:8790380..8798796	8416	chr5	14175231	14176758	1527	sens	1484	70	AT1G25054	VANDAL5(AT5TE50620);VANDAL5A(AT5TE50615);VANDA	*
3	chr1:14308590..14311281	2691	chr2	6659232	6659518	286	sens	275	48	.	ARNOLDY1(AT2TE27160)	0
3	chr1:14308838..14310174	1336	chr5	14524088	14524772	684	sens	637	92	.	ARNOLDY2(AT5TE52230)	3
3	chr1:14308889..14311281	2392	chr2	3194886	3195377	491	inverse	478	50	.	ATREP4(AT2TE13800)	4
3	chr1:14309262..14311729	2467	chr3	12466107	12467084	977	sens	844	54	.	ARNOLDY1(AT3TE51470);ATENSPM2(AT3TE51460)	0
4	chr1:15041442..15041806	364	chr5	11723512	11724603	1091	inverse	950	54	.	.	*
5	chr1:15191076..15191722	646	chr3	13588840	13592529	3689	inverse	3605	148	.	.	*
6	chr1:16511251..16519338	8087	chr2	3605745	3607837	2092	sens	2014	144	.	.	5
6	chr1:16512122..16518063	5941	chr1	15041294	15043026	1732	inverse	1465	84	.	ATHILA6B(AT1TE49260)	0
6	chr1:16512961..16521199	8238	chr4	4780196	4781800	1604	inverse	1425	94	.	ATHILA2(AT4TE19915)	57
6	chr1:16518766..16519288	522	chr2	3624476	3627786	3310	inverse	3295	250	.	.	17
6	chr1:16518843..16519201	358	chr4	3950604	3955922	5318	inverse	4999	252	.	.	9
6	chr1:16519314..16519371	57	chr3	13586989	13592511	5522	inverse	5475	592	.	.	*
6	chr1:16519748..16520280	532	chr2	3623378	3627870	4492	inverse	3371	220	.	.	16
6	chr1:16519781..16520125	344	chr3	13586989	13592587	5598	inverse	5551	618	.	.	7
6	chr1:16519781..16520628	847	chr4	3950548	3956015	5467	sens	4044	378	.	.	9
6	chr1:16519798..16519828	30	chr2	3605697	3607837	2140	sens	2086	216	.	.	5
6	chr1:16520167..16520186	19	chr5	11729334	11734963	5629	inverse	4106	314	.	.	*
6	chr1:16520338..16520772	434	chr2	3625795	3627516	1721	sens	1483	140	.	.	14
6	chr1:16520653..16522591	1938	chr2	3605681	3607823	2142	inverse	2079	114	.	.	5
7	chr1:16527921..16527949	28	chr4	3950555	3956013	5458	sens	3566	220	.	.	*
8	chr1:19449261..19455470	6209	chr3	1855904	1856012	108	sens	102	608	AT1G52220;AT1G52230	.	0
8	chr1:19449261..19455470	6209	chr4	3655374	3655990	616	inverse	583	634	AT1G52220;AT1G52230	.	0
8	chr1:19449261..19455470	6209	chr5	12710434	12710658	224	sens	216	632	AT1G52220;AT1G52230	ATENSPM5(AT5TE45000)	0
8	chr1:19449261..19455470	6209	chr5	18311649	18311778	129	inverse	120	430	AT1G52220;AT1G52230	ATCOPIA68(AT5TE65910)	2
8	chr1:19449261..19455470	6209	chr5	971392	971583	191	inverse	182	540	AT1G52220;AT1G52230	.	0
9	chr2:2890..4586	1696	chr5	3253128	3253276	148	inverse	147	5728	AT2G01010	.	1854
9	chr2:3352..3353	1	chr3	14199592	14204946	5354	sens	2600	570	.	.	870
10	chr2:6214..6854	640	chr1	15081959	15084804	2845	inverse	2464	144	AT2G01021	.	2
10	chr2:6505..6854	349	chr3	13586999	13592500	5501	inverse	5100	222	AT2G01021	.	19
10	chr2:6521..9530	3009	chr5	3253128	3253276	148	inverse	146	276	AT2G01021	.	1867
10	chr2:6871..7330	459	chr5	11141160	11143531	2371	sens	2189	106	.	ATGP2(AT5TE40350)	1
10	chr2:6871..7609	738	chr3	11419078	11421448	2370	inverse	1928	94	.	ATGP2(AT3TE47595)	0
11	chr2:9572..10253	681	chr5	3253128	3253276	148	inverse	146	1486	AT2G01023	.	*
12	chr2:10310..17886	7576	chr3	13588856	13592510	3654	inverse	3092	148	.	.	0
12	chr2:16914..17221	307	chr4	3950879	3955810	4931	sens	1531	236	.	.	43
13	chr2:21957..22699	742	chr1	15081964	15082392	428	inverse	424	50	.	.	0

13	chr2:21957..24376	2419	chr3	13588925	13592505	3580	inverse	3176	166	.	.	0
13	chr2:22576..23019	443	chr4	3950607	3955774	5167	sens	4374	294	.	.	53
13	chr2:23099..23253	154	chr4	3950607	3955774	5167	sens	4374	288	.	.	*
14	chr2:48845..58759	9914	chr3	13588963	13592505	3542	inverse	3189	230	.	.	0
14	chr2:49608..58812	9204	chr1	15081964	15085371	3407	inverse	3351	140	.	.	1
15	chr2:2563473..2564102	629	chr3	13588913	13592496	3583	inverse	3190	172	.	.	1
15	chr2:2563741..2564115	374	chr4	3950836	3955756	4920	sens	3910	258	.	.	7
16	chr2:2676246..2684100	7854	chr1	14592407	14592671	264	sens	251	102	.	ATENSPM5(AT1TE48035)	*
17	chr2:3226348..3226496	148	chr3	13588913	13592506	3593	inverse	3189	148	.	.	*
18	chr2:3226562..3227140	578	chr4	3950897	3955778	4881	sens	4049	242	.	.	*
19	chr2:3487630..3489942	2312	chr1	16360267	16360935	668	inverse	609	82	AT2G07806	ATGP1(AT1TE53915)	0
19	chr2:3487683..3488238	555	chr2	6263375	6271083	7708	inverse	7454	446	AT2G07806	ATGP1(AT2TE25540)	0
19	chr2:3487737..3488238	501	chr3	12613559	12620854	7295	inverse	7218	544	AT2G07806	ATGP1(AT3TE51930)	1
19	chr2:3487737..3489544	1807	chr4	5271738	5272576	838	sens	809	72	AT2G07806	ATGP1(AT4TE21910)	0
19	chr2:3488378..3488459	81	chr2	1628600	1639040	10440	inverse	3980	2078	AT2G07806	ATGP1(AT2TE07550);ATLINE1A(AT2TE07540);ATGP1(AT2	521
20	chr2:3616530..3621030	4500	chr3	13588788	13592577	3789	sens	3441	294	AT2G07981;AT2G08986	.	54
20	chr2:3618101..3618432	331	chr1	13180997	13181888	891	inverse	883	58	AT2G08986	ATGP7(AT1TE43030);ATGP7(AT1TE43035)	1
21	chr2:3621886..3622428	542	chr3	13588788	13592596	3808	sens	3787	362	AT2G08986	.	*
22	chr2:3623418..3626944	3526	chr1	15191683	15192850	1167	sens	955	60	AT2G08986	ATHILA2(AT1TE49850)	1
22	chr2:3623526..3624521	995	chr4	3950523	3955967	5444	inverse	3437	278	AT2G08986	.	28
23	chr2:3954621..3955159	538	chr3	13587014	13592582	5568	sens	4929	304	.	.	*
24	chr2:3955278..3955731	453	chr3	13587014	13592582	5568	sens	4929	308	.	.	*
25	chr2:3958520..3964949	6429	chr1	15082074	15086027	3953	sens	3322	170	.	.	*
26	chr2:4242822..4243445	623	chr3	13590043	13592496	2453	sens	2220	186	.	.	*
27	chr3:5285507..5285698	191	chr5	11686031	11697334	11303	sens	1634	534	.	ATHILA6A(AT5TE42135)	2
27	chr3:5285583..5285698	115	chr5	11320537	11331499	10962	sens	2181	538	.	ATHILA6A(AT5TE41070)	4
27	chr3:5285583..5285698	115	chr5	11810286	11820558	10272	sens	1633	466	.	ATHILA6A(AT5TE42440)	0
27	chr3:5285583..5285698	115	chr5	11935199	11945441	10242	inverse	1634	586	.	ATHILA6A(AT5TE42710)	0
28	chr3:13399068..13407524	8456	chr4	3655371	3655992	621	sens	538	438	.	.	2
28	chr3:13399068..13407524	8456	chr5	12710435	12710660	225	inverse	214	424	.	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	1
28	chr3:13399068..13407524	8456	chr5	18311649	18311776	127	sens	118	292	.	ATCOPIA68(AT5TE65910)	3
28	chr3:13399068..13407524	8456	chr5	971392	971583	191	sens	172	370	.	.	0
28	chr3:13403093..13407524	4431	chr1	7024969	7025327	358	inverse	344	420	.	ATENSPM6(AT1TE22710)	0
29	chr3:13589504..13590290	786	chr2	45063	47136	2073	inverse	1686	90	.	ATGP2(AT2TE00075)	0
29	chr3:13589524..13590239	715	chr2	3605678	3607335	1657	inverse	1629	212	.	.	8
29	chr3:13589918..13590287	369	chr2	3229047	3230228	1181	inverse	959	66	.	ATLINE1_2(AT2TE13950);ATGP2(AT2TE13955)	0
29	chr3:13589943..13589945	2	chr4	3950555	3955922	5367	inverse	1652	352	.	.	180
30	chr3:13590656..13591064	408	chr2	3626031	3626427	396	sens	347	44	.	.	*
31	chr3:13591720..13592268	548	chr5	11731584	11732789	1205	sens	1205	84	.	.	*
32	chr3:13592428..13592455	27	chr4	3950519	3955915	5396	inverse	4543	370	.	.	*
33	chr3:14194759..14201495	6736	chr2	9898	10163	265	sens	236	238	AT3G41761;AT3G41762;A	.	407
33	chr3:14196757..14202119	5362	chr4	3952675	3953600	925	inverse	846	50	AT3G41768;AT3G41979	.	2
33	chr3:14196993..14197827	834	chr5	3253128	3253276	148	inverse	147	5944	AT3G41768	.	2298
33	chr3:14197050..14200407	3357	chr2	7190932	7191445	513	sens	506	1180	AT3G41768;AT3G41979	.	159

33	chr3:14197139..14199592	2453	chr2	2594	3501	907	sens	834	666	AT3G41768	.	450
33	chr3:14200092..14200825	733	chr2	3626804	3627883	1079	sens	950	52	.	.	0
33	chr3:14200805..14201497	692	chr2	45575	47032	1457	sens	1191	74	.	ATGP2(AT2TE00075)	2
33	chr3:14200805..14201580	775	chr5	11141724	11143019	1295	sens	1212	92	.	ATGP2(AT5TE40350)	1
33	chr3:14201032..14201771	739	chr2	3625854	3627889	2035	inverse	1635	86	.	.	0
34	chr3:15712942..15719996	7054	chr2	3626923	3628054	1131	inverse	944	58	.	.	*
35	chr4:1801221..1804023	2802	chr3	13588913	13592514	3601	sens	2999	156	.	.	*
36	chr4:3737498..3743589	6091	chr3	13588799	13592431	3632	sens	3412	150	.	.	*
37	chr4:3868185..3868267	82	chr4	3950530	3955841	5311	inverse	5220	332	.	.	*
38	chr4:3953626..3954335	709	chr2	58715	60312	1597	sens	1411	110	.	ATGP2(AT2TE00110)	3
38	chr4:3953690..3954340	650	chr5	11138591	11143495	4904	sens	2782	224	.	ATGP2(AT5TE40350)	0
38	chr4:3953775..3954368	593	chr2	24365	25762	1397	sens	1250	68	.	ATGP2(AT2TE00015)	1
39	chr4:3955915..3965688	9773	chr3	13588791	13592580	3789	inverse	3428	580	.	.	*
40	chr4:3968769..3970314	1545	chr2	5565110	5567972	2862	sens	2706	156	.	ATHILA2(AT2TE22505)	*
41	chr4:3973033..3976039	3006	chr1	15832515	15834390	1875	inverse	1731	80	.	ATHILA2(AT1TE52125);ATHILA2(AT1TE52105)	9
41	chr4:3973840..3976053	2213	chr1	15179522	15181767	2245	sens	2213	120	.	ATHILA6B(AT1TE49845)	0
41	chr4:3974058..3980942	6884	chr2	5565119	5567910	2791	sens	2549	176	.	ATHILA2(AT2TE22505)	5
41	chr4:3974707..3974767	60	chr3	15718449	15722694	4245	inverse	2006	176	.	ATHILA2(AT3TE63540)	10
41	chr4:3978054..3984304	6250	chr3	15719800	15722640	2840	inverse	2508	114	.	ATHILA2(AT3TE63540)	0
41	chr4:3978900..3983305	4405	chr2	5577309	5578961	1652	sens	1488	80	.	ATHILA2(AT2TE22520)	0
41	chr4:3979156..3981352	2196	chr1	15191299	15192977	1678	sens	1620	144	.	ATHILA2(AT1TE49850)	3
41	chr4:3980958..3984425	3467	chr1	14450561	14450936	375	sens	349	44	.	ATHILA2(AT1TE47570)	0
42	chr4:3985148..3985581	433	chr3	15720789	15722008	1219	inverse	982	88	.	ATHILA2(AT3TE63540)	*
43	chr4:3988849..3989070	221	chr3	13589948	13591692	1744	inverse	1738	164	.	.	*
44	chr4:3996586..3998099	1513	chr1	15180375	15182061	1686	inverse	1430	78	.	ATHILA6B(AT1TE49845)	0
44	chr4:3997434..3998184	750	chr2	5565374	5568155	2781	inverse	2698	112	.	ATHILA2(AT2TE22505)	1
44	chr4:3998156..3999297	1141	chr3	15719688	15722168	2480	sens	2162	98	.	ATHILA2(AT3TE63540)	0
44	chr4:3998156..3999307	1151	chr1	15191603	15192871	1268	inverse	1154	62	.	ATHILA2(AT1TE49850)	0
45	chr4:4001021..4004177	3156	chr1	15179921	15181899	1978	sens	1677	86	.	ATHILA6B(AT1TE49845)	0
45	chr4:4002287..4006185	3898	chr2	5577300	5578907	1607	sens	1335	76	.	ATHILA2(AT2TE22520)	0
45	chr4:4003572..4006218	2646	chr3	15719817	15722472	2655	inverse	2231	146	.	ATHILA2(AT3TE63540)	0
45	chr4:4004072..4006154	2082	chr1	15191604	15192906	1302	sens	1104	128	.	ATHILA2(AT1TE49850)	0
45	chr4:4004074..4007296	3222	chr2	5565096	5567939	2843	sens	2373	174	.	ATHILA2(AT2TE22505)	4
46	chr4:4026735..4027262	527	chr4	3950519	3955539	5020	inverse	4146	434	.	.	*
47	chr4:4781567..4781670	103	chr1	16511488	16511864	376	inverse	357	76	.	ATHILA5(AT1TE54480)	*
48	chr4:9320217..9322362	2145	chr1	3360540	3361297	757	inverse	704	156	AT4G16550	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
48	chr4:9320271..9320618	347	chr1	7714761	7722341	7580	sens	4327	1018	AT4G16550	ATCOPIA78(AT1TE24850)	1
48	chr4:9320271..9320624	353	chr3	13369245	13373901	4656	sens	4349	810	AT4G16550	ATCOPIA78(AT3TE54550)	0
48	chr4:9320284..9320298	14	chr1	18012374	18018389	6015	sens	5167	1150	AT4G16550	ATCOPIA78(AT1TE59755)	0
48	chr4:9320308..9320618	310	chr1	3780812	3785713	4901	sens	4750	1282	AT4G16550	ATCOPIA78(AT1TE12295)	0
48	chr4:9320384..9320617	233	chr1	21525362	21529345	3983	inverse	3931	1414	AT4G16550	ATCOPIA78(AT1TE71045)	14
48	chr4:9320384..9320617	233	chr3	22059578	22064296	4718	sens	4661	1362	AT4G16550	ATCOPIA78(AT3TE89830)	1
48	chr4:9320384..9320617	233	chr3	22695600	22700082	4482	inverse	4104	1564	AT4G16550	ATCOPIA78(AT3TE92525)	0
48	chr4:9320384..9320617	233	chr5	4208083	4212977	4894	inverse	4780	1602	AT4G16550	ATCOPIA78(AT5TE15240)	0

49	chr4:10898815..10906343	7528	chr1	6706695	6707896	1201	sens	1190	80	AT4G20170	META1(AT1TE21710)	0
49	chr4:10898815..10906984	8169	chr3	14958425	14959625	1200	inverse	1169	78	AT4G20170;AT4G20190	META1(AT3TE61000)	0
49	chr4:10898815..10906984	8169	chr5	14471140	14472289	1149	sens	1133	66	AT4G20170;AT4G20190	META1(AT5TE52075)	0
49	chr4:10898815..10906985	8170	chr3	15815690	15816823	1133	sens	1115	60	AT4G20170;AT4G20190	META1(AT3TE63950)	0
49	chr4:10899121..10906343	7222	chr1	4407440	4408642	1202	inverse	1191	126	AT4G20170	META1(AT1TE14315)	0
49	chr4:10899121..10906343	7222	chr3	18736756	18737956	1200	sens	1189	98	AT4G20170	META1(AT3TE76010)	0
49	chr4:10899121..10906343	7222	chr5	10402175	10403365	1190	inverse	1173	92	AT4G20170	META1(AT5TE37975)	0
49	chr4:10899121..10906984	7863	chr1	6540896	6542087	1191	sens	1169	106	AT4G20170;AT4G20190	META1(AT1TE21210)	0
49	chr4:10899121..10907648	8527	chr3	16040467	16041003	536	inverse	525	74	AT4G20170;AT4G20190	META1(AT3TE64915)	0
49	chr4:10899121..10907971	8850	chr3	17364511	17364984	473	inverse	425	52	AT4G20170;AT4G20190	META1(AT3TE70410)	0
49	chr4:10903550..10906343	2793	chr2	911724	912937	1213	sens	1169	106	.	META1(AT2TE04085)	0
50	chr5:11322475..11322496	21	chr5	11730399	11733141	2742	sens	2379	162	.	.	*
51	chr5:11695482..11702306	6824	chr2	3607076	3607373	297	sens	257	50	.	.	*
52	chr5:11705150..11706982	1832	chr1	15191722	15192773	1051	sens	904	48	.	ATHILA2(AT1TE49850)	0
52	chr5:11705295..11705628	333	chr3	13588795	13592574	3779	inverse	3378	150	.	.	36
53	chr5:11707004..11708195	1191	chr3	13590204	13592574	2370	inverse	2094	96	.	.	*
54	chr5:11722019..11726258	4239	chr2	3618931	3619838	907	sens	867	62	.	.	*
55	chr5:11729458..11729855	397	chr2	3606159	3607812	1653	inverse	1620	90	.	.	13
55	chr5:11729704..11730394	690	chr3	13590219	13591773	1554	sens	1494	184	.	.	52
55	chr5:11729722..11730402	680	chr5	11320535	11325759	5224	sens	2004	252	.	ATHILA6A(AT5TE41070)	12
56	chr5:11730484..11733085	2601	chr2	5565335	5566980	1645	inverse	1419	70	.	ATHILA2(AT2TE22505)	1
56	chr5:11730729..11732156	1427	chr4	3951867	3953756	1889	inverse	1551	210	.	.	16
56	chr5:11731109..11731583	474	chr2	4238034	4243084	5050	sens	2866	268	.	ATGP2(AT2TE17875)	0
57	chr5:11733568..11734639	1071	chr3	13590217	13591834	1617	sens	1361	146	.	.	40
57	chr5:11733702..11734351	649	chr2	4237780	4242919	5139	sens	1078	236	.	ATGP2(AT2TE17875)	0
57	chr5:11733781..11734352	571	chr4	3950953	3955907	4954	inverse	4630	200	.	.	40
57	chr5:21318081..21319038	957	chr1	22872360	22873743	1383	sens	1210	58	AT5G52530	ATDNA2T9A(AT1TE75405)	0
57	chr5:21318180..21318482	302	chr4	2639584	2640353	769	sens	691	156	AT5G52530	ATDNA2T9A(AT4TE12345)	0
57	chr5:21318180..21318482	302	chr4	5471630	5472400	770	inverse	760	144	AT5G52530	VANDALNX2(AT4TE22920)	2
57	chr5:21318180..21318998	818	chr2	11983184	11983905	721	inverse	708	116	AT5G52530	ATDNA2T9B(AT2TE52165)	0
57	chr5:21318183..21318482	299	chr1	19728522	19729289	767	sens	761	184	AT5G52530	ATDNA2T9B(AT1TE65210);HELITRON2(AT1TE65205)	0
57	chr5:21318183..21318482	299	chr1	7960489	7961255	766	inverse	762	210	AT5G52530	ATDNA2T9B(AT1TE25615)	3
57	chr5:21318183..21318482	299	chr3	18631477	18632069	592	inverse	584	174	AT5G52530	ATDNA2T9A(AT3TE75575)	0
57	chr5:21318183..21318482	299	chr4	10538843	10539610	767	inverse	762	180	AT5G52530	ATDNA2T9A(AT4TE48190)	0
57	chr5:21318183..21318482	299	chr4	11080494	11081258	764	sens	688	186	AT5G52530	ATDNA2T9A(AT4TE50860)	0
57	chr5:21318183..21318890	707	chr2	12739524	12740244	720	inverse	716	150	AT5G52530	ATDNA2T9A(AT2TE55690)	0
57	chr5:21318906..21318998	92	chr2	12738700	12740168	1468	inverse	1397	170	AT5G52530	ATDNA2T9A(AT2TE55690)	2
57	chr5:21318906..21319259	353	chr1	26375155	26375954	799	sens	768	102	AT5G52530	ATDNA2T9A(AT1TE86375)	13
57	chr5:21318906..21319821	915	chr4	11081717	11082065	348	sens	343	52	AT5G52530	ATDNA2T9A(AT4TE50860)	2
57	chr5:21318906..21319958	1052	chr4	10537177	10537413	236	inverse	227	48	AT5G52530	ATDNA2T9A(AT4TE48175)	0

Supplementary Table 9: TE-Tracker output for wt2

	Acceptor		Donor									
ID	position	size	chr	start	stop	size	orientation	overlap	read count	acceptor annotation (TAIR10)	donor annotation (Quesneville natural transposons)	score
1	chr1:19449402..19455487	6085	chr5	971392	971584	192	inverse	189	664	AT1G52220;AT1G52230	.	0
1	chr1:19449665..19455487	5822	chr3	1855904	1856012	108	sens	107	804	AT1G52220;AT1G52230	.	0
1	chr1:19449665..19455487	5822	chr4	3655371	3655994	623	inverse	616	854	AT1G52220;AT1G52230	.	0
1	chr1:19449665..19455487	5822	chr5	12710435	12710663	228	sens	223	846	AT1G52220;AT1G52230	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	0
1	chr1:19449665..19455848	6183	chr5	18311648	18311776	128	inverse	127	438	AT1G52220;AT1G52230;A	ATCOPIA68(AT5TE65910)	6
2	chr2:2840..3355	515	chr3	14201455	14201779	324	sens	289	770	.	.	616
3	chr2:3425..7903	4478	chr3	14201455	14205200	3745	sens	3310	954	AT2G01010;AT2G01020;A	.	987
3	chr2:6560..6709	149	chr5	3253128	3253276	148	inverse	138	7020	AT2G01021	.	2317
4	chr2:10229..14196	3967	chr3	13586988	13587787	799	sens	782	164	.	.	*
5	chr2:22627..23220	593	chr1	15081961	15082410	449	inverse	427	84	.	.	*
6	chr2:49461..58764	9303	chr1	15081958	15082410	452	inverse	434	96	.	.	*
7	chr2:2676247..2680383	4136	chr1	14592409	14592675	266	sens	257	126	.	ATENSPM5(AT1TE48035)	*
8	chr2:3487345..3490343	2998	chr1	16360270	16360609	339	inverse	315	78	AT2G07806	ATGP1(AT1TE53915)	0
8	chr2:3487345..3490343	2998	chr4	5127353	5127670	317	inverse	292	86	AT2G07806	ATGP1(AT4TE21295)	0
8	chr2:3487345..3490399	3054	chr1	16012719	16013116	397	inverse	392	78	AT2G07806	ATGP1(AT1TE52795)	0
8	chr2:3487352..3490271	2919	chr3	15318310	15318947	637	sens	575	126	AT2G07806	ATGP1(AT3TE62245)	0
8	chr2:3488361..3488460	99	chr2	1628549	1638839	10290	inverse	3992	2456	AT2G07806	ATGP1(AT2TE07550);ATLINE1A(AT2TE07540);ATGP1(AT2	678
9	chr2:3950347..3950883	536	chr1	15700398	15700542	144	sens	137	50	.	ATGP1(AT1TE51700);ATHILA(AT1TE51705)	*
10	chr3:13398979..13403083	4104	chr1	15214383	15214506	123	inverse	112	262	.	ATENSPM5(AT1TE49955)	4
10	chr3:13398979..13407128	8149	chr4	3655373	3655993	620	sens	607	624	.	.	0
10	chr3:13398979..13407128	8149	chr5	12710433	12710663	230	inverse	220	618	.	ATENSPM5(AT5TE45000);ATHILA3(AT5TE45005)	2
10	chr3:13398979..13407128	8149	chr5	18311646	18311776	130	sens	121	320	.	ATCOPIA68(AT5TE65910)	2
10	chr3:13398979..13407128	8149	chr5	971393	971584	191	sens	187	494	.	.	0
10	chr3:13400684..13403083	2399	chr1	7024972	7025327	355	inverse	351	542	.	ATENSPM6(AT1TE22710)	0
10	chr3:13403191..13407128	3937	chr1	7024972	7025327	355	inverse	351	540	.	ATENSPM6(AT1TE22710)	*
11	chr3:13590649..13591011	362	chr2	3607052	3607359	307	inverse	296	58	.	.	*
12	chr3:13591661..13592276	615	chr1	15083651	15083766	115	sens	111	68	.	.	*
13	chr3:14192165..14194010	1845	chr2	9897	10145	248	sens	212	58	.	.	*
14	chr3:14194254..14195888	1634	chr5	3253128	3253276	148	inverse	148	7768	AT3G41762	.	*
15	chr3:14196014..14201467	5453	chr2	9863	10265	402	sens	353	680	AT3G41761;AT3G41762;A	.	886
15	chr3:14196645..14197681	1036	chr5	3253128	3253276	148	inverse	148	7660	AT3G41768	.	3581
15	chr3:14197127..14201455	4328	chr2	2594	3427	833	sens	831	998	AT3G41768;AT3G41979	.	773
15	chr3:14198110..14199474	1364	chr2	3626751	3627059	308	sens	287	78	AT3G41768	.	0
15	chr3:14198575..14200417	1842	chr2	3627218	3627517	299	sens	243	62	AT3G41768;AT3G41979	.	1
15	chr3:14200531..14200680	149	chr5	3253128	3253276	148	inverse	138	7690	.	.	2317
16	chr3:14201668..14202112	444	chr2	3355	3504	149	sens	136	110	.	.	*
17	chr4:3955708..3965624	9916	chr3	13587609	13587779	170	sens	149	142	.	.	*
18	chr4:3993111..3993607	496	chr1	15700256	15700543	287	inverse	277	136	.	ATGP1(AT1TE51700);ATHILA(AT1TE51705)	*

19	chr4:4781456..4781584	128	chr2	3965162	3965458	296	inverse	282	88	.	ATHILA2(AT2TE16860)	10
19	chr4:4781474..4781584	110	chr1	16511489	16511863	374	inverse	367	124	.	ATHILA5(AT1TE54480)	5
19	chr4:4781474..4781584	110	chr3	15300673	15300942	269	inverse	257	124	.	ATHILA5(AT3TE62215)	0
20	chr4:9318587..9320729	2142	chr3	14401608	14401734	126	sens	124	60	AT4G16550	ATCOPIA78(AT3TE59075)	0
20	chr4:9320322..9320397	75	chr1	18013200	18017637	4437	sens	4246	1794	AT4G16550	ATCOPIA78(AT1TE59755)	1
20	chr4:9320322..9320397	75	chr1	3780986	3785374	4388	sens	4215	1894	AT4G16550	ATCOPIA78(AT1TE12295)	0
20	chr4:9320322..9320397	75	chr3	22059574	22063883	4309	sens	4105	1612	AT4G16550	ATCOPIA78(AT3TE89830)	1
20	chr4:9320322..9320635	313	chr5	4208080	4213027	4947	inverse	4886	1988	AT4G16550	ATCOPIA78(AT5TE15240)	0
20	chr4:9320398..9320635	237	chr1	18013206	18018386	5180	sens	2114	1088	AT4G16550	ATCOPIA78(AT1TE59755)	1
20	chr4:9320398..9320635	237	chr1	21524992	21529471	4479	inverse	4453	1992	AT4G16550	ATCOPIA78(AT1TE71045)	25
20	chr4:9320398..9320635	237	chr1	3780809	3785720	4911	sens	2397	1146	AT4G16550	ATCOPIA78(AT1TE12295)	0
20	chr4:9320398..9320635	237	chr1	7714750	7722166	7416	sens	4344	1754	AT4G16550	ATCOPIA78(AT1TE24850)	0
20	chr4:9320398..9320635	237	chr1	7714750	7722242	7492	sens	4344	1748	AT4G16550	ATCOPIA78(AT1TE24850)	0
20	chr4:9320398..9320635	237	chr3	13369258	13373897	4639	sens	4383	1304	AT4G16550	ATCOPIA78(AT3TE54550)	0
20	chr4:9320398..9320635	237	chr3	22059574	22064322	4748	sens	2365	1038	AT4G16550	ATCOPIA78(AT3TE89830)	2
20	chr4:9320398..9320638	240	chr3	22695566	22700023	4457	inverse	4379	1698	AT4G16550	ATCOPIA78(AT3TE92525)	1
20	chr4:9320398..9322463	2065	chr1	3360537	3361307	770	inverse	749	246	AT4G16550	ATCOPIA78(AT1TE10930);ATCOPIA78(AT1TE10935)	0
21	chr5:11723043..11728127	5084	chr2	3607065	3607370	305	inverse	246	64	.	.	*
22	chr5:11730807..11732215	1408	chr4	3953385	3953793	408	inverse	382	114	.	.	*
23	chr5:21318059..21318840	781	chr2	11983185	11983910	725	inverse	724	172	AT5G52530	ATDNA2T9B(AT2TE52165)	0
23	chr5:21318226..21318840	614	chr1	19728522	19729288	766	sens	725	202	AT5G52530	ATDNA2T9B(AT1TE65210);HELITRON2(AT1TE65205)	0
23	chr5:21318226..21318840	614	chr1	7960490	7961249	759	inverse	726	218	AT5G52530	ATDNA2T9B(AT1TE25615)	4
23	chr5:21318226..21318840	614	chr2	12739524	12740249	725	inverse	638	194	AT5G52530	ATDNA2T9A(AT2TE55690)	0
23	chr5:21318226..21318840	614	chr3	18631300	18632066	766	inverse	725	214	AT5G52530	ATDNA2T9A(AT3TE75575)	0
23	chr5:21318226..21318840	614	chr4	10538843	10539610	767	inverse	726	212	AT5G52530	ATDNA2T9A(AT4TE48190)	0
23	chr5:21318226..21318888	662	chr4	5471631	5472397	766	inverse	725	174	AT5G52530	VANDALNX2(AT4TE22920)	0
23	chr5:21318801..21318840	39	chr4	11080492	11081972	1480	sens	1422	278	AT5G52530	ATDNA2T9A(AT4TE50860)	1
23	chr5:21318801..21318840	39	chr4	2639587	2640987	1400	sens	1345	276	AT5G52530	ATDNA2T9A(AT4TE12345)	13
24	chr5:23131700..23141264	9564	chr4	6070954	6071091	137	inverse	132	46	.	ATCOPIA49(AT4TE25590)	*

Annexe 2 : Liste des nouvelles insertions d'ET dans les épiRIL et le mutant *ddm1* utilisées pour les analyses de la partie2.

ID	site accepteur	site donneur			orientation	taille	annotation	score
		chr	start	stop				
epiRIL 71								
1	chr1:10273684..10273901	chr2	10001510	10009406	sens	7896	VANDAL21(AT2TE42810)	193
2	chr1:12179484..12182377	chr3	15283665	15286605	inverse	2940	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
3	chr1:12619333..12622822	chr3	15283665	15286610	sens	2945	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
14	chr1:26862708..26862790	chr2	10001527	10009597	inverse	8070	VANDAL21(AT2TE42810)	219
38	chr2:10507577..10512809	chr4	4358136	4360292	inverse	2156	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	131
39	chr2:14584556..14584704	chr2	10001479	10009518	sens	8039	VANDAL21(AT2TE42810)	388
40	chr3:8764398..8764653	chr3	22059532	22064329	inverse	4797	ATCOPIA78(AT3TE89830)	27
40	chr3:8764398..8764653	chr5	4208080	4213088	sens	5008	ATCOPIA78(AT5TE15240)	87
56	chr3:21599345..21599515	chr2	10001493	10009641	sens	8148	VANDAL21(AT2TE42810)	414
57	chr4:1151499..1151597	chr2	10001542	10009526	inverse	7984	VANDAL21(AT2TE42810)	405
76	chr4:16336419..16336578	chr2	10001536	10009556	sens	8020	VANDAL21(AT2TE42810)	209
77	chr5:619577..619658	chr2	10001480	10009596	inverse	8116	VANDAL21(AT2TE42810)	256
epiRIL 122								
1	chr1:2557186..2557278	chr5	5629984	5635305	sens	5321	ATCOPIA93(AT5TE20395)	225
2	chr1:2675675..2675678	chr2	4900906	4909185	sens	8279	ATENSPM3(AT2TE20205)	ND
3	chr1:4025667..4025679	chr5	5630035	5635273	sens	5238	ATCOPIA93(AT5TE20395)	99
4	chr1:4261921..4266051	chr5	9272155	9274530	inverse	2375	ATCOPIA63(AT5TE33540)	41
5	chr1:8103484..8103749	chr2	4904241	4909004	inverse	4763	ATENSPM3(AT2TE20205)	ND
6	chr1:8107658..8107871	chr2	4900912	4909167	sens	8255	ATENSPM3(AT2TE20205)	ND
7	chr1:9291297..9291381	chr2	4900821	4909273	sens	8452	ATENSPM3(AT2TE20205)	ND
19	chr1:18205766..18207653	chr3	15283677	15286611	sens	2934	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
20	chr1:19226964..19227197	chr2	4900859	4909216	sens	8357	ATENSPM3(AT2TE20205)	ND
22	chr1:19631371..19631702	chr2	4900850	4909255	sens	8405	ATENSPM3(AT2TE20205)	ND
23	chr1:20368518..20368605	chr2	4900819	4909254	sens	8435	ATENSPM3(AT2TE20205)	ND
24	chr1:21118871..21119063	chr5	5629980	5635298	sens	5318	ATCOPIA93(AT5TE20395)	163
25	chr1:21486398..21486523	chr5	5629988	5635284	inverse	5296	ATCOPIA93(AT5TE20395)	104
26	chr1:21880280..21880535	chr2	4900852	4909200	sens	8348	ATENSPM3(AT2TE20205)	ND
27	chr1:22224159..22224350	chr2	4900833	4909225	sens	8392	ATENSPM3(AT2TE20205)	231
28	chr1:22502051..22502110	chr5	5630015	5635303	sens	5288	ATCOPIA93(AT5TE20395)	ND
29	chr1:22502246..22502343	chr5	5630015	5635295	sens	5280	ATCOPIA93(AT5TE20395)	207
30	chr1:24215952..24216067	chr5	5629980	5635304	sens	5324	ATCOPIA93(AT5TE20395)	182
31	chr1:28365836..28372197	chr3	15285378	15286606	sens	1228	ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
41	chr2:164018..164174	chr2	4900831	4909043	inverse	8212	ATENSPM3(AT2TE20205)	ND
42	chr2:2392850..2392951	chr5	5629982	5635303	sens	5321	ATCOPIA93(AT5TE20395)	227
48	chr2:3327959..3328066	chr5	5629995	5635281	sens	5286	ATCOPIA93(AT5TE20395)	64
50	chr2:3618988..3625684	chr1	12755009	12757090	inverse	2081	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
53	chr2:4089159..4089373	chr5	5630033	5635306	inverse	5273	ATCOPIA93(AT5TE20395)	109
58	chr2:5842409..5846866	chr1	17203937	17206203	sens	2266	ATCOPIA63(AT1TE57025)	18
59	chr2:6706456..6706687	chr5	5629977	5635296	sens	5319	ATCOPIA93(AT5TE20395)	95
60	chr2:7519681..7519804	chr2	4900821	4909279	inverse	8458	ATENSPM3(AT2TE20205)	ND
61	chr2:7539527..7539636	chr2	4900862	4909204	inverse	8342	ATENSPM3(AT2TE20205)	350
62	chr2:9257368..9257476	chr5	5629988	5635277	inverse	5289	ATCOPIA93(AT5TE20395)	199
63	chr2:12035104..12035286	chr2	4900848	4909068	inverse	8220	ATENSPM3(AT2TE20205)	ND
64	chr2:15217064..15217240	chr5	5629984	5635293	sens	5309	ATCOPIA93(AT5TE20395)	208
65	chr2:16153132..16153223	chr5	5629978	5635289	inverse	5311	ATCOPIA93(AT5TE20395)	220

66	chr2:19059955..19060407	chr5	5629989	5635305	inverse	5316	ATCOPIA93(AT5TE20395)	242
66	chr2:19061865..19062014	chr5	5629989	5635229	inverse	5240	ATCOPIA93(AT5TE20395)	242
67	chr3:2164..2377	chr3	11664977	11669779	sens	4802	ATCOPIA63(AT3TE48480)	ND
68	chr3:661653..661796	chr2	4900839	4909236	sens	8397	ATENSPM3(AT2TE20205)	ND
69	chr3:670719..670832	chr2	4900837	4909259	inverse	8422	ATENSPM3(AT2TE20205)	333
70	chr3:2329309..2329593	chr5	5629986	5635301	sens	5315	ATCOPIA93(AT5TE20395)	109
71	chr3:4094839..4095121	chr5	5629982	5635247	sens	5265	ATCOPIA93(AT5TE20395)	111
72	chr3:4109778..4109893	chr2	4900799	4909243	inverse	8444	ATENSPM3(AT2TE20205)	196
73	chr3:4874425..4874593	chr5	5629982	5635287	sens	5305	ATCOPIA93(AT5TE20395)	112
74	chr3:5871163..5871305	chr2	4900855	4905609	inverse	4754	ATENSPM3(AT2TE20205)	145
75	chr3:5986019..5986088	chr5	5629977	5635302	inverse	5325	ATCOPIA93(AT5TE20395)	226
76	chr3:7178351..7178456	chr5	5629981	5635306	inverse	5325	ATCOPIA93(AT5TE20395)	112
77	chr3:8173986..8175983	chr4	5541892	5545537	inverse	3645	ATREP4(AT4TE23205) HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	95
78	chr3:9589610..9589716	chr2	4900802	4909178	sens	8376	ATENSPM3(AT2TE20205)	ND
79	chr3:10142083..10142241	chr5	5630190	5635246	inverse	5056	ATCOPIA93(AT5TE20395)	89
82	chr3:12191989..12192218	chr5	5629977	5635276	sens	5299	ATCOPIA93(AT5TE20395)	77
86	chr3:12948560..12948836	chr5	5630002	5635302	inverse	5300	ATCOPIA93(AT5TE20395)	94
87	chr3:12949514..12949525	chr1	12753550	12758600	inverse	5050	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
95	chr3:15322342..15322512	chr2	4900814	4909279	sens	8465	ATENSPM3(AT2TE20205)	ND
98	chr3:16372709..16372794	chr5	5629992	5635298	sens	5306	ATCOPIA93(AT5TE20395)	81
99	chr3:16397815..16398032	chr5	5630032	5635299	sens	5267	ATCOPIA93(AT5TE20395)	90
100	chr3:20887413..20887592	chr5	5629980	5635284	inverse	5304	ATCOPIA93(AT5TE20395)	76
101	chr3:22059229..22059406	chr5	5629977	5635304	inverse	5327	ATCOPIA93(AT5TE20395)	124
102	chr3:22314242..22314331	chr2	10001496	10009650	sens	8154	VANDAL21(AT2TE42810)	366
103	chr3:22815614..22815966	chr2	4389269	4390599	sens	1330	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	ND
104	chr4:1151512..1151580	chr2	10001487	10009562	inverse	8075	VANDAL21(AT2TE42810)	372
107	chr4:3408081..3408155	chr5	5629986	5635307	sens	5321	ATCOPIA93(AT5TE20395)	203
108	chr4:3408289..3410647	chr1	12753506	12755464	sens	1958	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	ND
117	chr4:4410835..4410929	chr5	5629979	5635305	sens	5326	ATCOPIA93(AT5TE20395)	194
119	chr4:9452593..9452670	chr5	5629977	5635305	sens	5328	ATCOPIA93(AT5TE20395)	181
120	chr4:9650042..9650113	chr5	5629989	5635305	sens	5316	ATCOPIA93(AT5TE20395)	230
121	chr4:9759914..9760044	chr5	5629977	5635299	inverse	5322	ATCOPIA93(AT5TE20395)	223
122	chr4:9818962..9819054	chr5	5630007	5635305	sens	5298	ATCOPIA93(AT5TE20395)	219
123	chr4:10312269..10312377	chr2	4900809	4909263	inverse	8454	ATENSPM3(AT2TE20205)	192
125	chr4:14401743..14401859	chr5	5629993	5635295	sens	5302	ATCOPIA93(AT5TE20395)	110
126	chr4:17266360..17266587	chr2	4902566	4909274	sens	6708	ATENSPM3(AT2TE20205)	ND
127	chr5:443390..443484	chr5	5629978	5635301	sens	5323	ATCOPIA93(AT5TE20395)	110
128	chr5:2167738..2167753	chr5	5630001	5635249	inverse	5248	ATCOPIA93(AT5TE20395)	70
129	chr5:2173158..2173261	chr2	4900844	4909222	sens	8378	ATENSPM3(AT2TE20205)	423
131	chr5:6086800..6086869	chr5	5629980	5634918	sens	4938	ATCOPIA93(AT5TE20395)	140
132	chr5:6996529..6996662	chr2	4900837	4909168	inverse	8331	ATENSPM3(AT2TE20205)	ND
133	chr5:8062273..8062410	chr5	5630025	5634821	sens	4796	ATCOPIA93(AT5TE20395)	81
134	chr5:10991586..10991754	chr2	4900808	4909200	sens	8392	ATENSPM3(AT2TE20205)	ND
135	chr5:11346508..11346594	chr5	5629991	5635143	inverse	5152	ATCOPIA93(AT5TE20395)	203
136	chr5:11631606..11631706	chr2	4900873	4909270	inverse	8397	ATENSPM3(AT2TE20205)	305
139	chr5:12757154..12757265	chr5	5629983	5634975	sens	4992	ATCOPIA93(AT5TE20395)	235
140	chr5:13539181..13539302	chr5	5629977	5634973	sens	4996	ATCOPIA93(AT5TE20395)	237
141	chr5:14701504..14701594	chr5	5629987	5634963	sens	4976	ATCOPIA93(AT5TE20395)	210
142	chr5:14702214..14704490	chr1	12753516	12755468	sens	1952	ATCOPIA93(AT1TE41580)	ND
143	chr5:14942997..14943109	chr5	5629978	5634969	inverse	4991	ATCOPIA93(AT5TE20395)	108
144	chr5:16011085..16011311	chr2	4900948	4909268	sens	8320	ATENSPM3(AT2TE20205)	ND
145	chr5:16269221..16269393	chr2	4904016	4909280	inverse	5264	ATENSPM3(AT2TE20205)	354
146	chr5:16270332..16270496	chr2	4386374	4391617	sens	5243	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	ND
147	chr5:18596063..18596141	chr5	5629977	5634970	inverse	4993	ATCOPIA93(AT5TE20395)	232
149	chr5:22557952..22558062	chr2	4900811	4909247	sens	8436	ATENSPM3(AT2TE20205)	407
150	chr5:23357867..23358212	chr2	4901423	4905858	inverse	4435	ATENSPM3(AT2TE20205)	ND

151	chr5:24009437..24009586	chr5	5629979	5634967	sens	4988	ATCOPIA93(AT5TE20395)	220
152	chr5:24048187..24048288	chr5	5629985	5634941	sens	4956	ATCOPIA93(AT5TE20395)	120
153	chr5:24723654..24727520	chr1	17203925	17206036	sens	2111	ATCOPIA63(AT1TE57025)	45
154	chr5:26093508..26093564	chr5	5629978	5634892	inverse	4914	ATCOPIA93(AT5TE20395)	122
155	chr5:26093584..26093813	chr5	5629978	5634968	inverse	4990	ATCOPIA93(AT5TE20395)	122

epiRIL 137

52	chr4:3440756..3440922	chr2	4900814	4909280	sens	8466	ATENSPM3(AT2TE20205)	371
65	chr4:10767041..10767121	chr2	4900846	4909265	sens	8419	ATENSPM3(AT2TE20205)	395
73	chr5:26286862..26286967	chr2	4900809	4909199	inverse	8390	ATENSPM3(AT2TE20205)	0

epiRIL 408

1	chr1:845287..845531	chr5	5630000	5635291	sens	5291	ATCOPIA93(AT5TE20395)	125
2	chr1:3577685..3577994	chr5	5629984	5635284	inverse	5300	ATCOPIA93(AT5TE20395)	113
3	chr1:4071314..4071562	chr2	4900813	4909264	inverse	8451	ATENSPM3(AT2TE20205)	ND
4	chr1:6621624..6622220	chr1	18013252	18018358	inverse	5106	ATCOPIA78(AT1TE59755)	45
27	chr2:5936758..5936842	chr2	4901198	4906698	inverse	5500	ATENSPM3(AT2TE20205)	361
28	chr2:16526250..16526371	chr2	4900854	4909150	sens	8296	ATENSPM3(AT2TE20205)	436
29	chr3:6725383..6725575	chr5	5629978	5635296	sens	5318	ATCOPIA93(AT5TE20395)	115
51	chr4:9048444..9048510	chr2	4903098	4909276	inverse	6178	ATENSPM3(AT2TE20205)	ND
52	chr4:9074071..9074148	chr2	4900812	4909143	inverse	8331	ATENSPM3(AT2TE20205)	ND
53	chr4:9080459..9080603	chr2	4900861	4906437	sens	5576	ATENSPM3(AT2TE20205)	414
55	chr4:9410551..9410644	chr2	4900821	4909279	inverse	8458	ATENSPM3(AT2TE20205)	509
60	chr5:14751432..14751665	chr5	5629985	5634962	inverse	4977	ATCOPIA93(AT5TE20395)	110
61	chr5:15490215..15490395	chr5	5629984	5635285	sens	5301	ATCOPIA93(AT5TE20395)	120
62	chr5:16694712..16694760	chr5	5629984	5634885	sens	4901	ATCOPIA93(AT5TE20395)	187
63	chr5:16694981..16695448	chr5	5630428	5634869	sens	4441	ATCOPIA93(AT5TE20395)	182
64	chr5:21186218..21186309	chr2	4900813	4909199	sens	8386	ATENSPM3(AT2TE20205)	306
67	chr5:26570358..26574862	chr1	3360541	3361316	sens	775	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
67	chr5:26572113..26573806	chr1	3780771	3785274	inverse	4503	ATCOPIA78(AT1TE12295)	0
67	chr5:26570492..26575187	chr1	7717452	7720103	inverse	2651	ATCOPIA78(AT1TE24850)	0
67	chr5:26572113..26573806	chr1	7717362	7720112	inverse	2750	ATCOPIA78(AT1TE24850)	0
67	chr5:26572037..26573806	chr1	18013168	18015606	inverse	2438	ATCOPIA78(AT1TE59755)	0
67	chr5:26572037..26573806	chr1	21524996	21529461	sens	4465	ATCOPIA78(AT1TE71045)	0
67	chr5:26571903..26573806	chr3	13369246	13371580	inverse	2334	ATCOPIA78(AT3TE54550)	0
67	chr5:26572037..26573806	chr3	22059541	22062055	inverse	2514	ATCOPIA78(AT3TE89830)	0
67	chr5:26572037..26574298	chr3	22695567	22699983	sens	4416	ATCOPIA78(AT3TE92525)	1
67	chr5:26570636..26575065	chr4	9320306	9320550	inverse	244	ATCOPIA78(AT4TE42010)	0
67	chr5:26572037..26573806	chr5	4208084	4212523	sens	4439	ATCOPIA78(AT5TE15240)	0
68	chr5:25213272..25213604	chr1	12753524	12759818	sens	6294	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
68	chr5:25213272..25213604	chr5	5629979	5634965	sens	4986	ATCOPIA93(AT5TE20395)	0

epiRIL 439

1	chr1:6622219..6622410	chr1	18013167	18018387	inverse	5220	ATCOPIA78(AT1TE59755)	134
2	chr1:11211722..11211794	chr5	5629978	5635310	inverse	5332	ATCOPIA93(AT5TE20395)	172
4	chr1:24091117..24091194	chr2	4900803	4909285	sens	8482	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	621
5	chr1:24334475..24334558	chr2	4900805	4909284	inverse	8479	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	637
6	chr1:28631078..28631250	chr2	10001501	10009630	inverse	8129	VANDAL21(AT2TE42810)	577
12	chr2:79850..80054	chr2	10001483	10009651	inverse	8168	VANDAL21(AT2TE42810)	578
18	chr2:10510085..10512441	chr4	4358130	4360294	inverse	2164	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	427
19	chr2:11505759..11505852	chr2	4900801	4909274	sens	8473	ATENSPM3(AT2TE20205)	656
20	chr3:8764431..8764434	chr3	22059538	22064329	inverse	4791	ATCOPIA78(AT3TE89830)	36
20	chr3:8764355..8764535	chr5	4208084	4213082	sens	4998	ATCOPIA78(AT5TE15240)	83
25	chr3:21970792..21970868	chr5	5629978	5635281	sens	5303	ATCOPIA93(AT5TE20395)	185

26	chr3:22615405..22615551	chr3	22695569	22700520	inverse	4951	ATCOPIA78(AT3TE92525)	ND
34	chr4:10739764..10739875	chr3	22059542	22063886	inverse	4344	ATCOPIA78(AT3TE89830)	40
34	chr4:10739764..10739875	chr3	22695564	22700474	sens	4910	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	26
36	chr5:969696..971395	chr4	4358131	4360292	sens	2161	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
37	chr5:6591834..6592026	chr5	5629983	5634958	sens	4975	ATCOPIA93(AT5TE20395)	127
38	chr5:7815404..7815488	chr2	4900819	4909279	inverse	8460	ATENSPM3(AT2TE20205)	681

epiRIL 454

1	chr1:3115..3249	chr5	5629979	5635291	sens	5312	ATCOPIA93(AT5TE20395)	137
3	chr1:11191545..11191795	chr5	5629978	5635286	inverse	5308	ATCOPIA93(AT5TE20395)	109
6	chr1:15083368..15083679	chr2	4904200	4908816	sens	4616	ATENSPM3(AT2TE20205)	ND
13	chr1:25653718..25654122	chr5	5629983	5635303	inverse	5320	ATCOPIA93(AT5TE20395)	205
14	chr1:29616254..29616578	chr5	5630010	5635273	inverse	5263	ATCOPIA93(AT5TE20395)	77
35	chr2:11496506..11496664	chr5	5629976	5635276	sens	5300	ATCOPIA93(AT5TE20395)	94
36	chr3:807190..807542	chr5	5629989	5635283	inverse	5294	ATCOPIA93(AT5TE20395)	75
37	chr3:1737581..1737875	chr5	5629978	5635293	sens	5315	ATCOPIA93(AT5TE20395)	123
38	chr3:3083438..3083594	chr5	5629986	5635305	sens	5319	ATCOPIA93(AT5TE20395)	214
39	chr3:4492935..4493027	chr5	5629985	5635297	inverse	5312	ATCOPIA93(AT5TE20395)	83
40	chr3:5595062..5595173	chr2	4900805	4909282	sens	8477	ATENSPM3(AT2TE20205)	425
41	chr3:8764369..8764724	chr3	22059537	22064320	inverse	4783	ATCOPIA78(AT3TE89830)	24
41	chr3:8764369..8764724	chr5	4208093	4213081	sens	4988	ATCOPIA78(AT5TE15240)	77
51	chr3:16580047..16580428	chr5	5629982	5635269	sens	5287	ATCOPIA93(AT5TE20395)	104
52	chr3:16894460..16894609	chr5	5629981	5635267	inverse	5286	ATCOPIA93(AT5TE20395)	107
53	chr3:21980027..21980291	chr5	5629983	5635265	inverse	5282	ATCOPIA93(AT5TE20395)	105
54	chr3:22228074..22228854	chr5	5630052	5635073	sens	5021	ATCOPIA93(AT5TE20395)	25
55	chr3:22556611..22556699	chr5	5629975	5635308	inverse	5333	ATCOPIA93(AT5TE20395)	193
56	chr4:1151475..1151604	chr2	10001512	10009495	inverse	7983	VANDAL21(AT2TE42810)	373
57	chr4:2199494..2202716	chr5	9640968	9644562	sens	3594	ARNOLD2(AT5TE34985) (AT5TE34980)	ND
58	chr4:3651904..3652219	chr5	5629980	5635304	inverse	5324	ATCOPIA93(AT5TE20395)	99
70	chr4:5006317..5006790	chr5	5629981	5635277	inverse	5296	ATCOPIA93(AT5TE20395)	105
71	chr4:7002644..7003242	chr1	12755000	12759346	sens	4346	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	1
72	chr4:7003389..7003625	chr5	5629977	5635296	sens	5319	ATCOPIA93(AT5TE20395)	240
73	chr4:7083238..7083461	chr5	5629981	5635293	inverse	5312	ATCOPIA93(AT5TE20395)	109
74	chr4:8862711..8862828	chr2	10001494	10009607	inverse	8113	VANDAL21(AT2TE42810)	326
76	chr4:10739334..10739851	chr3	22059537	22064209	inverse	4672	ATCOPIA78(AT3TE89830)	32
77	chr4:13252197..13252644	chr1	18013164	18017782	inverse	4618	ATCOPIA78(AT1TE59755)	36
78	chr4:16682304..16682584	chr5	5629982	5635298	sens	5316	ATCOPIA93(AT5TE20395)	107
79	chr5:729472..729675	chr5	5630049	5635303	sens	5254	ATCOPIA93(AT5TE20395)	99
80	chr5:8895666..8895869	chr5	5629980	5634870	sens	4890	ATCOPIA93(AT5TE20395)	178
86	chr5:13333017..13333469	chr5	5629978	5635101	inverse	5123	ATCOPIA93(AT5TE20395)	96
88	chr5:22921922..22922268	chr5	5629987	5634921	inverse	4934	ATCOPIA93(AT5TE20395)	62
89	chr5:23913821..23913992	chr5	5629982	5634892	sens	4910	ATCOPIA93(AT5TE20395)	145
89	chr5:25834575..25834843	chr5	5630011	5634975	sens	4964	ATCOPIA93(AT5TE20395)	145

epiRIL 277

5	chr1:25746302..25746355	chr2	4385855	4393464	sens	7609	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	ND
6	chr1:25746408..25746525	chr2	4900804	4909283	inverse	8479	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	484
17	chr2:10507408..10513173	chr4	4358131	4360279	inverse	2148	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
18	chr2:14295264..14295420	chr2	4900802	4909192	inverse	8390	ATENSPM3(AT2TE20205)	510
19	chr3:8764449..8764901	chr3	22059537	22064331	inverse	4794	ATCOPIA78(AT3TE89830)	26
19	chr3:8764449..8764726	chr5	4208082	4213083	sens	5001	ATCOPIA78(AT5TE15240)	78
31	chr4:8143015..8143115	chr2	4900807	4909276	inverse	8469	ATENSPM3(AT2TE20205)	461
33	chr5:6361915..6362503	chr5	4208084	4212620	inverse	4536	ATCOPIA78(AT5TE15240)	28

38	chr5:26570618..26575666	chr1	3360539	3361316	sens	777	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
38	chr5:26571044..26574113	chr1	3780764	3783282	inverse	2518	ATCOPIA78(AT1TE12295)	0
38	chr5:26570454..26575890	chr1	7717812	7720033	inverse	2221	ATCOPIA78(AT1TE24850)	0
38	chr5:26570590..26575341	chr1	7717355	7720108	inverse	2753	ATCOPIA78(AT1TE24850)	0
38	chr5:26571044..26574113	chr1	18013162	18015679	inverse	2517	ATCOPIA78(AT1TE59755)	0
38	chr5:26570615..26574849	chr1	21527338	21529542	sens	2204	ATCOPIA78(AT1TE71045)	0
38	chr5:26571044..26575114	chr3	13369245	13371624	inverse	2379	ATCOPIA78(AT3TE54550)	0
38	chr5:26571044..26574113	chr3	22059534	22062052	inverse	2518	ATCOPIA78(AT3TE89830)	0
38	chr5:26570539..26575890	chr3	22698005	22700113	sens	2108	ATCOPIA78(AT3TE92525)	0
38	chr5:26570615..26575890	chr3	22698005	22700147	sens	2142	ATCOPIA78(AT3TE92525)	0
38	chr5:26571044..26575114	chr4	9320299	9320549	inverse	250	ATCOPIA78(AT4TE42010)	0

epiRIL 52

1	chr1:1919672..1919813	chr2	10001499	10009663	sens	8164	VANDAL21(AT2TE42810)	ND
15	chr2:5566776..5577645	chr1	12945139	12946169	inverse	1030	ATENSPM3(AT1TE42210)	ND
20	chr3:15713149..15721050	chr2	4903198	4904649	sens	1451	ATENSPM3(AT2TE20205)	10
21	chr4:1151513..1151582	chr2	10001484	10009567	inverse	8083	VANDAL21(AT2TE42810)	594
26	chr4:10739750..10740065	chr3	22059536	22064330	inverse	4794	ATCOPIA78(AT3TE89830)	28
26	chr4:10739674..10740093	chr3	22695566	22700410	sens	4844	ATCOPIA78(AT3TE92525)	8
28	chr4:13252448..13252476	chr1	18013170	18018262	inverse	5092	ATCOPIA78(AT1TE59755)	57
29	chr4:16482538..16482624	chr2	4900804	4909278	sens	8474	ATENSPM3(AT2TE20205)	636
30	chr4:16530417..16534012	chr3	15283663	15286612	inverse	2949	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
31	chr5:6361622..6362874	chr5	4208083	4212592	inverse	4509	ATCOPIA78(AT5TE15240)	35

epiRIL 260

1	chr1:2668987..2669079	chr2	10001503	10009633	inverse	8130	VANDAL21(AT2TE42810)	ND
2	chr1:6622186..6622366	chr1	18013162	18018417	inverse	5255	ATCOPIA78(AT1TE59755)	161
13	chr1:15397477..15406559	chr4	2068968	2071132	sens	2164	ATCOPIA20(AT4TE09805)	ND
16	chr1:17995264..17995337	chr2	10001478	10009656	sens	8178	VANDAL21(AT2TE42810)	775
18	chr1:24397651..24399160	chr4	5541884	5545529	inverse	3645	HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	ND
47	chr2:15999918..16000066	chr2	10001477	10009538	inverse	8061	VANDAL21(AT2TE42810)	339
48	chr3:9793171..9793239	chr2	10001477	10009661	inverse	8184	VANDAL21(AT2TE42810)	763
49	chr3:11238043..11238129	chr2	10001488	10009658	inverse	8170	VANDAL21(AT2TE42810)	756
50	chr3:11299566..11299642	chr2	10001508	10009661	sens	8153	VANDAL21(AT2TE42810)	635
68	chr3:22615394..22615664	chr3	22695567	22700517	inverse	4950	ATCOPIA78(AT3TE92525)	ND
90	chr4:10739361..10739936	chr3	22059534	22064165	inverse	4631	ATCOPIA78(AT3TE89830)	39
90	chr4:10739332..10739936	chr3	22695564	22700438	sens	4874	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	11
92	chr5:1405136..1405215	chr2	10001495	10009661	inverse	8166	VANDAL21(AT2TE42810)	704
114	chr5:20047910..20047987	chr2	10001484	10009518	inverse	8034	VANDAL21(AT2TE42810)	254
117	chr5:26571017..26574804	chr1	3360536	3361305	sens	769	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
117	chr5:26571662..26574273	chr1	3780773	3785278	inverse	4505	ATCOPIA78(AT1TE12295)	2
117	chr5:26570903..26574423	chr1	7717813	7722131	inverse	4318	ATCOPIA78(AT1TE24850)	0
117	chr5:26571180..26574273	chr1	7714735	7715037	inverse	302	ATCOPIA78(AT1TE24850)	0
117	chr5:26571738..26574273	chr1	18013170	18017637	inverse	4467	ATCOPIA78(AT1TE59755)	0
117	chr5:26571180..26574273	chr1	21524991	21529463	sens	4472	ATCOPIA78(AT1TE71045)	0
117	chr5:26571662..26574273	chr3	13369245	13373734	inverse	4489	ATCOPIA78(AT3TE54550)	0
117	chr5:26571662..26574273	chr3	22059532	22064137	inverse	4605	ATCOPIA78(AT3TE89830)	0
117	chr5:26571662..26574273	chr3	22695562	22700079	sens	4517	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	0
117	chr5:26570979..26574273	chr4	9320297	9320549	inverse	252	ATCOPIA78(AT4TE42010)	0
117	chr5:26571738..26574423	chr5	4208079	4212619	sens	4540	ATCOPIA78(AT5TE15240)	0

epiRIL 144

22	chr2:3623811..3624481	chr2	4903216	4909106	sens	5890	ATENSPM3(AT2TE20205)	82
29	chr2:18228415..18228522	chr2	10001484	10009580	inverse	8096	VANDAL21(AT2TE42810)	246
30	chr3:3453438..3453531	chr2	10001495	10009648	sens	8153	VANDAL21(AT2TE42810)	289
32	chr3:5606823..5606957	chr2	4900810	4909280	sens	8470	ATENSPM3(AT2TE20205)	279
61	chr4:10739647..10740104	chr3	22059535	22064328	inverse	4793	ATCOPIA78(AT3TE89830)	13
62	chr4:13251469..13253104	chr1	18013162	18016892	inverse	3730	ATCOPIA78(AT1TE59755)	33
75	chr5:15446312..15446414	chr2	10001496	10009554	sens	8058	VANDAL21(AT2TE42810)	260
76	chr5:26570719..26575276	chr1	3360542	3361311	sens	769	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
76	chr5:26573171..26574037	chr1	3780765	3783278	inverse	2513	ATCOPIA78(AT1TE12295)	1
76	chr5:26572553..26574512	chr1	7717353	7720093	inverse	2740	ATCOPIA78(AT1TE24850)	0
76	chr5:26573171..26574037	chr1	18013162	18015675	inverse	2513	ATCOPIA78(AT1TE59755)	0
76	chr5:26573171..26574037	chr1	21524997	21529466	sens	4469	ATCOPIA78(AT1TE71045)	0
76	chr5:26572553..26574037	chr3	13369254	13371630	inverse	2376	ATCOPIA78(AT3TE54550)	0
76	chr5:26573171..26574037	chr3	22059532	22062036	inverse	2504	ATCOPIA78(AT3TE89830)	0
76	chr5:26573171..26574037	chr3	22695568	22700071	sens	4503	ATCOPIA78(AT3TE92525)	0
76	chr5:26573171..26574037	chr5	4208085	4212611	sens	4526	ATCOPIA78(AT5TE15240)	0

epiRIL 393

1	chr1:1010635..1010853	chr5	5629999	5635270	inverse	5271	ATCOPIA93(AT5TE20395)	87
6	chr1:27309649..27309906	chr5	5629980	5635298	inverse	5318	ATCOPIA93(AT5TE20395)	118
15	chr3:1808197..1808485	chr5	5629986	5635291	sens	5305	ATCOPIA93(AT5TE20395)	140
16	chr3:6121581..6121708	chr5	5629978	5635307	sens	5329	ATCOPIA93(AT5TE20395)	120
17	chr3:7051040..7051352	chr5	5629980	5635282	sens	5302	ATCOPIA93(AT5TE20395)	157
18	chr3:8764172..8764592	chr3	22059538	22064330	inverse	4792	ATCOPIA78(AT3TE89830)	28
18	chr3:8764172..8764807	chr5	4208083	4213065	sens	4982	ATCOPIA78(AT5TE15240)	85
21	chr4:672061..672150	chr5	5629979	5635307	inverse	5328	ATCOPIA93(AT5TE20395)	226
22	chr4:2213280..2213621	chr5	5629992	5635307	sens	5315	ATCOPIA93(AT5TE20395)	97
29	chr4:13252366..13252730	chr1	18013162	18018262	inverse	5100	ATCOPIA78(AT1TE59755)	60
30	chr4:14628521..14628900	chr5	5629986	5635299	sens	5313	ATCOPIA93(AT5TE20395)	70
31	chr5:1789418..1789595	chr5	5629979	5635307	inverse	5328	ATCOPIA93(AT5TE20395)	137
33	chr5:13858490..13858575	chr1	12753529	12760094	inverse	6565	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
33	chr5:13858490..13858575	chr5	5629982	5635011	inverse	5029	ATCOPIA93(AT5TE20395)	0

epiRIL 344

1	chr1:5064214..5068166	chr3	15283683	15286611	inverse	2928	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
2	chr1:6621944..6622218	chr1	18013182	18018387	inverse	5205	ATCOPIA78(AT1TE59755)	130
11	chr1:28364592..28372739	chr3	15285397	15286612	sens	1215	ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
27	chr2:11145638..11148991	chr4	5541892	5545526	inverse	3634	HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	280
28	chr3:8763977..8764586	chr3	22059532	22064327	inverse	4795	ATCOPIA78(AT3TE89830)	49
28	chr3:8764315..8764721	chr5	4208083	4213088	sens	5005	ATCOPIA78(AT5TE15240)	100
35	chr3:22615290..22615752	chr3	22695569	22700520	inverse	4951	ATCOPIA78(AT3TE92525)	ND
46	chr4:10739116..10740452	chr3	22059541	22064209	inverse	4668	ATCOPIA78(AT3TE89830)	34
46	chr4:10739116..10740452	chr3	22695587	22700103	sens	4516	ATCOPIA78(AT3TE92525)	8

epiRIL 377

1	chr1:6981105..6981246	chr5	5629990	5635306	sens	5316	ATCOPIA93(AT5TE20395)	ND
2	chr1:11097396..11097567	chr5	5630039	5635302	sens	5263	ATCOPIA93(AT5TE20395)	130
4	chr1:12824462..12824497	chr1	12753513	12758378	sens	4865	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
6	chr1:22238462..22238556	chr5	5629980	5635305	sens	5325	ATCOPIA93(AT5TE20395)	134
7	chr1:22650750..22650922	chr5	5629982	5635245	sens	5263	ATCOPIA93(AT5TE20395)	345
8	chr1:23720275..23721152	chr5	5629980	5635294	sens	5314	ATCOPIA93(AT5TE20395)	76
9	chr1:23954169..23954285	chr5	5629984	5635301	inverse	5317	ATCOPIA93(AT5TE20395)	282

10	chr1:25432395..25432538	chr5	5629979	5635305	sens	5326	ATCOPIA93(AT5TE20395)	107
11	chr1:26275918..26276042	chr5	5629978	5635292	sens	5314	ATCOPIA93(AT5TE20395)	119
12	chr1:28005602..28006116	chr5	18142163	18146897	inverse	4734	ATCOPIA21(AT5TE65370)	ND
13	chr2:2092..2186	chr5	5629979	5635203	sens	5224	ATCOPIA93(AT5TE20395)	ND
16	chr2:715496..717274	chr4	5541906	5545509	sens	3603	HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	135
17	chr2:2360192..2360243	chr5	5629976	5635165	inverse	5189	ATCOPIA93(AT5TE20395)	107
18	chr2:2360268..2360365	chr5	5629976	5635237	inverse	5261	ATCOPIA93(AT5TE20395)	107
21	chr2:3353399..3353534	chr5	5629978	5635303	sens	5325	ATCOPIA93(AT5TE20395)	102
23	chr2:10507681..10510711	chr4	4358146	4360291	inverse	2145	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
24	chr2:14267849..14267946	chr2	10001482	10009609	inverse	8127	VANDAL21(AT2TE42810)	524
25	chr2:16605589..16605762	chr5	5630003	5635245	sens	5242	ATCOPIA93(AT5TE20395)	150
25	chr2:16605890..16605977	chr5	5630003	5635302	sens	5299	ATCOPIA93(AT5TE20395)	150
26	chr2:19346318..19346405	chr1	12753544	12759328	inverse	5784	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
27	chr2:19347043..19347143	chr5	5629992	5635298	inverse	5306	ATCOPIA93(AT5TE20395)	123
28	chr3:5914061..5914159	chr5	5629982	5635297	inverse	5315	ATCOPIA93(AT5TE20395)	210
29	chr3:6218403..6218553	chr5	5629981	5635304	sens	5323	ATCOPIA93(AT5TE20395)	120
30	chr3:12434921..12435046	chr5	5629978	5635292	sens	5314	ATCOPIA93(AT5TE20395)	151
31	chr3:12435078..12435111	chr1	12753526	12758629	sens	5103	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
32	chr3:12435154..12437862	chr1	12753526	12755451	sens	1925	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
34	chr3:14196059..14196153	chr5	5629979	5635286	sens	5307	ATCOPIA93(AT5TE20395)	129
35	chr3:14356885..14357025	chr5	5629978	5635287	sens	5309	ATCOPIA93(AT5TE20395)	153
36	chr3:14507559..14509454	chr1	12753512	12755466	inverse	1954	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
37	chr3:14509887..14510050	chr5	5629978	5635307	inverse	5329	ATCOPIA93(AT5TE20395)	145
38	chr3:14561537..14561665	chr5	5629982	5635299	inverse	5317	ATCOPIA93(AT5TE20395)	105
39	chr3:20783239..20783367	chr5	5629988	5635303	inverse	5315	ATCOPIA93(AT5TE20395)	144
40	chr4:176626..176758	chr5	5629996	5635303	sens	5307	ATCOPIA93(AT5TE20395)	186
41	chr4:1151493..1151631	chr2	10001488	10009585	inverse	8097	VANDAL21(AT2TE42810)	503
42	chr4:1228876..1228947	chr2	10001495	10009565	sens	8070	VANDAL21(AT2TE42810)	463
43	chr4:3064592..3064800	chr5	5629977	5635263	inverse	5286	ATCOPIA93(AT5TE20395)	207
44	chr4:3140937..3141009	chr5	5629982	5635299	inverse	5317	ATCOPIA93(AT5TE20395)	222
45	chr4:3141031..3141441	chr1	12755070	12760306	inverse	5236	ATCOPIA93(AT1TE41580)	ND
46	chr4:4248075..4248280	chr5	5629980	5635307	sens	5327	ATCOPIA93(AT5TE20395)	180
47	chr4:4446527..4448129	chr3	15283668	15286594	sens	2926	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
51	chr4:7937971..7938116	chr5	5630000	5635284	inverse	5284	ATCOPIA93(AT5TE20395)	89
53	chr4:10739742..10739931	chr3	22059536	22063979	inverse	4443	ATCOPIA78(AT3TE89830)	25
53	chr4:10739666..10739931	chr3	22695571	22700444	sens	4873	ATCOPIA78(AT3TE92525)	19
55	chr4:11746947..11747022	chr5	5629991	5635300	inverse	5309	ATCOPIA93(AT5TE20395)	190
56	chr4:13772011..13772196	chr5	5629986	5635294	sens	5308	ATCOPIA93(AT5TE20395)	241
57	chr4:14137385..14137523	chr5	5629976	5635306	inverse	5330	ATCOPIA93(AT5TE20395)	247
58	chr4:16317918..16317992	chr5	5629982	5635296	sens	5314	ATCOPIA93(AT5TE20395)	201
59	chr4:18365339..18367807	chr1	12753537	12755467	inverse	1930	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
60	chr4:18367839..18368049	chr5	5629979	5635308	inverse	5329	ATCOPIA93(AT5TE20395)	178
61	chr5:3388841..3388913	chr5	5629977	5635294	inverse	5317	ATCOPIA93(AT5TE20395)	122
62	chr5:4462148..4462273	chr5	5629983	5635297	inverse	5314	ATCOPIA93(AT5TE20395)	68
64	chr5:8839337..8839474	chr5	5629994	5634972	inverse	4978	ATCOPIA93(AT5TE20395)	80
65	chr5:12799106..12799231	chr5	5629977	5635271	sens	5294	ATCOPIA93(AT5TE20395)	109
66	chr5:18302605..18302684	chr5	5629980	5634887	sens	4907	ATCOPIA93(AT5TE20395)	173
67	chr5:18788963..18789087	chr5	5629981	5634946	inverse	4965	ATCOPIA93(AT5TE20395)	137
68	chr5:20590822..20590941	chr5	5630036	5635301	inverse	5265	ATCOPIA93(AT5TE20395)	114
69	chr5:20624127..20624283	chr5	5629983	5635292	inverse	5309	ATCOPIA93(AT5TE20395)	96
71	chr5:21346254..21346378	chr5	5629983	5634887	sens	4904	ATCOPIA93(AT5TE20395)	187
71	chr5:23271828..23271897	chr5	5630008	5634970	inverse	4962	ATCOPIA93(AT5TE20395)	187

1	chr1:2560099..2560196	chr5	5629995	5635304	inverse	5309	ATCOPIA93(AT5TE20395)	163
2	chr1:3184297..3188055	chr3	15283670	15286601	inverse	2931	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
3	chr1:5025929..5026028	chr5	5629987	5635304	sens	5317	ATCOPIA93(AT5TE20395)	209
4	chr1:8518660..8518810	chr5	5629985	5635306	inverse	5321	ATCOPIA93(AT5TE20395)	103
5	chr1:8574200..8574346	chr5	5629978	5635286	sens	5308	ATCOPIA93(AT5TE20395)	112
6	chr1:8704699..8709655	chr4	4358138	4360286	inverse	2148	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
7	chr1:9223529..9229007	chr4	4358136	4360288	sens	2152	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
8	chr1:11488135..11488279	chr2	4900814	4909274	sens	8460	ATENSPM3(AT2TE20205)	218
9	chr1:11584650..11584836	chr5	5629982	5635303	sens	5321	ATCOPIA93(AT5TE20395)	96
17	chr1:23440928..23441100	chr2	4900816	4909168	sens	8352	ATENSPM3(AT2TE20205)	237
18	chr1:23638799..23638884	chr5	5629981	5635307	inverse	5326	ATCOPIA93(AT5TE20395)	124
19	chr1:25662069..25662212	chr5	5629988	5635288	inverse	5300	ATCOPIA93(AT5TE20395)	98
20	chr1:26405975..26412312	chr4	4358188	4360236	inverse	2048	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
21	chr1:28721700..28721723	chr5	5629986	5635305	sens	5319	ATCOPIA93(AT5TE20395)	191
22	chr1:29356596..29356682	chr5	5629985	5635276	inverse	5291	ATCOPIA93(AT5TE20395)	134
43	chr2:14470927..14471287	chr5	5629983	5635295	inverse	5312	ATCOPIA93(AT5TE20395)	76
44	chr3:4095710..4095803	chr5	5629979	5635295	inverse	5316	ATCOPIA93(AT5TE20395)	184
45	chr3:5086750..5086916	chr5	5629986	5635286	inverse	5300	ATCOPIA93(AT5TE20395)	106
46	chr3:9280024..9280127	chr5	5629977	5635221	inverse	5244	ATCOPIA93(AT5TE20395)	113
55	chr3:22070517..22070536	chr5	5629986	5635307	inverse	5321	ATCOPIA93(AT5TE20395)	101
56	chr3:22070593..22070966	chr5	5629986	5635290	inverse	5304	ATCOPIA93(AT5TE20395)	101
57	chr4:968246..968401	chr5	5629978	5635299	sens	5321	ATCOPIA93(AT5TE20395)	161
58	chr4:1151510..1151582	chr2	10001505	10009642	inverse	8137	VANDAL21(AT2TE42810)	399
69	chr4:6495674..6495774	chr2	4900834	4909268	inverse	8434	ATENSPM3(AT2TE20205)	222
70	chr4:7209836..7210023	chr2	4900815	4909284	sens	8469	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	131
71	chr4:8675742..8675855	chr5	5629991	5635268	inverse	5277	ATCOPIA93(AT5TE20395)	82
73	chr4:9488290..9488414	chr5	5630034	5635273	inverse	5239	ATCOPIA93(AT5TE20395)	66
74	chr4:9649967..9650147	chr5	5629982	5635285	sens	5303	ATCOPIA93(AT5TE20395)	222
76	chr4:11140256..11144312	chr4	4358134	4360286	inverse	2152	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
77	chr4:12724239..12724324	chr2	4900900	4909263	sens	8363	ATENSPM3(AT2TE20205)	310
78	chr4:13337464..13337555	chr2	4900818	4909270	sens	8452	ATENSPM3(AT2TE20205)	213
79	chr4:13463657..13463864	chr5	5629989	5635297	inverse	5308	ATCOPIA93(AT5TE20395)	114
80	chr4:13463982..13464112	chr5	5629989	5635303	inverse	5314	ATCOPIA93(AT5TE20395)	130
81	chr4:14691795..14694289	chr2	4906168	4907001	inverse	833	ATENSPM3(AT2TE20205)	23
82	chr4:16396570..16396862	chr5	5629978	5635294	inverse	5316	ATCOPIA93(AT5TE20395)	83
83	chr4:17188662..17193777	chr4	4358153	4360294	inverse	2141	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
84	chr5:3321377..3321483	chr5	5629990	5635311	sens	5321	ATCOPIA93(AT5TE20395)	144
85	chr5:4938147..4939585	chr4	4358139	4360284	inverse	2145	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
86	chr5:7003537..7003755	chr5	5629992	5634871	inverse	4879	ATCOPIA93(AT5TE20395)	70
87	chr5:8173673..8173799	chr5	5630007	5634973	sens	4966	ATCOPIA93(AT5TE20395)	79
90	chr5:20091363..20093145	chr1	12944720	12946097	inverse	1377	ATENSPM3(AT1TE42210)	0
90	chr5:20091363..20091542	chr2	4900837	4907492	inverse	6655	ATENSPM3(AT2TE20205)	0
90	chr5:20091336..20092959	chr2	4390835	4393295	sens	2460	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	0

1	chr1:2165215..2165304	chr5	5629978	5635309	sens	5331	ATCOPIA93(AT5TE20395)	201
5	chr1:19830847..19830968	chr5	5629975	5635282	inverse	5307	ATCOPIA93(AT5TE20395)	208
6	chr1:28367981..28370315	chr3	15283664	15286611	sens	2947	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
7	chr1:30167279..30167428	chr5	5629983	5635292	inverse	5309	ATCOPIA93(AT5TE20395)	153
15	chr2:18351223..18351278	chr5	5629985	5635293	inverse	5308	ATCOPIA93(AT5TE20395)	178
15	chr2:18351299..18353464	chr5	5629985	5633354	inverse	3369	ATCOPIA93(AT5TE20395)	146
16	chr3:659331..659424	chr5	5629978	5635306	sens	5328	ATCOPIA93(AT5TE20395)	168
17	chr3:2573757..2573836	chr5	5629980	5635287	inverse	5307	ATCOPIA93(AT5TE20395)	207
20	chr3:14302838..14302841	chr5	5629982	5635236	inverse	5254	ATCOPIA93(AT5TE20395)	122
21	chr3:14303020..14303031	chr5	5629982	5635273	inverse	5291	ATCOPIA93(AT5TE20395)	122
22	chr3:14965645..14965825	chr5	5629982	5635278	inverse	5296	ATCOPIA93(AT5TE20395)	165
23	chr3:22615397..22615486	chr3	22695566	22700516	inverse	4950	ATCOPIA78(AT3TE92525)	ND
29	chr4:9650044..9650118	chr5	5629976	5635310	sens	5334	ATCOPIA93(AT5TE20395)	390
31	chr4:14566389..14566499	chr5	5629984	5635300	inverse	5316	ATCOPIA93(AT5TE20395)	178

epiRIL 98

1	chr1:12177658..12182006	chr3	15283682	15286600	inverse	2918	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
2	chr1:12617476..12622888	chr3	15283685	15286604	sens	2919	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
5	chr1:18832464..18832965	chr5	5629977	5635304	inverse	5327	ATCOPIA93(AT5TE20395)	200
7	chr1:22623115..22623766	chr5	5629977	5635305	sens	5328	ATCOPIA93(AT5TE20395)	152
8	chr1:23828998..23829693	chr5	5629978	5635283	sens	5305	ATCOPIA93(AT5TE20395)	164
9	chr1:24927442..24928047	chr5	5629982	5635303	sens	5321	ATCOPIA93(AT5TE20395)	159
10	chr1:26229693..26230176	chr5	5629981	5635287	inverse	5306	ATCOPIA93(AT5TE20395)	174
17	chr2:3297324..3297439	chr5	5629977	5635305	inverse	5328	ATCOPIA93(AT5TE20395)	114
22	chr2:9963786..9964184	chr1	12753516	12759323	inverse	5807	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
23	chr2:9964427..9965153	chr5	5629980	5635276	inverse	5296	ATCOPIA93(AT5TE20395)	146
24	chr2:9966532..9967013	chr1	12753511	12760283	inverse	6772	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
25	chr2:10506856..10513566	chr4	4358131	4360279	inverse	2148	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
26	chr3:6658949..6659050	chr2	10001507	10009569	inverse	8062	VANDAL21(AT2TE42810)	387
30	chr3:20004977..20005209	chr2	10001536	10009566	sens	8030	VANDAL21(AT2TE42810)	246
31	chr4:3430033..3430181	chr2	10001488	10009548	sens	8060	VANDAL21(AT2TE42810)	657
37	chr4:9649853..9650191	chr5	5629977	5635304	sens	5327	ATCOPIA93(AT5TE20395)	356
39	chr4:12164751..12164898	chr5	5629976	5635291	inverse	5315	ATCOPIA93(AT5TE20395)	336
40	chr4:16482577..16482641	chr2	4900805	4909106	sens	8301	ATENSPM3(AT2TE20205)	790
40	chr4:16482653..16483442	chr2	4904963	4909273	sens	4310	ATENSPM3(AT2TE20205)	658
41	chr5:5548906..5548976	chr2	10001481	10009637	inverse	8156	VANDAL21(AT2TE42810)	631
43	chr5:14253108..14253220	chr2	10001510	10009656	inverse	8146	VANDAL21(AT2TE42810)	649
44	chr5:26570900..26575083	chr1	3360540	3361315	sens	775	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
44	chr5:26573102..26573475	chr1	3780764	3785273	inverse	4509	ATCOPIA78(AT1TE12295)	0
44	chr5:26570033..26575715	chr1	7717814	7720038	inverse	2224	ATCOPIA78(AT1TE24850)	0
44	chr5:26571004..26573475	chr1	7714733	7715146	inverse	413	ATCOPIA78(AT1TE24850)	0
44	chr5:26573102..26573475	chr1	7714733	7722130	inverse	7397	ATCOPIA78(AT1TE24850)	0
44	chr5:26571004..26573475	chr1	18013162	18015683	inverse	2521	ATCOPIA78(AT1TE59755)	0
44	chr5:26573102..26573475	chr1	18013162	18017641	inverse	4479	ATCOPIA78(AT1TE59755)	0
44	chr5:26573102..26573475	chr1	21524991	21529470	sens	4479	ATCOPIA78(AT1TE71045)	0
44	chr5:26573102..26573475	chr3	13369244	13373729	inverse	4485	ATCOPIA78(AT3TE54550)	0
44	chr5:26573102..26573475	chr3	22059534	22063882	inverse	4348	ATCOPIA78(AT3TE89830)	0
44	chr5:26573102..26573475	chr3	22695566	22700079	sens	4513	ATCOPIA78(AT3TE92525)	0
44	chr5:26570643..26575482	chr4	9320299	9320553	inverse	254	ATCOPIA78(AT4TE42010)	0
44	chr5:26573102..26573475	chr5	4208083	4212619	sens	4536	ATCOPIA78(AT5TE15240)	0

epiRIL 60

1	chr1:6261367..6261460	chr2	10001481	10009628	sens	8147	VANDAL21(AT2TE42810)	574
2	chr1:12178270..12181765	chr3	15283668	15286610	inverse	2942	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
6	chr1:25956040..25956309	chr2	4900837	4909258	sens	8421	ATENSPM3(AT2TE20205)	ND
11	chr2:10321820..10321904	chr2	4900801	4909201	inverse	8400	ATENSPM3(AT2TE20205)	641
12	chr2:12660632..12660713	chr2	10001497	10009624	inverse	8127	VANDAL21(AT2TE42810)	666
13	chr2:13660418..13663919	chr3	15283667	15286611	sens	2944	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
14	chr2:15677093..15677345	chr5	5629994	5635309	sens	5315	ATCOPIA93(AT5TE20395)	180
15	chr2:17097260..17097372	chr2	4900803	4909259	sens	8456	ATENSPM3(AT2TE20205)	372
16	chr2:19390375..19394443	chr3	15283666	15286591	inverse	2925	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
17	chr3:3464127..3467214	chr4	4358131	4360296	inverse	2165	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
18	chr3:8764387..8764606	chr3	22059532	22064309	inverse	4777	ATCOPIA78(AT3TE89830)	29
18	chr3:8764557..8764584	chr5	4208082	4213088	sens	5006	ATCOPIA78(AT5TE15240)	110
22	chr3:15773104..15778796	chr4	4358130	4360295	sens	2165	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
23	chr3:16946117..16951602	chr4	4358128	4360296	sens	2168	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
24	chr3:18274281..18277593	chr3	15283667	15286610	inverse	2943	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
25	chr3:22615358..22615562	chr3	22695570	22700517	inverse	4947	ATCOPIA78(AT3TE92525)	ND
26	chr4:1151516..1151621	chr2	10001482	10009649	inverse	8167	VANDAL21(AT2TE42810)	636
29	chr4:5859205..5859488	chr2	10001510	10006094	inverse	4584	VANDAL21(AT2TE42810)	ND
30	chr4:5876408..5876685	chr2	10001505	10009634	inverse	8129	VANDAL21(AT2TE42810)	ND
32	chr4:9650038..9650112	chr5	5629979	5635286	sens	5307	ATCOPIA93(AT5TE20395)	412
35	chr5:16740222..16745361	chr4	4358148	4360294	sens	2146	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
36	chr5:17050129..17053285	chr4	4358190	4360292	sens	2102	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	0

epiRIL 53

51	chr5:6362046..6362908	chr5	4208183	4212680	inverse	4497	ATCOPIA78(AT5TE15240)	14

epiRIL8

4	chr1:13388612..13389923	chr4	5541898	5545527	inverse	3629	HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	144
11	chr1:15113574..15114120	chr2	3084118	3088980	inverse	4862	ATCOPIA20(AT2TE13385)	ND
12	chr1:15133809..15134182	chr2	3084118	3088887	inverse	4769	ATCOPIA20(AT2TE13385)	ND
15	chr1:15231427..15231956	chr2	3084149	3088980	inverse	4831	ATCOPIA20(AT2TE13385)	ND
16	chr1:15251356..15251782	chr2	3084057	3089394	inverse	5337	ATCOPIA20(AT2TE13385)	ND
17	chr1:15328623..15329106	chr2	3083919	3089301	inverse	5382	ATCOPIA20(AT2TE13385)	22
21	chr1:16836997..16837095	chr2	4900829	4909283	inverse	8454	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	358
23	chr1:20152246..20154468	chr3	15283669	15286590	sens	2921	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
24	chr1:28368128..28369087	chr3	15283666	15286613	sens	2947	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
52	chr2:9119833..9119990	chr2	10001489	10009656	sens	8167	VANDAL21(AT2TE42810)	318

53	chr2:12760655..12763305	chr2	4900800	4902452	inverse	1652	ATENSPM3(AT2TE20205)	113
54	chr2:13914006..13914213	chr5	5629998	5635231	sens	5233	ATCOPIA93(AT5TE20395)	89
56	chr3:5857582..5857776	chr2	10001487	10009537	sens	8050	VANDAL21(AT2TE42810)	343
102	chr5:7799010..7799202	chr5	5629986	5634948	inverse	4962	ATCOPIA93(AT5TE20395)	85
114	chr5:18949531..18949660	chr2	10002842	10009401	inverse	6559	VANDAL21(AT2TE42810)	290

epiRIL 480

30	chr3:8764309..8764554	chr3	22059540	22064311	inverse	4771	ATCOPIA78(AT3TE89830)	8
30	chr3:8764389..8764554	chr5	4208107	4213080	sens	4973	ATCOPIA78(AT5TE15240)	36
40	chr3:22615077..22615488	chr3	22695596	22700496	inverse	4900	ATCOPIA78(AT3TE92525)	ND
51	chr5:6361590..6363275	chr5	4209164	4212689	inverse	3525	ATCOPIA78(AT5TE15240)	18

epiRIL 466

1	chr1:3774883..3774985	chr2	10001482	10009630	sens	8148	VANDAL21(AT2TE42810)	ND
2	chr1:9460314..9460488	chr2	4900812	4909276	sens	8464	ATENSPM3(AT2TE20205)	ND
3	chr1:12177699..12182061	chr3	15283668	15286607	inverse	2939	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
4	chr1:12612105..12615963	chr3	15283667	15286611	inverse	2944	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
5	chr1:17447997..17452720	chr2	4900803	4903590	inverse	2787	ATENSPM3(AT2TE20205)	ND
7	chr1:22871543..22875324	chr3	15283670	15286607	sens	2937	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
14	chr2:15040636..15040717	chr2	10001503	10009612	inverse	8109	VANDAL21(AT2TE42810)	ND
15	chr2:19125583..19125713	chr2	10001479	10009651	sens	8172	VANDAL21(AT2TE42810)	ND
16	chr3:8115103..8115195	chr2	10001486	10009627	inverse	8141	VANDAL21(AT2TE42810)	ND
17	chr3:8764184..8764439	chr3	22059542	22064326	inverse	4784	ATCOPIA78(AT3TE89830)	25
17	chr3:8764184..8764855	chr5	4208084	4213082	sens	4998	ATCOPIA78(AT5TE15240)	68
21	chr4:9320171..9320627	chr3	13369258	13373908	sens	4650	ATCOPIA78(AT3TE54550)	1
21	chr4:9320171..9320627	chr3	22059570	22064215	sens	4645	ATCOPIA78(AT3TE89830)	1
21	chr4:9320171..9320627	chr3	22695570	22700483	inverse	4913	ATCOPIA78(AT3TE92525)	0
22	chr4:10739483..10740160	chr3	22059535	22063963	inverse	4428	ATCOPIA78(AT3TE89830)	31
22	chr4:10739483..10740160	chr3	22695564	22700522	sens	4958	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	9
23	chr4:10750276..10752734	chr3	15283671	15286599	inverse	2928	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
24	chr4:12270990..12271153	chr2	4901200	4907239	inverse	6039	ATENSPM3(AT2TE20205)	286
25	chr4:13393608..13395146	chr3	15283671	15286530	sens	2859	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND

epiRIL 467

1	chr1:12177780..12181721	chr3	15283669	15286611	inverse	2942	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
2	chr1:12618490..12622275	chr3	15283666	15286608	sens	2942	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
7	chr3:4713956..4714064	chr5	5629976	5635296	sens	5320	ATCOPIA93(AT5TE20395)	221
14	chr4:9649930..9650166	chr5	5629991	5635275	sens	5284	ATCOPIA93(AT5TE20395)	115
15	chr5:6361527..6363224	chr1	3360542	3361305	inverse	763	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
15	chr5:6362141..6362441	chr1	3780766	3785344	sens	4578	ATCOPIA78(AT1TE12295)	0
15	chr5:6361554..6362885	chr1	7714739	7715144	sens	405	ATCOPIA78(AT1TE24850)	0
15	chr5:6362065..6362441	chr1	7714739	7721408	sens	6669	ATCOPIA78(AT1TE24850)	0
15	chr5:6362141..6362441	chr1	18013163	18017054	sens	3891	ATCOPIA78(AT1TE59755)	0
15	chr5:6362141..6362441	chr1	21524997	21529546	inverse	4549	ATCOPIA78(AT1TE71045)	0
15	chr5:6362141..6362885	chr3	13369245	13372967	sens	3722	ATCOPIA78(AT3TE54550)	0
15	chr5:6362141..6362441	chr3	22059536	22064141	sens	4605	ATCOPIA78(AT3TE89830)	0

15	chr5:6362065..6362441	chr3	22695568	22700062	inverse	4494	ATCOPIA78(AT3TE92525)	0
15	chr5:6362065..6362441	chr5	4208085	4212602	inverse	4517	ATCOPIA78(AT5TE15240)	0

epiRIL 437

1	chr1:13710676..13710839	chr5	5629979	5635286	sens	5307	ATCOPIA93(AT5TE20395)	148
6	chr1:17407431..17408766	chr3	15283669	15286608	inverse	2939	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
14	chr2:448804..449071	chr5	5629985	5635302	sens	5317	ATCOPIA93(AT5TE20395)	144
17	chr2:9535972..9536251	chr5	5629977	5635286	sens	5309	ATCOPIA93(AT5TE20395)	146
18	chr2:16544762..16544879	chr5	5629980	5635301	sens	5321	ATCOPIA93(AT5TE20395)	109
19	chr3:909854..910412	chr5	5629982	5635297	inverse	5315	ATCOPIA93(AT5TE20395)	141
25	chr3:22615137..22615232	chr3	22695575	22700519	inverse	4944	ATCOPIA78(AT3TE92525)	ND
26	chr3:23215849..23216153	chr1	12754967	12760304	sens	5337	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
27	chr3:23216319..23216353	chr5	5629982	5635306	sens	5324	ATCOPIA93(AT5TE20395)	283
28	chr4:3473737..3473991	chr5	5629979	5635269	sens	5290	ATCOPIA93(AT5TE20395)	137
31	chr4:6497941..6498264	chr5	5629978	5635283	sens	5305	ATCOPIA93(AT5TE20395)	286
34	chr4:13252376..13252909	chr1	18013163	18017953	inverse	4790	ATCOPIA78(AT1TE59755)	58
35	chr4:15139416..15139620	chr5	5629978	5635306	inverse	5328	ATCOPIA93(AT5TE20395)	132
36	chr5:7827932..7828113	chr5	5629978	5634970	sens	4992	ATCOPIA93(AT5TE20395)	264

epiRIL 458

5	chr2:10508544..10511015	chr4	4358138	4360286	inverse	2148	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
6	chr2:18286844..18287065	chr5	5629979	5635291	sens	5312	ATCOPIA93(AT5TE20395)	143
9	chr3:17888839..17888923	chr5	5629977	5635304	sens	5327	ATCOPIA93(AT5TE20395)	213
10	chr3:22615319..22615504	chr3	22695568	22700513	inverse	4945	ATCOPIA78(AT3TE92525)	ND
11	chr4:129610..129794	chr5	5629977	5635292	sens	5315	ATCOPIA93(AT5TE20395)	195
12	chr4:9320298..9320621	chr1	18013198	18019244	sens	6046	ATCOPIA30(AT1TE59760)	0
13	chr4:9650042..9650133	chr5	5629979	5635278	sens	5299	ATCOPIA93(AT5TE20395)	202
15	chr4:11302881..11303075	chr2	10001495	10009584	sens	8089	VANDAL21(AT2TE42810)	346
16	chr4:11749997..11750140	chr5	5629978	5635306	inverse	5328	ATCOPIA93(AT5TE20395)	224
17	chr4:12690209..12690424	chr5	5629984	5635307	inverse	5323	ATCOPIA93(AT5TE20395)	94
18	chr5:10192629..10192724	chr2	10001480	10009501	inverse	8021	VANDAL21(AT2TE42810)	0
18	chr5:10192629..10193652	chr3	11546388	11546865	sens	477	VANDAL21(AT3TE48070)	0

epiRIL 371

1	chr1:1817223..1817293	chr2	10001491	10009487	sens	7996	VANDAL21(AT2TE42810)	ND
2	chr1:10035463..10035546	chr2	10001508	10009623	inverse	8115	VANDAL21(AT2TE42810)	523
9	chr1:21919979..21920097	chr2	10001491	10009656	inverse	8165	VANDAL21(AT2TE42810)	ND
10	chr1:25760058..25760196	chr5	5629977	5635288	sens	5311	ATCOPIA93(AT5TE20395)	174
17	chr2:3883263..3883443	chr2	10001555	10009635	sens	8080	VANDAL21(AT2TE42810)	ND
22	chr3:5595012..5595211	chr2	4900829	4909282	sens	8453	ATENSPM3(AT2TE20205)	267
23	chr3:8788662..8788731	chr2	10001483	10009621	inverse	8138	VANDAL21(AT2TE42810)	478
26	chr4:1151517..1151582	chr2	10001480	10009603	inverse	8123	VANDAL21(AT2TE42810)	259
33	chr4:5797218..5797350	chr2	10001540	10009654	sens	8114	VANDAL21(AT2TE42810)	244
35	chr4:9649965..9650115	chr5	5629984	5635282	sens	5298	ATCOPIA93(AT5TE20395)	321
37	chr4:11509807..11509887	chr2	10001491	10009651	sens	8160	VANDAL21(AT2TE42810)	533
41	chr5:26571824..26574128	chr1	3360538	3361313	sens	775	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
41	chr5:26572104..26574128	chr1	3780765	3783283	inverse	2518	ATCOPIA78(AT1TE12295)	0
41	chr5:26571177..26574438	chr1	7714734	7715148	inverse	414	ATCOPIA78(AT1TE24850)	0
41	chr5:26572104..26574128	chr1	7717356	7720110	inverse	2754	ATCOPIA78(AT1TE24850)	0
41	chr5:26572028..26574128	chr1	18013162	18015604	inverse	2442	ATCOPIA78(AT1TE59755)	0
41	chr5:26571391..26574349	chr1	21524993	21525650	sens	657	ATCOPIA78(AT1TE71045)	0
41	chr5:26572104..26574128	chr1	21527337	21529542	sens	2205	ATCOPIA78(AT1TE71045)	0
41	chr5:26571748..26574438	chr3	13369245	13371628	inverse	2383	ATCOPIA78(AT3TE54550)	0
41	chr5:26572028..26574128	chr3	22059535	22061977	inverse	2442	ATCOPIA78(AT3TE89830)	0
41	chr5:26572104..26574128	chr3	22059535	22062053	inverse	2518	ATCOPIA78(AT3TE89830)	0

41	chr5:26572028..26574128	chr3	22698004	22700151	sens	2147	ATCOPIA78(AT3TE92525)	0
41	chr5:26571353..26574842	chr4	9320300	9320554	inverse	254	ATCOPIA78(AT4TE42010)	0

epiRIL 14

1	chr1:6621988..6622697	chr1	18013163	18018342	inverse	5179	ATCOPIA78(AT1TE59755)	96
6	chr1:17642563..17646058	chr4	4358132	4360293	inverse	2161	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
18	chr2:10508830..10513341	chr4	4358157	4360295	inverse	2138	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
19	chr3:8764227..8764745	chr3	22059534	22064329	inverse	4795	ATCOPIA78(AT3TE89830)	36
19	chr3:8764151..8764745	chr5	4208084	4213084	sens	5000	ATCOPIA78(AT5TE15240)	100
24	chr4:2805669..2805826	chr2	10001510	10009654	sens	8144	VANDAL21(AT2TE42810)	465
29	chr4:10739677..10739982	chr3	22059534	22063974	inverse	4440	ATCOPIA78(AT3TE89830)	25
29	chr4:10739601..10739982	chr3	22695564	22700438	sens	4874	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	9
31	chr4:13251484..13253188	chr1	18013162	18016899	inverse	3737	ATCOPIA78(AT1TE59755)	73
32	chr5:6361600..6362606	chr5	4208082	4212620	inverse	4538	ATCOPIA78(AT5TE15240)	42
36	chr5:26570588..26575161	chr1	3360544	3361302	sens	758	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
36	chr5:26573050..26575161	chr1	3780765	3785273	inverse	4508	ATCOPIA78(AT1TE12295)	1
36	chr5:26568063..26578029	chr1	7714753	7715142	inverse	389	ATCOPIA78(AT1TE24850)	0
36	chr5:26571322..26575473	chr1	7717812	7722131	inverse	4319	ATCOPIA78(AT1TE24850)	0
36	chr5:26573050..26575161	chr1	18013162	18017637	inverse	4475	ATCOPIA78(AT1TE59755)	0
36	chr5:26573050..26575161	chr1	21524999	21529471	sens	4472	ATCOPIA78(AT1TE71045)	0
36	chr5:26570739..26575386	chr3	13369257	13373734	inverse	4477	ATCOPIA78(AT3TE54550)	0
36	chr5:26573050..26575161	chr3	22059535	22064141	inverse	4606	ATCOPIA78(AT3TE89830)	0
36	chr5:26573050..26575161	chr3	22695570	22700072	sens	4502	ATCOPIA78(AT3TE92525)	0
36	chr5:26573050..26575386	chr4	9320300	9320552	inverse	252	ATCOPIA78(AT4TE42010)	0
36	chr5:26573050..26575386	chr5	4208087	4212612	sens	4525	ATCOPIA78(AT5TE15240)	0

epiRIL 46

6	chr1:16836944..16837093	chr2	4900802	4909284	inverse	8482	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	474
7	chr1:28366625..28369902	chr3	15283671	15286595	sens	2924	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
19	chr2:14560239..14568159	chr4	5544625	5545513	inverse	888	ATMU11(AT4TE23200) ATMU5(AT4TE23195)	ND
33	chr4:10739288..10740019	chr3	22059541	22064313	inverse	4772	ATCOPIA78(AT3TE89830)	28
33	chr4:10739288..10740359	chr3	22695577	22700078	sens	4501	ATCOPIA78(AT3TE92525)	12
35	chr5:6361861..6362990	chr5	4208082	4212616	inverse	4534	ATCOPIA78(AT5TE15240)	22

epiRIL 36

1	chr1:1718953..1719040	chr2	10001490	10009649	inverse	8159	VANDAL21(AT2TE42810)	426
11	chr1:28366429..28369568	chr3	15283675	15286603	sens	2928	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
26	chr2:3666915..3667127	chr5	5629987	5635304	inverse	5317	ATCOPIA93(AT5TE20395)	95
35	chr2:9010659..9010878	chr5	5629988	5635300	sens	5312	ATCOPIA93(AT5TE20395)	140
36	chr2:10299883..10300015	chr2	10001486	10009601	sens	8115	VANDAL21(AT2TE42810)	398
37	chr2:10625142..10625256	chr2	10001499	10009641	inverse	8142	VANDAL21(AT2TE42810)	442
38	chr2:15200684..15202970	chr4	5541956	5545526	sens	3570	HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	120
40	chr3:7750885..7750970	chr2	10001522	10009646	inverse	8124	VANDAL21(AT2TE42810)	362
64	chr4:10075184..10075412	chr5	5629978	5635299	sens	5321	ATCOPIA93(AT5TE20395)	240
66	chr4:11478281..11478351	chr2	10001490	10009654	inverse	8164	VANDAL21(AT2TE42810)	416

67	chr4:14388760..14391905	chr4	5541890	5545532	inverse	3642	HELITRONY1B(AT4TE23185) ATMU11(AT4TE23200) ATMU5(AT4TE23195) ATMU3(AT4TE23190)	93
68	chr5:6122467..6122670	chr5	5629987	5634960	inverse	4973	ATCOPIA93(AT5TE20395)	119
76	chr5:26570676..26575368	chr1	3360542	3361312	sens	770	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
76	chr5:26568446..26575532	chr1	3785065	3785337	inverse	272	ATCOPIA78(AT1TE12295)	0
76	chr5:26570968..26573815	chr1	3780767	3785261	inverse	4494	ATCOPIA78(AT1TE12295)	1
76	chr5:26570676..26575233	chr1	7714737	7715144	inverse	407	ATCOPIA78(AT1TE24850)	0
76	chr5:26571044..26573815	chr1	7717452	7722124	inverse	4672	ATCOPIA78(AT1TE24850)	0
76	chr5:26570676..26573815	chr1	18013164	18015678	inverse	2514	ATCOPIA78(AT1TE59755)	0
76	chr5:26570676..26575233	chr1	21524997	21525651	sens	654	ATCOPIA78(AT1TE71045)	0
76	chr5:26570968..26573815	chr1	21524997	21529469	sens	4472	ATCOPIA78(AT1TE71045)	0
76	chr5:26571044..26573815	chr3	13369247	13373736	inverse	4489	ATCOPIA78(AT3TE54550)	0
76	chr5:26571044..26573815	chr3	22059537	22064142	inverse	4605	ATCOPIA78(AT3TE89830)	0
76	chr5:26570968..26573815	chr3	22695568	22700075	sens	4507	ATCOPIA78(AT3TE92525)	0
76	chr5:26571044..26573815	chr3	22698006	22700151	sens	2145	ATCOPIA78(AT3TE92525)	0
76	chr5:26570573..26575451	chr4	9320302	9320547	inverse	245	ATCOPIA78(AT4TE42010)	0
76	chr5:26570968..26573815	chr5	4208085	4212615	sens	4530	ATCOPIA78(AT5TE15240)	0

epiRIL 24

14	chr3:8764445..8764544	chr3	22059539	22064326	inverse	4787	ATCOPIA78(AT3TE89830)	33
14	chr3:8764444..8764544	chr5	4208087	4213081	sens	4994	ATCOPIA78(AT5TE15240)	64
20	chr3:22615330..22615513	chr3	22695574	22700517	inverse	4943	ATCOPIA78(AT3TE92525)	ND
27	chr4:10739712..10739945	chr3	22059535	22064170	inverse	4635	ATCOPIA78(AT3TE89830)	23
27	chr4:10739712..10739854	chr3	22695564	22700439	sens	4875	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	5

epiRIL 108

1	chr1:6622263..6622354	chr1	18013164	18018380	inverse	5216	ATCOPIA78(AT1TE59755)	167
2	chr1:11085173..11085183	chr5	5629986	5635302	sens	5316	ATCOPIA93(AT5TE20395)	418
3	chr1:11085320..11085393	chr5	5629980	5635302	sens	5322	ATCOPIA93(AT5TE20395)	418
5	chr1:23332320..23332396	chr5	5629977	5635286	inverse	5309	ATCOPIA93(AT5TE20395)	206
9	chr3:5595095..5595167	chr2	4900811	4909273	sens	8462	ATENSPM3(AT2TE20205)	785
15	chr4:1151507..1151588	chr2	10001482	10009575	inverse	8093	VANDAL21(AT2TE42810)	763
21	chr4:10739880..10739933	chr3	22059536	22063887	inverse	4351	ATCOPIA78(AT3TE89830)	62
21	chr4:10739880..10739933	chr3	22695565	22700100	sens	4535	ATCOPIA78(AT3TE92525)	21
23	chr4:13252048..13252303	chr1	18013162	18016901	inverse	3739	ATCOPIA78(AT1TE59755)	94
24	chr5:6362150..6362478	chr5	4208083	4212616	inverse	4533	ATCOPIA78(AT5TE15240)	43

epiRIL 425

1	chr1:2204577..2204740	chr5	5629978	5635303	sens	5325	ATCOPIA93(AT5TE20395)	256
2	chr1:2554524..2554616	chr5	5629979	5635304	sens	5325	ATCOPIA93(AT5TE20395)	279
3	chr1:2631164..2631251	chr5	5629992	5635306	inverse	5314	ATCOPIA93(AT5TE20395)	268
4	chr1:3280721..3280794	chr5	5629978	5635303	inverse	5325	ATCOPIA93(AT5TE20395)	399
5	chr1:3803879..3803951	chr2	10001480	10009603	sens	8123	VANDAL21(AT2TE42810)	ND
6	chr1:11102525..11102620	chr5	5630000	5635307	inverse	5307	ATCOPIA93(AT5TE20395)	249
7	chr1:12179755..12181629	chr3	15283671	15286607	inverse	2936	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
8	chr1:12618703..12621412	chr3	15283666	15286612	sens	2946	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
9	chr1:14220145..14220235	chr5	5629983	5635301	inverse	5318	ATCOPIA93(AT5TE20395)	414
10	chr1:16679896..16680029	chr5	5630011	5635250	inverse	5239	ATCOPIA93(AT5TE20395)	119
11	chr1:16695190..16695272	chr5	5629984	5635307	sens	5323	ATCOPIA93(AT5TE20395)	327
14	chr1:23560932..23561025	chr5	5629980	5635309	inverse	5329	ATCOPIA93(AT5TE20395)	416
15	chr1:23889302..23889427	chr5	5629978	5635281	sens	5303	ATCOPIA93(AT5TE20395)	ND
16	chr1:25978237..25978406	chr5	5629994	5635268	inverse	5274	ATCOPIA93(AT5TE20395)	ND
17	chr1:29781800..29781907	chr5	5629978	5635228	sens	5250	ATCOPIA93(AT5TE20395)	402

18	chr1:30236411..30236507	chr5	5629980	5635279	sens	5299	ATCOPIA93(AT5TE20395)	182
22	chr2:6325160..6325237	chr2	10001526	10009643	inverse	8117	VANDAL21(AT2TE42810)	ND
23	chr2:10042905..10042979	chr5	5629990	5635304	inverse	5314	ATCOPIA93(AT5TE20395)	367
24	chr2:18269663..18269821	chr5	5629979	5635299	inverse	5320	ATCOPIA93(AT5TE20395)	409
25	chr2:18857685..18857770	chr5	5629984	5635305	sens	5321	ATCOPIA93(AT5TE20395)	175
26	chr3:3730268..3730365	chr5	5629977	5635304	inverse	5327	ATCOPIA93(AT5TE20395)	159
27	chr3:5440413..5440539	chr5	5629977	5635303	inverse	5326	ATCOPIA93(AT5TE20395)	184
28	chr3:5826247..5826322	chr5	5629981	5635300	sens	5319	ATCOPIA93(AT5TE20395)	369
29	chr3:7473159..7473232	chr5	5629980	5635310	inverse	5330	ATCOPIA93(AT5TE20395)	244
30	chr3:8764476..8764557	chr3	22059533	22064284	inverse	4751	ATCOPIA78(AT3TE89830)	53
30	chr3:8764455..8764523	chr5	4208094	4213087	sens	4993	ATCOPIA78(AT5TE15240)	132
31	chr3:11048849..11048955	chr5	5629978	5635291	sens	5313	ATCOPIA93(AT5TE20395)	213
36	chr3:17330899..17331042	chr5	5629985	5635305	sens	5320	ATCOPIA93(AT5TE20395)	221
37	chr3:19455696..19455810	chr5	5629979	5635298	sens	5319	ATCOPIA93(AT5TE20395)	211
38	chr3:21098230..21098347	chr5	5629992	5635259	sens	5267	ATCOPIA93(AT5TE20395)	187
39	chr3:23229881..23229962	chr5	5629981	5635309	sens	5328	ATCOPIA93(AT5TE20395)	236
40	chr4:2524711..2524808	chr5	5629978	5635289	inverse	5311	ATCOPIA93(AT5TE20395)	192
44	chr4:9516404..9516607	chr5	5629976	5635237	sens	5261	ATCOPIA93(AT5TE20395)	194
46	chr4:13408237..13408333	chr5	5629981	5635292	sens	5311	ATCOPIA93(AT5TE20395)	219
47	chr4:13692037..13692196	chr5	5629979	5635257	inverse	5278	ATCOPIA93(AT5TE20395)	255
48	chr4:15633264..15633348	chr5	5630007	5635294	inverse	5287	ATCOPIA93(AT5TE20395)	197
49	chr4:16482435..16482614	chr2	4900801	4909205	sens	8404	ATENSPM3(AT2TE20205)	798
50	chr4:18519777..18519887	chr5	5629986	5635306	sens	5320	ATCOPIA93(AT5TE20395)	249
51	chr5:874947..875029	chr5	5629980	5635286	sens	5306	ATCOPIA93(AT5TE20395)	200
52	chr5:8153522..8153677	chr5	5629977	5634956	inverse	4979	ATCOPIA93(AT5TE20395)	369
53	chr5:8239360..8239449	chr5	5629980	5634964	inverse	4984	ATCOPIA93(AT5TE20395)	166
54	chr5:10478018..10478122	chr5	5629976	5634958	sens	4982	ATCOPIA93(AT5TE20395)	164
55	chr5:10640096..10640204	chr5	5629983	5634946	sens	4963	ATCOPIA93(AT5TE20395)	173
57	chr5:10934625..10934697	chr2	10001486	10009654	sens	8168	VANDAL21(AT2TE42810)	659
58	chr5:11090130..11090852	chr1	12755092	12759424	inverse	4332	ATCOPIA93(AT1TE41580)	ND
59	chr5:11145404..11145503	chr5	5629988	5634881	inverse	4893	ATCOPIA93(AT5TE20395)	149
62	chr5:20508511..20508673	chr5	5629981	5634916	inverse	4935	ATCOPIA93(AT5TE20395)	169
63	chr5:20771654..20771996	chr5	5629985	5634973	inverse	4988	ATCOPIA93(AT5TE20395)	366
64	chr5:20772134..20772381	chr1	12753512	12759742	inverse	6230	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	1
65	chr5:21197562..21197822	chr5	5629977	5634889	inverse	4912	ATCOPIA93(AT5TE20395)	380
66	chr5:21197826..21198008	chr5	5629977	5634965	inverse	4988	ATCOPIA93(AT5TE20395)	380
69	chr5:26571232..26574771	chr1	3360539	3361316	sens	777	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
69	chr5:26568812..26574724	chr1	3785066	3785353	inverse	287	ATCOPIA78(AT1TE12295)	0
69	chr5:26571669..26573792	chr1	3780767	3783281	inverse	2514	ATCOPIA78(AT1TE12295)	1
69	chr5:26571131..26574837	chr1	7717839	7720032	inverse	2193	ATCOPIA78(AT1TE24850)	0
69	chr5:26571745..26574837	chr1	7717358	7720032	inverse	2674	ATCOPIA78(AT1TE24850)	0
69	chr5:26571669..26573792	chr1	18013164	18015602	inverse	2438	ATCOPIA78(AT1TE59755)	0
69	chr5:26571745..26573792	chr1	21527339	21529474	sens	2135	ATCOPIA78(AT1TE71045)	0
69	chr5:26571745..26573792	chr1	21527339	21529547	sens	2208	ATCOPIA78(AT1TE71045)	0
69	chr5:26568812..26574724	chr3	13373452	13373809	inverse	357	ATCOPIA78(AT3TE54550)	0
69	chr5:26571669..26574627	chr3	13369244	13371548	inverse	2304	ATCOPIA78(AT3TE54550)	0
69	chr5:26571669..26573792	chr3	22059537	22062051	inverse	2514	ATCOPIA78(AT3TE89830)	0
69	chr5:26571745..26573792	chr3	22698006	22700141	sens	2135	ATCOPIA78(AT3TE92525)	0
69	chr5:26571290..26574627	chr4	9320406	9320551	inverse	145	ATCOPIA78(AT4TE42010)	0
69	chr5:25458219..25458304	chr1	12753507	12758488	sens	4981	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	0
69	chr5:25458219..25458741	chr5	5629977	5634891	sens	4914	ATCOPIA93(AT5TE20395)	0

epiRIL 70

1	chr1:195259..195368	chr5	5629978	5635292	inverse	5314	ATCOPIA93(AT5TE20395)	130
4	chr1:24748365..24748469	chr5	5629978	5635281	sens	5303	ATCOPIA93(AT5TE20395)	117
5	chr1:29712841..29714602	chr5	18144132	18146872	inverse	2740	ATCOPIA21(AT5TE65370)	ND

18	chr4:13036989..13037159	chr5	5629983	5635271	inverse	5288	ATCOPIA93(AT5TE20395)	96
19	chr5:11835926..11836129	chr2	10001503	10009635	inverse	8132	VANDAL21(AT2TE42810)	ND
20	chr5:11845546..11845669	chr2	10001522	10009399	inverse	7877	VANDAL21(AT2TE42810)	ND
21	chr5:19502451..19502763	chr1	11148090	11152533	inverse	4443	ATCOPIA25(AT1TE36030) ATCOPIA51(AT1TE36035)	45
22	chr5:19502776..19503033	chr1	11148090	11152612	inverse	4522	ATCOPIA25(AT1TE36030) ATCOPIA51(AT1TE36035)	0

epiRIL 276

4	chr1:9180828..9181649	chr5	5629988	5635303	inverse	5315	ATCOPIA93(AT5TE20395)	118
8	chr1:15180684..15191931	chr2	4903430	4904980	inverse	1550	ATENSPM3(AT2TE20205)	1
45	chr2:5860301..5860320	chr5	5629977	5635305	sens	5328	ATCOPIA93(AT5TE20395)	135
46	chr2:5860377..5860476	chr1	12753525	12760289	sens	6764	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
48	chr2:7987319..7987730	chr5	5629979	5635278	inverse	5299	ATCOPIA93(AT5TE20395)	104
49	chr2:8318491..8318574	chr2	10001492	10009552	sens	8060	VANDAL21(AT2TE42810)	261
50	chr2:10506927..10513811	chr4	4358168	4360294	inverse	2126	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
51	chr3:3106727..3106822	chr2	10001487	10009634	sens	8147	VANDAL21(AT2TE42810)	346
53	chr3:11073585..11073713	chr2	10001500	10009647	inverse	8147	VANDAL21(AT2TE42810)	283
79	chr4:3055002..3055416	chr5	13962514	13967458	sens	4944	VANDAL21(AT5TE49755)	ND
96	chr4:10740005..10740627	chr3	22059538	22064322	inverse	4784	ATCOPIA78(AT3TE89830)	14
96	chr4:10740005..10740627	chr3	22695574	22700514	sens	4940	ATCOPIA78(AT3TE92525)	10
115	chr5:12816882..12823634	chr2	10004237	10006534	sens	2297	VANDAL21(AT2TE42810)	ND
119	chr5:15850013..15850084	chr1	12753507	12758442	sens	4935	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	0
119	chr5:15850013..15850084	chr5	5629979	5634973	sens	4994	ATCOPIA93(AT5TE20395)	0

epiRIL 218

10	chr1:28365448..28371759	chr3	15285382	15286586	sens	1204	ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
36	chr2:10507579..10512052	chr4	4358136	4360286	inverse	2150	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
37	chr3:8764367..8764429	chr3	22059535	22064306	inverse	4771	ATCOPIA78(AT3TE89830)	32
37	chr3:8764521..8764625	chr5	4208085	4213072	sens	4987	ATCOPIA78(AT5TE15240)	58
45	chr3:22615410..22615543	chr3	22695617	22700500	inverse	4883	ATCOPIA78(AT3TE92525)	ND
46	chr4:1151497..1151585	chr2	10001488	10009584	inverse	8096	VANDAL21(AT2TE42810)	446
56	chr4:10739770..10739860	chr3	22059540	22064330	inverse	4790	ATCOPIA78(AT3TE89830)	33
56	chr4:10739770..10739860	chr3	22695566	22700479	sens	4913	ATCOPIA78(AT3TE92525)	7
58	chr4:13252074..13252207	chr1	18013165	18016864	inverse	3699	ATCOPIA78(AT1TE59755)	54
59	chr5:6361957..6362465	chr5	4208080	4212617	inverse	4537	ATCOPIA78(AT5TE15240)	25

epiRIL 252

1	chr1:5194117..5194253	chr2	10001499	10009632	sens	8133	VANDAL21(AT2TE42810)	311
11	chr1:15135655..15135836	chr2	3084996	3089194	inverse	4198	ATCOPIA20(AT2TE13385)	ND
18	chr1:15251284..15251298	chr2	3083995	3089373	inverse	5378	ATCOPIA20(AT2TE13385)	10
21	chr1:15337556..15341909	chr2	3084820	3088679	inverse	3859	ATCOPIA20(AT2TE13385)	ND
22	chr1:15349381..15349395	chr2	3083995	3088982	inverse	4987	ATCOPIA20(AT2TE13385)	11
29	chr1:22057038..22057125	chr2	10001514	10009535	sens	8021	VANDAL21(AT2TE42810)	151
53	chr2:3618983..3619853	chr2	4904164	4908841	sens	4677	ATENSPM3(AT2TE20205)	32
70	chr3:8764669..8764729	chr3	22059561	22064327	inverse	4766	ATCOPIA78(AT3TE89830)	12
70	chr3:8764669..8764826	chr5	4208086	4213057	sens	4971	ATCOPIA78(AT5TE15240)	26
130	chr4:10739915..10740392	chr3	22059536	22064163	inverse	4627	ATCOPIA78(AT3TE89830)	18
130	chr4:10739498..10739769	chr3	22695565	22700081	sens	4516	ATCOPIA78(AT3TE92525)	10
132	chr4:13252481..13252629	chr1	18013183	18016875	inverse	3692	ATCOPIA78(AT1TE59755)	35
133	chr5:6362244..6362801	chr5	4208084	4212682	inverse	4598	ATCOPIA78(AT5TE15240)	20
161	chr5:26570703..26575364	chr1	3360556	3361309	sens	753	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0

161	chr5:26571010..26572899	chr1	3780765	3783287	inverse	2522	ATCOPIA78(AT1TE12295)	0
161	chr5:26570817..26574549	chr1	7717356	7720114	inverse	2758	ATCOPIA78(AT1TE24850)	0
161	chr5:26570817..26575364	chr1	7717818	7719298	inverse	1480	ATCOPIA78(AT1TE24850)	0
161	chr5:26571010..26572899	chr1	18013162	18015684	inverse	2522	ATCOPIA78(AT1TE59755)	0
161	chr5:26570934..26572899	chr1	21525117	21529243	sens	4126	ATCOPIA78(AT1TE71045)	0
161	chr5:26570934..26572899	chr3	13369245	13371543	inverse	2298	ATCOPIA78(AT3TE54550)	0
161	chr5:26571010..26572899	chr3	22059535	22062057	inverse	2522	ATCOPIA78(AT3TE89830)	0
161	chr5:26570934..26572899	chr3	22695582	22699910	sens	4328	ATCOPIA78(AT3TE92525)	0
161	chr5:26570601..26572899	chr5	4208099	4208763	sens	664	ATCOPIA78(AT5TE15240)	0
161	chr5:26573024..26574549	chr5	4208190	4212450	sens	4260	ATCOPIA78(AT5TE15240)	0

epiRIL 222

11	chr1:16680331..16680447	chr1	21833196	21838007	inverse	4811	ATRE1(AT1TE72060)	8
12	chr1:16694340..16694888	chr1	21796942	21801730	sens	4788	ATRE1(AT1TE71950)	ND
13	chr1:16837018..16837131	chr2	4900810	4909262	inverse	8452	ATENSPM3(AT2TE20205)	302
15	chr1:20150954..20155688	chr3	15283671	15286609	sens	2938	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
52	chr3:22615112..22615883	chr3	22695579	22700517	inverse	4938	ATCOPIA78(AT3TE92525)	ND

epiRIL 166

42	chr3:8764206..8764420	chr3	22059540	22064326	inverse	4786	ATCOPIA78(AT3TE89830)	28
42	chr3:8764521..8764651	chr5	4208087	4213067	sens	4980	ATCOPIA78(AT5TE15240)	82
72	chr4:13252630..13252804	chr1	18013167	18016819	inverse	3652	ATCOPIA78(AT1TE59755)	63
73	chr5:6362214..6363149	chr5	4209160	4212692	inverse	3532	ATCOPIA78(AT5TE15240)	14
84	chr5:26570717..26575205	chr1	3360540	3361314	sens	774	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
84	chr5:26571205..26573519	chr1	3780763	3783211	inverse	2448	ATCOPIA78(AT1TE12295)	0
84	chr5:26571205..26575580	chr1	7717358	7720114	inverse	2756	ATCOPIA78(AT1TE24850)	0
84	chr5:26571279..26573519	chr1	7717354	7720114	inverse	2760	ATCOPIA78(AT1TE24850)	0
84	chr5:26571279..26573519	chr1	18013162	18015608	inverse	2446	ATCOPIA78(AT1TE59755)	0
84	chr5:26571074..26574907	chr1	21524995	21525635	sens	640	ATCOPIA78(AT1TE71045)	0
84	chr5:26571279..26573519	chr1	21527333	21529535	sens	2202	ATCOPIA78(AT1TE71045)	0
84	chr5:26571279..26574907	chr3	13369245	13371624	inverse	2379	ATCOPIA78(AT3TE54550)	0
84	chr5:26571205..26573519	chr3	22059535	22061981	inverse	2446	ATCOPIA78(AT3TE89830)	0
84	chr5:26571279..26573519	chr3	22698000	22700149	sens	2149	ATCOPIA78(AT3TE92525)	0
84	chr5:26571150..26574907	chr4	9320407	9320553	inverse	146	ATCOPIA78(AT4TE42010)	0

epiRIL 150

46	chr2:5875795..5876101	chr5	5629993	5635288	sens	5295	ATCOPIA93(AT5TE20395)	110
47	chr3:1772503..1772580	chr2	10001518	10009631	inverse	8113	VANDAL21(AT2TE42810)	170
95	chr5:26570943..26574741	chr1	3360544	3361250	sens	706	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
95	chr5:26571515..26574453	chr1	3780784	3783260	inverse	2476	ATCOPIA78(AT1TE12295)	0
95	chr5:26571287..26575459	chr1	7717446	7720087	inverse	2641	ATCOPIA78(AT1TE24850)	0
95	chr5:26571515..26574453	chr1	18013181	18015657	inverse	2476	ATCOPIA78(AT1TE59755)	0
95	chr5:26570921..26575298	chr1	21527360	21529547	sens	2187	ATCOPIA78(AT1TE71045)	0
95	chr5:26571287..26574453	chr3	13369248	13371614	inverse	2366	ATCOPIA78(AT3TE54550)	0
95	chr5:26571515..26574453	chr3	22059554	22062030	inverse	2476	ATCOPIA78(AT3TE89830)	0
95	chr5:26570921..26575298	chr3	22698027	22700150	sens	2123	ATCOPIA78(AT3TE92525)	0
95	chr5:26572493..26574453	chr5	4208087	4208749	sens	662	ATCOPIA78(AT5TE15240)	0

epiRIL 99

1	chr1:6622179..6622460	chr1	18013167	18018308	inverse	5141	ATCOPIA78(AT1TE59755)	136
6	chr1:17407762..17410302	chr3	15283670	15286610	inverse	2940	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
29	chr3:10449120..10454462	chr4	4358133	4360295	inverse	2162	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND

58	chr4:10739954..10740020	chr3	22059541	22063884	inverse	4343	ATCOPIA78(AT3TE89830)	48
58	chr4:10739423..10739849	chr3	22695568	22700499	sens	4931	ATCOPIA78(AT3TE92525)	10
60	chr4:13251838..13252877	chr1	18013165	18016882	inverse	3717	ATCOPIA78(AT1TE59755)	75
63	chr5:26570785..26575057	chr1	3360538	3361315	sens	777	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
63	chr5:26572216..26574504	chr1	3780768	3783287	inverse	2519	ATCOPIA78(AT1TE12295)	0
63	chr5:26570877..26575212	chr1	7717812	7719878	inverse	2066	ATCOPIA78(AT1TE24850)	0
63	chr5:26572216..26574504	chr1	7717359	7720114	inverse	2755	ATCOPIA78(AT1TE24850)	0
63	chr5:26572216..26574504	chr1	7717812	7720038	inverse	2226	ATCOPIA78(AT1TE24850)	0
63	chr5:26572216..26574504	chr1	18013165	18015684	inverse	2519	ATCOPIA78(AT1TE59755)	0
63	chr5:26572140..26574504	chr1	21525117	21529397	sens	4280	ATCOPIA78(AT1TE71045)	0
63	chr5:26571160..26574504	chr3	13369244	13371623	inverse	2379	ATCOPIA78(AT3TE54550)	0
63	chr5:26572140..26574504	chr3	22059538	22061981	inverse	2443	ATCOPIA78(AT3TE89830)	0
63	chr5:26572216..26574504	chr3	22698000	22700154	sens	2154	ATCOPIA78(AT3TE92525)	0
63	chr5:26570999..26574586	chr4	9320303	9320551	inverse	248	ATCOPIA78(AT4TE42010)	0

epiRIL 69

10	chr1:15231324..15231980	chr2	3083905	3089405	inverse	5500	ATCOPIA20(AT2TE13385)	ND
11	chr1:15250843..15251291	chr2	3083908	3089462	inverse	5554	ATCOPIA20(AT2TE13385)	ND
12	chr1:15252746..15258505	chr2	3083908	3087675	inverse	3767	ATCOPIA20(AT2TE13385)	ND
13	chr1:15351119..15357262	chr2	3083908	3087521	inverse	3613	ATCOPIA20(AT2TE13385)	ND

epiRIL 503

3	chr1:697955..697963	chr5	5630015	5635292	sens	5277	ATCOPIA93(AT5TE20395)	130
5	chr1:6622161..6622347	chr1	18013174	18018313	inverse	5139	ATCOPIA78(AT1TE59755)	120
6	chr1:6622701..6623652	chr1	3780765	3783624	inverse	2859	ATCOPIA78(AT1TE12295)	0
6	chr1:6622701..6623652	chr3	22695565	22700515	sens	4950	ATCOPIA78(AT3TE92525)	0
7	chr1:12480166..12480609	chr5	5629997	5634495	sens	4498	ATCOPIA93(AT5TE20395)	137
16	chr1:23254120..23255147	chr3	15283681	15286611	sens	2930	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
17	chr1:23255445..23255788	chr3	15283681	15286611	sens	2930	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
18	chr1:27998340..27998485	chr5	5629978	5635069	inverse	5091	ATCOPIA93(AT5TE20395)	298
18	chr1:27998512..27999175	chr5	5629994	5634186	inverse	4192	ATCOPIA93(AT5TE20395)	263
19	chr1:30343731..30343795	chr5	5629983	5635297	sens	5314	ATCOPIA93(AT5TE20395)	225
36	chr2:3798899..3798964	chr2	10001483	10009621	inverse	8138	VANDAL21(AT2TE42810)	531
41	chr2:14339361..14339438	chr5	5629978	5635304	sens	5326	ATCOPIA93(AT5TE20395)	277
42	chr2:14449114..14449204	chr5	5629977	5635298	sens	5321	ATCOPIA93(AT5TE20395)	329
43	chr3:2303622..2303774	chr5	5629996	5635291	inverse	5295	ATCOPIA93(AT5TE20395)	ND
44	chr3:2303866..2306003	chr5	5629996	5633002	inverse	3006	ATCOPIA93(AT5TE20395)	101
45	chr3:7632055..7632138	chr2	4900800	4909275	sens	8475	ATENSPM3(AT2TE20205)	342
46	chr3:11173260..11174753	chr1	12753541	12755437	inverse	1896	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
47	chr3:11175109..11175279	chr5	5629978	5635203	inverse	5225	ATCOPIA93(AT5TE20395)	147
58	chr4:2328881..2328978	chr2	4900804	4909175	sens	8371	ATENSPM3(AT2TE20205)	ND
70	chr4:9649457..9649812	chr1	12755065	12758677	sens	3612	ATCOPIA93(AT1TE41580)	0
71	chr4:9649962..9649985	chr5	5629983	5634798	sens	4815	ATCOPIA93(AT5TE20395)	310
71	chr4:9650038..9650114	chr5	5629981	5635282	sens	5301	ATCOPIA93(AT5TE20395)	310
73	chr4:13577984..13578065	chr2	10001479	10009625	sens	8146	VANDAL21(AT2TE42810)	575
74	chr4:14805684..14805908	chr5	5629981	5634203	sens	4222	ATCOPIA93(AT5TE20395)	138
75	chr4:14806166..14806793	chr5	5630216	5635139	sens	4923	ATCOPIA93(AT5TE20395)	166
79	chr5:18926834..18926958	chr1	12754711	12759079	sens	4368	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
80	chr5:18927162..18927279	chr5	5629986	5634802	sens	4816	ATCOPIA93(AT5TE20395)	202
81	chr5:18927625..18929396	chr1	12753537	12755468	sens	1931	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
82	chr5:20157838..20157927	chr2	4900819	4909281	inverse	8462	ATENSPM3(AT2TE20205)	586
83	chr5:21197409..21197694	chr5	5629982	5634895	sens	4913	ATCOPIA93(AT5TE20395)	107

84	chr5:21197807..21197819	chr1	12753523	12758407	sens	4884	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
85	chr5:21197901..21198018	chr5	5630388	5634589	sens	4201	ATCOPIA93(AT5TE20395)	ND
86	chr5:21198387..21200600	chr1	12753523	12755453	sens	1930	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0

epiRIL 92

1	chr1:2771295..2771456	chr2	10001479	10009459	sens	7980	VANDAL21(AT2TE42810)	146
2	chr1:12179521..12181235	chr3	15283673	15286606	inverse	2933	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
3	chr1:12618315..12621314	chr3	15283669	15286591	sens	2922	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
4	chr1:14253311..14253399	chr5	5629986	5635294	sens	5308	ATCOPIA93(AT5TE20395)	131
6	chr1:15646282..15646508	chr5	5630043	5635106	inverse	5063	ATCOPIA93(AT5TE20395)	54
9	chr1:18071267..18071455	chr2	10001481	10009655	sens	8174	VANDAL21(AT2TE42810)	288
28	chr3:9054093..9054191	chr5	5629998	5635276	inverse	5278	ATCOPIA93(AT5TE20395)	80
37	chr3:20370100..20370177	chr2	10001483	10009620	inverse	8137	VANDAL21(AT2TE42810)	299
38	chr3:21071720..21071822	chr2	10001522	10009613	sens	8091	VANDAL21(AT2TE42810)	294
39	chr4:1151516..1151668	chr2	10001501	10009584	inverse	8083	VANDAL21(AT2TE42810)	296
40	chr4:3559681..3559779	chr2	10001483	10009655	sens	8172	VANDAL21(AT2TE42810)	255
50	chr4:9650004..9650164	chr5	5629978	5635258	sens	5280	ATCOPIA93(AT5TE20395)	76
51	chr4:9650214..9650322	chr1	12753517	12758543	sens	5026	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
52	chr4:10739766..10739868	chr3	22059535	22063926	inverse	4391	ATCOPIA78(AT3TE89830)	9
54	chr4:12909808..12910004	chr5	5629991	5635220	inverse	5229	ATCOPIA93(AT5TE20395)	72
55	chr4:14153731..14153834	chr5	5629983	5635135	inverse	5152	ATCOPIA93(AT5TE20395)	91
56	chr5:8316788..8316888	chr2	5852423	5857323	inverse	4900	ATCOPIA12(AT2TE23850) ATCOPIA13(AT2TE23855)	ND
62	chr5:26571124..26574965	chr1	3360543	3361313	sens	770	ATCOPIA78(AT1TE10930) ATCOPIA78(AT1TE10935)	0
62	chr5:26571520..26574601	chr1	3780764	3783281	inverse	2517	ATCOPIA78(AT1TE12295)	0
62	chr5:26571462..26574635	chr1	7714737	7714982	inverse	245	ATCOPIA78(AT1TE24850)	0
62	chr5:26571573..26574601	chr1	7717355	7720113	inverse	2758	ATCOPIA78(AT1TE24850)	0
62	chr5:26571520..26574601	chr1	18013172	18015678	inverse	2506	ATCOPIA78(AT1TE59755)	0
62	chr5:26571497..26574601	chr1	21525123	21529363	sens	4240	ATCOPIA78(AT1TE71045)	0
62	chr5:26571444..26574635	chr3	13369251	13371545	inverse	2294	ATCOPIA78(AT3TE54550)	0
62	chr5:26571497..26574635	chr3	22059534	22062051	inverse	2517	ATCOPIA78(AT3TE89830)	0
62	chr5:26571573..26574601	chr3	22698001	22700131	sens	2130	ATCOPIA78(AT3TE92525)	0

epiRIL 421

2	chr1:13876141..13876312	chr2	10001483	10009652	inverse	8169	VANDAL21(AT2TE42810)	217
5	chr1:15916738..15916827	chr5	5629997	5635253	inverse	5256	ATCOPIA93(AT5TE20395)	36
9	chr1:28367825..28370153	chr3	15283675	15286608	sens	2933	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
10	chr1:28961091..28961458	chr5	5629990	5635306	sens	5316	ATCOPIA93(AT5TE20395)	59
20	chr2:3359985..3360177	chr5	5629977	5635294	sens	5317	ATCOPIA93(AT5TE20395)	61
20	chr2:3360254..3360624	chr5	5629977	5635306	sens	5329	ATCOPIA93(AT5TE20395)	61
35	chr2:5598233..5598410	chr2	10001492	10009656	sens	8164	VANDAL21(AT2TE42810)	221
36	chr2:10760215..10760303	chr2	10001478	10009643	inverse	8165	VANDAL21(AT2TE42810)	224
37	chr2:17748581..17748734	chr5	5629991	5635287	inverse	5296	ATCOPIA93(AT5TE20395)	58
38	chr2:18188774..18189346	chr5	5629979	5635282	inverse	5303	ATCOPIA93(AT5TE20395)	61
39	chr2:18189466..18189563	chr1	12754415	12758535	inverse	4120	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
40	chr3:442474..444097	chr1	12753527	12755468	inverse	1941	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
41	chr3:444491..444645	chr5	5629976	5635296	inverse	5320	ATCOPIA93(AT5TE20395)	137
53	chr3:22615338..22615797	chr3	22695569	22700508	inverse	4939	ATCOPIA78(AT3TE92525)	ND
54	chr4:1268178..1268529	chr5	5629985	5635299	sens	5314	ATCOPIA93(AT5TE20395)	53

55	chr4:1268649..1268768	chr1	12754967	12758404	sens	3437	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
56	chr4:1400269..1400607	chr5	5629978	5635246	sens	5268	ATCOPIA93(AT5TE20395)	73
74	chr4:8209068..8209156	chr5	5629977	5635285	sens	5308	ATCOPIA93(AT5TE20395)	81
76	chr4:10739747..10740123	chr3	22059541	22064328	inverse	4787	ATCOPIA78(AT3TE89830)	20
76	chr4:10739747..10740123	chr3	22695568	22700091	sens	4523	ATCOPIA78(AT3TE92525)	9
78	chr4:11227421..11227611	chr5	5629994	5635294	inverse	5300	ATCOPIA93(AT5TE20395)	63
79	chr4:11500072..11500101	chr5	5629982	5635295	sens	5313	ATCOPIA93(AT5TE20395)	64
80	chr4:11500148..11502715	chr1	12753605	12755454	sens	1849	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
81	chr4:13252152..13252508	chr1	18013165	18018167	inverse	5002	ATCOPIA78(AT1TE59755)	24
82	chr4:16919036..16919188	chr5	5630007	5635280	sens	5273	ATCOPIA93(AT5TE20395)	50
83	chr5:1603424..1603589	chr5	5629980	5635307	sens	5327	ATCOPIA93(AT5TE20395)	128
84	chr5:2514803..2514939	chr5	5629986	5635300	sens	5314	ATCOPIA93(AT5TE20395)	127
85	chr5:3558546..3558874	chr5	5629976	5635294	inverse	5318	ATCOPIA93(AT5TE20395)	126
86	chr5:4172558..4172832	chr5	5630017	5635286	inverse	5269	ATCOPIA93(AT5TE20395)	146
102	chr5:20215599..20215924	chr5	5629990	5634795	sens	4805	ATCOPIA93(AT5TE20395)	39
104	chr5:21740634..21740932	chr5	5629990	5634890	inverse	4900	ATCOPIA93(AT5TE20395)	83
104	chr5:22902812..22902973	chr5	5629983	5634972	sens	4989	ATCOPIA93(AT5TE20395)	83

epiRIL 193

1	chr1:665520..665630	chr5	5629994	5635298	sens	5304	ATCOPIA93(AT5TE20395)	115
2	chr1:4716146..4716263	chr5	5629982	5635275	sens	5293	ATCOPIA93(AT5TE20395)	118
3	chr1:12698467..12698690	chr2	10001502	10009640	inverse	8138	VANDAL21(AT2TE42810)	ND
16	chr2:3239418..3241442	chr1	12753516	12755441	inverse	1925	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
17	chr2:3241769..3241909	chr5	5629977	5635282	inverse	5305	ATCOPIA93(AT5TE20395)	105
18	chr2:3253951..3253955	chr5	5629977	5635297	inverse	5320	ATCOPIA93(AT5TE20395)	54
19	chr2:3288776..3288850	chr5	5629993	5635264	sens	5271	ATCOPIA93(AT5TE20395)	59
24	chr2:4728499..4728539	chr5	5629976	5635312	sens	5336	ATCOPIA93(AT5TE20395)	214
25	chr2:5299521..5299637	chr5	5629978	5635307	inverse	5329	ATCOPIA93(AT5TE20395)	177
26	chr2:13578211..13578374	chr5	5629980	5635284	sens	5304	ATCOPIA93(AT5TE20395)	185
27	chr2:14419822..14420030	chr5	5629982	5635300	sens	5318	ATCOPIA93(AT5TE20395)	204
28	chr2:16859073..16859150	chr5	5629985	5635305	sens	5320	ATCOPIA93(AT5TE20395)	193
29	chr2:17875275..17875346	chr5	5629982	5635293	inverse	5311	ATCOPIA93(AT5TE20395)	123
30	chr2:19662586..19662665	chr2	10001532	10009567	inverse	8035	VANDAL21(AT2TE42810)	388
31	chr3:2444777..2444935	chr2	10001528	10009613	sens	8085	VANDAL21(AT2TE42810)	377
32	chr3:10233828..10233970	chr5	5630155	5635264	inverse	5109	ATCOPIA93(AT5TE20395)	100
34	chr3:13282197..13282275	chr2	10001484	10009645	inverse	8161	VANDAL21(AT2TE42810)	368
37	chr3:19665311..19665401	chr5	5629978	5635292	sens	5314	ATCOPIA93(AT5TE20395)	118
38	chr3:19665434..19666029	chr1	12753750	12758292	sens	4542	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
39	chr3:22520389..22520530	chr5	5629983	5635295	inverse	5312	ATCOPIA93(AT5TE20395)	71
44	chr4:4434605..4434761	chr5	5630004	5635249	sens	5245	ATCOPIA93(AT5TE20395)	66
46	chr4:6023826..6023904	chr2	10001513	10006770	sens	5257	VANDAL21(AT2TE42810)	95
47	chr4:6395707..6395888	chr5	5629981	5635266	sens	5285	ATCOPIA93(AT5TE20395)	61
49	chr4:9650034..9650133	chr5	5629981	5635261	sens	5280	ATCOPIA93(AT5TE20395)	210
51	chr4:11113631..11113650	chr5	5630010	5635301	sens	5291	ATCOPIA93(AT5TE20395)	64
52	chr4:11113707..11116124	chr1	12753516	12755442	sens	1926	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
53	chr4:14738643..14738724	chr5	5629980	5635296	sens	5316	ATCOPIA93(AT5TE20395)	225
54	chr5:152775..153733	chr5	5629984	5635305	sens	5321	ATCOPIA93(AT5TE20395)	203
57	chr5:14416367..14416412	chr5	5630056	5635223	sens	5167	ATCOPIA93(AT5TE20395)	68
58	chr5:14475616..14475738	chr5	5630025	5635036	sens	5011	ATCOPIA93(AT5TE20395)	72
60	chr5:22803125..22803243	chr1	12753508	12759002	sens	5494	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	0
61	chr5:23867347..23867391	chr1	12753507	12758422	inverse	4915	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	ND
62	chr5:23867423..23867545	chr5	5630000	5634899	inverse	4899	ATCOPIA93(AT5TE20395)	202

epiRIL 340

6	chr1:15082840..15083684	chr2	4904187	4909162	sens	4975	ATENSPM3(AT2TE20205)	56
7	chr1:15117435..15117862	chr2	3084106	3089200	inverse	5094	ATCOPIA20(AT2TE13385)	ND
11	chr1:15253786..15258219	chr2	3084129	3088124	inverse	3995	ATCOPIA20(AT2TE13385)	ND
12	chr1:15328167..15329104	chr2	3083909	3089204	inverse	5295	ATCOPIA20(AT2TE13385)	ND
72	chr2:10508917..10512622	chr4	4358130	4360283	inverse	2153	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	102
103	chr4:267868..268084	chr2	10001548	10009571	sens	8023	VANDAL21(AT2TE42810)	190
132	chr4:12642779..12642874	chr2	10003457	10008199	inverse	4742	VANDAL21(AT2TE42810)	227

epiRIL 356

1	chr1:2014889..2015359	chr5	5629992	5635300	inverse	5308	ATCOPIA93(AT5TE20395)	68
2	chr1:4995803..4995995	chr5	5629979	5635233	inverse	5254	ATCOPIA93(AT5TE20395)	65
3	chr1:8347908..8348097	chr5	5629989	5635296	sens	5307	ATCOPIA93(AT5TE20395)	139
4	chr1:8735007..8735082	chr5	5630009	5635307	sens	5298	ATCOPIA93(AT5TE20395)	57
7	chr1:10067576..10068012	chr5	5629977	5635303	inverse	5326	ATCOPIA93(AT5TE20395)	165
8	chr1:10068161..10068641	chr5	5629977	5635303	inverse	5326	ATCOPIA93(AT5TE20395)	165
9	chr1:12682098..12682429	chr1	12753689	12760347	inverse	6658	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
13	chr1:14371264..14371593	chr5	5629977	5635300	sens	5323	ATCOPIA93(AT5TE20395)	159
14	chr1:14416833..14417246	chr5	5629977	5635300	sens	5323	ATCOPIA93(AT5TE20395)	164
15	chr1:14478243..14478759	chr5	5629997	5635300	sens	5303	ATCOPIA93(AT5TE20395)	ND
19	chr1:15082338..15082557	chr5	5630063	5634681	inverse	4618	ATCOPIA93(AT5TE20395)	ND
20	chr1:15083158..15083675	chr5	5630063	5635200	inverse	5137	ATCOPIA93(AT5TE20395)	33
23	chr1:15084734..15085271	chr5	5633046	5635276	inverse	2230	ATCOPIA93(AT5TE20395)	13
28	chr1:16836970..16836991	chr2	4385865	4394211	sens	8346	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	ND
29	chr1:16836992..16837109	chr2	4900864	4909273	inverse	8409	ATENSPM3(AT2TE20205)	ND
31	chr1:25355546..25355884	chr5	5629993	5635297	inverse	5304	ATCOPIA93(AT5TE20395)	105
32	chr1:26611068..26611288	chr5	5630004	5635286	inverse	5282	ATCOPIA93(AT5TE20395)	83
33	chr1:28094323..28094700	chr5	5629998	5635308	sens	5310	ATCOPIA93(AT5TE20395)	94
42	chr2:2121215..2121546	chr5	5629987	5635298	sens	5311	ATCOPIA93(AT5TE20395)	65
51	chr2:3606693..3607683	chr5	5629981	5635200	sens	5219	ATCOPIA93(AT5TE20395)	4
53	chr2:3617436..3618094	chr5	5629999	5632593	inverse	2594	ATCOPIA93(AT5TE20395)	28
53	chr2:3619179..3619916	chr5	5629999	5635293	inverse	5294	ATCOPIA93(AT5TE20395)	46
53	chr2:3619977..3620652	chr5	5630038	5635217	inverse	5179	ATCOPIA93(AT5TE20395)	46
53	chr2:3623845..3624465	chr5	5630024	5635290	inverse	5266	ATCOPIA93(AT5TE20395)	46
55	chr2:3627222..3627231	chr5	5630111	5635034	sens	4923	ATCOPIA93(AT5TE20395)	ND
67	chr2:7436026..7436320	chr5	5629997	5635249	sens	5252	ATCOPIA93(AT5TE20395)	53
68	chr2:12758364..12763299	chr1	12942912	12944479	inverse	1567	ATENSPM3(AT1TE42210)	ND
69	chr2:19076440..19076568	chr5	5629986	5635281	sens	5295	ATCOPIA93(AT5TE20395)	84
70	chr2:19577135..19577149	chr5	5629981	5635299	inverse	5318	ATCOPIA93(AT5TE20395)	106
71	chr3:1106199..1108706	chr1	12753512	12755590	inverse	2078	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
72	chr3:1109382..1109669	chr5	5629977	5635275	inverse	5298	ATCOPIA93(AT5TE20395)	99
74	chr3:9934856..9935176	chr5	5629988	5635302	inverse	5314	ATCOPIA93(AT5TE20395)	83
82	chr3:13589641..13589940	chr5	5630016	5635303	inverse	5287	ATCOPIA93(AT5TE20395)	137
83	chr3:13591269..13591414	chr5	5630184	5635249	inverse	5065	ATCOPIA93(AT5TE20395)	137
83	chr3:13591660..13592053	chr5	5629981	5635200	inverse	5219	ATCOPIA93(AT5TE20395)	137
89	chr3:16988111..16988533	chr5	5629993	5635270	sens	5277	ATCOPIA93(AT5TE20395)	85
90	chr3:18367051..18367562	chr5	5629978	5635299	sens	5321	ATCOPIA93(AT5TE20395)	78
91	chr3:20604822..20605207	chr5	5629975	5635284	sens	5309	ATCOPIA93(AT5TE20395)	83
92	chr3:20893906..20894097	chr5	5629977	5635298	inverse	5321	ATCOPIA93(AT5TE20395)	96
93	chr3:21605776..21606184	chr5	5629992	5635251	sens	5259	ATCOPIA93(AT5TE20395)	104
94	chr3:21606217..21609519	chr1	12753511	12755464	sens	1953	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
102	chr4:3953540..3954368	chr5	5630052	5635200	sens	5148	ATCOPIA93(AT5TE20395)	12
108	chr4:7153178..7153352	chr5	5629976	5635268	inverse	5292	ATCOPIA93(AT5TE20395)	109
109	chr4:7153395..7153941	chr1	12753835	12759334	inverse	5499	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
111	chr4:10801492..10801581	chr5	5629986	5635290	sens	5304	ATCOPIA93(AT5TE20395)	75

113	chr4:11323784..11324307	chr5	5629981	5635307	inverse	5326	ATCOPIA93(AT5TE20395)	114
114	chr4:13736669..13736886	chr5	5629983	5635293	sens	5310	ATCOPIA93(AT5TE20395)	116
115	chr4:13737002..13739877	chr1	12753519	12755405	sens	1886	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
116	chr4:15759989..15760193	chr5	5629995	5635284	inverse	5289	ATCOPIA93(AT5TE20395)	82
117	chr4:16945101..16945375	chr5	5630004	5635271	sens	5267	ATCOPIA93(AT5TE20395)	107
118	chr5:1759478..1760406	chr5	5629983	5635302	sens	5319	ATCOPIA93(AT5TE20395)	87
120	chr5:11114646..11115105	chr5	5630002	5634960	inverse	4958	ATCOPIA93(AT5TE20395)	117
125	chr5:11705347..11707112	chr5	5630070	5633046	sens	2976	ATCOPIA93(AT5TE20395)	7
128	chr5:11728517..11729215	chr5	5630052	5635200	inverse	5148	ATCOPIA93(AT5TE20395)	21
129	chr5:11732018..11732025	chr1	12754392	12759687	inverse	5295	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
131	chr5:11732383..11733078	chr5	5629981	5635293	inverse	5312	ATCOPIA93(AT5TE20395)	21
138	chr5:18705863..18706035	chr1	12753516	12758392	sens	4876	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
138	chr5:18705863..18706035	chr5	5629980	5634970	sens	4990	ATCOPIA93(AT5TE20395)	0

epiRIL 363

1	chr1:4605590..4605777	chr5	5629979	5635278	inverse	5299	ATCOPIA93(AT5TE20395)	98
2	chr1:7226667..7226684	chr5	5629993	5635290	inverse	5297	ATCOPIA93(AT5TE20395)	125
3	chr1:7916453..7916613	chr5	5629981	5635297	inverse	5316	ATCOPIA93(AT5TE20395)	131
17	chr1:16837016..16837017	chr2	4900832	4909278	inverse	8446	ATENSPM3(AT2TE20205)	374
19	chr1:20150892..20155331	chr3	15283696	15286612	sens	2916	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
20	chr1:21431629..21431643	chr1	12753526	12758412	inverse	4886	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
21	chr1:21431732..21431893	chr5	5629985	5635298	inverse	5313	ATCOPIA93(AT5TE20395)	111
22	chr1:23025585..23025680	chr5	5629985	5635303	inverse	5318	ATCOPIA93(AT5TE20395)	113
23	chr1:26408789..26408796	chr5	5629983	5635306	sens	5323	ATCOPIA93(AT5TE20395)	194
24	chr1:26737177..26737295	chr5	5629989	5635303	inverse	5314	ATCOPIA93(AT5TE20395)	90
25	chr1:28366431..28370204	chr3	15283671	15286611	sens	2940	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
26	chr1:28692030..28692034	chr5	5629978	5635298	inverse	5320	ATCOPIA93(AT5TE20395)	197
26	chr1:28692079..28692189	chr5	5629978	5635297	inverse	5319	ATCOPIA93(AT5TE20395)	197
27	chr1:28692308..28692540	chr1	12755060	12759800	inverse	4740	ATCOPIA93(AT1TE41580)	ND
53	chr2:12453788..12454059	chr5	5629983	5635292	inverse	5309	ATCOPIA93(AT5TE20395)	79
55	chr3:10517923..10518228	chr5	5629977	5635286	sens	5309	ATCOPIA93(AT5TE20395)	112
56	chr3:10518299..10518585	chr1	12753524	12758594	sens	5070	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
92	chr4:16144598..16144728	chr5	5629983	5635303	inverse	5320	ATCOPIA93(AT5TE20395)	188
101	chr5:21833522..21833660	chr1	12753512	12758389	sens	4877	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
101	chr5:21833291..21833660	chr5	5630001	5634970	sens	4969	ATCOPIA93(AT5TE20395)	0

epiRIL 366

1	chr1:13528978..13529279	chr5	5630010	5635256	inverse	5246	ATCOPIA93(AT5TE20395)	55
2	chr1:13529377..13529399	chr5	5630008	5635152	inverse	5144	ATCOPIA93(AT5TE20395)	57
6	chr1:15084666..15085274	chr2	4903667	4908097	sens	4430	ATENSPM3(AT2TE20205)	ND
16	chr1:20148478..20148570	chr2	4900843	4909058	sens	8215	ATENSPM3(AT2TE20205)	99
53	chr2:11332320..11332620	chr5	5630027	5635281	inverse	5254	ATCOPIA93(AT5TE20395)	50
56	chr3:7062386..7062570	chr5	5629983	5635301	sens	5318	ATCOPIA93(AT5TE20395)	136
57	chr3:8153734..8154178	chr5	5629989	5635302	inverse	5313	ATCOPIA93(AT5TE20395)	57
58	chr3:8154693..8154905	chr1	12755063	12760218	inverse	5155	ATCOPIA93(AT1TE41580)	ND
59	chr3:9291305..9291835	chr5	5629992	5635250	sens	5258	ATCOPIA93(AT5TE20395)	56
60	chr3:10178810..10179244	chr5	5630158	5635260	inverse	5102	ATCOPIA93(AT5TE20395)	61
78	chr3:16050321..16050385	chr5	5629997	5635291	sens	5294	ATCOPIA93(AT5TE20395)	62
79	chr3:17132409..17132855	chr5	5630016	5635291	inverse	5275	ATCOPIA93(AT5TE20395)	51
80	chr3:17132862..17136151	chr5	5630041	5632410	inverse	2369	ATCOPIA93(AT5TE20395)	21
81	chr3:17225672..17225889	chr5	5629986	5635284	sens	5298	ATCOPIA93(AT5TE20395)	82
82	chr3:19072813..19073029	chr5	5629982	5635293	sens	5311	ATCOPIA93(AT5TE20395)	50

91	chr4:3954049..3954371	chr2	4903917	4908338	inverse	4421	ATENSPM3(AT2TE20205)	19
101	chr4:5132228..5132691	chr5	5629990	5635304	sens	5314	ATCOPIA93(AT5TE20395)	49
103	chr4:16027518..16027934	chr5	5630084	5635154	sens	5070	ATCOPIA93(AT5TE20395)	40
104	chr5:2698477..2698807	chr5	5629980	5635272	sens	5292	ATCOPIA93(AT5TE20395)	56
105	chr5:9415367..9415432	chr5	5630021	5634854	sens	4833	ATCOPIA93(AT5TE20395)	44
110	chr5:11545238..11545316	chr2	4903027	4909097	inverse	6070	ATENSPM3(AT2TE20205)	202
112	chr5:11728519..11729402	chr2	4903917	4908446	sens	4529	ATENSPM3(AT2TE20205)	ND
115	chr5:11732381..11733090	chr2	4903917	4908338	sens	4421	ATENSPM3(AT2TE20205)	13
121	chr5:14207486..14207786	chr1	12753511	12759819	sens	6308	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	0
121	chr5:14207486..14207786	chr5	5629994	5634950	sens	4956	ATCOPIA93(AT5TE20395)	0

epiRIL 362

1	chr1:2404809..2404938	chr5	5630007	5635247	inverse	5240	ATCOPIA93(AT5TE20395)	117
2	chr1:2443482..2443699	chr5	5630015	5635290	inverse	5275	ATCOPIA93(AT5TE20395)	48
3	chr1:3338040..3338078	chr2	10001488	10009646	sens	8158	VANDAL21(AT2TE42810)	211
3	chr1:3338116..3338216	chr2	10001488	10009522	sens	8034	VANDAL21(AT2TE42810)	210
6	chr1:5236431..5236618	chr5	5629979	5635291	sens	5312	ATCOPIA93(AT5TE20395)	39
7	chr1:7290905..7290908	chr1	12753516	12760285	sens	6769	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
8	chr1:7290981..7291074	chr5	5629979	5635297	sens	5318	ATCOPIA93(AT5TE20395)	117
9	chr1:7449652..7449968	chr5	5630011	5634915	sens	4904	ATCOPIA93(AT5TE20395)	34
10	chr1:9106548..9106611	chr5	5630013	5634835	sens	4822	ATCOPIA93(AT5TE20395)	ND
11	chr1:9106614..9106996	chr5	5630358	5635292	sens	4934	ATCOPIA93(AT5TE20395)	79
12	chr1:9180731..9181403	chr1	12756349	12760388	inverse	4039	ATCOPIA93(AT1TE41580)	ND
13	chr1:9181662..9181805	chr5	5629988	5635303	inverse	5315	ATCOPIA93(AT5TE20395)	107
14	chr1:10089470..10089687	chr5	5630007	5635031	sens	5024	ATCOPIA93(AT5TE20395)	ND
15	chr1:10743880..10744139	chr5	5629990	5635270	sens	5280	ATCOPIA93(AT5TE20395)	63
16	chr1:11950446..11952396	chr1	12753535	12755247	inverse	1712	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
17	chr1:11952617..11952771	chr5	5630000	5635199	inverse	5199	ATCOPIA93(AT5TE20395)	65
18	chr1:11952932..11953087	chr5	5630000	5635199	inverse	5199	ATCOPIA93(AT5TE20395)	65
23	chr1:16992379..16992404	chr5	5629984	5635258	inverse	5274	ATCOPIA93(AT5TE20395)	84
24	chr1:16992472..16993426	chr5	5630027	5635258	inverse	5231	ATCOPIA93(AT5TE20395)	84
29	chr1:21782288..21782470	chr5	5629987	5635246	inverse	5259	ATCOPIA93(AT5TE20395)	46
30	chr1:21818544..21818728	chr5	5629987	5635246	inverse	5259	ATCOPIA93(AT5TE20395)	45
31	chr1:22149063..22149063	chr5	18490618	18495235	inverse	4617	ATCOPIA31(AT5TE66565)	ND
32	chr1:22149342..22149527	chr1	11789507	11793170	sens	3663	ATCOPIA31(AT1TE38210)	13
33	chr1:24876291..24876727	chr5	5630028	5635231	sens	5203	ATCOPIA93(AT5TE20395)	107
34	chr1:24876805..24880501	chr5	5634080	5635287	sens	1207	ATCOPIA93(AT5TE20395)	ND
35	chr1:27293113..27293332	chr5	5629981	5635187	sens	5206	ATCOPIA93(AT5TE20395)	ND
36	chr1:27293471..27293616	chr5	5630390	5635224	sens	4834	ATCOPIA93(AT5TE20395)	39
37	chr1:28080003..28080007	chr5	5630642	5635182	inverse	4540	ATCOPIA93(AT5TE20395)	ND
54	chr2:13020608..13020904	chr5	5630039	5635307	inverse	5268	ATCOPIA93(AT5TE20395)	65
55	chr2:13021221..13021240	chr1	12753544	12760314	inverse	6770	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
56	chr2:13021297..13021445	chr1	12755174	12759859	inverse	4685	ATCOPIA93(AT1TE41580)	ND
57	chr2:13862956..13865523	chr5	5632406	5633596	sens	1190	ATCOPIA93(AT5TE20395)	7
58	chr2:18782913..18782959	chr5	5629985	5634716	sens	4731	ATCOPIA93(AT5TE20395)	111
59	chr2:18783112..18783112	chr5	5629985	5635218	sens	5233	ATCOPIA93(AT5TE20395)	ND
60	chr2:18783188..18783822	chr5	5630879	5635218	sens	4339	ATCOPIA93(AT5TE20395)	89
61	chr2:19037777..19041143	chr5	5633656	5634964	sens	1308	ATCOPIA93(AT5TE20395)	ND
62	chr3:330626..330903	chr5	5630118	5635275	inverse	5157	ATCOPIA93(AT5TE20395)	ND
63	chr3:330949..330976	chr1	12753522	12760282	inverse	6760	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
64	chr3:331025..331896	chr1	12755079	12758662	inverse	3583	ATCOPIA93(AT1TE41580)	ND
65	chr3:8764089..8764334	chr3	22059568	22064246	inverse	4678	ATCOPIA78(AT3TE89830)	21
65	chr3:8764436..8764734	chr3	22059539	22064318	inverse	4779	ATCOPIA78(AT3TE89830)	21
65	chr3:8764436..8764621	chr5	4208083	4213081	sens	4998	ATCOPIA78(AT5TE15240)	34
70	chr3:19868997..19869180	chr5	5630005	5635298	sens	5293	ATCOPIA93(AT5TE20395)	98
71	chr3:20352137..20352155	chr5	5629981	5634924	sens	4943	ATCOPIA93(AT5TE20395)	ND
72	chr3:20352213..20352367	chr5	5629981	5635248	sens	5267	ATCOPIA93(AT5TE20395)	78
73	chr3:20352623..20353237	chr5	5631200	5635287	sens	4087	ATCOPIA93(AT5TE20395)	64

74	chr3:22056110..22056219	chr5	5630107	5635276	inverse	5169	ATCOPIA93(AT5TE20395)	48
86	chr4:6459257..6459283	chr5	5630133	5635236	inverse	5103	ATCOPIA93(AT5TE20395)	ND
87	chr4:6459288..6459589	chr5	5630133	5635181	inverse	5048	ATCOPIA93(AT5TE20395)	59
88	chr4:7161256..7161311	chr5	5630019	5635304	sens	5285	ATCOPIA93(AT5TE20395)	61
89	chr4:7355510..7355648	chr5	5629989	5635211	sens	5222	ATCOPIA93(AT5TE20395)	61
90	chr4:7355743..7357351	chr1	12753539	12755446	sens	1907	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
91	chr4:9320401..9320630	chr3	22059572	22063995	sens	4423	ATCOPIA78(AT3TE89830)	0
92	chr4:9650030..9650044	chr5	5629981	5635307	sens	5326	ATCOPIA93(AT5TE20395)	126
93	chr4:12856117..12856156	chr5	5630024	5635020	inverse	4996	ATCOPIA93(AT5TE20395)	87
94	chr4:12856488..12857428	chr5	5630101	5635020	inverse	4919	ATCOPIA93(AT5TE20395)	87
96	chr4:13252659..13253079	chr1	18013165	18016859	inverse	3694	ATCOPIA78(AT1TE59755)	30
96	chr4:13252659..13253351	chr1	18013165	18016671	inverse	3506	ATCOPIA78(AT1TE59755)	25
97	chr4:13340110..13341930	chr1	12756672	12758464	inverse	1792	ATCOPIA93(AT1TE41580)	ND
98	chr4:15341484..15341550	chr5	5630003	5635215	sens	5212	ATCOPIA93(AT5TE20395)	61
98	chr4:15341560..15341810	chr5	5629996	5635215	sens	5219	ATCOPIA93(AT5TE20395)	61
98	chr4:15341889..15342004	chr5	5630027	5635285	sens	5258	ATCOPIA93(AT5TE20395)	61
99	chr4:15342122..15344476	chr1	12753528	12755444	sens	1916	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
100	chr4:16528493..16528680	chr5	5629978	5635247	inverse	5269	ATCOPIA93(AT5TE20395)	101
101	chr4:16528890..16529640	chr5	5629978	5633981	inverse	4003	ATCOPIA93(AT5TE20395)	86
106	chr5:8231932..8231954	chr5	5630002	5634891	sens	4889	ATCOPIA93(AT5TE20395)	79
107	chr5:9025343..9025623	chr5	5630158	5634874	sens	4716	ATCOPIA93(AT5TE20395)	32
108	chr5:9025667..9026152	chr5	5630158	5634664	sens	4506	ATCOPIA93(AT5TE20395)	31
110	chr5:11468942..11469156	chr5	5629982	5634969	inverse	4987	ATCOPIA93(AT5TE20395)	124
111	chr5:11469483..11469585	chr1	12755071	12759071	inverse	4000	ATCOPIA93(AT1TE41580)	ND
114	chr5:15901533..15901609	chr2	10001521	10009484	inverse	7963	VANDAL21(AT2TE42810)	ND
115	chr5:18033802..18034012	chr5	5630059	5634833	sens	4774	ATCOPIA93(AT5TE20395)	38
116	chr5:18299747..18300422	chr5	5629997	5634855	inverse	4858	ATCOPIA93(AT5TE20395)	35
117	chr5:18300639..18300764	chr5	5630056	5634931	inverse	4875	ATCOPIA93(AT5TE20395)	35
118	chr5:19653026..19653270	chr5	5629981	5634850	sens	4869	ATCOPIA93(AT5TE20395)	65
119	chr5:19653425..19655485	chr1	12753521	12755414	sens	1893	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
120	chr5:23909725..23909968	chr5	5630005	5634922	inverse	4917	ATCOPIA93(AT5TE20395)	56

epiRIL 559

36	chr3:20848404..20848562	chr2	10001484	10009576	sens	8092	VANDAL21(AT2TE42810)	472

epiRIL 159

1	chr1:310834..310907	chr5	5629983	5635306	sens	5323	ATCOPIA93(AT5TE20395)	612
2	chr1:2816544..2816631	chr5	5629981	5635302	inverse	5321	ATCOPIA93(AT5TE20395)	318
3	chr1:10823529..10823605	chr5	5629977	5635306	inverse	5329	ATCOPIA93(AT5TE20395)	654
4	chr1:12178819..12180424	chr3	15283667	15286610	inverse	2943	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
5	chr1:12618580..12620328	chr3	15283666	15286612	sens	2946	ATENSPM3(AT3TE62120) ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
10	chr1:23702919..23703108	chr5	5629978	5635306	sens	5328	ATCOPIA93(AT5TE20395)	258
11	chr1:23744319..23744502	chr5	5629980	5635302	sens	5322	ATCOPIA93(AT5TE20395)	ND
12	chr1:26556571..26556735	chr5	5629979	5635304	sens	5325	ATCOPIA93(AT5TE20395)	ND
30	chr2:7673068..7673223	chr5	5629988	5635306	sens	5318	ATCOPIA93(AT5TE20395)	ND
31	chr2:7674228..7675028	chr1	12753510	12755446	sens	1936	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	ND
33	chr2:11727250..11727268	chr5	5629982	5635306	sens	5324	ATCOPIA93(AT5TE20395)	348
34	chr2:14654227..14654294	chr2	10001489	10009568	inverse	8079	VANDAL21(AT2TE42810)	ND
35	chr2:17227288..17228175	chr1	12755070	12758673	sens	3603	ATCOPIA93(AT1TE41580)	ND
36	chr2:17228255..17228329	chr5	5629977	5635307	sens	5330	ATCOPIA93(AT5TE20395)	400
37	chr2:19234233..19234339	chr5	5629981	5635308	sens	5327	ATCOPIA93(AT5TE20395)	343
38	chr3:5586658..5586730	chr2	4900802	4909201	sens	8399	ATENSPM3(AT2TE20205)	1180
39	chr3:10768190..10768368	chr5	5629978	5635306	inverse	5328	ATCOPIA93(AT5TE20395)	ND

41	chr3:12662819..12663011	chr5	5629978	5635304	inverse	5326	ATCOPIA93(AT5TE20395)	256
48	chr3:17037926..17037926	chr5	5629984	5635303	sens	5319	ATCOPIA93(AT5TE20395)	ND
49	chr3:17612904..17613039	chr5	5629978	5635301	inverse	5323	ATCOPIA93(AT5TE20395)	645
50	chr3:17613753..17614031	chr1	12755064	12758676	inverse	3612	ATCOPIA93(AT1TE41580)	ND
51	chr3:22056613..22056671	chr5	5629977	5635307	sens	5330	ATCOPIA93(AT5TE20395)	329
52	chr4:898922..898985	chr5	5629977	5635305	inverse	5328	ATCOPIA93(AT5TE20395)	317
55	chr4:2669092..2669221	chr5	5629977	5635310	sens	5333	ATCOPIA93(AT5TE20395)	ND
67	chr4:5801186..5801251	chr2	10001493	10009645	sens	8152	VANDAL21(AT2TE42810)	1156
69	chr4:10736583..10740042	chr3	22062806	22063898	inverse	1092	ATCOPIA78(AT3TE89830)	61
69	chr4:10739723..10739871	chr3	22059536	22063898	inverse	4362	ATCOPIA78(AT3TE89830)	61
69	chr4:10739723..10739871	chr3	22695568	22700077	sens	4509	ATCOPIA78(AT3TE92525)	21
71	chr4:13252238..13252625	chr1	18013163	18016822	inverse	3659	ATCOPIA78(AT1TE59755)	121
71	chr4:13252238..13252625	chr1	18013163	18016898	inverse	3735	ATCOPIA78(AT1TE59755)	121
71	chr4:13252238..13252625	chr1	18013163	18018107	inverse	4944	ATCOPIA78(AT1TE59755)	121
73	chr5:1161803..1165749	chr1	12754333	12754819	sens	486	ATCOPIA93(AT1TE41575)	0
73	chr5:1161915..1161990	chr1	12753511	12760310	sens	6799	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	1
73	chr5:1161991..1162095	chr1	12753522	12758675	sens	5153	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	1
73	chr5:1161915..1161990	chr5	5629981	5635307	sens	5326	ATCOPIA93(AT5TE20395)	613
84	chr5:13596099..13596177	chr1	12753507	12758495	inverse	4988	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	189
85	chr5:17586067..17589123	chr1	12757258	12758396	inverse	1138	ATCOPIA93(AT1TE41580)	0
85	chr5:17587924..17590860	chr1	12755080	12756503	inverse	1423	ATCOPIA93(AT1TE41580)	0
85	chr5:17585869..17587760	chr1	12753506	12755462	inverse	1956	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	0
86	chr5:18459177..18459331	chr5	5629980	5634969	inverse	4989	ATCOPIA93(AT5TE20395)	262
88	chr5:23866866..23866906	chr1	12753510	12758419	sens	4909	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41570) ATCOPIA93(AT1TE41575)	0

epiRIL 55

1	chr1:10549..13848	chr4	4358130	4360287	2157	2081	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
2	chr1:1392803..1393537	chr5	5630359	5635230	4871	4346	ATCOPIA93(AT5TE20395)	ND
3	chr1:3990167..3990279	chr5	5629997	5635305	5308	5217	ATCOPIA93(AT5TE20395)	116
4	chr1:12169241..12169264	chr5	5630036	5635281	5245	5215	ATCOPIA93(AT5TE20395)	ND
5	chr1:12169447..12169578	chr5	5630036	5635281	5245	5215	ATCOPIA93(AT5TE20395)	123
29	chr1:20281062..20281407	chr5	5630087	5635202	5115	4309	ATCOPIA93(AT5TE20395)	ND
30	chr1:20679800..20679956	chr5	5629977	5635293	5316	5304	ATCOPIA93(AT5TE20395)	158
31	chr1:21486789..21486913	chr2	4900800	4909264	8464	2923	ATENSPM3(AT2TE20205)	683
32	chr1:21603943..21604210	chr2	4901064	4909286	8222	3081	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	226
33	chr1:25815758..25815886	chr5	5629983	5635302	5319	5192	ATCOPIA93(AT5TE20395)	117
34	chr1:27101182..27103667	chr2	4900803	4903338	2535	2522	ATENSPM3(AT2TE20205)	369
35	chr1:28367370..28371506	chr3	15285379	15286610	1231	1202	ATENSPM3(AT3TE62130) HELITRON2(AT3TE62100)	ND
36	chr2:1797..1882	chr3	11544047	11549107	5060	1563	VANDAL21(AT3TE48070)	0
36	chr2:1765..1882	chr5	13963334	13967691	4357	908	VANDAL21(AT5TE49755)	0
55	chr2:3215890..3216071	chr5	5629997	5635284	5287	5211	ATCOPIA93(AT5TE20395)	130
58	chr2:3275722..3275965	chr5	5629977	5635296	5319	5295	ATCOPIA93(AT5TE20395)	108
59	chr2:3371620..3371975	chr5	5630013	5635274	5261	5153	ATCOPIA93(AT5TE20395)	58
61	chr2:3505788..3506083	chr5	5629993	5635311	5318	5071	ATCOPIA93(AT5TE20395)	95
73	chr2:4171056..4171205	chr5	5629977	5635294	5317	5232	ATCOPIA93(AT5TE20395)	139
79	chr2:8272168..8272305	chr5	5629981	5635296	5315	5302	ATCOPIA93(AT5TE20395)	316
80	chr2:19475806..19475903	chr5	5629983	5635304	5321	5313	ATCOPIA93(AT5TE20395)	174
81	chr3:3351852..3356952	chr4	4358136	4360281	2145	2050	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
82	chr3:3549888..3550083	chr2	4900810	4909268	8458	4038	ATENSPM3(AT2TE20205)	202
83	chr3:3573618..3573970	chr5	5630194	5635156	4962	4256	ATCOPIA93(AT5TE20395)	ND
84	chr3:4589949..4590097	chr5	5629982	5635305	5323	5273	ATCOPIA93(AT5TE20395)	168

86	chr3:8036011..8036108	chr5	5630001	5635297	5296	5241	ATCOPIA93(AT5TE20395)	134
87	chr3:9019495..9019591	chr2	4900831	4909285	8454	2854	ATENSPM3(AT2TE20205) ATENSPM7(AT2TE20210)	277
88	chr3:9043438..9043531	chr2	4900808	4909219	8411	3572	ATENSPM3(AT2TE20205)	651
89	chr3:9142495..9142632	chr2	4901288	4906335	5047	2257	ATENSPM3(AT2TE20205)	ND
90	chr3:10647627..10647765	chr2	4900823	4909265	8442	2950	ATENSPM3(AT2TE20205)	366
105	chr3:14195764..14195849	chr2	10001490	10007272	5782	3420	VANDAL21(AT2TE42810)	271
110	chr3:14457168..14457296	chr2	4903490	4909015	5525	2484	ATENSPM3(AT2TE20205)	22
113	chr3:16618765..16619206	chr5	5630176	5635179	5003	4493	ATCOPIA93(AT5TE20395)	ND
114	chr3:16767338..16767437	chr2	4900833	4909247	8414	3467	ATENSPM3(AT2TE20205)	401
115	chr3:17341732..17341847	chr2	5852427	5857334	4907	4877	ATCOPIA12(AT2TE23850) ATCOPIA13(AT2TE23855)	ND
134	chr4:8561572..8561742	chr5	5629980	5635289	5309	5223	ATCOPIA93(AT5TE20395)	159
136	chr4:9650040..9650151	chr5	5629977	5635306	5329	5292	ATCOPIA93(AT5TE20395)	388
137	chr4:9650230..9651158	chr5	5629980	5635282	5302	5284	ATCOPIA93(AT5TE20395)	ND
138	chr4:10739659..10739988	chr3	22059539	22064250	4711	4404	ATCOPIA78(AT3TE89830)	42
138	chr4:10739659..10739988	chr3	22695564	22700155	4591	4568	ATCOPIA78(AT3TE92520) ATCOPIA78(AT3TE92525)	16
140	chr4:12413348..12416017	chr4	4358132	4360293	2161	1988	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
142	chr4:13252525..13253326	chr1	18013165	18016901	3736	3698	ATCOPIA78(AT1TE59755)	77
143	chr4:13668518..13673852	chr4	4358131	4360262	2131	2111	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
144	chr4:17154860..17155612	chr5	5630033	5635134	5101	4976	ATCOPIA93(AT5TE20395)	ND
145	chr4:17512208..17512273	chr2	10001502	10009608	8106	3457	VANDAL21(AT2TE42810)	593
146	chr5:913859..914446	chr1	11304563	11309787	5224	194	ATENSPM3(AT1TE36570)	ND
147	chr5:915339..915383	chr2	4386308	4392316	6008	380	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	ND
148	chr5:2123289..2123790	chr2	4901371	4909242	7871	1390	ATENSPM3(AT2TE20205)	ND
149	chr5:2610008..2612736	chr4	4358137	4360243	2106	2066	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
150	chr5:5980263..5980983	chr4	4358131	4360292	2161	2144	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	212
151	chr5:9624750..9624842	chr2	4900822	4909218	8396	3259	ATENSPM3(AT2TE20205)	296
158	chr5:11728605..11729400	chr2	4386593	4391184	4591	274	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	0
158	chr5:11731495..11732251	chr2	4386461	4391184	4723	151	ATHILA3(AT2TE18410) ATENSPM3(AT2TE18415)	0
159	chr5:11733791..11734524	chr2	4904048	4908562	4514	411	ATENSPM3(AT2TE20205)	23
165	chr5:12485252..12486832	chr2	4904534	4906825	2291	1886	ATENSPM3(AT2TE20205)	ND
166	chr5:12603975..12604164	chr5	5629977	5634971	4994	4936	ATCOPIA93(AT5TE20395)	165
167	chr5:12949683..12950233	chr2	10002100	10009589	7489	2202	VANDAL21(AT2TE42810)	ND
168	chr5:14550492..14550576	chr5	5629986	5634975	4989	4928	ATCOPIA93(AT5TE20395)	160
169	chr5:19519546..19519749	chr5	5629984	5634967	4983	4913	ATCOPIA93(AT5TE20395)	179
172	chr5:23373982..23374373	chr1	12753862	12758435	4573	3039	ATCOPIA93(AT1TE41580) ATCOPIA93(AT1TE41575)	ND
173	chr5:23864824..23869728	chr4	4358136	4360231	2095	1996	ATENSPM3(AT4TE18510) ATLANTYS1(AT4TE18500) ATENSPM3(AT4TE18505)	ND
174	chr5:24028648..24028773	chr2	4900855	4909267	8412	3551	ATENSPM3(AT2TE20205)	351
175	chr5:24196449..24197993	chr1	12755359	12759037	3678	3046	ATCOPIA93(AT1TE41580)	ND
176	chr5:25775460..25775570	chr2	10001516	10009548	8032	3671	VANDAL21(AT2TE42810)	653

ces insertions sont considérées comme une seule car les sites accepteurs proposés sont en fait une même séquence répétée

ces insertions sont considérées comme une seule car les sites accepteurs sont trop proches les un des autres pour être en théorie discrminés par TE tracker

insertion d'*ATCOPIA93* attribuée au donneur *AT1TE41580*

Remarque : lorsque le score ne permet pas de discriminer entre plusieurs donneurs, ils sont tous présentés

ddm1

8	chr3:7610..10284	chr5	18142173	18144272	1927	2099	ATCOPIA21(AT5TE65370)	138
8	chr3:8094..13048	chr5	18145928	18146895	892	967	ATCOPIA21(AT5TE65370)	49
15	chr4:12903883..12905371	chr2	18709080	18710778	1696	1698	ATCOPIA57(AT2TE84975);ATCOPIA57(AT2TE84970);ATCOPIA57(AT2TE84980)	257
16	chr5:7997182..7997325	chr1	3780764	3785352	4384	4588	ATCOPIA78(AT1TE12295)	1
16	chr5:7997140..7997325	chr1	7717844	7722201	4131	4357	ATCOPIA78(AT1TE24850)	0
16	chr5:7997140..7999213	chr1	7717844	7719473	1555	1629	ATCOPIA78(AT1TE24850)	0
16	chr5:7997106..7997325	chr1	18013164	18017548	4346	4384	ATCOPIA78(AT1TE59755)	0
16	chr5:7997182..7997325	chr1	21524994	21529467	4315	4473	ATCOPIA78(AT1TE71045)	0
16	chr5:7997182..7997325	chr3	13369254	13373802	4265	4548	ATCOPIA78(AT3TE54550)	0
16	chr5:7997182..7997411	chr3	22059534	22063885	4223	4351	ATCOPIA78(AT3TE89830)	0
16	chr5:7997182..7997325	chr3	22695565	22700092	4415	4527	ATCOPIA78(AT3TE92525)	0
16	chr5:7997182..7997325	chr5	4208082	4212632	4376	4550	ATCOPIA78(AT5TE15240)	0
17	chr5:9423718..9423970	chr1	12942887	12947012	4021	4125	7(AT1TE42215);ATENSPM3(A	153

Annexe 3 : Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. (Colomé-Tatché et al. 2012)

Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation

Maria Colomé-Tatché[a,1], Sandra Cortijo[b,1], René Wardenaar[a], Lionel Morgado[a], Benoit Lahouze[b], Alexis Sarazin[b], Mathilde Etcheverry[b], Antoine Martin[b], Suhua Feng[c,d], Evelyne Duvernois-Berthet[b], Karine Labadie[e], Patrick Wincker[e], Steven E. Jacobsen[c,d], Ritsert C. Jansen[a], Vincent Colot[b,2], and Frank Johannes[a,2]

[a]Groningen Bioinformatics Centre, University of Groningen, 9747 AG Groningen, The Netherlands; [b]Institut de Biologie de l'Ecole Normale Supérieure, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8197, Institut National de la Santé et de la Recherche Médicale (INSERM) Unité 1024, Paris F-75005, France; [c]Howard Hughes Medical Institute and [d]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095; and [e]Genoscope, Institut de Génomique, Commissariat à l'Energie Atomique, Evry F-91057, France

Edited by David C. Baulcombe, University of Cambridge, Cambridge, United Kingdom, and approved August 16, 2012 (received for review July 27, 2012)

The rate of meiotic crossing over (CO) varies considerably along chromosomes, leading to marked distortions between physical and genetic distances. The causes underlying this variation are being unraveled, and DNA sequence and chromatin states have emerged as key factors. However, the extent to which the suppression of COs within the repeat-rich pericentromeric regions of plant and mammalian chromosomes results from their high level of DNA polymorphisms and from their heterochromatic state, notably their dense DNA methylation, remains unknown. Here, we test the combined effect of removing sequence polymorphisms and repeat-associated DNA methylation on the meiotic recombination landscape of an *Arabidopsis* mapping population. To do so, we use genome-wide DNA methylation data from a large panel of isogenic epigenetic recombinant inbred lines (epiRILs) to derive a recombination map based on 126 meiotically stable, differentially methylated regions covering 81.9% of the genome. We demonstrate that the suppression of COs within pericentromeric regions of chromosomes persists in this experimental setting. Moreover, suppression is reinforced within 3-Mb regions flanking pericentromeric boundaries, and this effect appears to be compensated by increased recombination activity in chromosome arms. A direct comparison with 17 classical *Arabidopsis* crosses shows that these recombination changes place the epiRILs at the boundary of the range of natural variation but are not severe enough to transgress that boundary significantly. This level of robustness is remarkable, considering that this population represents an extreme with key recombination barriers having been forced to a minimum.

decrease in DNA methylation 1 | epigenetic inheritance | DNA methylome | epi-haplotype

Meiotic recombination is a fundamental process in genetics whereby maternally and paternally inherited homologous chromosomes exchange material, either nonreciprocally by gene conversion or reciprocally by crossing over (CO). COs are not distributed uniformly along the genome but occur more often in chromosome arms and are strongly suppressed in pericentromeric regions (1–3), partly as a result of sequence and chromatin determinants (1, 4–8). It is commonly believed that in plants and mammals high levels of DNA sequence polymorphisms as well as heterochromatic features associated with repeats, notably dense DNA methylation and transcriptional silencing, play a central role in this suppression (1, 4).

Suppression of COs by dense DNA methylation has been demonstrated experimentally in the fungus *Ascobolus* (7). Specifically, COs were reduced when the recombination interval was methylated on one homolog and were abolished almost completely when methylated on both homologs. In *Arabidopsis*, two recent mapping studies analyzed F_2 progeny derived from crosses between Columbia *ddm1* and *met1* [Col(*ddm1*),Col(*met1*)] DNA methylation mutants and wild-type Landsberg [Ler(WT)] accessions and showed that loss of DNA methylation could not alleviate the suppression of COs in pericentromeric regions of chromosomes (9, 10). However, as pointed out by the authors, this experimental design could not rule out an inhibitory effect of sequence divergence between Col and Ler on COs.

An ideal design would use crosses between isogenic individuals, with one of the crossing partners having decreased DNA methylation levels throughout the genome (9). Melamed-Bessudo and Levy (9) implemented such an approach by crossing Col(*ddm1*) mutant to Col(WT). Using two fluorescent markers spanning a 16-centimorgan (cM) interval on the arm of chromosome 3, they detected increased CO rates in F_2 plants derived from these parents relative to plants derived from a Col (WT)×Col(WT) control cross and concluded that COs in euchromatic regions can be up-regulated by loss of DNA methylation. A similar approach at a genome-wide scale and with high mapping resolution, particularly in pericentromeric regions, has not been attempted because of a lack of appropriate molecular and genetic tools. Hence, the combined effect of DNA methylation and sequence variation on COs has not been tested comprehensively in *Arabidopsis* or in any other higher eukaryote.

We previously reported the construction of a large population of epigenetic recombinant inbred lines (epiRILs) in *Arabidopsis* (11, 12), which provides a powerful experimental system to conduct such a test. These epiRILs were obtained by first crossing a fourth-generation plant homozygous for the recessive *ddm1-2* mutation with a near-isogenic WT individual. The *ddm1-2* mutation mostly affects transposable elements (TEs) and other repeats, which lose DNA methylation and become transcriptionally reactivated in a transmissible manner in many instances (11–14). However, transposition events are relatively rare (15).

Author contributions: V.C. and F.J. designed research; M.C.-T., S.C., M.E., A.M., S.F., K.L., P.W., S.E.J., V.C., and F.J. performed research; M.C.-T., S.C., R.W., A.S., M.E., A.M., S.F., E.D.-B., K.L., P.W., S.E.J., R.C.J., V.C., and F.J. contributed new reagents/analytic tools; M.C.-T., S.C., R.W., L.M., B.L., A.S., A.M., E.D.-B., V.C., and F.J. analyzed data; and M.C.-T., V.C., and F.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The genome-wide data reported in this paper have been deposited in the Gene Expression Omnibus public functional genomics data repository, http://www.ncbi.nlm.nih.gov/geo/ [GEO accession nos. GSE37284 (methylomes), GSE37106 (transcriptomes), and GSE37285 (methylomes of intermediate generations)].

[1]M.C.-T. and S.C. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: colot@biologie.ens.fr or f.johannes@rug.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1212955109/-/DCSupplemental.

Thus, F_1 individuals can be considered homozygous throughout the genome, except at the *DDM1* locus and at the few loci affected by TE mobilization, but have chromosome pairs that differ markedly in their DNA methylation levels and transcriptional activity over TEs and other repeats (11, 16). A single F_1 *DDM1*/*ddm1* individual was backcrossed to the WT parental line, and after the progeny homozygous for the WT *DDM1* allele were selected, the epiRILs were propagated through seven rounds of selfing. In this design, more than 85% of all informative recombination events occur in the first two inbreeding generations (F_1 and backcross), with fewer informative events being contributed by each subsequent generation (17).

Previous targeted analysis indicated that many of the parental differences in DNA methylation and transcriptional activity of repeats are inherited stably in the epiRILs (11, 12). Regions with segregating DNA methylation states therefore can serve as physical markers to detect the frequency and distribution of recombination events along chromosomes even though the two homologs have nearly identical DNA sequences.

In this study we report the construction of a recombination map using genome-wide DNA methylation data from 123 epiRILs. This map was derived from 126 meiotically stable differentially methylated regions (DMRs) covering 81.9% of the total genome. Estimates of the genetic length for each chromosome revealed that global recombination rates are comparable with those of classical *Arabidopsis* crosses. On a local scale, we demonstrate that suppressed recombination activity within repeat-rich, pericentromeric regions of chromosomes is maintained robustly even after the removal of sequence polymorphisms and repeat-associated DNA methylation. Furthermore, we were able to identify 3-Mb regions flanking pericentromeric boundaries that appear to be subject to additional suppression and show that this effect is accompanied by increased recombination activity in chromosome arms. A direct comparison with 17 classical *Arabidopsis* crosses reveals that these recombination changes place the epiRILs at the boundary of the range of natural variation but appear not to be severe enough to transgress that boundary significantly.

Results

Construction of a Recombination Map Using Transgenerationally Stable DMRs. To demonstrate that transgenerationally stable DMRs can be used for the construction of a recombination map in an isogenic population, we carried out methylated DNA immunoprecipitation followed by hybridization to a whole genome DNA tiling array (MeDIP-chip) on 123 epiRILs and on the two parental lines (256 array experiments including replicates). The 123 epiRILs originally were chosen using a selective (epi)genotyping strategy for two uncorrelated complex traits, flowering time and root length. We used a three-state Hidden Markov Model (HMM) to classify tiling array signals into three underlying DNA methylation states (18): unmethylated (U), intermediate methylation (I), or methylated (M). Benchmarking of these HMM calls against whole-genome bisulphite sequencing data (~30×) for six epiRILs confirmed that both the MeDIP protocol and the analysis method performed well (*SI Appendix*, Fig. S1 and Table S1). Comparison of the two parental DNA methylomes revealed 2,611 DMRs representing clear instances of methylation loss in *ddm1* (transitions from M to U). These DMRs (median length: 1,211 bp; range: 318–24,624 bp) were distributed throughout the genome but, as expected, were more abundant in pericentromeric regions (Fig. 1*A* and *SI Appendix*, Table S2) (19).

We examined the DNA methylation state at all parental DMRs in each of the 123 epiRILs and inferred their parent of origin (i.e., epigenotypes). Segregation was not compatible with stable inheritance of *ddm1*-induced DNA hypomethylation for 1,744 (66.8%) of the parental DMRs, and in most of these cases our data pointed to fully or partially penetrant reversion to WT DNA methylation. In contrast, 867 (33.2%) of the parental DMRs segregated in the expected 3:1 Mendelian ratio (*SI Appendix*, Fig. S2 and Table S3). Stable DMRs were associated with a comparatively lower abundance of siRNAs in the WT and *ddm1* parental

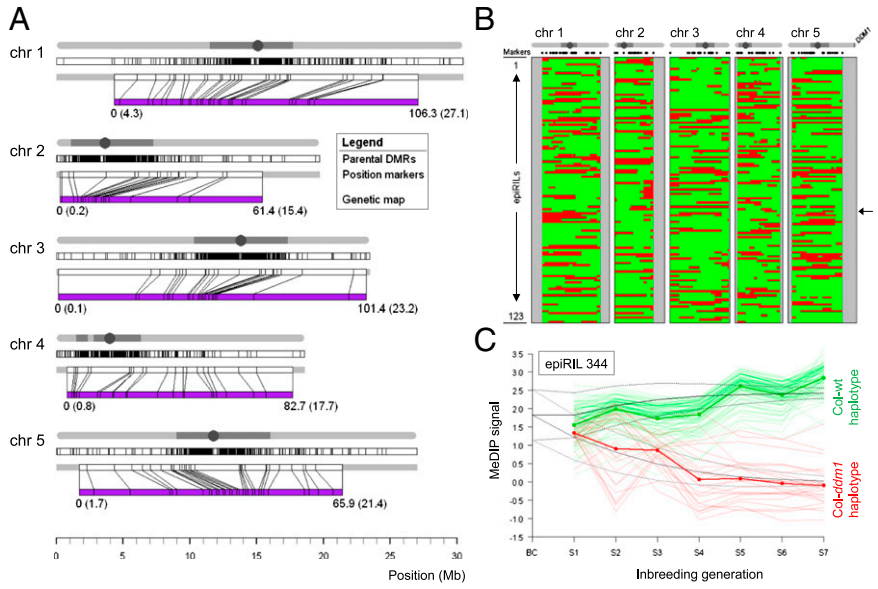


Fig. 1. Recombination map construction. (*A*) Genome-wide distribution of the 2,611 parental DMRs (*Top*) and the 126 DMRs (i.e., markers; *Middle*) retained for construction of the recombination map (purple, *Bottom*) for each of the five *Arabidopsis* chromosomes. The mapping between physical and genetic positions of markers is shown. (*B*) Inference of inherited WT (green) and ddm1 (red) haplotypes along the genome (*x*-axis) as inferred from the recombination map for each of the 123 epiRILs (*y*-axis) (*SI Appendix*, Table S5). Chromosome extremities not covered by the genetic map are indicated in gray. The genome of epiRIL 344 is indicated by an arrow. A schematic representation of each chromosome is plotted above the map with the physical location of the *DDM1* gene shown at the end of chromosome 5. (*C*) Transgenerational methylation data for epiRIL 344. Shown are the average methylation signals for the 126 markers, with regions that are predicted to become fixed for the *ddm1* haplotypes (thin red lines) and the WT haplotypes (thin green lines) after seven selfing generations. The average signals (red and green thick solid lines) are in agreement with Mendelian inbreeding theory (black solid lines).

lines (*SI Appendix*, Fig. S3). These findings are in agreement with previous analyses (11, 12) and indicated that the 867 stable DMRs are not efficient targets of siRNA-mediated DNA remethylation, even after eight rounds of meiosis. These stable DMRs therefore could serve as physical markers in an extension of the Lander–Green algorithm (20) to derive a genetic map. After application of the algorithm and removal of mainly genetically redundant markers (i.e., markers located less than 0.0001 cM apart), 126 of the original 867 markers were retained (Fig. 1 *A* and *B* and *SI Appendix*, Fig. S2 and Table S4). These 126 markers covered ∼81.9% of the total genome (74.7, 77.0, 98.4, 91.1, and 73.0% of chromosomes 1, 2, 3, 4, and 5, respectively).

Many of the 126 markers contained TE sequences, consistent with the targeted effect of *ddm1* on these and other repeats (*SI Appendix*, Fig. S4). However, in a vast majority of cases, markers included only TE relics, which likely have lost their capacity to be mobilized (*SI Appendix*, Table S6). Indeed, both comparative genomic hybridization (*SI Appendix*, Fig. S5) and preliminary whole-genome resequencing suggested that none of the 126 DMRs contain sequences that were mobilized in the parental *ddm1* line or the epiRILs (*SI Appendix*, Table S6). Consistent with this finding, pair-wise recombination fractions between the 126 markers indicated a well-behaved and robust genetic map, reminiscent of those typically seen in classical crosses involving DNA sequence markers, with high correlation among linked loci and virtually no correlations among loci in different chromosomes (*SI Appendix*, Fig. S6). Moreover, all inferred *ddm*1-inherited non-recombinant pericentromeric haplotypes contained significantly less DNA methylation and were more actively transcribed than their WT counterparts (*SI Appendix*, Figs. S7 and S8).

To test further the transgenerational stability of the 126 markers as well as our inference of the parental epigenotypes at these marker locations, we performed genome-wide DNA methylation analysis for one selected line (epiRIL 344) for each of its seven selfing generations (7 × 2 replicates = 14 array experiments). Fixation occurred for the predicted parental epigenotype in each case, and the rate of approach toward fixation was consistent with Mendelian inbreeding theory for a backcross-derived RIL (19) (Fig. 1*C*). Taken together, these results rule out any ambiguity in the actual location or DNA methylation state of the stable DMRs used for constructing the genetic map.

Total Genetic Length in the epiRILs Does Not Diverge Significantly from the Natural Range. One approach for evaluating the epiRILs recombination map is by comparison with a Col (WT)×Col(WT)-derived reference cross. In this set-up, changes in recombination patterns can be attributed directly to DNA methylation loss. However, tracking recombination events in such a reference is experimentally challenging. It requires a system akin to the fluorescent marker reporters used by Melamed-Bessudo and Levy (9), which does not easily scale genome-wide. An alternative approach is to evaluate the epiRILs in the context of natural variation. In terms of DNA sequence and DNA methylome divergence of its founder parental lines, the epiRILs can be viewed as representing an extreme situation with key barriers to recombination having been forced to a minimum. An important question therefore is how genome-wide recombination patterns in this population compare with those seen in crosses derived from different pairs of natural accessions.

We estimated the genetic length for each of the five epiRIL chromosomes using Haldane's map function. The lengths were 106.3, 61.4, 101.4, 82.7, and 65.9 cM for chromosomes 1–5, respectively, and correlated positively with physical chromosome length (*SI Appendix*, Fig. S9). The total length of the genetic map was 417.7 cM, yielding an average marker spacing of ∼0.804 Mb (3.45 cM). These estimates are similar to those previously reported for genetic maps based on classical *Arabidopsis* crosses (21–24). The use of other map functions that account for CO interference, such as the Kosambi or Carter and Falconer functions, yielded very similar results (*SI Appendix*, Fig. S10). To perform a more direct comparison between the epiRIL map and those of classical *Arabidopsis* crosses, we reanalyzed recombination data obtained for 17 F_2 populations (24) that were derived from pairs of 18 distinct natural accessions. In total, these populations consisted of 7,045 plants (∼410 plants per cross; range: 235–462 plants), which were genotyped at 235 markers on average (range: 215–257 markers) (24). To facilitate a meaningful comparison, we constructed a consensus map using 83 markers that were shared across populations (*SI Appendix*, Fig. S11 and Table S7). Thorough testing showed that the reduction to 83 markers in the epiRIL and F_2 maps led to no significant loss of information in capturing the linkage structure along chromosomes (*SI Appendix*, Figs. S12 and S13), and the 83 markers therefore were deemed appropriate for this comparative analysis.

Estimates of the genetic length of each of the five chromosomes revealed substantial natural variation among the F_2 populations (Figs. 2*B* and 3*A*). However, the genetic lengths of the

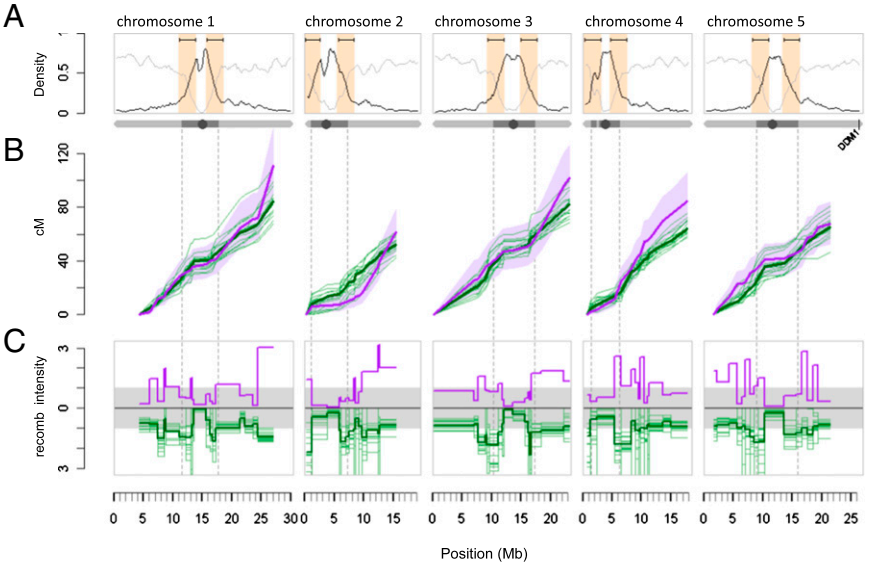


Fig. 2. Comparison of global and local recombination patterns in the epiRILs and the 17 F_2 populations (24). (*A*) Chromosome-wide gene (light gray line) and transposon (dark line) density distribution. The 3-Mb windows bracketing the intersection points between transposon- and gene-dense regions are indicated in orange. (*B*) Cumulative cM lengths of the epiRILs (thick purple line) and each of the F_2 populations (thin green lines) using the consensus map. Purple shading shows the ±95% confidence interval (CI).The thick green line denotes the average F_2 cumulative length (in cM). The dotted vertical lines define the pericentromeric regions of each chromosome. (*C*) The distribution of normalized recombination intensities (cM/Mb of a given marker interval divided by the cM/Mb chromosome average) shows suppression of recombination within pericentromeric regions and elevation at its boundaries. Color coding is as in *B*.

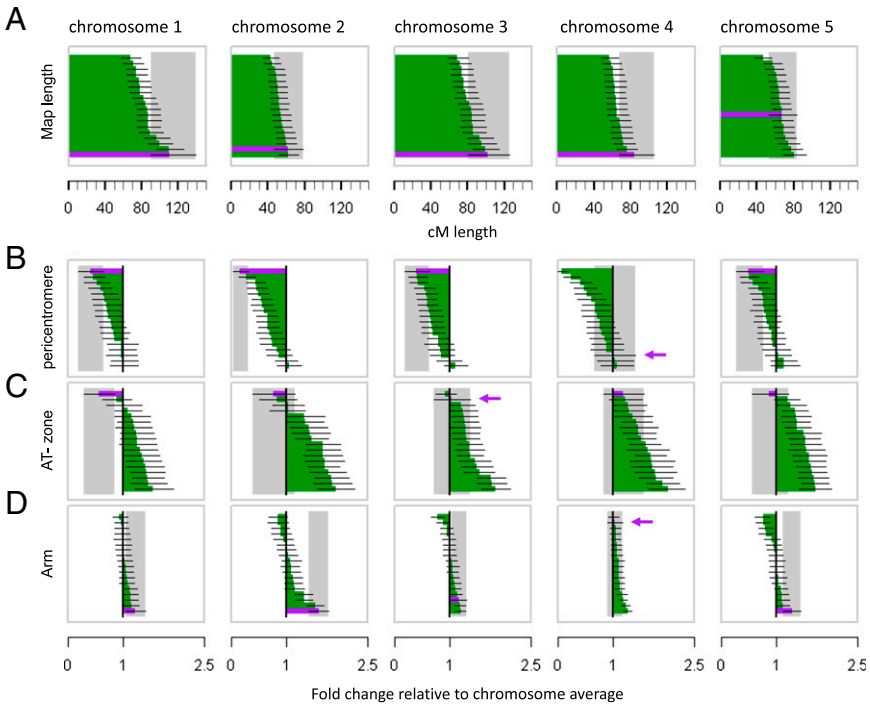


Fig. 3. Estimated genetic lengths and fold-change recombination intensities. (*A*) Estimated genetic lengths (±95% CI) of the epiRILs (purple) and each of the 17 F_2 populations (green) (24). (*B*–*D*) Fold change in recombination intensity [(cM/Mb) region/(cM/Mb) chromosome average] ±95% CI in pericentromeric regions (*B*), AT zones defined by a 3-Mb window bracketing the intersection point between transposon- and gene-dense regions at pericentromeric boundaries (*C*), and chromosome arms (*D*). Purple arrows indicate the location of the epiRILs when applicable. The values presented in each panel are ordered to highlight trends in the epiRILs recombination landscape. The identifiers of individual F_2 crosses corresponding to this ordering can be found in *SI Appendix*, Table S11.

epiRIL chromosomes did not diverge significantly from the natural range (Figs. 2*B* and 3*A*). The exception was chromosome 1, where we observed a significant increase relative to five of the F_2 crosses. Overall, therefore, our data indicate that the global recombination rate in the epiRILs is not altered drastically. Nonetheless, we noted a clear, but nonsignificant, trend toward longer genetic lengths for chromosomes 1–4 as compared with the F_2 populations (Fig. 3*A*); this trend is at least partly consistent with DNA methylation and DNA sequence polymorphisms acting as barriers to the global recombination rate in *Arabidopsis*.

Suppression of Pericentromeric Recombination Persists in the epiRILs and Shows a Trend Toward Additional Reinforcement. To explore the relationship between the epiRILs map and those of the different F_2 crosses at a subchromosomal scale, we examined in more detail the distribution of recombination intensities, expressed as cM/Mb, for each marker interval along the genome (Fig. 2*C*). All populations, including the epiRILs, had clearly suppressed recombination activity across pericentromeric regions relative to the chromosome averages (Figs. 2*C* and 3*B*). The exception to this trend was chromosome 4, for which the epiRILs showed a slight increase of recombination intensity (Fig. 3*B*). However, the presence of the heterochromatic knob on chromosome 4 in the Columbia accession, but not in other accessions, makes this result difficult to interpret (10).

Specifically, recombination intensities in pericentromeric regions of epiRIL chromosomes 1, 2, 3, and 5 were, respectively, 2.50, 6.88, 2.53, and 2.01 times lower than the chromosome average, which compares to 1.27 (range: 0.97–2.15), 1.51 (range: 0.95–3.68), 1.52 (range: 0.90–2.48), and 1.20 (range: 0.87–1.98) in the F_2 populations (Fig. 3*B* and *SI Appendix*, Table S8). This persistent suppression effect in the epiRIL agrees with the results of Melamed-Bessudo and Levy (9) and Mirouze et al. (10), who examined mapping populations derived from a Col(*ddm1*)×Ler (WT) and a Col(*met1*)×Ler(WT) cross, respectively. Hence, loss of DNA methylation appears to be insufficient to release pericentromeric suppression of recombination, even in the absence of DNA sequence polymorphisms. On the contrary, we found a clear trend toward enhanced suppression in the epiRILs: Recombination intensities in this population were consistently at the bottom of the natural range compared with the F_2 populations, even though chromosome-wide recombination rates were comparatively large. Enhanced suppressive effects also were reported by Melamed-Bessudo and Levy (9) and Mirouze et al. (10), thus highlighting an unexpected and complex relationship between DNA methylation and the suppression of recombination in pericentromeric regions of *Arabidopsis* chromosomes.

Reinforced Suppression of Recombination Extends to Pericentromeric Boundaries in the epiRILs and Appears to Be Compensated by Increased Recombination in Chromosome Arms. In contrast to core pericentromeric regions, recombination intensities in the F_2 populations increase rapidly at pericentromeric boundaries with chromosome arms (Figs. 2*C* and 3*C*). An important property of these regions is that they correspond to major transitions in genome content from TE-rich to gene-rich sequences (Fig. 2*A*) and also have been described recently as hotspots of historical recombination activity at the species level (*SI Appendix*, Fig. S14) (25). We found that nearly 40% of all detected recombination breakpoints in the F_2 populations mapped within a 3-Mb window bracketing the intersection point in these transition zones (henceforth referred to as "annotation transition zones"; AT zones), yielding local recombination intensities that were consistently above the chromosome averages (Fig. 3*C* and *SI Appendix*, Fig. S15 and Table S8).

This finding differs strongly from the situation seen in the epiRILs: AT zones accounted for only 25.31% of all detected recombinants in this population, and recombination intensities were close to the chromosome average (in chromosomes 3 and 4) or even below it (in chromosomes 1, 2, and 5) (Fig. 3*C* and *SI Appendix*, Table S8). These results suggest that the enhanced suppression of recombination seen in the epiRIL pericentromeric regions (see above) is driven at least in part by the more localized reduction of recombination within AT zones, which cover (on average) only 63.4% of the pericentromeric regions on either side of the centromeres. The two previous studies using mapping populations derived from crosses between Col (*ddm1*) and Ler(WT) (9) and between Col(*met1*)and Ler(WT) (10) were not able to delineate these local effects, most likely

because of the sparsity of their genetic markers (two to three markers per pericentromeric region). Marker density in the epiRIL map, in contrast, was relatively high within AT zones and even permitted fine mapping of shared and nonshared recombination breakpoints to a resolution as low as 4 kb (*SI Appendix*, Figs. S16 and S17 and Tables S9 and S10).

Furthermore, our data indicate that suppression of recombination within AT zones in the epiRILs is accompanied by increased recombination in chromosome arms (Fig. 3*D* and *SI Appendix*, Fig. S18). This apparent compensatory effect reconciles the enhanced local suppression seen in the epiRILs with the earlier observation that chromosome-wide recombination rates are relatively large compared with the F_2 populations. This effect was most pronounced on epiRIL chromosomes 1, 2, and 5 (the chromosomes with the strongest suppression in the AT zone), with recombination intensities being 1.23, 1.6, and 1.3 times above the chromosomes' average (*SI Appendix*, Table S8). We failed to identify a similar trend in the F_2 populations (*SI Appendix*, Fig. S18 and Table S8), suggesting that this effect is a specific feature of the epiRIL recombination landscape.

Discussion

In this study we demonstrate that stable DNA methylation differences can be used as physical markers to derive genome-wide recombination patterns in a near isogenic population of epiRILs. We find that recombination suppression is maintained robustly in pericentromeric regions of the epiRILs, despite the extensive loss of sequence variation and of DNA methylation and transcriptional silencing over repeats. This observation indicates that these factors do not play a major role in the suppression of pericentromeric COs. This finding is contrary to common belief and is particularly intriguing given the interplay between recombination and transcription observed in yeast and the mouse (26, 27). Whether mechanisms exist in *Arabidopsis* that actively sequester the recombination machinery away from gene-promoter regions or other genomic elements, as in the mouse (27), remains to be determined.

Nonetheless, our results indicate that loss of DNA methylation over repeat sequences can lead to a local reinforcement of recombination suppression in pericentromeric regions and to increased recombination activity along chromosome arms. Similar results were reported by Melamed-Bessudo and Levy (9) and Mirouze et al. (10) using genetically divergent populations. Therefore we conclude that the absence of sequence polymorphisms is insufficient to counteract the enhanced suppressive effects induced by the loss of DNA methylation in pericentromeric regions. On the other hand, the lack of sequence polymorphisms still may be partly responsible for the increased recombination rates observed in chromosome extremities (9).

Melamed-Bessudo and Levy (9) demonstrated that *ddm1*-induced demethylation of only one homolog produces the same recombination changes seen when both homologs are demethylated. Our results and conclusions therefore should be generalizable to the two-homolog situation. However, it has been shown in *Ascobolus* that DNA methylation of a known recombination hotspot inhibits COs more severely when both homologs are methylated (7). Similar localized dosage effects may therefore also be present in *Arabidopsis*.

Our study and those of Melamed-Bessudo and Levy (9) and Mirouze et al. (10) have used well-characterized *ddm1* and *met1* DNA methylation mutants as a tool to perturb genome-wide methylation levels experimentally. Both *ddm1* and *met1* experience a nearly 70% reduction in DNA methylation levels genome-wide. This drastic loss probably sets an upper limit to the amount of demethylation that can be incurred in nature. Indeed, it is difficult to conceive of mechanisms that would elicit similar or more severe changes under natural settings, unless they involve spontaneous mutations in genes important for DNA methylation control, such as *ddm1* or *met1*. Interestingly, a recent analysis of *Arabidopsis* mutation accumulation lines showed that drastic alterations in the methylome of one outlier line were likely caused by a spontaneous mutation in a methyl-transferase gene (28, 29), which must have arisen during just 30 generations of selfing. This observation suggests that similar events are certainly plausible under natural conditions.

An assessment of whether strong methylation loss can elicit recombination changes at magnitudes that are sufficient to drive genome evolution in this species has been lacking. Our study is an initial step in providing such an assessment. Our analysis of the 17 F_2 populations derived from 18 natural accessions (24) allowed us to quantify the magnitude of the recombination changes observed in the epiRILs in the context of natural variation. Although we find that the epiRILs nearly always are situated at the boundary of the natural range, there is no strong evidence that local and global recombination patterns in this population markedly transgress the natural range. Indeed, in many cases, several of the F_2 populations displayed even more extreme divergence from the F_2 population average than did the epiRILs. These findings lead us to conclude that severe losses of DNA methylation along *Arabidopsis* chromosomes have no drastic implications for recombination-mediated genome evolution. This high level of robustness raises questions concerning the precise mechanisms that have shaped the recombination landscape in this species in the first place.

Of course, severe depletion of DNA methylation can drive other important events, such as large-scale structural rearrangements and polyploidization, which may impact the course of genome evolution. In addition, natural epigenetic variation, such as that associated with differential DNA methylation, can act on complex traits that are under natural selection (30), thereby changing linkage disequilibrium relations within and across chromosomes. However, understanding and documenting the impact of epigenetic variants on complex traits is challenging, mainly because of the technical difficulties in ruling out the confounding effect of DNA sequence polymorphisms (31). Because of this limitation, it has been argued that the epiRILs constitute an ideal system for the study of epigenetic inheritance in *Arabidopsis* (17, 32–34). We and others have shown recently that many adaptive phenotypes, such as plant height, flowering time, and growth rate, are highly heritable in this population (12, 35, 36). Segregating phenotypic effects also have been observed in another epiRIL population which was obtained from a cross between Col(*met*1) and Col(WT) (37).

A logical next step in the analysis of these populations is to map and characterize the epigenetic basis of these complex traits. The linkage map reported here (Fig. 1*B*) can be used in conjunction with classical quantitative trait-locus mapping methods to achieve this characterization in the *ddm*1-derived epiRILs. Ultimately, such efforts should contribute significantly to our understanding of epigenetics in adaptive evolution.

Materials and Methods

Methylome Analysis. MeDIP was carried out as previously described (18) followed by hybridization to a custom NimbleGen tiling array covering the Arabidopsis genome at 165 nt resolution (38). Including dye-swaps, we performed a total of 256 array experiments (*SI Appendix*, section 1). For each array, probe signals were classified into three underlying methylation states, methylated (M), intermediate (I), or unmethylated (U), using the HMM presented previously (*SI Appendix*, section 2) (18). These inferred methylation states were cross-validated against whole-genome bisulphite-sequencing data of six epiRILs (*SI Appendix*, section 3, Fig. S1, and Table S1).

Definition of Parental DMRs. We conducted a probe-level comparison of the HMM calls between the *ddm1* and WT parents (*SI Appendix*, section 4). Probe-level methylation calls were denoted as polymorphic when the parents differed (e.g., I in *ddm1* and M in WT) and as nonpolymorphic when they

were identical (e.g., U in *ddm1* and U in WT). Neighboring probes reporting the same polymorphic state were collapsed into single regions. Hence, parental DMRs were defined as regions of at least three consecutive probes that reported the same extreme polymorphic state (i.e., transitions from M in WT to U in *ddm1* or vice versa). We found 2,611 DMRs, all of which were U in the *ddm1*. Detailed summary statistics are given in *SI Appendix*, Table S2.

Calling of Parental Origin of DMRs in the epiRILs. For any given epiRIL the parental origin of each DMR (i.e., epigenotype) was determined using an HMM-based inference method (*SI Appendix*, section 4).

Mendelian Segregation Criterion. Under the assumption that DMRs were stable for eight generations of breeding, both WT- and *ddm1*-like parental states should appear according to Mendelian segregation ratios in the epiRILs. The sampling variation around these ratios was calculated from a binomial distribution taking into account the sample size (n = 123), the cross design, and the 8% F_2 contamination previously reported (12). DMRs in the epiRILs showing a percentage of WT-like states between 62.7% and 83.3% (the expected value being 73%) were taken as putative transgenerationally stable markers. In total 867 parental DMRs fulfilled this criterion and were used subsequently as a starting point for map construction (*SI Appendix*, section 5, Fig. S2, and Table S3).

Extension of Lander–Green Algorithm. Derivation of a genetic map using DMRs was carried out through an extension of the Lander-Green algorithm (20), which was designed to accommodate marker and individual specific error rates. Our implementation of this algorithm is detailed in *SI Appendix*, section 6.

Transcriptome Analysis of epiRILs and *ddm1* Seedlings. Whole-genome expression profiling was performed using a custom NimbleGen tiling array (37). For experimental details, see *SI Appendix*, section 7.

Transgenerational Analysis of DMRs. MeDIP-chip was carried out for epiRIL 344 for seven generations of selfing after the backcross, following the protocol described above. At each generation, DNA from five siblings was pooled. The expected signal behavior was derived using a Markov Chain strategy, considering the Mendelian inheritance of the marker probes (*SI Appendix*, section 8).

Construction of the Consensus Map. To facilitate a meaningful comparison of the epiRILs map with those of the 17 different F_2 populations, we constructed a consensus map (*SI Appendix*, section 9 and Fig. S11) by using the epiRILs map as a reference and selecting from each of the F_2 maps the SNPs closest to the reference, allowing a maximum distance of $\pm$1.39 Mb. The average distance from reference was $\pm$0.17 Mb, which led to little loss of information in capturing the recombination structure along the genome (*SI Appendix*, Figs. S12 and S13). Markers deemed too distant were not included in the consensus map. This process resulted in 83 markers (*SI Appendix*, Table S7).

Recombination Intensities at Major Annotation Transitions. Fig. 2 *A* and *C* shows that the recombination intensity increases rapidly at the pericentromeric boundaries, which also coincide with major transitions in genome content from genes to transposons. To find the area where the recombination intensity is maximal, we implemented a sliding window approach (*SI Appendix*, section 10 and Fig. S15).

Note Added in Proof. During the reviewing process, Yelina et al (Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, et al. (2012) Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants. PLoS Genet 8(8): e1002844. doi:10.1371/journal.pgen.1002844) reported elevated centromere-proximal COs, coincident with pericentromeric decreases and distal increases in met1 mutants. However, total numbers of CO events were found to be similar between wild type and met1. These results support the trends observed in the epiRIL population.

ACKNOWLEDGMENTS. We thank Tony Heitkam for help with the TE analysis. This work was funded by grants from the Ministère de la Recherche et de l'Enseignement Supérieur (to S.C., B.L., and M.E.); Agence Nationale de La Recherche TAG and MEIOMETH projects (to V.C.) and EPIMOBILE project (to V.C. and P.W.); European Union Seventh Framework Programme Network of Excellence EpiGeneSys (Award 257082; to VC); Netherlands Organisation for Scientific Research (NWO) (to F.J., M.L., and M.C.-T.); Consortium for Improving Plant Yield (CIPY) (to M.C.-T.); Netherlands Bioinformatics Centre (NBIC) (to R.W.); and EURATRANS (to R.C.J.). Work in the S.E.J. laboratory is supported by National Institutes of Health Grant GM60398. S.F. is a Special Fellow of the Leukemia & Lymphoma Society. S.E.J. is an investigator of the Howard Hughes Medical Institute.

1. Lichten M, de Massy B (2011) The impressionistic landscape of meiotic recombination. *Cell* 147:267–270.
2. Mézard C, Vignard J, Drouaud J, Mercier R (2007) The road to crossovers: Plants have their say. *Trends Genet* 23:91–99.
3. Muyt AD, Mercier R, Mézard C, Grelon M (2009) Meiotic recombination and crossovers in plants. *Genome Dyn* 5:14–25.
4. Edlinger B, Schlögelhofer P (2011) Have a break: Determinants of meiotic DNA double strand break (DSB) formation and processing in plants. *J Exp Bot* 62:1545–1563.
5. Chen W, Jinks-Robertson S (1999) The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics* 151:1299–1313.
6. Emmanuel E, Yehuda E, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA (2006) The role of *AtMSH2* in homologous recombination in *Arabidopsis thaliana*. *EMBO Rep* 7:100–105.
7. Maloisel L, Rossignol JL (1998) Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev* 12:1381–1389.
8. Shi J, et al. (2010) Widespread gene conversion in centromere cores. *PLoS Biol* 8:e1000327.
9. Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. *Proc Natl Acad Sci USA* 109:E981–E988.
10. Mirouze M, et al. (2012) Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. *Proc Natl Acad Sci USA* 109:5880–5885.
11. Teixeira FK, et al. (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* 323:1600–1604.
12. Johannes F, et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* 5:e1000530.
13. Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* 260:1926–1928.
14. Kakutani T, Munakata K, Richards EJ, Hirochika H (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of *Arabidopsis thaliana*. *Genetics* 151:831–838.
15. Tsukahara S, et al. (2009) Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* 461:423–426.
16. Lippman Z, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476.
17. Johannes F, Colomé-Tatché M (2011) Quantitative epigenetics through epigenomic perturbation of isogenic lines. *Genetics* 188:215–227.
18. Cortijo S, Wardenaar R, Colomé-Tatché M, Johannes F, Colot V (2012) Genome-wide analysis of DNA methylation in Arabidopsis using MeDIP-chip. *Plant Epigenome: Understanding and Analysis*, eds McKeown P, Spillane C (Humana Press/Springer, New York).
19. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* 3:e3156.
20. Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 84:2363–2367.
21. Giraut L, et al. (2011) Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. *PLoS Genet* 7:e1002354.
22. Drouaud J, et al. (2007) Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. *PLoS Genet* 3:e106.
23. Drouaud J, et al. (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots" *Genome Res* 16:106–114.
24. Salomé PA, et al. (2012) The recombination landscape in *Arabidopsis thaliana* F2 populations. *Heredity (Edinb)* 108:447–455.
25. Horton MW, et al. (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* 44:212–216.
26. Pan J, et al. (2011) A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144:719–731.
27. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV (2012) Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485:642–645.
28. Schmitz RJ, et al. (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334:369–373.
29. Becker C, et al. (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480:245–249.
30. Richards EJ (2008) Population epigenetics. *Curr Opin Genet Dev* 18:221–226.
31. Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: A quantitative genetics perspective. *Nat Rev Genet* 9:883–890.
32. Richards EJ (2009) Quantitative epigenetics: DNA sequence variation need not apply. Genes Dev 15;23(14):1601–1605.
33. Weigel D (2012) Natural variation in *Arabidopsis thaliana*: From molecular genetics to ecological genomics. *Plant Physiol* 158:2–22.
34. Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends Plant Sci* 17:149–154.
35. Roux F, et al. (2011) Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics* 188:1015–1017.
36. Latzel V, Zhang Y, Karlsson Moritz K, Fischer M, Bossdorf O (2012) Epigenetic variation in plant responses to defense hormones. *Ann Bot (Lond)*, 10.1093/aob/mcs088.
37. Reinders J, et al. (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev* 23:939–950.
38. Roudier F, et al. (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J* 30:1928–1938.

Annexe 4 : Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. (Fu et al., 2013)

The EMBO Journal (2013) 32, 2407–2417
www.embojournal.org

SOURCE DATA | TRANSPARENT PROCESS

THE EMBO JOURNAL

Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor

Yu Fu[1,2], Akira Kawabe[3], Mathilde Etcheverry[4], Tasuku Ito[1,5], Atsushi Toyoda[6], Asao Fujiyama[6], Vincent Colot[4], Yoshiaki Tarutani[1,2] and Tetsuji Kakutani[1,2,5,*]

[1]Department of Integrated Genetics, National Institute of Genetics, Shizuoka, Japan, [2]Department of Genetics, School of Life science, The Graduate University for Advanced Studies (SOKENDAI), Shizuoka, Japan, [3]Department of Bioresource and Environmental Sciences, Faculty of Life Sciences, Kyoto Sangyo University, Kyoto, Japan, [4]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Centre National de la Recherche Scientifique (CNRS), UMR 8197, Institut national de la santé et de la recherche médicale (INSERM), U1024, Paris, France, [5]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Hongo, Tokyo, Japan and [6]Center for Genetic Resource Information, National Institute of Genetics, Shizuoka, Japan

Transposable elements (TEs) have a major impact on genome evolution, but they are potentially deleterious, and most of them are silenced by epigenetic mechanisms, such as DNA methylation. Here, we report the characterization of a TE encoding an activity to counteract epigenetic silencing by the host. In *Arabidopsis thaliana*, we identified a mobile copy of the *Mutator*-like element (*MULE*) with degenerated terminal inverted repeats (TIRs). This TE, named *Hiun* (*Hi*), is silent in wild-type plants, but it transposes when DNA methylation is abolished. When a *Hi* transgene was introduced into the wild-type background, it induced excision of the endogenous *Hi* copy, suggesting that *Hi* is the autonomously mobile copy. In addition, the transgene induced loss of DNA methylation and transcriptional activation of the endogenous *Hi*. Most importantly, the *trans*-activation of *Hi* depends on a *Hi*-encoded protein different from the conserved transposase. Proteins related to this anti-silencing factor, which we named VANC, are widespread in the non-TIR *MULE*s and may have contributed to the recent success of these TEs in natural *Arabidopsis* populations.

The EMBO Journal (2013) **32,** 2407–2417. doi:10.1038/emboj.2013.169; Published online 30 July 2013
Subject Categories: chromatin & transcription; plant biology
Keywords: DNA methylation; epigenetics; evolution; transposon

Introduction

Control of transposable elements (TEs) has been extensively studied in plants. A pioneering early observation in maize is that TE activity often changes between active and inactive states in heritable but reversible manners (McClintock, 1951, 1958). The changes in the TE activity are generally correlated with the DNA methylation status; inactive TEs tend to be more methylated than active TEs (Chandler and Walbot, 1986; Brettell and Dennis, 1991; Fedoroff, 1996; Martienssen, 1996). The importance of DNA methylation in TE control has also been demonstrated using mutants of Arabidopsis; in Arabidopsis mutants with reduced genomic DNA methylation, a variety of silent TEs are de-repressed and mobilized (Miura *et al*, 2001; Singer *et al*, 2001; Kato *et al*, 2003; Lippman *et al*, 2004; Mirouze *et al*, 2009; Tsukahara *et al*, 2009).

Intriguingly, some TEs have mechanisms to counteract DNA methylation and silencing by the host. For example, McClintock's *Suppressor-mutator* (*Spm*) element in maize encodes a protein TnpA, which induces loss of DNA methylation in regions controlling transcript formation in *Spm* (Schläppi *et al*, 1994, 1996; Cui and Fedoroff, 2002). An active *Spm* transiently activates silent *Spm* copies *in trans*, a process likely to be mediated by TnpA (Cui and Fedoroff, 2002).

Robertson's *Mutator*, another well-characterized TE in maize, spontaneously changes its activity and DNA methylation in a coordinated manner (Chandler and Walbot, 1986; Martienssen and Baron, 1994; Martienssen, 1996). Like *Spm*, a silent *Mutator* element loses DNA methylation when an active *Mutator* is present in the same genome (Brown and Sundaresan, 1992; Lisch *et al*, 1995, 1999). *MuDR*, an autonomously mobile copy of maize *Mutator* family, contains two genes, *mudrA* and *mudrB*. The *mudrA* encodes the MURA protein, which is structurally similar to known transposases of other TEs (Eisen *et al*, 1994; Lisch, 2002). In addition, *mudrA* is sufficient for excision of *Mutator*, further suggesting that MURA functions as a transposase (Lisch *et al*, 1999). TEs similar to the maize *Mutator* are widespread in eukaryotes and they are referred to as *Mutator*-like elements (*MULE*s) (Jiang *et al*, 2004). ORFs related to *mudrA* are generally found in autonomous *MULE*s. Some of the autonomous *MULE*s also have additional ORF(s), such as *mudrB* in *MuDR*, but the structures of the proteins encoded in these ORFs are diverse and their functions remain largely unknown. In addition, some of the *MULE*s carry fragments of cellular genes, but their impacts on the TE dynamics and host fitness are still elusive (Talbert and Chandler, 1988; Yu *et al*, 2000; Jiang *et al*, 2004; Hoen *et al*, 2006).

Most of the class II (DNA-type) TEs have long terminal inverted repeat (TIR), but there are a few exceptions (Wicker, 2007). Although the maize *Mutator* elements have relatively long TIRs of almost identical sequences, subgroups of *MULE*s with extensively degenerated TIR have been found in the Arabidopsis genome and they are classified as non-TIR *MULE*s (Le *et al*, 2000; Yu *et al*, 2000). Although the theoretical sequence analyses of Arabidopsis genome suggest movement of these non-TIR *MULE*s in the past (Yu *et al*, 2000), direct evidence for *de novo* movements is limited (Hoen *et al*, 2006; Tsukahara *et al*, 2009).

We have previously reported that a group of non-TIR *MULE*s, called *VANDAL21*, seem to transpose in a back-

*Corresponding author. Department of Integrated Genetics, National Institute of Genetics, Yata 1111, Mishima, Shizuoka, Japan. Tel.: +81 55 981 6801; Fax: +81 55 981 6804; E-mail: tkakutan@lab.nig.ac.jp

Received: 14 March 2013; accepted: 4 July 2013; published online: 30 July 2013

©2013 European Molecular Biology Organization

ground of reduced genomic DNA methylation (Tsukahara *et al*, 2009). Here, we identified an autonomously mobile copy of *VANDAL21*, which we renamed *Hiun* (*Hi*). Despite the degeneration of TIRs, *Hi* is competent to excise and transpose in precise manners. Interestingly, a *Hi* transgene induced loss of DNA methylation, transcriptional activation, and excision of the endogenous *Hi* copy. Most importantly, these *trans*-acting effects of *Hi* do not depend on the protein related to MURA-type transposase but instead depend on another protein encoded by *Hi*. The function of this novel anti-silencing protein, which we named VANC, will be discussed in the context of TE evolution.

Results

Identification of mobile VANDAL21 copies

The genome sequence of wild-type Col (http://www.arabidopsis.org/) suggests seven copies of *VANDAL21* elements with relatively similar sequences. Other copies of *VANDAL21* are distant in the sequences. Consistent with that, Southern analysis revealed seven bands for that group of *VANDAL21* (Tsukahara *et al*, 2009). We have previously shown that additional bands emerge in Arabidopsis plants derived from several rounds of self-pollinations in *ddm1* (*decrease in DNA methylation 1*) mutant backgrounds, suggesting mobility of one or more copies of the *VANDAL21* members (Tsukahara *et al*, 2009). Arabidopsis *ddm1* mutation generally induces loss of DNA methylation in TEs, which causes mobilization of diverse TEs (Miura *et al*, 2001; Singer *et al*, 2001; Lippman *et al*, 2004; Mirouze *et al*, 2009; Tsukahara *et al*, 2009). In order to know which of *VANDAL21* copies are mobile, we used two methods: suppression PCR and whole-genome re-sequencing (see Materials and methods for details).

In total, we identified 72 *de novo* insertions of *VANDAL21*s (61 by genome re-sequencing and 14 by suppression PCR with 3 overlaps) in the self-pollinated *ddm1* lines (Supplementary Table S1). Of these 72 insertions, 69 correspond to one copy (Figure 1A; *AT2TE42810*) of *VANDAL21* element. The remain-

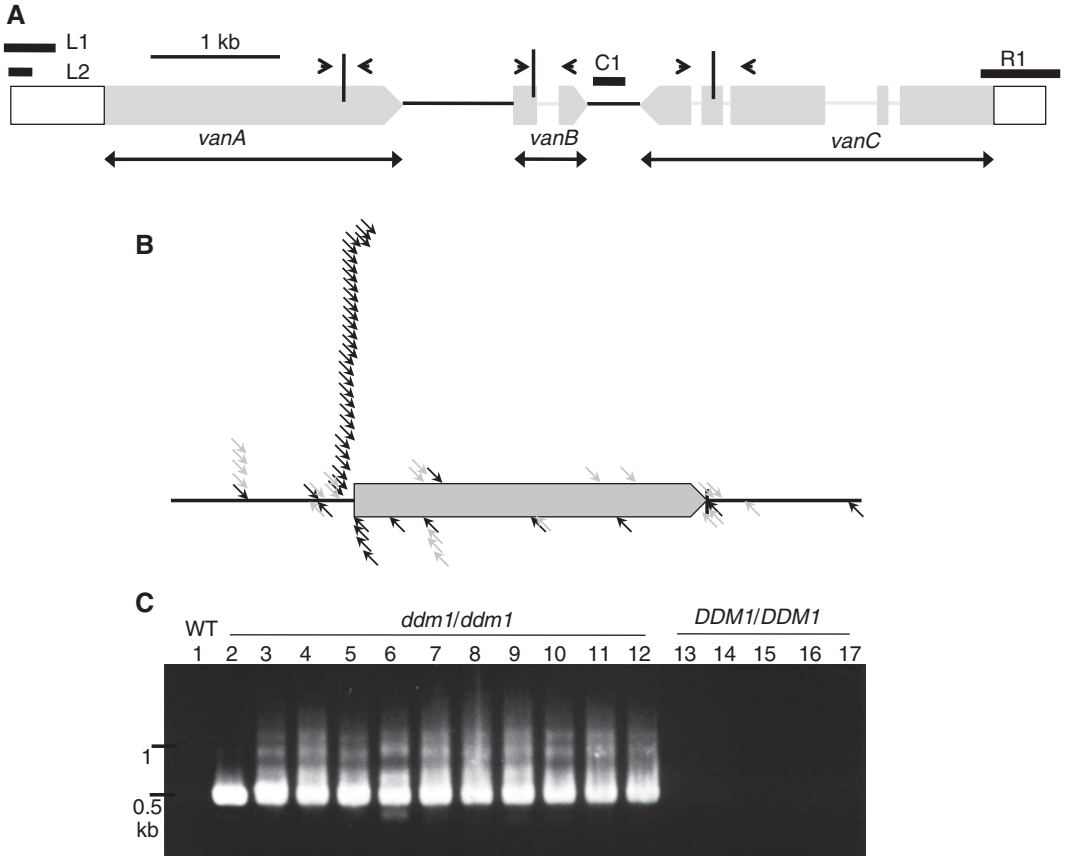


Figure 1 Mobilization of *Hi* in *ddm1* mutant. (**A**) Schematic diagram for structure of *Hi*. Terminal regions, exons, introns, and intergenic regions are shown by white bars, grey bars, grey lines, and black lines, respectively. Regions examined by bisulphite sequencing are shown by L1 and R1 with thick black lines. Region examined by McrBC-qPCR is shown by L2. Region examined for copy number quantification is shown by C1. In most of the transgene constructs, silent mutation is introduced for each ORF, so that the transcripts from the transgene and endogenous copy could be distinguished between. The sites of the silent mutations are shown by vertical bars, with surrounding arrowheads showing regions amplified by RT–PCR. Regions for ORFs deleted in each of the deletion constructs are shown by horizontal bars with two arrowheads. (**B**) *De novo* integration sites of *Hi* in relation to flanking transcription units. The position of integration is normalized by length of the flanking transcription unit. Rightward and leftward arrows indicate insertions with 5′ to 3′ and 3′ to 5′ orientations of *Hi*, respectively. Insertions flanking pseudogenes and transposon genes are shown by grey arrows, and those flanking canonical genes by black arrows. Sequences of the integration sites are shown in Supplementary Table S1. Four out of the sixty-nine insertions are not included in this figure, because they are further away from transcription units. Genomic locations of all 69 transpositions are shown in Supplementary Figure S2. (**C**) Excision of *Hi* in *ddm1* plants detected by PCR. Genomic DNA of 11 *ddm1* plants (lane numbers from 2 to 12) and 5 wild-type sibling plants (lane numbers from 13 to 17) was used to analyse excision of endogenous *Hi* by nested PCR. These lines are derived from segregating population in self-pollinated progeny of a *DDM1/ddm1-1* heterozygote. Sequences of primers used are shown in Supplementary Table S2. Source data for this figure is available on the online supplementary information page.

 ©2013 European Molecular Biology Organization

ing three insertions correspond to another copy (*AT4TE15615*). For the other five copies of related *VANDAL21*, no new insertion has been identified. In the following parts, we concentrate on the most active copy, *AT2TE42810*, which we renamed *Hiun* (*Hi*, Japanese for 'a flying cloud').

Structure of Hi

Hi is 8177-bp long and includes three ORFs: *At2g23500*, *At2g23490*, and *At2g23480* (Figure 1A). One ORF (*At2g23500*; called *vanA*) encodes a protein with high sequence similarities with MURA-type transposases, which are generally found in *MULE*s. Proteins encoded by two other ORFs (*vanB* and *vanC*) do not have sequence similarity to any characterized proteins. An unorthodox feature is that, unlike other typical mobile DNA-type TEs, the TIRs of this TE are extensively degenerated (Supplementary Figure S1A), showing the characteristics of non-TIR *MULE*s (Yu *et al*, 2000).

Integration and excision of Hi

Hi is transposed throughout the genome, although insertions may be more concentrated near the original locus than in unlinked regions (Supplementary Figure S2). *Mutator* elements in maize preferentially transpose into 5′ region of genes (Hardeman and Chandler, 1989; Dietrich *et al*, 2002; Liu *et al*, 2009). That was also the case for *Hi*; most of the integration sites are localized around transcription start sites of genes (Figure 1B). Interestingly, integration of *Hi* there had bias in the orientation (Figure 1B). Such a bias in the orientation has not been reported for the maize *Mutator* (Brown *et al*, 1989). The bias in the orientation of *Hi* integration might be related to the asymmetry in its terminal sequences. We could detect 9 bp of target site duplication (TSD) for most of the insertions examined (Supplementary Figure S1B), as is the case for integrations of other *MULE*s (Yu *et al*, 2000).

Many DNA-type TEs often transpose with loss of the original copy. Because *Hi* does not have typical structure of DNA-type TE, an interesting question would be whether *Hi* excision occurs at defined termini or not. In order to detect somatic excision events, we used PCR with primers for both of the flanking regions of the original *Hi* locus (details in Materials and methods). Using this assay, we could detect *Hi* excision in all independent *ddm1* lines examined (Figure 1C). We examined the mode of the excisions by sequencing the PCR products (Supplementary Figure S3A). Interestingly, despite the degeneration of TIRs, many of the excision products showed excision around the terminal sites predicted from the integrated copies. Even TSDs are lost in a significant part of the excision product.

In summary, these observations suggest that long perfect TIRs are dispensable not only for integration, but also for reasonably precise excision of this element in the defined termini. The transposition of *Hi* occurred in a manner comparable to that of typical *MULE*s with long TIR.

Mobilization of endogenous Hi by transgene

As the *ddm1* mutation results in transcriptional de-repression of many repeat sequences (Lippmann *et al*, 2004; Tsukahara *et al*, 2009), it is formally possible that mobilization of *Hi* in *ddm1* is triggered by de-repression of other sequence(s). In order to test if expression of proteins encoded by *Hi* is sufficient for its mobilization, we introduced a cloned *Hi* copy into wild-type plants by Agrobacterium-mediated transformation. In these transgenic *Hi* lines, we could detect transcripts corresponding to their three ORFs (Supplementary Figure S4). In these lines, the *Hi* transgene induced excision of the original copy (Figure 2A).

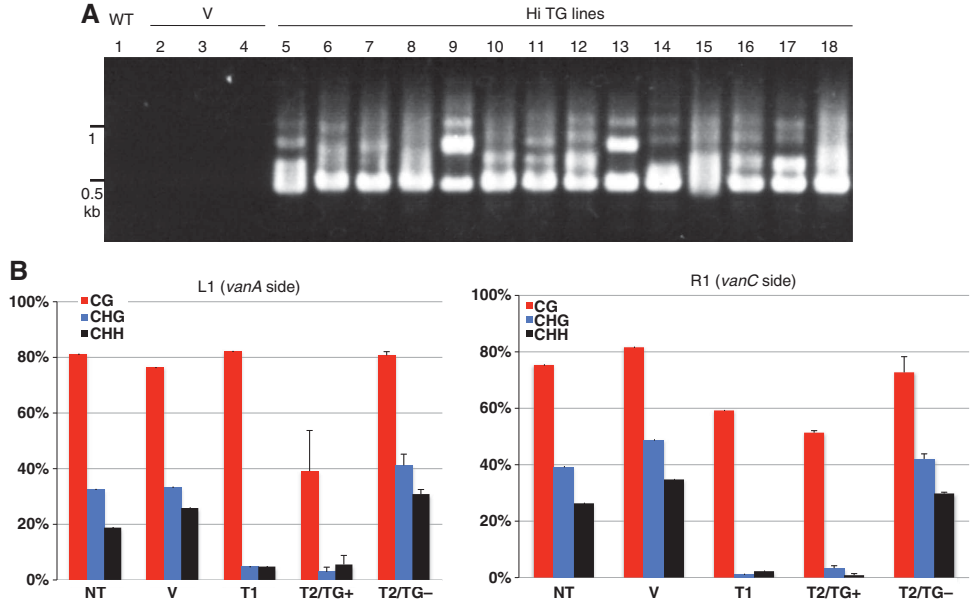


Figure 2 Introduction of *Hi* transgene induces loss of DNA methylation and excision of endogenous *Hi* copy. (**A**) Excision of endogenous *Hi* induced by transgene for *Hi* (*Hi* TG: lanes 5–18). Lanes 1 and 2–4 are non-transgenic plant (wt) and transformant lines with empty vector (V) used as negative controls, respectively. Excision of *Hi* copy in the transgene was also detected in some of the transgenic lines (Supplementary Figure S5). (**B**) DNA methylation status of *Hi* termini in the transgenic line and progeny. T1 transformant with *Hi* transgene showed reduction in DNA methylation in both termini, compared to non-transgenic plant (NT) and transformants with empty vector (V). T2/TG + and T2/TG − are self-pollinated progeny of the T1 with and without transgene, respectively. In both classes, averages and standard deviations of three segregants are shown. We also obtained essentially the same results for a segregating T3 family (Supplementary Figure S7). Regions L1 (upstream of *vanA*) and R1 (upstream of *vanC*) were examined (shown in Figure 1). At least 11 clones were examined for each plant. Source data for this figure is available on the online supplementary information page.

©2013 European Molecular Biology Organization

Control transformants with empty vectors did not show excision of the *Hi*, confirming that *Hi* transgene triggered mobilization of the endogenous copy. We also examined *Hi* activity in self-pollinated progeny of one of the *Hi* transgenic lines. We could detect *Hi* excision in most of 15 progeny plants that inherited the transgene (Supplementary Figure S6). On the other hand, we could not detect the excision in any of six progeny plants that lost the transgene by segregation, further confirming that the *Hi* transgene is responsible for the mobilization.

DNA demethylation induced by Hi in trans

The observations described above suggest that expression of *Hi* transgene induces mobilization of the endogenous copy. Interestingly, in the presence of the transgene, DNA methylation levels were reduced in both termini of the endogenous *Hi* (Figure 2B). The loss of methylation is more extensive in non-CpG sites than in CpG sites. When the transgene was segregated away in the self-pollinated progeny of the transgenic line (T2/TG − in Figure 2B), *Hi* was remethylated in the terminal regions, which was associated with its loss of excision activity.

We also tested the *trans*-acting demethylation effect for transposed *Hi* copies, using epi-Recombinant Inbred Lines (epi-RILs) from *ddm1* mutant (Johannes *et al*, 2009). The epi-RILs are originated from *ddm1* mutant backcrossed to parental wild-type Col to generate *DDM1/DDM1* homozygotes, and subsequent self-pollination to fix the heritable epigenetic defects induced by *ddm1*. The methylation status of *Hi* at the original locus was examined by PCR after digestion by methylation-sensitive restriction enzyme. As expected from the crossing scheme used to generate the

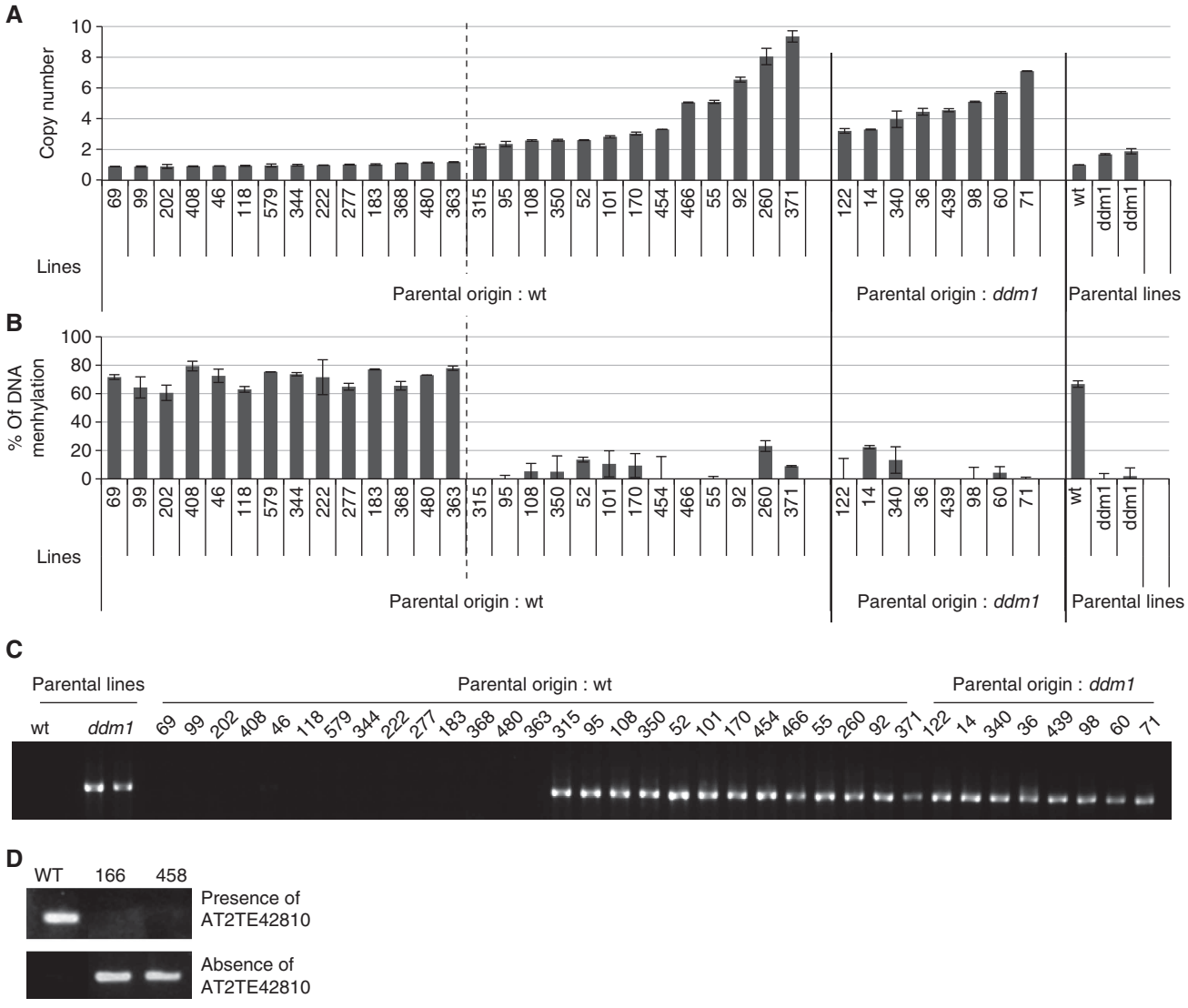


Figure 3 Transposed *Hi* induces loss of DNA methylation and excision of *Hi* in the original locus. (**A**) Copy number of *Hi* in each of the epi-RILs. The copy number was estimated by quantitative PCR using region C1 in Figure 1. Average and standard deviation of two technical replicates are shown in this and next panel. Parental origin of the original *Hi* locus (wt or *ddm1*) was determined by methylation status of the linked region (Colomé-Tatché *et al*, 2012). (**B**) DNA methylation status in the 5′ region (L2 in Figure 1) of the original *Hi* locus was estimated by McrBC digestion and subsequent qPCR. Details are described in Materials and methods. *Hi* in the original locus showed loss of DNA methylation when extra copies of *Hi* exist. (**C**) Excision analysis of original *Hi* copy by nested PCR. (**D**) Two of the epi-RILs showed germinal transmission of the excised *Hi* allele. Origin of the *Hi* locus is wild-type *DDM1* for line 166 and *ddm1* mutant for line 458. Presence of *Hi* in the original locus was examined by PCR in the 5′ border of *Hi* (the primer sequences are shown in Supplementary Table S2). Lack of the signal suggests fixation of the empty allele. Source data for this figure is available on the online supplementary information page.

 ©2013 European Molecular Biology Organization

epi-RILs (Johannes *et al*, 2009), approximately three quarters of the epi-RILs tested had inherited the original *Hi* copy from wild-type *DDM1* parent. Some of these lines have additional transposed *Hi* copies *in trans* (Figure 3A). In those lines, *Hi* at the original locus showed loss of methylation (Figure 3B), despite its wild-type origin. On the other hand, in epi-RILs that did not carry the additional *Hi*, *Hi* at the original locus remained methylated to the level comparable to parental wild type. Together, these results suggest that the stable demethylation of *Hi* at the original locus is due to *trans*-acting effect of the transposed *Hi* copies. In addition to its demethylation, the *Hi* copies present at the original locus showed excision (Figure 3C), and that happened only when the additional *trans*-acting copies exist. These *trans*-acting effects of the transposed *Hi* are consistent with the *trans*-acting demethylation and mobilization by the *Hi* transgene.

Expression of vanC is sufficient for the trans-activation and mobilization of endogenous Hi

The results shown above demonstrate that *Hi* transgenes induced mobilization of endogenous copy *in trans*, which is associated with loss of DNA methylation in the terminal regions. In order to further dissect the role for each of the ORFs in *Hi*, we generated transgenes with deletion in each ORF. Transgene with deletion of the central small ORF (*vanB*) still caused loss of methylation in both termini of *Hi* (Figure 4). By contrast, deletion of 3′ ORF (*vanC*) abolished the demethylation effect for both termini, suggesting that this ORF is essential for the demethylation. Transgene with deletion of 5′ ORF (*vanA*), which is structurally similar to transposase, still caused loss of methylation in 3′ (*vanC* side) terminal region of *Hi*, although the demethylation effect in the 5′ (*vanA* side) region was less complete than that in the full-length *Hi* transgenic lines. These results suggest that *vanC* is important for the demethylation of both of the terminal regions.

We then examined the effect of these deletion constructs on excision of endogenous copies (Figure 5A). Very importantly, transgenic lines with deletion in *vanA* (Δ*A*-TG) still induced excision of endogenous *Hi* (Figure 5A). This was surprising because *vanA* encodes the putative transposase. We then examined expression of endogenous *vanA* gene in the presence of Δ*A*-TG. In most of the Δ*A*-TG lines, endogenous *vanA* was de-repressed and transcribed, although the expression level was generally lower and less robust than that of the *Hi* transgene keeping *vanA* (Figure 5B; Supplementary Figure S8).

The results above suggest that expression of *vanB* and/or *vanC* can cause de-repression of *vanA*. We could also detect expression of *vanB* in Δ*B*-TG lines, but *vanC* transcript was undetectable in Δ*C*-TG lines (Supplementary Figure S8), suggesting possible role of *vanC* in the *trans*-activation. In addition, deletion of *vanB* from the transgene did not affect excision of endogenous *Hi* by the transgene, while the excision tended to be less robust in Δ*C*-TG (Figure 5A; Supplementary Figure S9). In order to know if *vanB* is dispensable for the *trans*-activation of *Hi*, we examined the effect of transgene with deletion of both *vanA* and *vanB* (Δ*AB*-TG). Only the *vanC* ORF remains in the Δ*AB*-TG construct. The Δ*AB*-TG also induced demethylation of 3′ (*vanC* side) terminal regions to the level comparable to the Δ*A*-TG (Figure 4). In most of the Δ*AB*-TG lines, we could detect excision of *Hi* and transcription of *vanA* and *vanB* (Figure 6A and B). In T2 generation that originated from self-pollination of a T1 Δ*AB*-TG plant, all T2 plants with the transgene showed excision, but none of T2 plants without transgene showed excision, confirming that the Δ*AB* transgene induces the excisions of endogenous *Hi* (Figure 6C). Taken together, these results demonstrate the key role of *vanC* for the *trans*-acting anti-silencing of *Hi*.

Proteins related to the anti-silencing factor are widespread in non-TIR MULEs but not found in TIR MULEs

Non-TIR *MULE*s in *A. thaliana* genome are consisted of multiple *VANDAL* and *ARNOLD* families (Yu *et al*, 2000; Figure 7A). Interestingly, the non-TIR *MULE*s seem to be very successful in the recent proliferation. The phylogenetic analyses revealed recent proliferations in multiple subfamilies of non-TIR *MULE*s; each of the subfamilies shows terminal proliferations after separation of *A. thaliana* and *A. lyrata* lineages. The proliferation rates of the non-TIR *MULE* clusters are significantly higher than those of TIR-*MULE* in both *A. thaliana* and *A. lyrata* lineages (Figure 7A; Table I).

Notably, all major branches of these proliferating non-TIR *MULE*s contain TEs with ORFs encoding proteins similar to the anti-silencing factor VANC (protein product of

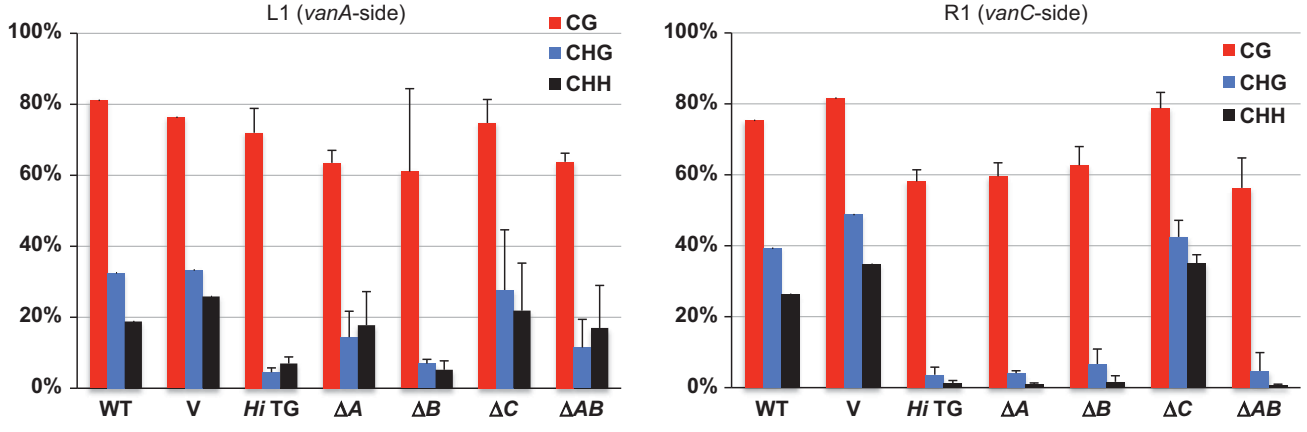


Figure 4 DNA methylation status of endogenous *Hi* after introduction of *Hi* transgene and its deletion derivatives. For each of the deletion derivatives, averages and standard deviations of four independent transgenic plant lines are shown. WT and V are from Figure 2.

©2013 European Molecular Biology Organization

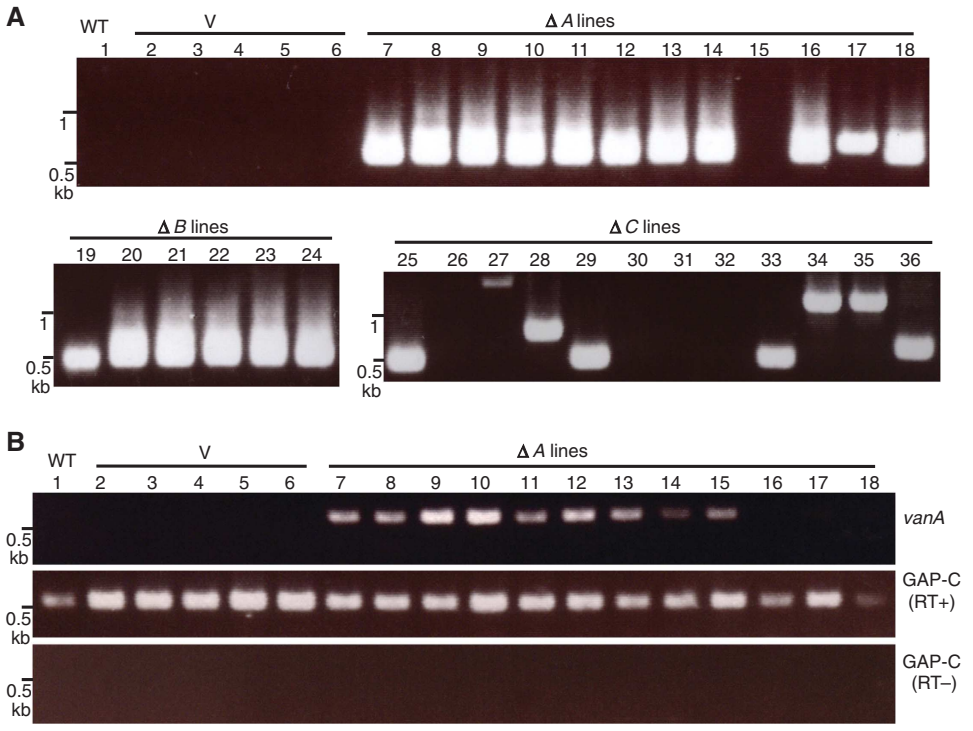


Figure 5 *Trans*-activation by *Hi* transgene without putative transposase. (**A**) Excision of endogenous *Hi* induced by Δ*A* transgene. Lanes 1 and 2–6 are non-transgenic plant (wt) and transformant lines with empty vector (V) used as negative controls, respectively. Excision of endogenous *Hi* induced by Δ*B* and Δ*C* transgene is also shown below. The results using additional Δ*C* lines are shown in Supplementary Figure S9. (**B**) Transcriptional activation of *vanA* induced by Δ*A* transgene. Materials for the same lane number in (**A**) and (**B**) are from the same plant, although the DNA and RNA are prepared from different leaves. Source data for this figure is available on the online supplementary information page.

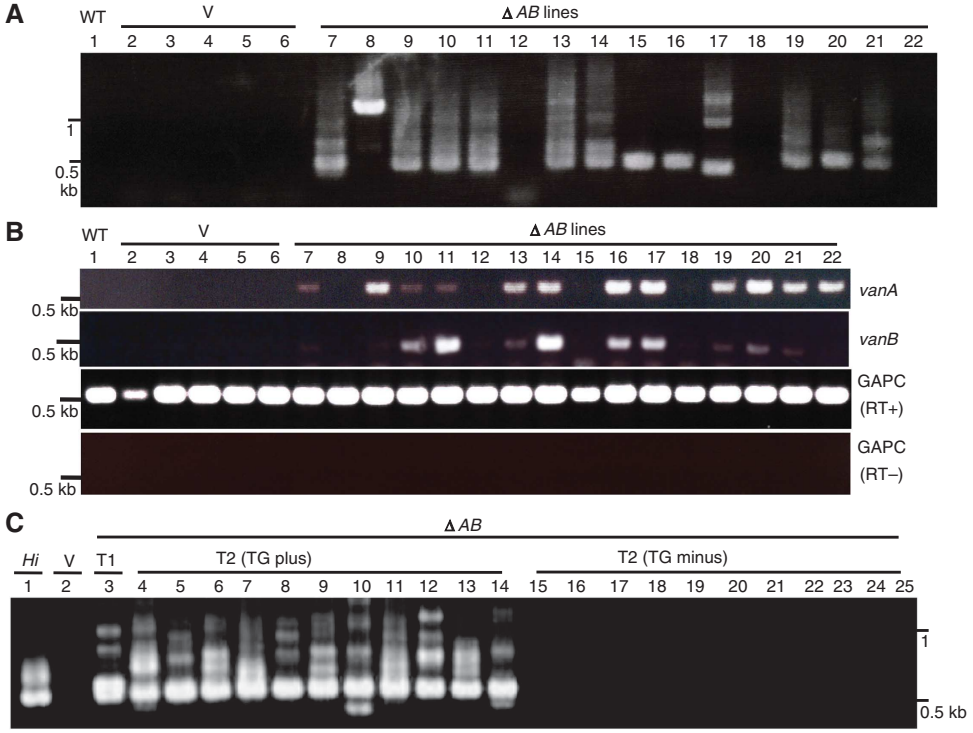


Figure 6 Expression of *vanC* is sufficient for the *trans*-activation. (**A**) Excision of endogenous *Hi* induced by Δ*AB* transgene. (**B**) Transcriptional activation of *vanA* and *vanB* genes induced by Δ*AB* transgene. Materials for the same lane number in (**A**) and (**B**) are from the same plant, although the DNA and RNA are prepared from different leaves. (**C**) Excision of endogenous *Hi* induced by Δ*AB* transgene in the T2 generation. T2 plants from self-pollinated progeny of a T1 (the plant shown in lane 9 of **A**) were examined after determining the presence/absence of the transgene. Source data for this figure is available on the online supplementary information page.

 ©2013 European Molecular Biology Organization

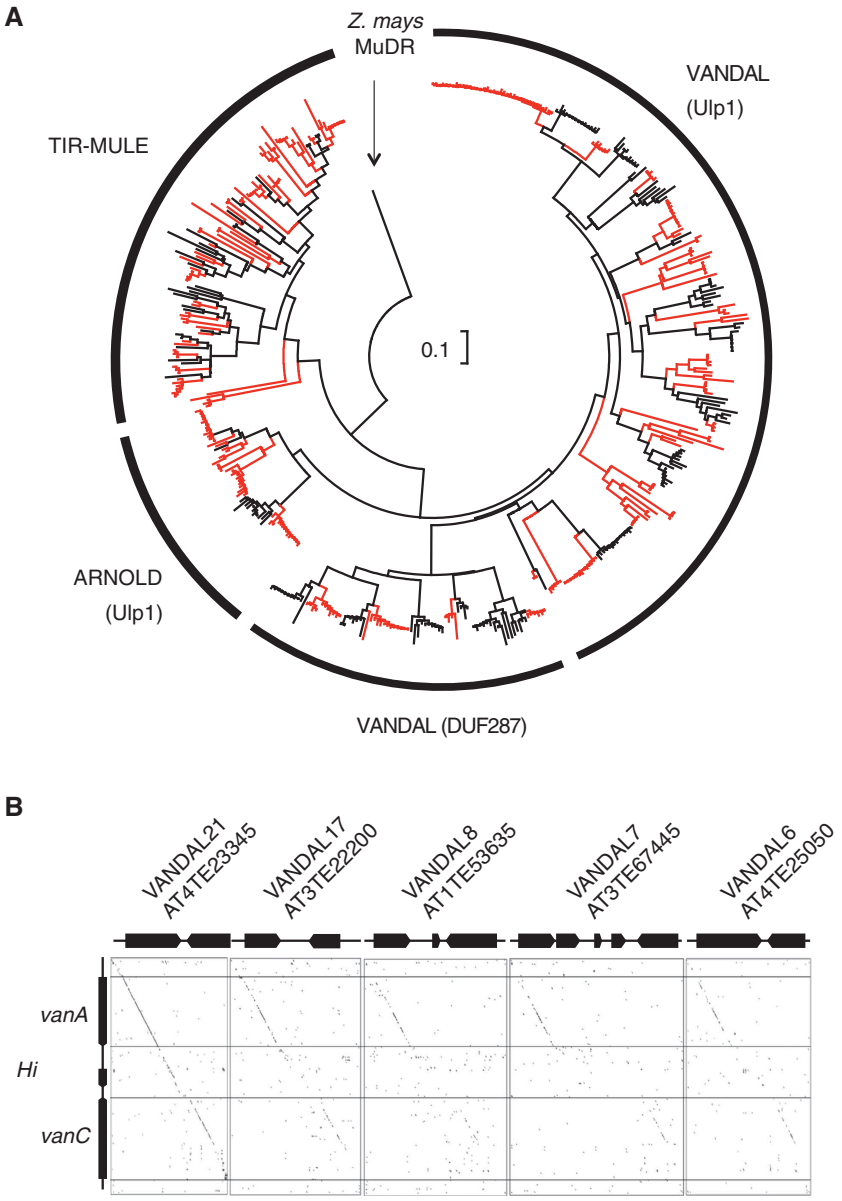


Figure 7 Evolution and proliferation of TIR and non-TIR *MULE* families. (**A**) Phylogenetic relationship among *MULE* families in genomes of *A. thaliana* and *A. lyrata*. *A. lyrata*-specific lineages are shown by red lines. An NJ tree made by *p*-distance is shown. Scale bar is shown in the centre of the tree. The families containing DUF287 or Ulp1 protease domain are indicated in the parentheses. Names of *VANDAL* members are shown in phylogenetic tree in Supplementary Figure S10. (**B**) Dot-plot (Harr-plot) analyses among *VANDAL* families. Regions with nucleotide identities of 15 out of 20 or more are shown by dots. Copies with typical structure were chosen from *VANDAL21*, *17*, *8*, *7* and *6* families. From *VANDAL21* family, one sequence each from two clusters was used. Coding regions are indicated by pointed thick lines.

vanC). All the VANC-related proteins contain the domain DUF1985 (domain of unknown function; Supplementary Figure S11). Some of them, including VANC, also contain the domain DUF287 (Figure 7; Supplementary Figure S11). Other *VANDAL* members, as well as *ARNOLD* members, have ORFs encoding Ulp1 proteases (Figure 7A; Supplementary Figure S11). Both the Ulp1 and DUF287 types of proteins contain DUF1985, generating a wide distribution of VANC-related proteins. In contrast, such VANC-related proteins were not found in any of the TIR-*MULE* members. Therefore, association between the VANC-related proteins and the absence of long TIRs is very tight. The non-TIR *MULE*s seem to keep VANC-related proteins during evolution.

While transposases of the MURA class are generally found in autonomous *MULE*s, other ORFs in *MULE*s tend to be diverse. Consistent with that, nucleotide divergence is higher in *vanC*-related genes than in *vanA*-related genes, especially between distant groups (Figures 7B and 8). However, that does not seem to be due to weaker constraint on the amino-acid sequence of the VANC-related proteins, because non-synonymous mutation rates are comparable between both families of genes and the overall high divergence seems to reflect the divergence in synonymous sites (Figure 8). The conserved amino-acid sequences in the VANC-related proteins suggest advantage in their functions, at least in short term.

Genome-wide effects of Hi transgene on DNA methylation

Hi transgene affects endogenous *Hi* sequence. We then examined the effect of *Hi* transgene genome-wide. DNA methy-

©2013 European Molecular Biology Organization

lation was examined, because that could be accessed relatively precisely using whole-genome bisulphite sequencing. In the whole-genome bisulphite sequencing, deep sequencing reads are mapped to the reference genome sequence. In order to reduce possible noise due to repetitive nature of TEs, we used only uniquely mapped reads; reads mapped to multiple loci were not used for the analyses. The analyses revealed that many of the *VANDAL21* members showed reduced DNA methylation in the transgenic line (Figure 9A and B; Supplementary Figures S12–S14; Supplementary Table S3). Interestingly, not only terminal regions but also internal regions were affected for both *Hi* and other *VANDAL21* copies (Figure 9B; Supplementary Figure S14). As is the case for the *Hi* sequence, non-CpG sites tend to be more affected than CpG sites. On the other hand, the demethylating effect was much less in the other TEs and genes (Figure 9A; Supplementary Figure S12), suggesting that the *trans*-acting effects of *Hi* are highly specific.

Table I Comparison of copy numbers per cluster between non-TIR and TIR *MULE*s

	#Seq[a]	#Cluster[b]	Seq/cluster	*t*
A. thaliana				
TIR *MULE*s	27	24 (21)	1.125	4.42
Non-TIR *MULE*s	143	34 (13)	4.206	$P<0.001$
A. lyrata				
TIR *MULE*s	85	39 (21)	2.179	2.38
Non-TIR *MULE*s	219	49 (20)	4.347	$P<0.02$

Analyses were performed using transposase genes. Therefore, copies without the conserved transposase gene are not included in the analyses. Details for the analyses are described in Materials and methods.
[a]Number of sequences analysed.
[b]Number of cluster with <0.1 *p*-distances. In the parentheses, number of clusters with only one sequence is shown.

Discussion

Trans-activation as a counteraction against genome defense by the host

Here, we report identification of an autonomously mobile copy of *MULE* that lacks long TIRs. Among three ORFs *Hi* contains, *vanA* encodes for a putative transposase conserved among other TEs. Most importantly, *vanC*, another ORF of *Hi*, plays a key role in the transcriptional activation of the other ORFs in *Hi*, and in mobilization of *Hi*. These *trans*-acting effects represent efficient means for counteracting DNA methylation and silencing by the host.

Importantly, the anti-silencing mechanism of *Hi* seems to function specifically, rather than globally (Figure 9). In contrast, some viruses have evolved mechanisms to globally inhibit RNA interference (RNAi)/post-transcriptional gene silencing (PTGS). RNAi/PTGS is an important mechanism for defense against viruses, and the global anti-silencing would facilitate proliferation of the viruses (Zamore, 2004; Bivalkar-Mehla *et al*, 2011). In contrast to the global anti-silencing mechanisms of these viruses, *Hi* seems to target related TEs specifically. The specific (rather than global) anti-silencing can be reasonable for a TE. If a TE has global anti-silencing mechanisms, such as genome-wide demethylation, then that would activate diverse TEs, reducing fitness of the host (Lippman *et al*, 2004; Tsukahara *et al*, 2012). Reduction of host fitness would be deleterious for survival of the TE itself, given that the TE is an obligatory intra-genomic parasite/symbiont. The strategy of a TE should be different from that of viruses, which can propagate horizontally away from damaged hosts.

Mechanisms for the trans-acting effects

VANC induces both transcriptional activation and loss of DNA methylation *in trans*. Genome-wide analyses of transgenic plants revealed that the *trans*-acting demethylation

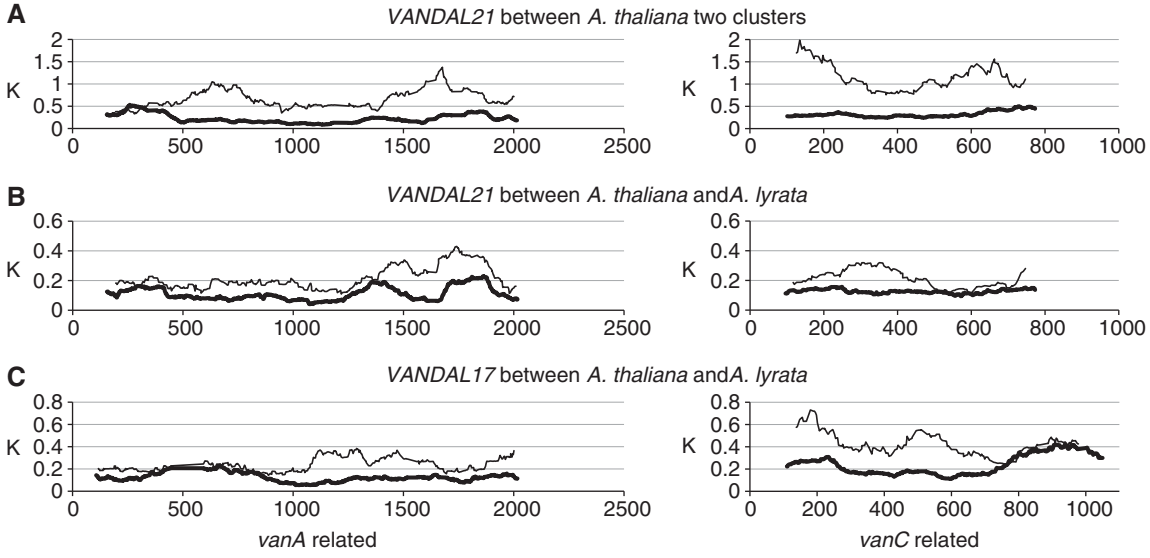


Figure 8 Sliding window plot analyses of *VANDAL21* and the most related family *VANDAL17*. (**A**) Divergence between two clusters (seven copies and three copies) of *VANDAL21* family in *A. thaliana*. (**B**) Divergence between *VANDAL21* copies in *A. thaliana* (three copies, which are orthologous to the *A. lyrata* copies) and *A. lyrata* (six copies). (**C**) Divergence between *VANDAL17* family in *A. thaliana* (four copies) and *A. lyrata* (three copies). Level of divergence for synonymous and non-synonymous sites in a 200-bp window was plotted in 1 bp intervals. Only coding sequences of complete structures were used. Thin lines and thick lines indicate synonymous and non-synonymous divergences, respectively.

 ©2013 European Molecular Biology Organization

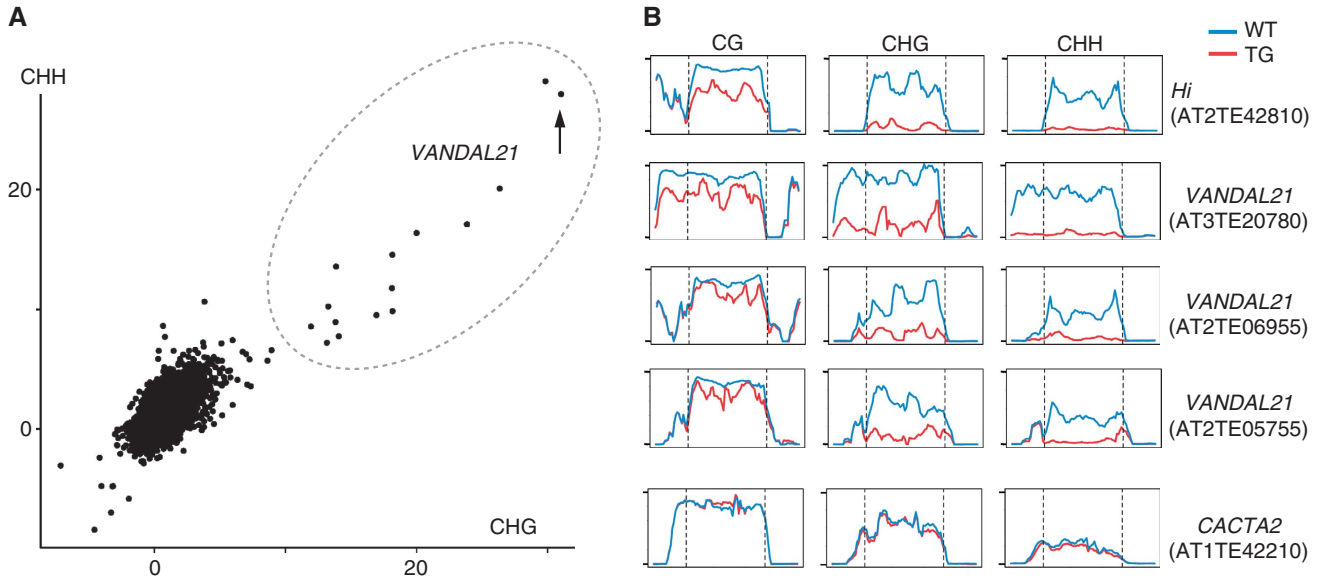


Figure 9 Demethylation of *VANDAL21* members induced by *Hi* transgene. (**A**) Effects of *Hi* transgene on DNA methylation in TEs. Changes in methylation at CpHpG sites and CpHpH sites are plotted. For each of them, significance of decrease in methylation was accessed by the value $(Mn/Cn - Mt/Ct)/(1/\sqrt{Cn} + 1/\sqrt{Ct})$, where *Mn*, *Cn*, *Mt* and *Ct* are methylated cytosine (*M*) and total cytosine (*C*) counts mapped for each TE in the non-transgenic (*n*) and transgenic (*t*) plants, respectively. Each of these values is shown in Supplementary Table S3, with the corresponding values for CpG sites. This figure contains 24 282 TEs plotted, while 6910 TEs are not plotted due to lack of mapped cytosine in one or more of the three contexts in non-transgenic or transgenic plant. The lack is mainly because we did not use reads mapped multiple loci. Some of the *VANDAL21* family members (surrounded by broken ellipse) showed most significant reduction for both CpHpG and CpHpH sites. The dot for *Hi* is indicated by an arrow. Similar analyses done on genes are shown in Supplementary Figure S12. (**B**) Effect of *Hi* transgene on DNA methylation across each of the *VANDAL21* members. Left and right terminals are shown by broken lines for each element. Each point represents proportion of methylated cytosine for a sliding window with seven fractions after separating each TE for 100 fractions. Right and left flanking regions are also analysed by the same conditions. Scale bars for CpG, CpHpG, and CpHpH sites in each panel indicate 1, 0.8, and 0.4, respectively. For some of the *VANDAL21* copies, reduced methylation in the terminal region is also confirmed by conventional bisulphite sequencing using primers within and flanking the TE (Supplementary Figure S13). Results of DNA methylation across six additional *VANDAL21* members are shown in Supplementary Figure S14. *CACTA2* is shown as a negative control.

occurs specifically in *Hi* and other *VANDAL21* members. How is the specific loss of methylation induced? In the case of *Tam3* transposon of snapdragon, loss of methylation is associated with binding of the transposase (Hashida *et al*, 2006). It would be informative to know localization of VANC protein in the genome.

Another possible pathway could be demethylation through transcription. Transcription can induce loss of methylation in lysine 9 of histone H3 (H3K9) through the function of histone demethylase IBM1 (Inagaki *et al*, 2010). The loss of H3K9 methylation induces loss of non-CpG methylation. That would account for the effect of *Hi* on non-CpG methylation in the internal regions of TEs (Figure 9; Supplementary Figure S14). Generally, DNA methylation in both CpG and non-CpG sites could be involved in silencing of TEs (Johnson *et al*, 2003; Kato *et al*, 2003). Considering that active/inactive states of TEs can be stabilized by positive feedback loops, possible interactions between transcription and DNA methylation at CpG and non-CpG sites would be important for understanding TE dynamics (Saze and Kakutani, 2011; Inagaki and Kakutani, 2013; further discussion in Supplementary Figure S15).

Role of VANC and loss of TIR

Non-TIR *MULE*s found within *A. thaliana* genome seem to be derived from ancestral TIR *MULE*s (Yu *et al*, 2000). Interestingly, evolution of non-TIR *MULE*s is tightly associated with the VANC-related proteins (Supplementary Figure S11). How has the transition from TIR type to non-TIR type of *MULE*s occurred? That could be due to loss of restraint to keep long TIRs, possibly reflecting evolution of transposition machinery. Despite degeneration of TIRs, most of the *Hi* excisions occurred in a manner to keep the original sequence before integration. That was found for both *Hi* mobilized in *ddm1* mutation (Supplementary Figure S3A) and by the *Hi* transgene (Supplementary Figure S3B). We could detect excision of endogenous *Hi* in Δ*C*-TG, but the efficiency tends to be low compared with the full-length TG (Figure 5A; Supplementary Figures S3C and S9). VANC might also have a role in efficient transposition with degenerated TIR. It is also possible that degeneration of the TIR had some advantage for the TE. One possible advantage of losing the TIR would be escape from siRNA-based silencing mechanisms. Long TIRs of *MULE*s are often associated with siRNA, and generation of siRNA might be less efficient in non-TIR *MULE*s. Irrespective of the range of functions, the VANC-related proteins are tightly associated with the non-TIR *MULE* lineages, and the efficient recent proliferations of these TEs (Figure 7A; Table I) may depend on these proteins.

Materials and methods

Plant materials

The *ddm1-1* mutant allele was used throughout, except for epi-RILs, which are originated from *ddm1-2* mutant. Details for self-pollination of *ddm1-1/ddm1-1* mutants and wild-type *DDM1/DDM1* siblings were described previously (Kakutani *et al*, 1996). Generation of epi-RILs is described in Johannes *et al* (2009). Mutant alleles used for experiments in Supplementary Figure S15 are those described in Sasaki *et al* (2012).

©2013 European Molecular Biology Organization

Identification of mobile VANDAL21 copies

Conditions for suppression PCR were described previously (Miura *et al*, 2001). To detect *de novo* insertions from whole-genome sequence data of self-pollinated *ddm1* plants (Tsukahara *et al*, 2012), reads containing terminal sequences of *VANDAL21* copies were selected first, and the flanking sequences were subsequently extracted from those reads. In both approaches, each of the *VANDAL21* copies can be distinguished by polymorphisms in the terminal regions. Raw sequence data for the self-pollinated *ddm1* plants were deposited in the DDBJ (DNA Data Bank of Japan) Sequence Read Archive (DRA; accession nos. DRA000420–000424).

DNA preparation and copy number quantification

DNA was extracted using DNeasy kit (QIAGEN). Copy number of *Hi* in the genome was quantified at region C1 (Figure 1) by quantitative PCR using Light Cycler 480 machine (Roche) and SYBR green I MasterMix (Roche) and normalized with signals for two single-copy loci, *AT5G13440* and *AT5G36220*. For this and all other experiments, sequences of primers used are listed in Supplementary Table S2.

Detecting excision of endogenous Hi

Excision of *Hi* was detected using nested PCR with the following conditions. First PCR: 94°C for 2 min, 25 cycles of (94°C for 30 s; 55°C for 30 s; 72°C for 30 s), and 72°C for 2 min. The products diluted to 20 times by H_2O were used for second PCR: 94°C for 2 min, 30 cycles of (94°C for 30 s; 55°C for 30 s; 72°C for 30 s), and 72°C for 2 min. The sequence after excision of endogenous *Hi* was determined after PCR and subsequent cloning by TA-cloning kit (Invitrogen).

DNA methylation analyses

Conventional bisulphite sequence analysis was performed as described previously (Tsukahara *et al*, 2009). For genome-wide bisulphite sequencing analyses, sequencing libraries (insert size: 300–400 bp) were prepared using TruSeq DNA LT Sample Prep Kit (Illumina) and subjected to bisulphite conversion using MethylCode Bisulfite Conversion Kit (Life Technologies). Bisulphite-treated DNA molecules were PCR amplified with 10 cycles using KAPA HiFi HotStart Uracil+ ReadyMix (2 ×) (Kapa Biosystems) and purified with Agencourt AMPure XP (Beckman Coulter). Conditions of the sequencing are as described previously (Tsukahara *et al*, 2012). Raw sequence data were deposited in the DDBJ (DNA Data Bank of Japan) Sequence Read Archive (DRA; accession nos. DRA001060 and DRA001061). Reads were mapped to the reference genome (Release 10 of the Arabidopsis Information Resources) using the Bowtie alignment algorithm (Langmead *et al*, 2009) with conditions described by Chen *et al* (2010). Only uniquely mapped reads were used; reads mapped more than once were not used. Annotation of TEs is obtained from TAIR (http://www.arabidopsis.org/), which is based on Buisine *et al* (2008).

Measurement of DNA methylation was also performed by McrBC digestion and subsequent quantitative PCR as described previously (Teixeria *et al*, 2009). McrBC (New England Biolabs) digestion was performed on 200 ng of genomic DNA. Digested and undigested DNA samples were quantified as described above. The methylation status was estimated by the loss of long DNA after McrBC digestion.

Transcription analysis

Total RNA was isolated by Promega SV Total RNA Isolation System (cat. #Z3100). Reverse transcription reaction was performed using Takara RNA PCR Kit (RR019A) following the manufacturer's instructions. Oligo dT-Adaptor primer was used to reverse transcribe predicted products of three coding genes of *Hi*. GAPC was used as a control. qRT–PCR was performed using SYBR Premix Ex Taq II (Takara) on Thermal Cycler Dice_Real Time System TP800 (Takara), with the following conditions; 95°C for 30′, (95°C for 5′, 60°C for 30′ and 72°C for 30′) for 40 times. The *UBC* gene (*At5g25760*) was used as an internal control.

Transgene construction

Full-length *Hi* was recovered by PCR and cloned into vector PZP2H-lac (Fuse *et al*, 2001). In most constructs, silent mutations were introduced for each of the three ORFs, so that transcripts from the transgene and endogenous copies could be distinguished between (Figure 1). Transgenes with deletion in each of the three ORFs were generated by PCR using the full-length *Hi* as the template. Primer sequences for generating these constructs are shown in Supplementary Table S2.

Molecular evolutionary analyses

BLAST searches of the *A. lyrata* genome were conducted with *A. thaliana MULE* family sequences as the query against assembled genomes by Phytozome ver 8.0 (Goodstein *et al*, 2012). A conserved transposase region, analysed by Yu *et al* (2000), was used for estimation of entire *MULE* family phylogeny. The neighbour-joining trees were constructed by the *p*-distance in MEGA 5 (Tamura *et al*, 2011). To analyse the proliferation rate, copy numbers in recently proliferated clusters were estimated. The *p*-distance of 0.1 was used as a threshold value to classify each copy to clusters and then numbers of copies in each cluster were counted. Average numbers of copies per cluster were compared between non-TIR MULE families and TIR MULE families by *t*-test. The sliding window analyses of synonymous and non-synonymous distances between subcluster and between species for *VANDAL21* and *VANDAL17* families were conducted by DnaSP ver. 5 (Librado and Rozas, 2009). Conserved domain was searched by NCBI conserved domain search site (Marchler-Bauer *et al*, 2011; http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Akiko Terui and Kazuya Takashima for technical assistance. Special thanks to Damon Lisch for valuable advice about *Mutator* and TIR, Yasuhiko Sekine for transposition mechanisms, and Yasushi Hiromi for comments on the manuscript. Supported by grants from Mitsubishi Foundation (to TK), Takeda Science Foundation (to TK), Japanese Ministry of Education, Culture, Sports, Science and Technology (19207002 and 19060014, to TK), Systems Functional Genetics Project of the Transdisciplinary Research Integration Center, ROIS, Japan (to AT, AF, YT, and TK), the Agence Nationale de la Recherche (Project EPIMOBILE, NT09_501422 to VC), and the European Union Network of Excellence Epigenesys (to VC). ME was supported by a PhD studentship from the French Ministry of Research.

Author contributions: AK designed and performed sequence analyses (Figures 7 and 8; Supplementary Figures S10 and S11). ME and VC designed and performed experiments using epi-RILs (Figure 3). YF, TI, AT, AF, YT, and TK designed, performed, and analysed whole-genome bisulphite sequencing and genome re-sequencing experiments. YF, YT, and TK designed and performed all other experiments. YF and TK wrote the paper with incorporating comments from other authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Bivalkar-Mehla S, Vakharia J, Mehla R, Abreha M, Kanwar JR, Tikoo A, Cahuhan A (2011) Viral RNA silencing suppressors (RSS): novel strategy of viruses to ablate the host RNA interference (RNAi) defense system. *Virus Res* **155:** 1–9

Brettell RI, Dennis ES (1991) Reactivation of a silent Ac following tissue culture is associated with heritable alterations in its methylation. *Mol Gen Genet* **229:** 365–372

Brown WE, Robertson DS, Bennetzen JL (1989) Molecular analysis of multiple Mutator-derived alleles of the Bronze locus of maize. *Genetics* **122:** 439–445

Brown J, Sundaresan V (1992) Genetic study of the loss and restoration of Mutator transposon activity in maize: evidence against dominant-negative regulator associated with loss of activity. *Genetics* **130:** 889–898

 ©2013 European Molecular Biology Organization

Buisine N, Quesneville H, Colot V (2008) Improved detection and annotation of transposable elements in sequenced genomes using multiple reference sequence sets. *Genomics* **91:** 467–475

Chandler VL, Walbot V (1986) DNA modification of a maize transposable element correlates with loss of activity. *Proc Natl Acad Sci USA* **83:** 1767–1771

Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics* **11:** 203

Colomé-Tatché M, Cortijo S, Wardenaar R, Morgado L, Lahouze B, Sarazin A, Etcheverry M, Martin A, Feng S, Duvernois-Berthet E, Labadie K, Wincker P, Jacobsen SE, Jansen RC, Colot V & Johannes F (2012) Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc Natl Acad Sci USA* **109:** 16240–16245

Cui H, Fedoroff NV (2002) Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein. *Plant Cell* **14:** 2883–2899

Dietrich CR, Cui F, Packila ML, Li J, Ashlock DA, Nikolau BJ, Schnable PS (2002) Maize Mu transposons are targeted to the 5′ untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics* **160:** 697–716

Eisen JA, Benito MI, Walbot V (1994) Sequence similarity of putative transposases links the maize Mutator autonomous element and a group of bacterial insertion sequences. *Nucleic Acids Res* **22:** 2634–2636

Fedoroff N (1996) Epigenetic regulation of the maize *Spm* transposable element. In *Epigenetic Mechanisms of Gene Regulation*, Riggs A, Martienssen R, Russo V (eds), pp 575–592. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Fuse T, Sasaki T, Yano M (2001) Ti-plasmid vectors useful for functional analysis of rice genes. *Plant Biotechnol* **18:** 219–222

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40:** D1178–D1186

Hashida S, Uchiyama T, Martin C, Kishima Y, Sano Y, Mikami T (2006) The temperature dependent change in Methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. *Plant Cell* **18:** 104–118

Hardeman KJ, Chandler VL (1989) Characterization of bz1 mutants isolated from mutator stocks with high and low numbers of Mu1 elements. *Dev Genet* **10:** 460–472

Hoen DR, Park KC, Elrouby N, Yu Z, Mohabir N, Cowan RK, Bureau TE (2006) Transposon-mediated expansion and diversification of a family of ULP-like genes. *Mol Biol Evol* **23:** 1254–1268

Inagaki S, Miura A, Nakamura K, Lu F, Cui X, Cao X, Kimura H, Saze H, Kakutani T (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J* **29:** 3496–3506

Inagaki S, Kakutani T (2013) What triggers differential DNA methylation of genes and TEs: contribution of body methylation? *Cold Spring Harb Symp Quant Biol* **77:** 155–160

Jiang N, Bao Z, Zhang X, Eddy SE, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431:** 569–573

Johnson L, Cao X, Jacobsen S (2003) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* **12:** 1360–1367

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P, Bouchez D, Dillmann C, Guerche P, Hospital F, Colot V (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* **5:** e1000530

Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc Natl Acad Sci USA* **93:** 12406–12411

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T (2003) Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Curr Biol* **13:** 421–426

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25

Le QH, Wright S, Yu Z, Bureau T (2000) Transposon diversity in Arabidopsis thaliana. *Proc Natl Acad Sci USA* **97:** 7376–7381

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430:** 810–813

Liu S, Yeh CT, Ji T, Ying K, Wu H, Tang HM, Fu Y, Nettleton D, Schnable PS (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet* **5:** e1000733

Lisch D (2002) Mutator transposons. *Trends Plant Sci* **7:** 498–504

Lisch D, Chomet P, Freeling M (1995) Genetic characterization of the Mutator system in maize: behavior and regulation of Mu transposons in a minimal line. *Genetics* **139:** 1777–1796

Lisch D, Girard L, Donlin M, Freeling M (1999) Functional analysis of deletion derivatives of the maize transposon MuDR delineates roles for the MURA and MURB proteins. *Genetics* **151:** 331–341

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL *et al* (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39,** (Database issue) D225–D229

Martienssen R (1996) Epigenetic silencing of *Mu* transposable elements in maize. In *Epigenetic Mechanisms of Gene Regulation*, Riggs A, Martienssen R, Russo V (eds), pp 593–608. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Martienssen R, Baron A (1994) Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. *Genetics* **136:** 1157–1170

McClintock B (1951) Chromosome organization and genic expression. *Cold Spring Harbor Symp Quant Biol* **16:** 13–47

McClintock B (1958) The suppressor-mutator system of control of gene action in maize. *Carnegie Institution of Washington Year Book* Vol. 57: 415–429

Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461:** 427–430

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature* **411:** 212–214

Saze H, Kakutani T (2011) Differentiation of epigenetic modifications between transposons and genes. *Curr Opin Plant Biol* **65:** 589–599

Sasaki T, Kobayashi A, Saze H, Kakutani T (2012) RNAi-independent *de novo* DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene DDM1. *Plant J* **70:** 750–758

Schläppi M, Raina R, Fedoroff N (1994) Epigenetic regulation of the maize Spm transposable element: novel activation of a methylated promoter by TnpA. *Cell* **77:** 427–437

Schläppi M, Raina R, Fedoroff N (1996) A highly sensitive plant hybrid protein assay system based on the Spm promoter and TnpA protein for detection and analysis of transcription activation domains. *Plant Mol Biol* **32:** 717–725

Singer T, Yordan C, Martienssen RA (2001) Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes Dev* **15:** 591–602

Talbert LE, Chandler VL (1988) Characterization of a highly conserved sequence related to *mutator* transposable elements in maize. *Mol Biol Evol* **5:** 519–529

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731–2739

Teixeria FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M, Colot V (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science* **323:** 1600–1604

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T (2009) Bursts of retrotransposition reproduced in Arabidopsis. *Nature* **461:** 423–426

Tsukahara S, Kawabe A, Kobayashi A, Ito T, Aizu T, Shin-i T, Toyoda A, Fujiyama A, Tarutani Y, Kakutani T (2012) Centromere-targeted *de novo* integrations of an LTR retrotransposon of Arabidopsis lyrata. *Genes Dev* **26:** 705–713

Wicker T (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8:** 973–982

Yu Z, Wright SI, Bureau TE (2000) Mutator-like elements in Arabidopsis thaliana. Structure, diversity and evolution. *Genetics* **156:** 2019–2031

Zamore PD (2004) Plant RNAi: How a viral silencing suppressor inactivate siRNA. *Curr Biol* **14:** R198–R200

©2013 European Molecular Biology Organization

Annexe 5 : Position des nouvelles insertions *d'ATCOPIA93, ATENSPM3* et *VANDAL21* par rapport aux gènes.

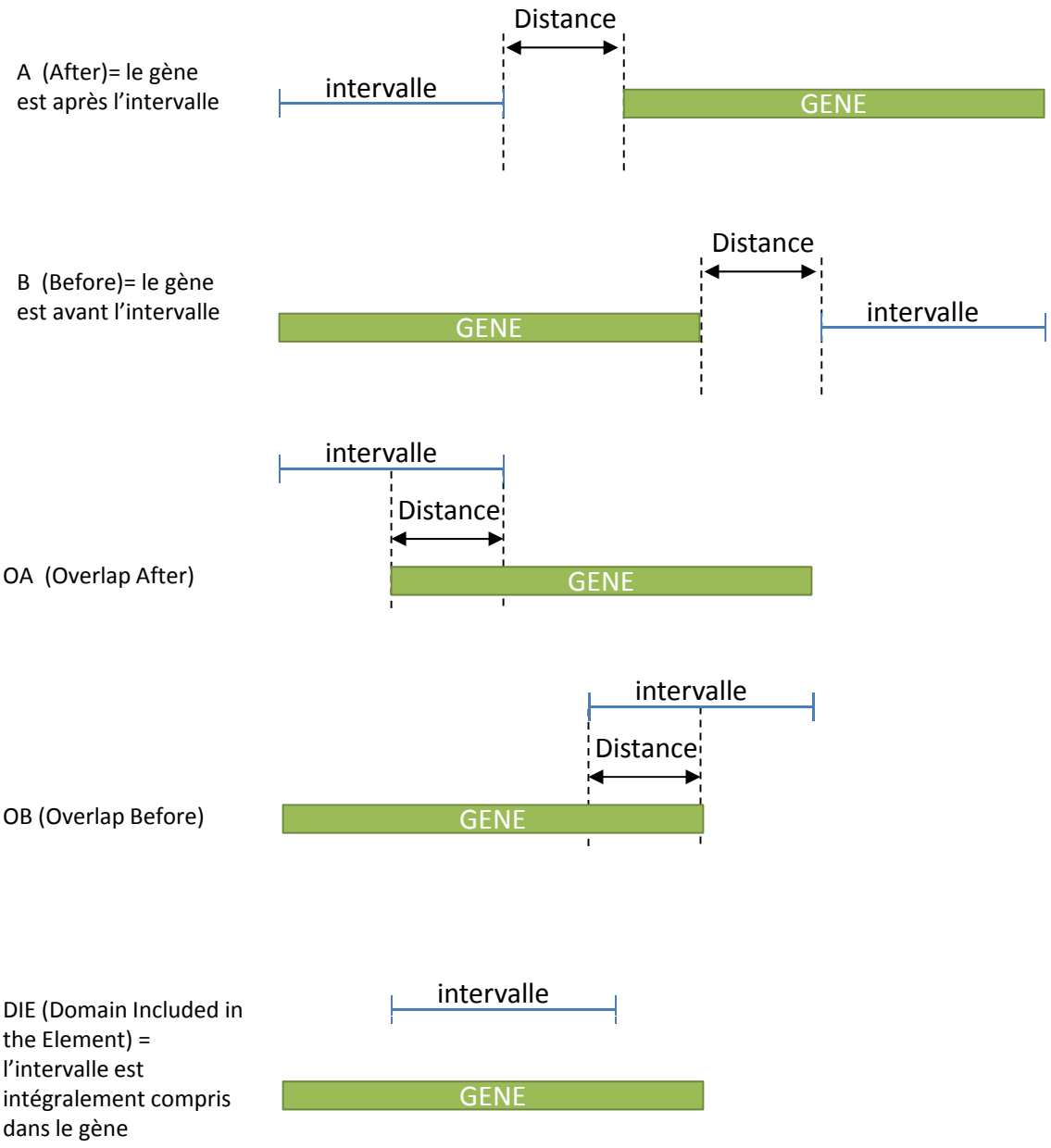


Schéma explicatif des informations présentées dans les tables de l'annexe 5.

insertions d'*ATCOPIA93*

épiRIL	ID	intervalle accepteur				gene			strand	position/gène	distance/gène
		CHR.	START	STOP	TAILLE	NOM	start	stop			
8	54	Chr2	13914006	13914213	207	NO_GENE_DATA					
8	102	Chr5	7799010	7799202	192	AT5G23170	7798197	7799448	-	DIE	*
36	26	Chr2	3666915	3667127	212	NO_GENE_DATA					
36	35	Chr2	9010659	9010878	219	AT2G20970	9009867	9011067	+	DIE	*
36	64	Chr4	10075184	10075412	228	AT4G18205	10073758	10075271	+	OB	87
36	68	Chr5	6122467	6122670	203	AT5G18460	6122867	6126721	+	A	197
55	2	Chr1	1392803	1393537	734	AT1G04920	1391502	1395864	-	DIE	*
55	3	Chr1	3990167	3990279	112	AT1G11810	3989763	3991066	+	DIE	*
55	4	Chr1	12169241	12169578	337	AT1G33560	12168768	12172005	+	DIE	*
55	29	Chr1	20281062	20281407	345	AT1G54330	20279567	20280952	-	B	110
55	30	Chr1	20679800	20679956	156	NO_GENE_DATA					
55	33	Chr1	25815758	25815886	128	AT1G68740	25812449	25816657	-	DIE	*
55	55	Chr2	3215890	3216071	181	NO_GENE_DATA					
55	58	Chr2	3275722	3275965	243	NO_GENE_DATA					
55	59	Chr2	3371620	3371975	355	AT2G07701	3371832	3372152	+	OA	143
55	61	Chr2	3505788	3506083	295	NO_GENE_DATA					
55	73	Chr2	4171056	4171205	149	NO_GENE_DATA					
55	79	Chr2	8272168	8272305	137	NO_GENE_DATA					
55	80	Chr2	19475806	19475903	97	NO_GENE_DATA					
55	83	Chr3	3573618	3573970	352	AT3G11402	3573361	3578125	+	DIE	*
55	84	Chr3	4589949	4590097	148	AT3G13900	4586151	4590681	+	DIE	*
55	86	Chr3	8036011	8036108	97	NO_GENE_DATA					
55	113	Chr3	16618765	16619206	441	AT3G45290	16617544	16621370	+	DIE	*
55	134	Chr4	8561572	8561742	170	AT4G14970	8553643	8561664	+	OB	92
						AT4G14980	8561876	8563639	-	A	134
55	144	Chr4	17154860	17155612	752	AT4G36260	17155556	17157245	-	OA	56
55	166	Chr5	12603975	12604164	189	AT5G33370	12602468	12604764	-	DIE	*
55	168	Chr5	14550492	14550576	84	NO_GENE_DATA					
55	169	Chr5	19519546	19519749	203	AT5G48140	19518770	19520528	-	DIE	*
55	172	Chr5	23373982	23374373	391	AT5G57690	23372063	23374469	-	DIE	*
60	14	Chr2	15677093	15677345	252	NO_GENE_DATA					
70	1	Chr1	195259	195368	109	NO_GENE_DATA					
70	4	Chr1	24748365	24748469	104	AT1G66350	24748195	24750043	+	DIE	*
70	18	Chr4	13036989	13037159	170	NO_GENE_DATA					
92	4	Chr1	14253311	14253399	88	NO_GENE_DATA					
92	6	Chr1	15646282	15646508	226	NO_GENE_DATA					
92	28	Chr3	9054093	9054191	98	AT3G24790	9052982	9054531	+	DIE	*
92	54	Chr4	12909808	12910004	196	AT4G25170	12909575	12911580	+	DIE	*
92	55	Chr4	14153731	14153834	103	AT4G28670	14151387	14153935	+	DIE	*
95	1	Chr1	2560099	2560196	97	AT1G08160	2559487	2560403	-	DIE	*
95	3	Chr1	5025929	5026028	99	AT1G14640	5023656	5025955	+	OB	26
95	4	Chr1	8518660	8518810	150	AT1G24070	8516238	8519734	-	DIE	*
95	5	Chr1	8574200	8574346	146	AT1G24210	8573549	8575156	-	DIE	*
95	9	Chr1	11584650	11584836	186	NO_GENE_DATA					
95	18	Chr1	23638799	23638884	85	NO_GENE_DATA					
95	19	Chr1	25662069	25662212	143	AT1G68450	25661501	25662239	-	DIE	*
95	21	Chr1	28721700	28721723	23	AT1G76540	28720328	28722521	-	DIE	*
95	22	Chr1	29356596	29356682	86	AT1G78070	29355022	29358993	+	DIE	*
95	43	Chr2	14470927	14471287	360	AT2G34280	14470201	14471837	-	DIE	*
95	44	Chr3	4095710	4095803	93	AT3G12880	4095535	4096236	+	DIE	*
95	45	Chr3	5086750	5086916	166	AT3G15115	5085992	5087489	-	DIE	*
95	46	Chr3	9280024	9280127	103	AT3G25560	9279513	9282860	-	DIE	*
95	55	Chr3	22070517	22070966	449	AT3G59750	22069855	22071821	-	DIE	*
95	57	Chr4	968246	968401	155	AT4G02190	967372	969351	+	DIE	*
95	71	Chr4	8675742	8675855	113	AT4G15215	8672070	8679176	+	DIE	*
95	73	Chr4	9488290	9488414	124	AT4G16860	9488466	9495816	-	A	52
95	79	Chr4	13463657	13464112	455	AT4G26700	13463116	13467601	+	DIE	*
95	82	Chr4	16396570	16396862	292	AT4G34250	16394708	16396495	+	B	75
95	84	Chr5	3321377	3321483	106	AT5G10520	3320370	3322910	-	DIE	*

95	86	Chr5	7003537	7003755	218	AT5G20690	7002453	7004551	+	DIE	*
95	87	Chr5	8173673	8173799	126	AT5G24150	8172316	8175431	-	DIE	*
98	5	Chr1	18832464	18832965	501	NO_GENE_DATA					
98	7	Chr1	22623115	22623766	651	AT1G61330	22622975	22624527	+	DIE	*
98	8	Chr1	23828998	23829693	695	AT1G64200	23828526	23830002	-	DIE	*
98	9	Chr1	24927442	24928047	605	AT1G66810	24927613	24928803	-	OA	434
98	10	Chr1	26229693	26230176	483	AT1G69730	26228632	26231339	-	DIE	*
98	17	Chr2	3297324	3297439	115	NO_GENE_DATA					
98	39	Chr4	12164751	12164898	147	AT4G23250	12162004	12167026	-	DIE	*
108	2	Chr1	11085173	11085393	220	AT1G31070	11084859	11088563	+	DIE	*
108	5	Chr1	23332320	23332396	76	AT1G62980	23331676	23333110	+	DIE	*
122	1	Chr1	2557186	2557278	92	AT1G08150	2556343	2559074	-	DIE	*
122	3	Chr1	4025667	4025679	12	NO_GENE_DATA					
122	24	Chr1	21118871	21119063	192	AT1G56410	21117147	21119241	+	DIE	*
122	25	Chr1	21486398	21486523	125	AT1G58050	21478590	21487839	-	DIE	*
122	28	Chr1	22502051	22502343	292	AT1G61093	22501885	22502941	-	DIE	*
122	30	Chr1	24215952	24216067	115	AT1G65190	24215644	24217328	-	DIE	*
122	42	Chr2	2392850	2392951	101	NO_GENE_DATA					
122	48	Chr2	3327959	3328066	107	NO_GENE_DATA					
122	53	Chr2	4089159	4089373	214	NO_GENE_DATA					
122	59	Chr2	6706456	6706687	231	NO_GENE_DATA					
122	62	Chr2	9257368	9257476	108	NO_GENE_DATA					
122	64	Chr2	15217064	15217240	176	NO_GENE_DATA					
122	65	Chr2	16153132	16153223	91	AT2G38620	16152505	16154086	+	DIE	*
122	66	Chr2	19059955	19060407	452	AT2G46430	19058130	19061436	+	DIE	*
122	70	Chr3	2329309	2329593	284	AT3G07310	2327818	2329801	-	DIE	*
122	71	Chr3	4094839	4095121	282	AT3G12870	4094417	4095037	-	OB	198
122	73	Chr3	4874425	4874593	168	NO_GENE_DATA					
122	75	Chr3	5986019	5986088	69	AT3G17490	5984827	5985993	+	B	26
122	76	Chr3	7178351	7178456	105	AT3G20555	7178200	7178614	-	DIE	*
122	79	Chr3	10142083	10142241	158	AT3G27400	10140323	10143225	+	DIE	*
122	82	Chr3	12191989	12192218	229	NO_GENE_DATA					
122	86	Chr3	12948560	12949525	965	NO_GENE_DATA					
122	98	Chr3	16372709	16372794	85	AT3G44840	16372499	16373759	-	DIE	*
122	99	Chr3	16397815	16398032	217	AT3G44920	16397038	16399725	-	DIE	*
122	100	Chr3	20887413	20887592	179	AT3G56320	20885894	20889656	-	DIE	*
122	101	Chr3	22059229	22059406	177	NO_GENE_DATA					
122	117	Chr4	4410835	4410929	94	NO_GENE_DATA					
122	119	Chr4	9452593	9452670	77	AT4G16790	9451585	9453254	-	DIE	*
122	121	Chr4	9759914	9760044	130	AT4G17500	9759325	9760340	+	DIE	*
122	122	Chr4	9818962	9819054	92	AT4G17615	9816452	9819312	+	DIE	*
122	125	Chr4	14401743	14401859	116	AT4G29210	14400920	14403519	+	DIE	*
122	127	Chr5	443390	443484	94	NO_GENE_DATA					
122	128	Chr5	2167738	2167753	15	AT5G06980	2167468	2169391	+	DIE	*
122	131	Chr5	6086800	6086869	69	AT5G18370	6085036	6088926	-	DIE	*
122	133	Chr5	8062273	8062410	137	AT5G23890	8058689	8063201	+	DIE	*
122	135	Chr5	11346508	11346594	86	NO_GENE_DATA					
122	139	Chr5	12757154	12757265	111	NO_GENE_DATA					
122	140	Chr5	13539181	13539302	121	NO_GENE_DATA					
122	143	Chr5	14942997	14943109	112	AT5G37620	14942396	14944354	-	DIE	*
122	147	Chr5	18596063	18596141	78	AT5G45840	18593917	18597995	-	DIE	*
122	151	Chr5	24009437	24009586	149	AT5G59590	24008983	24010649	-	DIE	*
122	152	Chr5	24048187	24048288	101	AT5G59680	24046792	24050801	+	DIE	*
122	154	Chr5	26093508	26093813	305	AT5G65290	26089556	26094548	+	DIE	*
139	2	Chr1	11211722	11211794	72	AT1G31319	11211807	11211914	-	A	13
139	25	Chr3	21970792	21970868	76	AT3G59440	21970355	21971122	+	DIE	*
139	37	Chr5	6591834	6592026	192	NO_GENE_DATA					
150	46	Chr2	5875795	5876101	306	NO_GENE_DATA					
159	1	Chr1	310834	310907	73	AT1G01900	310316	313130	+	DIE	*
159	2	Chr1	2816544	2816631	87	AT1G08800	2813061	2818008	-	DIE	*
159	3	Chr1	10823529	10823605	76	AT1G30550	10819554	10823659	-	DIE	*
159	10	Chr1	23702919	23703108	189	AT1G63860	23701484	23706005	-	DIE	*
159	11	Chr1	23744319	23744502	183	AT1G63990	23742842	23745374	-	DIE	*

159	12	Chr1	26556571	26556735	164	AT1G70460	26555846	26559284	+	DIE	*
159	33	Chr2	11727250	11727268	18	AT2G27420	11726144	11727614	-	DIE	*
159	37	Chr2	19234233	19234339	106	AT2G46790	19232649	19235087	+	DIE	*
159	39	Chr3	10768190	10768368	178	NO_GENE_DATA					
159	41	Chr3	12662819	12663011	192	NO_GENE_DATA					
159	48	Chr3	17037926	17037926	0	AT3G46350	17036427	17041680	+	DIE	*
159	51	Chr3	22056613	22056671	58	AT3G59710	22054937	22057156	-	DIE	*
159	52	Chr4	898922	898985	63	AT4G02050	898307	900870	-	DIE	*
159	55	Chr4	2669092	2669221	129	AT4G05170	2667990	2669364	-	DIE	*
159	73	Chr5	1161915	1161990	75	AT5G04230	1160634	1163898	+	DIE	*
159	84	Chr5	13596099	13596177	78	AT5G35380	13593429	13596293	-	DIE	*
159	86	Chr5	18459177	18459331	154	AT5G45540	18458154	18460821	-	DIE	*
159	88	Chr5	23866866	23866906	40	AT5G59120	23864730	23868052	-	DIE	*
193	1	Chr1	665520	665630	110	AT1G02950	665152	666452	-	DIE	*
						AT1G02952	665354	666367	+	DIE	*
193	2	Chr1	4716146	4716263	117	AT1G13750	4715394	4718235	-	DIE	*
193	19	Chr2	3288776	3288850	74	AT2G07681	3287552	3292366	+	DIE	*
193	24	Chr2	4728499	4728539	40	NO_GENE_DATA					
193	25	Chr2	5299521	5299637	116	NO_GENE_DATA					
193	26	Chr2	13578211	13578374	163	AT2G31930	13577555	13578248	+	OB	37
193	27	Chr2	14419822	14420030	208	AT2G34150	14419295	14423478	-	DIE	*
193	28	Chr2	16859073	16859150	77	AT2G40370	16858166	16860606	-	DIE	*
193	29	Chr2	17875275	17875346	71	AT2G42980	17875005	17876588	-	DIE	*
193	32	Chr3	10233828	10233970	142	AT3G27640	10232248	10235621	+	DIE	*
193	37	Chr3	19665311	19666029	718	AT3G53040	19664674	19666487	-	DIE	*
193	39	Chr3	22520389	22520530	141	NO_GENE_DATA					
193	44	Chr4	4434605	4434761	156	NO_GENE_DATA					
193	47	Chr4	6395707	6395888	181	AT4G10310	6391984	6395877	+	OB	170
193	53	Chr4	14738643	14738724	81	AT4G30140	14738387	14740676	-	DIE	*
193	54	Chr5	152775	153733	958	AT5G01370	152530	154467	+	DIE	*
193	57	Chr5	14416367	14416412	45	AT5G36690	14415185	14417288	-	DIE	*
193	58	Chr5	14475616	14475738	122	AT5G36780	14474485	14476588	-	DIE	*
193	60	Chr5	22803125	22803243	118	AT5G56310	22802316	22804377	+	DIE	*
193	61	Chr5	23867347	23867545	198	AT5G59120	23864730	23868052	-	DIE	*
276	45	Chr2	5860301	5860476	175	AT2G13960	5859585	5861311	+	DIE	*
276	48	Chr2	7987319	7987730	411	NO_GENE_DATA					
276	119	Chr5	15850013	15850084	71	NO_GENE_DATA					
356	1	Chr1	2014889	2015359	470	AT1G06580	2014003	2016054	-	DIE	*
356	2	Chr1	4995803	4995995	192	NO_GENE_DATA					
356	3	Chr1	8347908	8348097	189	AT1G23540	8346827	8349786	-	DIE	*
356	4	Chr1	8735007	8735082	75	AT1G24650	8734570	8737315	+	DIE	*
356	9	Chr1	12682098	12682429	331	NO_GENE_DATA					
356	13	Chr1	14371264	14371593	329	NO_GENE_DATA					
356	14	Chr1	14416833	14417246	413	NO_GENE_DATA					
356	15	Chr1	14478243	14478759	516	NO_GENE_DATA					
356	31	Chr1	25355546	25355884	338	NO_GENE_DATA					
356	32	Chr1	26611068	26611288	220	AT1G70570	26608625	26612309	+	DIE	*
356	33	Chr1	28094323	28094700	377	AT1G74770	28089571	28094834	-	DIE	*
356	42	Chr2	2121215	2121546	331	NO_GENE_DATA					
356	51	Chr2	3606693	3607683	990	NO_GENE_DATA					
356	53	Chr2	3617436	3618094	658	AT2G07981	3614986	3617716	+	OB	280
						AT2G08986	3618058	3626460	+	OA	36
356	55	Chr2	3627222	3627231	9	NO_GENE_DATA					
356	67	Chr2	7436026	7436320	294	AT2G17090	7435088	7437298	-	DIE	*
356	69	Chr2	19076440	19076568	128	AT2G46480	19076405	19078386	-	DIE	*
356	70	Chr2	19577135	19577149	14	AT2G47800	19574864	19580599	+	DIE	*
356	74	Chr3	9934856	9935176	320	AT3G26935	9932600	9935499	-	DIE	*
356	89	Chr3	16988111	16988533	422	AT3G46240	16985843	16988064	-	B	47
356	90	Chr3	18367051	18367562	511	NO_GENE_DATA					
356	91	Chr3	20604822	20605207	385	AT3G55560	20604616	20606380	-	DIE	*
356	92	Chr3	20893906	20894097	191	AT3G56350	20893946	20895625	-	OA	151
356	102	Chr4	3953540	3954368	828	NO_GENE_DATA					
356	108	Chr4	7153178	7153941	763	AT4G11900	7150241	7153571	-	OB	393

356	111	Chr4	10801492	10801581	89	NO_GENE_DATA					
356	113	Chr4	11323784	11324307	523	AT4G21240	11322411	11323664	+	B	120
356	116	Chr4	15759989	15760193	204	AT4G32670	15759243	15762925	-	DIE	*
356	117	Chr4	16945101	16945375	274	AT4G35770	16944941	16946192	+	DIE	*
356	118	Chr5	1759478	1760406	928	AT5G05840	1757098	1760666	-	DIE	*
356	120	Chr5	11114646	11115105	459	NO_GENE_DATA					
356	138	Chr5	18705863	18706035	172	AT5G46140	18704997	18706708	-	DIE	*
362	1	Chr1	2404809	2404938	129	AT1G07750	2404037	2405943	-	DIE	*
362	2	Chr1	2443482	2443699	217	AT1G07900	2442622	2443632	+	OB	150
362	6	Chr1	5236431	5236618	187	AT1G15210	5231348	5236573	-	OB	142
362	7	Chr1	7290905	7291074	169	AT1G20925	7289973	7292507	+	DIE	*
362	9	Chr1	7449652	7449968	316	NO_GENE_DATA					
362	10	Chr1	9106548	9106996	448	AT1G26320	9105038	9107189	+	DIE	*
362	14	Chr1	10089470	10089687	217	AT1G28710	10086600	10089772	-	DIE	*
						AT1G28720	10089437	10089508	+	OB	38
362	15	Chr1	10743880	10744139	259	AT1G30410	10738805	10747323	+	DIE	*
362	17	Chr1	11952617	11953087	470	AT1G32975	11953089	11953899	+	A	2
362	29	Chr1	21782288	21782470	182	AT1G58807	21780185	21784314	+	DIE	*
362	30	Chr1	21818544	21818728	184	AT1G59124	21816443	21820572	+	DIE	*
362	35	Chr1	27293113	27293616	503	NO_GENE_DATA					
362	37	Chr1	28080003	28080007	4	AT1G74730	28078857	28079987	+	B	16
						AT1G74740	28079946	28082644	-	DIE	*
362	55	Chr2	13021221	13021445	224	AT2G30575	13020397	13024208	-	DIE	*
362	58	Chr2	18782913	18783822	909	AT2G45580	18782325	18784347	-	DIE	*
362	63	Chr3	330949	331896	947	AT3G01990	330137	332262	+	DIE	*
362	70	Chr3	19868997	19869180	183	AT3G53590	19867379	19871651	-	DIE	*
362	74	Chr3	22056110	22056219	109	AT3G59710	22054937	22057156	-	DIE	*
362	86	Chr4	6459257	6459589	332	AT4G10440	6459728	6461932	-	A	139
362	88	Chr4	7161256	7161311	55	AT4G11920	7160276	7163382	-	DIE	*
362	106	Chr5	8231932	8231954	22	AT5G24240	8230942	8234150	-	DIE	*
362	107	Chr5	9025343	9026152	809	AT5G25880	9024549	9028377	+	DIE	*
362	110	Chr5	11468942	11469585	643	NO_GENE_DATA					
362	115	Chr5	18033802	18034012	210	AT5G44700	18032864	18036894	-	DIE	*
362	120	Chr5	23909725	23909968	243	AT5G59260	23907901	23909925	-	OB	200
363	1	Chr1	4605590	4605777	187	AT1G13420	4604841	4606053	+	DIE	*
363	2	Chr1	7226667	7226684	17	AT1G20795	7226310	7227566	-	DIE	*
363	3	Chr1	7916453	7916613	160	AT1G22420	7916076	7917518	-	DIE	*
363	20	Chr1	21431629	21431893	264	AT1G57870	21431051	21436103	-	DIE	*
363	22	Chr1	23025585	23025680	95	NO_GENE_DATA					
363	23	Chr1	26408789	26408796	7	AT1G70120	26408921	26409597	+	A	125
363	24	Chr1	26737177	26737295	118	AT1G70920	26735956	26738665	+	DIE	*
363	26	Chr1	28692030	28692540	510	AT1G76480	28692076	28693428	-	OA	464
363	53	Chr2	12453788	12454059	271	NO_GENE_DATA					
363	55	Chr3	10517923	10518585	662	AT3G28200	10518070	10519166	+	OA	515
363	92	Chr4	16144598	16144728	130	AT4G33600	16144391	16145884	-	DIE	*
363	101	Chr5	21833291	21833660	369	NO_GENE_DATA					
366	1	Chr1	13528978	13529399	421	AT1G36150	13528257	13529470	+	DIE	*
366	53	Chr2	11332320	11332620	300	AT2G26650	11331850	11336607	-	DIE	*
366	56	Chr3	7062386	7062570	184	AT3G20250	7058902	7063240	-	DIE	*
366	59	Chr3	9291305	9291835	530	AT3G25577	9291686	9292092	+	OA	149
366	60	Chr3	10178810	10179244	434	AT3G27490	10178100	10180148	-	DIE	*
366	78	Chr3	16050321	16050385	64	NO_GENE_DATA					
366	81	Chr3	17225672	17225889	217	NO_GENE_DATA					
366	82	Chr3	19072813	19073029	216	NO_GENE_DATA					
366	101	Chr4	5132228	5132691	463	NO_GENE_DATA					
366	103	Chr4	16027518	16027934	416	AT4G33230	16026591	16028754	-	DIE	*
366	104	Chr5	2698477	2698807	330	NO_GENE_DATA					
366	105	Chr5	9415367	9415432	65	NO_GENE_DATA					
366	121	Chr5	14207486	14207786	300	NO_GENE_DATA					
371	10	Chr1	25760058	25760196	138	AT1G68600	25759842	25762934	+	DIE	*
375	1	Chr1	2165215	2165304	89	AT1G07050	2164188	2165239	-	OB	24
375	5	Chr1	19830847	19830968	121	NO_GENE_DATA					
375	7	Chr1	30167279	30167428	149	NO_GENE_DATA					

375	15	Chr2	18351223	18351278	55	NO_GENE_DATA					
375	16	Chr3	659331	659424	93	NO_GENE_DATA					
375	17	Chr3	2573757	2573836	79	AT3G08490	2573944	2575215	-	A	108
375	20	Chr3	14302838	14303031	193	AT3G42140	14302060	14303018	-	OB	180
375	22	Chr3	14965645	14965825	180	NO_GENE_DATA					
375	31	Chr4	14566389	14566499	110	AT4G29740	14566265	14569332	+	DIE	*
377	1	Chr1	6981105	6981246	141	AT1G20132	6981264	6983627	+	A	18
377	2	Chr1	11097396	11097567	171	NO_GENE_DATA					
377	4	Chr1	12824462	12824497	35	NO_GENE_DATA					
377	6	Chr1	22238462	22238556	94	AT1G60340	22238734	22240399	-	A	178
377	7	Chr1	22650750	22650922	172	AT1G61390	22650338	22654395	-	DIE	*
377	8	Chr1	23720275	23721152	877	AT1G63910	23719783	23721774	-	DIE	*
377	9	Chr1	23954169	23954285	116	AT1G64500	23953233	23954492	+	DIE	*
377	10	Chr1	25432395	25432538	143	AT1G67830	25431586	25432990	-	DIE	*
377	11	Chr1	26275918	26276042	124	AT1G69800	26274230	26276329	-	DIE	*
377	13	Chr2	2092	2186	94	AT2G01008	1025	2810	+	DIE	*
377	17	Chr2	2360192	2360365	173	AT2G06050	2359114	2362302	-	DIE	*
377	21	Chr2	3353399	3353534	135	NO_GENE_DATA					
377	25	Chr2	16605589	16605762	173	NO_GENE_DATA					
377	26	Chr2	19346318	19347143	825	AT2G47115	19344668	19346921	-	OB	603
377	28	Chr3	5914061	5914159	98	NO_GENE_DATA					
377	29	Chr3	6218403	6218553	150	AT3G18150	6217929	6219500	+	DIE	*
377	34	Chr3	14196059	14196153	94	AT3G41762	14195741	14196603	-	DIE	*
						AT3G41761	14195917	14196078	+	OB	19
377	35	Chr3	14356885	14357025	140	NO_GENE_DATA					
377	38	Chr3	14561537	14561665	128	NO_GENE_DATA					
377	39	Chr3	20783239	20783367	128	AT3G55990	20780315	20783233	+	B	6
						AT3G56000	20783318	20785796	-	OA	49
377	40	Chr4	176626	176758	132	AT4G00400	174062	176853	-	DIE	*
377	43	Chr4	3064592	3064800	208	NO_GENE_DATA					
377	44	Chr4	3140937	3141441	504	NO_GENE_DATA					
377	46	Chr4	4248075	4248280	205	NO_GENE_DATA					
377	51	Chr4	7937971	7938116	145	AT4G13640	7936592	7938645	-	DIE	*
377	55	Chr4	11746947	11747022	75	AT4G22200	11746513	11750698	-	DIE	*
377	56	Chr4	13772011	13772196	185	AT4G27590	13771185	13772594	+	DIE	*
377	57	Chr4	14137385	14137523	138	AT4G28620	14135526	14137953	-	DIE	*
377	58	Chr4	16317918	16317992	74	AT4G34060	16314003	16319426	+	DIE	*
377	61	Chr5	3388841	3388913	72	AT5G10720	3385901	3390686	+	DIE	*
377	62	Chr5	4462148	4462273	125	AT5G13820	4460894	4464666	+	DIE	*
377	64	Chr5	8839337	8839474	137	AT5G25420	8838417	8841775	-	DIE	*
377	65	Chr5	12799106	12799231	125	NO_GENE_DATA					
377	66	Chr5	18302605	18302684	79	AT5G45230	18302147	18308303	-	DIE	*
377	67	Chr5	18788963	18789087	124	AT5G46320	18788246	18788888	-	B	75
377	68	Chr5	20590822	20590941	119	AT5G50600	20589702	20591689	-	DIE	*
377	69	Chr5	20624127	20624283	156	AT5G50700	20623024	20625034	-	DIE	*
377	71	Chr5	21346254	21346378	124	AT5G52610	21346423	21347478	-	A	45
393	1	Chr1	1010635	1010853	218	AT1G03940	1009542	1011056	-	DIE	*
393	6	Chr1	27309649	27309906	257	AT1G72520	27308515	27312754	+	DIE	*
393	15	Chr3	1808197	1808485	288	NO_GENE_DATA					
393	16	Chr3	6121581	6121708	127	AT3G17860	6119767	6123019	+	DIE	*
393	17	Chr3	7051040	7051352	312	AT3G20200	7047581	7051342	+	OB	302
393	21	Chr4	672061	672150	89	AT4G01540	670483	672789	-	DIE	*
393	22	Chr4	2213280	2213621	341	NO_GENE_DATA					
393	30	Chr4	14628521	14628900	379	AT4G29910	14627963	14630740	-	DIE	*
393	31	Chr5	1789418	1789595	177	AT5G05950	1788767	1789438	+	OB	20
393	33	Chr5	13858490	13858575	85	AT5G35680	13857921	13859013	-	DIE	*
408	1	Chr1	845287	845531	244	AT1G03410	844436	847683	-	DIE	*
408	2	Chr1	3577685	3577994	309	AT1G10750	3574807	3578474	-	DIE	*
408	29	Chr3	6725383	6725575	192	AT3G19400	6725468	6727000	+	OA	107
408	60	Chr5	14751432	14751665	233	AT5G37260	14751344	14753088	-	DIE	*
408	61	Chr5	15490215	15490395	180	AT5G38700	15489594	15490399	+	DIE	*
408	62	Chr5	16694712	16695448	736	AT5G41750	16693909	16698943	+	DIE	*
408	67	Chr5	25213272	25213604	332	AT5G62780	25213386	25214137	+	OA	218

421	5	Chr1	15916738	15916827	89	NO_GENE_DATA					
421	10	Chr1	28961091	28961458	367	AT1G77090	28960387	28961922	-	DIE	*
421	20	Chr2	3359985	3360177	192	AT2G07798	3357619	3361102	+	DIE	*
421	37	Chr2	17748581	17748734	153	AT2G42610	17747917	17750440	+	DIE	*
421	38	Chr2	18188774	18189563	789	AT2G43920	18189008	18191348	-	OA	555
421	54	Chr4	1268178	1268768	590	AT4G02850	1266463	1268612	-	OB	434
						AT4G02860	1268713	1270525	-	OA	55
421	56	Chr4	1400269	1400607	338	AT4G03165	1399378	1400291	-	OB	22
421	74	Chr4	8209068	8209156	88	AT4G14250	8208748	8213237	-	DIE	*
421	78	Chr4	11227421	11227611	190	AT4G21010	11226286	11227363	-	B	58
421	82	Chr4	16919036	16919188	152	AT4G35680	16917749	16920008	+	DIE	*
421	83	Chr5	1603424	1603589	165	AT5G05410	1602205	1603926	+	DIE	*
421	84	Chr5	2514803	2514939	136	AT5G07870	2513794	2515418	+	DIE	*
421	85	Chr5	3558546	3558874	328	AT5G11180	3557261	3561575	-	DIE	*
421	86	Chr5	4172558	4172832	274	AT5G13150	4172829	4174930	-	OA	3
421	102	Chr5	20215599	20215924	325	NO_GENE_DATA					
421	104	Chr5	21740634	21740932	298	AT5G53510	21738849	21742082	+	DIE	*
425	1	Chr1	2204577	2204740	163	AT1G07180	2204320	2206934	+	DIE	*
425	2	Chr1	2554524	2554616	92	AT1G08140	2552036	2555161	-	DIE	*
425	3	Chr1	2631164	2631251	87	AT1G08340	2631068	2632871	+	DIE	*
425	4	Chr1	3280721	3280794	73	AT1G10050	3279270	3283482	+	DIE	*
425	6	Chr1	11102525	11102620	95	NO_GENE_DATA					
425	9	Chr1	14220145	14220235	90	NO_GENE_DATA					
425	10	Chr1	16679896	16680029	133	NO_GENE_DATA					
425	11	Chr1	16695190	16695272	82	NO_GENE_DATA					
425	14	Chr1	23560932	23561025	93	AT1G63520	23559887	23562143	+	DIE	*
425	15	Chr1	23889302	23889427	125	AT1G64370	23888920	23889661	+	DIE	*
425	16	Chr1	25978237	25978406	169	AT1G69090	25977501	25978706	-	DIE	*
425	17	Chr1	29781800	29781907	107	NO_GENE_DATA					
425	18	Chr1	30236411	30236507	96	AT1G80420	30235164	30238535	-	DIE	*
425	23	Chr2	10042905	10042979	74	AT2G23600	10042199	10043673	-	DIE	*
425	24	Chr2	18269663	18269821	158	AT2G44180	18269243	18271860	-	DIE	*
425	25	Chr2	18857685	18857770	85	AT2G45790	18855775	18858018	+	DIE	*
						AT2G45800	18857941	18859189	-	A	171
425	26	Chr3	3730268	3730365	97	AT3G11820	3729305	3731149	-	DIE	*
425	27	Chr3	5440413	5440539	126	AT3G16030	5439609	5442802	+	DIE	*
425	28	Chr3	5826247	5826322	75	NO_GENE_DATA					
425	29	Chr3	7473159	7473232	73	NO_GENE_DATA					
425	31	Chr3	11048849	11048955	106	NO_GENE_DATA					
425	36	Chr3	17330899	17331042	143	AT3G47050	17328033	17330889	-	B	10
425	37	Chr3	19455696	19455810	114	AT3G52490	19455699	19458919	-	OA	111
425	38	Chr3	21098230	21098347	117	AT3G57020	21098315	21100360	-	OA	32
425	39	Chr3	23229881	23229962	81	AT3G62820	23229749	23230501	+	DIE	*
425	40	Chr4	2524711	2524808	97	AT4G04955	2522034	2525280	-	DIE	*
425	44	Chr4	9516404	9516607	203	AT4G16900	9511859	9516541	-	OB	137
						AT4G16915	9516724	9518697	+	A	117
425	46	Chr4	13408237	13408333	96	AT4G26570	13408422	13411070	-	A	89
425	47	Chr4	13692037	13692196	159	AT4G27350	13690821	13692796	-	DIE	*
425	48	Chr4	15633264	15633348	84	AT4G32378	15633112	15633222	-	B	42
425	50	Chr4	18519777	18519887	110	AT4G39940	18519717	18521568	+	DIE	*
425	51	Chr5	874947	875029	82	AT5G03495	873331	875251	-	DIE	*
425	52	Chr5	8153522	8153677	155	AT5G24110	8153115	8154709	-	DIE	*
425	53	Chr5	8239360	8239449	89	AT5G24270	8238534	8243357	-	DIE	*
425	54	Chr5	10478018	10478122	104	AT5G28500	10477771	10479384	+	DIE	*
425	55	Chr5	10640096	10640204	108	NO_GENE_DATA					
425	58	Chr5	11090130	11090852	722	NO_GENE_DATA					
425	59	Chr5	11145404	11145503	99	NO_GENE_DATA					
425	62	Chr5	20508511	20508673	162	NO_GENE_DATA					
425	63	Chr5	20771654	20772381	727	AT5G51090	20772237	20773194	-	OA	144
425	65	Chr5	21197562	21198008	446	AT5G52170	21196974	21199959	+	DIE	*
425	69	Chr5	25458219	25458741	522	AT5G63590	25456993	25458470	-	OB	251
437	1	Chr1	13710676	13710839	163	NO_GENE_DATA					
437	14	Chr2	448804	449071	267	NO_GENE_DATA					

437	17	Chr2	9535972	9536251	279	NO_GENE_DATA					
437	18	Chr2	16544762	16544879	117	AT2G39700	16543832	16545584	-	DIE	*
437	19	Chr3	909854	910412	558	AT3G03680	907500	910807	+	DIE	*
437	26	Chr3	23215849	23216353	504	AT3G62750	23214371	23217045	+	DIE	*
437	28	Chr4	3473737	3473991	254	NO_GENE_DATA					
437	31	Chr4	6497941	6498264	323	AT4G10510	6495955	6499010	+	DIE	*
437	35	Chr4	15139416	15139620	204	NO_GENE_DATA					
437	36	Chr5	7827932	7828113	181	AT5G23240	7826732	7828724	-	DIE	*
454	1	Chr1	3115	3249	134	NO_GENE_DATA					
454	3	Chr1	11191545	11191795	250	AT1G31290	11188293	11192317	+	DIE	*
454	13	Chr1	25653718	25654122	404	AT1G68420	25653893	25654156	+	OA	229
454	14	Chr1	29616254	29616578	324	AT1G78760	29615174	29616738	-	DIE	*
454	35	Chr2	11496506	11496664	158	AT2G26940	11496518	11497378	-	OA	146
454	36	Chr3	807190	807542	352	AT3G03380	799615	808522	+	DIE	*
454	37	Chr3	1737581	1737875	294	AT3G05830	1736677	1738664	+	DIE	*
454	38	Chr3	3083438	3083594	156	AT3G10010	3081814	3088195	-	DIE	*
454	39	Chr3	4492935	4493027	92	AT3G13700	4490628	4493737	-	DIE	*
						AT3G13710	4490628	4493831	-	DIE	*
454	51	Chr3	16580047	16580428	381	NO_GENE_DATA					
454	52	Chr3	16894460	16894609	149	NO_GENE_DATA					
454	53	Chr3	21980027	21980291	264	AT3G59470	21978847	21980753	-	DIE	*
454	54	Chr3	22228074	22228854	780	AT3G60160	22223803	22229204	-	DIE	*
454	55	Chr3	22556611	22556699	88	NO_GENE_DATA					
454	58	Chr4	3651904	3652219	315	NO_GENE_DATA					
454	70	Chr4	5006317	5006790	473	NO_GENE_DATA					
454	71	Chr4	7002644	7003625	981	AT4G11560	6999896	7003493	-	OB	849
454	73	Chr4	7083238	7083461	223	AT4G11770	7082279	7083469	-	DIE	*
454	78	Chr4	16682304	16682584	280	AT4G35050	16682580	16684848	-	OA	4
454	79	Chr5	729472	729675	203	NO_GENE_DATA					
454	80	Chr5	8895666	8895869	203	AT5G25550	8894179	8895480	+	B	186
454	86	Chr5	13333017	13333469	452	NO_GENE_DATA					
454	88	Chr5	22921922	22922268	346	NO_GENE_DATA					
454	89	Chr5	23913821	23913992	171	AT5G59280	23914141	23915139	-	A	149
458	6	Chr2	18286844	18287065	221	AT2G44230	18286525	18288400	+	DIE	*
458	9	Chr3	17888839	17888923	84	AT3G48310	17888101	17889814	+	DIE	*
458	11	Chr4	129610	129794	184	AT4G00300	126509	130399	+	DIE	*
458	16	Chr4	11749997	11750140	143	AT4G22200	11746513	11750698	-	DIE	*
458	17	Chr4	12690209	12690424	215	AT4G24580	12687879	12694248	-	DIE	*
467	7	Chr3	4713956	4714064	108	AT3G14200	4712574	4714653	-	DIE	*
503	3	Chr1	697955	697963	8	NO_GENE_DATA					
503	7	Chr1	12480166	12480609	443	NO_GENE_DATA					
503	18	Chr1	27998340	27998485	145	AT1G74500	27998175	27998821	-	DIE	*
503	19	Chr1	30343731	30343795	64	AT1G80740	30342394	30347048	+	DIE	*
503	41	Chr2	14339361	14339438	77	NO_GENE_DATA					
503	42	Chr2	14449114	14449204	90	AT2G34210	14445938	14450522	+	DIE	*

insertions d'*ATENSPM3*

	ID	intervalle accepteur				gene				position/gène	distance/gène
		CHR.	START	STOP	TAILLE	NOM	start	stop	strand		
55	31	Chr1	21486789	21486913	124	AT1G58050	21478590	21487839	-	DIE	*
55	32	Chr1	21603943	21604210	267	NO_GENE_DATA					
55	82	Chr3	3549888	3550083	195	AT3G11325	3549771	3551661	-	DIE	*
55	87	Chr3	9019495	9019591	96	NO_GENE_DATA					
55	88	Chr3	9043438	9043531	93	NO_GENE_DATA					
55	89	Chr3	9142495	9142632	137	NO_GENE_DATA					
55	90	Chr3	10647627	10647765	138	AT3G28415	10647123	10651540	-	DIE	*
55	110	Chr3	14457168	14457296	128	NO_GENE_DATA					
55	114	Chr3	16767338	16767437	99	NO_GENE_DATA					
						AT5G18065	5978678	5980376	+	OB	113
55	150	Chr5	5980263	5980983	720	AT5G18070	5981007	5982934	+	A	24
55	174	Chr5	24028648	24028773	125	NO_GENE_DATA					
60	6	Chr1	25956040	25956309	269	NO_GENE_DATA					
60	11	Chr2	10321820	10321904	84	AT2G24260	10319334	10322222	-	DIE	*
						AT2G40960	17095522	17097454	+	DIE	*
60	15	Chr2	17097260	17097372	112	AT2G40970	17097550	17098695	-	A	178
95	8	Chr1	11488135	11488279	144	AT1G31990	11488016	11489287	+	DIE	*
95	17	Chr1	23440928	23441100	172	AT1G63206	23439642	23440931	+	OB	3
95	69	Chr4	6495674	6495774	100	AT4G10510	6495955	6499010	+	A	181
95	70	Chr4	7209836	7210023	187	AT4G12020	7201656	7209648	+	B	188
95	77	Chr4	12724239	12724324	85	NO_GENE_DATA					
95	78	Chr4	13337464	13337555	91	NO_GENE_DATA					
95	90	Chr5	20091363	20091542	179	AT5G49520	20090776	20093346	+	DIE	*
122	2	Chr1	2675675	2675678	3	AT1G08465	2675813	2679781	+	A	135
						AT1G22890	8102818	8103532	+	OB	48
122	5	Chr1	8103484	8103749	265	AT1G22900	8103659	8104505	-	OA	90
122	6	Chr1	8107658	8107871	213	AT1G22910	8105589	8108164	+	DIE	*
122	7	Chr1	9291297	9291381	84	AT1G26815	9290870	9291282	-	B	15
122	20	Chr1	19226964	19227197	233	AT1G51810	19227119	19232359	-	OA	78
122	22	Chr1	19631371	19631702	331	AT1G52700	19631126	19633440	-	DIE	*
122	23	Chr1	20368518	20368605	87	AT1G54540	20367572	20368434	-	B	84
122	26	Chr1	21880280	21880535	255	AT1G59550	21879649	21880788	-	DIE	*
122	27	Chr1	22224159	22224350	191	AT1G60270	22221677	22224257	-	OB	98
122	41	Chr2	164018	164174	156	AT2G01340	163954	165290	+	DIE	*
122	60	Chr2	7519681	7519804	123	AT2G17290	7516415	7519633	+	B	48
122	61	Chr2	7539527	7539636	109	NO_GENE_DATA					
122	63	Chr2	12035104	12035286	182	AT2G28220	12033953	12037527	+	DIE	*
122	68	Chr3	661653	661796	143	AT3G02940	661873	664006	+	A	77
						AT3G02970	669103	670541	-	B	178
122	69	Chr3	670719	670832	113	AT3G02980	670897	672215	-	A	65
122	72	Chr3	4109778	4109893	115	AT3G12910	4109375	4110648	+	DIE	*
122	74	Chr3	5871163	5871305	142	NO_GENE_DATA					
122	78	Chr3	9589610	9589716	106	AT3G26200	9589282	9591179	+	DIE	*
122	95	Chr3	15322342	15322512	170	NO_GENE_DATA					
122	103	Chr3	22815614	22815966	352	AT3G61660	22816152	22816541	+	A	186
122	123	Chr4	10312269	10312377	108	AT4G18770	10311030	10313104	+	DIE	*
122	126	Chr4	17266360	17266587	227	AT4G36610	17265390	17267314	-	DIE	*
122	129	Chr5	2173158	2173261	103	NO_GENE_DATA					
122	132	Chr5	6996529	6996662	133	NO_GENE_DATA					
122	134	Chr5	10991586	10991754	168	NO_GENE_DATA					
122	136	Chr5	11631606	11631706	100	NO_GENE_DATA					
122	144	Chr5	16011085	16011311	226	AT5G40000	16011072	16012484	-	DIE	*
122	145	Chr5	16269221	16269393	172	AT5G40620	16269263	16269655	+	OA	130
122	146	Chr5	16270332	16270496	164	NO_GENE_DATA					
122	149	Chr5	22557952	22558062	110	AT5G55730	22558148	22560461	-	A	86
122	150	Chr5	23357867	23358212	345	NO_GENE_DATA					
122	155	Chr5	26130419	26130525	106	NO_GENE_DATA					
137	52	Chr4	3440756	3440922	166	NO_GENE_DATA					
137	65	Chr4	10767041	10767121	80	NO_GENE_DATA					

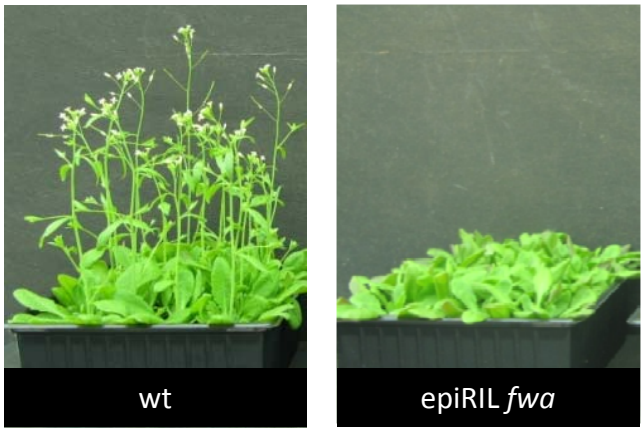
137	73	Chr5	26286862	26286967	105	NO_GENE_DATA					
144	22	Chr2	3623811	3624481	670	AT2G08986	3618058	3626460	+	DIE	*
144	32	Chr3	5606823	5606957	134	NO_GENE_DATA					
159	38	Chr3	5586658	5586730	72	AT3G16440	5585944	5587750	+	DIE	*
252	53	Chr2	3618983	3619853	870	AT2G08986	3618058	3626460	+	DIE	*
277	5	Chr1	25746302	25746355	53	NO_GENE_DATA					
277	6	Chr1	25746408	25746525	117	NO_GENE_DATA					
277	18	Chr2	14295264	14295420	156	AT2G33790	14293561	14295271	-	OB	7
277	31	Chr4	8143015	8143115	100	NO_GENE_DATA					
366	16	Chr1	20148478	20148570	92	AT1G53970	20148419	20148898	+	DIE	*
366	91	Chr4	3954049	3954371	322	NO_GENE_DATA					
366	110	Chr5	11545238	11545316	78	NO_GENE_DATA					
366	112	Chr5	11728519	11729402	883	NO_GENE_DATA					
366	115	Chr5	11732381	11733090	709	NO_GENE_DATA					
408	3	Chr1	4071314	4071562	248	AT1G12040	4070124	4072567	+	DIE	*
408	27	Chr2	5936758	5936842	84	NO_GENE_DATA					
408	28	Chr2	16526250	16526371	121	AT2G39640	16525086	16527012	-	DIE	*
408	51	Chr4	9048444	9048510	66	NO_GENE_DATA					
408	52	Chr4	9074071	9074148	77	AT4G16015	9073202	9074809	+	DIE	*
408	53	Chr4	9080459	9080603	144	NO_GENE_DATA					
408	55	Chr4	9410551	9410644	93	AT4G16740	9407768	9410877	+	DIE	*
408	64	Chr5	21186218	21186309	91	AT5G52140	21184522	21186872	-	DIE	*
439	4	Chr1	24091117	24091194	77	AT1G64830	24091271	24092566	-	A	77
439	5	Chr1	24334475	24334558	83	NO_GENE_DATA					
439	19	Chr2	11505759	11505852	93	NO_GENE_DATA					
439	38	Chr5	7815404	7815488	84	AT5G23212	7815405	7815882	-	OA	83
466	2	Chr1	9460314	9460488	174	AT1G27213	9460035	9460187	-	B	127
466	24	Chr4	12270990	12271153	163	NO_GENE_DATA					
503	17	Chr1	23255445	23255788	343	AT1G62800	23253703	23257463	-	DIE	*
503	45	Chr3	7632055	7632138	83	NO_GENE_DATA					
503	58	Chr4	2328881	2328978	97	AT4G04620	2328449	2330200	-	DIE	*
503	82	Chr5	20157838	20157927	89	NO_GENE_DATA					

insertions de *VANDAL21*

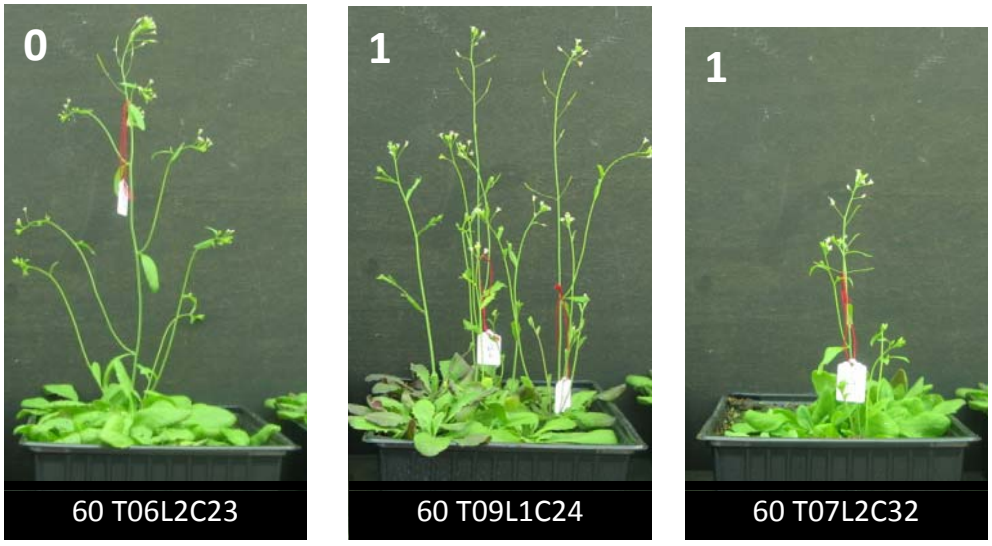
	ID	intervalle accepteur				gene				position/gène	distance/gène
		CHR.	START	STOP	TAILLE	NOM	start	stop	strand		
8	52	Chr2	9119833	9119990	157	AT2G21300	9114154	9119946	-	OB	113
8	56	Chr3	5857582	5857776	194	AT3G17180	5855861	5859157	-	DIE	*
8	114	Chr5	18949531	18949660	129	NO_GENE_DATA					
14	24	Chr4	2805669	2805826	157	AT4G05520	2804451	2807929	+	DIE	*
36	36	Chr2	10299883	10300015	132	AT2G24220	10300115	10301787	+	A	100
36	37	Chr2	10625142	10625256	114	AT2G24990	10625214	10628299	+	OA	42
36	40	Chr3	7750885	7750970	85	NO_GENE_DATA					
36	66	Chr4	11478281	11478351	70	AT4G21585	11477988	11479953	+	DIE	*
52	1	Chr1	1919672	1919813	141	AT1G06270	1918088	1919760	-	OB	88
60	1	Chr1	6261367	6261460	93	AT1G18190	6257851	6261457	-	OB	90
60	12	Chr2	12660632	12660713	81	NO_GENE_DATA					
60	29	Chr4	5859205	5859488	283	AT4G09200	5859257	5861803	+	OA	231
60	30	Chr4	5876408	5876685	277	AT4G09250	5876578	5878978	+	OA	107
70	19	Chr5	11835926	11836129	203	NO_GENE_DATA					
70	20	Chr5	11845546	11845669	123	NO_GENE_DATA					
92	1	Chr1	2771295	2771456	161	AT1G08700	2769815	2771477	-	DIE	*
						AT1G08710	2771633	2773460	+	A	177
92	9	Chr1	18071267	18071455	188	AT1G48860	18068707	18071466	-	DIE	*
92	37	Chr3	20370100	20370177	77	AT3G54970	20368168	20370098	-	B	2
						AT3G54980	20370234	20373069	+	A	57
92	38	Chr3	21071720	21071822	102	AT3G56920	21070648	21072702	+	DIE	*
92	40	Chr4	3559681	3559779	98	NO_GENE_DATA					
98	26	Chr3	6658949	6659050	101	AT3G19210	6652745	6658988	-	OB	39
						AT3G19220	6659164	6660852	+	A	114
98	30	Chr3	20004977	20005209	232	AT3G54010	20000861	20005183	+	OB	206
98	31	Chr4	3430033	3430181	148	NO_GENE_DATA					
98	41	Chr5	5548906	5548976	70	AT5G16870	5546811	5548736	-	B	170
						AT5G16880	5548914	5551463	+	OA	62
98	43	Chr5	14253108	14253220	112	AT5G36210	14247913	14253295	-	DIE	*
122	102	Chr3	22314242	22314331	89	AT3G60360	22312258	22314400	-	DIE	*
144	29	Chr2	18228415	18228522	107	AT2G44065	18228473	18231407	+	OA	49
144	30	Chr3	3453438	3453531	93	AT3G11010	3450496	3453683	-	DIE	*
144	75	Chr5	15446312	15446414	102	NO_GENE_DATA					
150	47	Chr3	1772503	1772580	77	AT3G05935	1771962	1772599	-	DIE	*
159	34	Chr2	14654227	14654294	67	AT2G34730	14654357	14657623	+	A	63
159	67	Chr4	5801186	5801251	65	NO_GENE_DATA					
193	3	Chr1	12698467	12698690	223	NO_GENE_DATA					
193	30	Chr2	19662586	19662665	79	AT2G48060	19651196	19662617	-	OB	31
						AT2G48070	19662817	19664465	+	A	152
193	31	Chr3	2444777	2444935	158	AT3G07660	2444778	2450561	-	OA	157
193	34	Chr3	13282197	13282275	78	NO_GENE_DATA					
193	46	Chr4	6023826	6023904	78	AT4G09510	6021164	6023873	-	OB	47
252	1	Chr1	5194117	5194253	136	AT1G15100	5193426	5194203	-	OB	86
252	29	Chr1	22057038	22057125	87	AT1G59910	22053951	22057052	-	OB	14
260	1	Chr1	2668987	2669079	92	AT1G08450	2667798	2671892	-	DIE	*
260	16	Chr1	17995264	17995337	73	AT1G48650	17989398	17995326	-	OB	62
260	47	Chr2	15999918	16000066	148	AT2G38185	15999926	16003169	+	OA	140
260	48	Chr3	9793171	9793239	68	AT3G26630	9791371	9792985	-	B	186
						AT3G26640	9793220	9794457	+	OA	19
260	49	Chr3	11238043	11238129	86	AT3G29280	11236610	11237995	-	B	48
						AT3G29290	11238111	11240287	+	OA	18
260	50	Chr3	11299566	11299642	76	AT3G29400	11297283	11299620	-	OB	54
260	92	Chr5	1405136	1405215	79	AT5G04840	1405294	1407934	+	A	79
260	114	Chr5	20047910	20047987	77	NO_GENE_DATA					
276	49	Chr2	8318491	8318574	83	AT2G19170	8314005	8318596	-	DIE	*
276	51	Chr3	3106727	3106822	95	AT3G10070	3103880	3106815	-	OB	88
276	53	Chr3	11073585	11073713	128	AT3G29090	11073703	11075478	+	OA	10
276	79	Chr4	3055002	3055416	414	NO_GENE_DATA					
340	103	Chr4	267868	268084	216	AT4G00630	261449	268050	-	OB	182

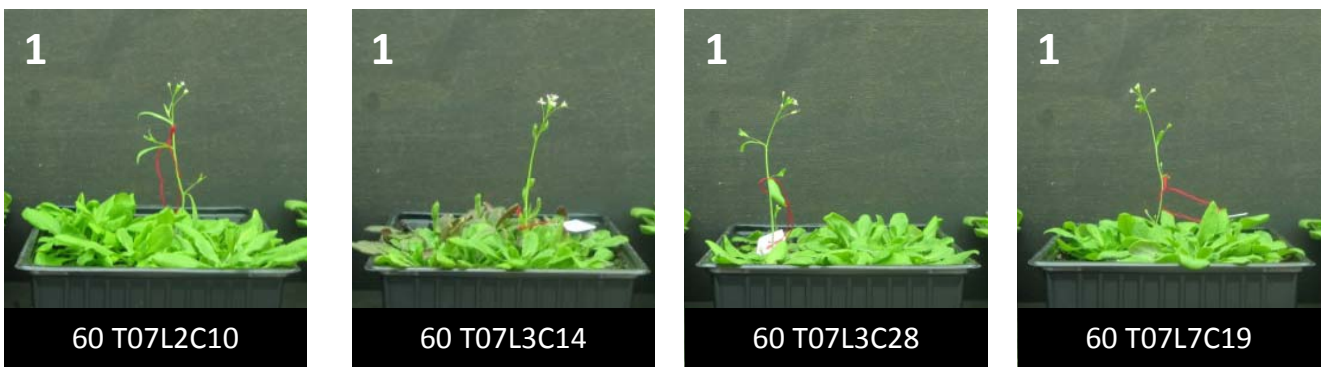
340	132	Chr4	12642779	12642874	95	AT4G24450	12635426	12642797	+	OB	18
						AT4G24460	12642854	12645501	+	OA	20
362	3	Chr1	3338040	3338216	176	AT1G10180	3338303	3341608	-	A	87
362	114	Chr5	15901533	15901609	76	AT5G39730	15901679	15902870	+	A	70
371	1	Chr1	1817223	1817293	70	AT1G05970	1816056	1817282	-	OB	59
						AT1G05980	1817400	1817471	+	A	107
371	2	Chr1	10035463	10035546	83	AT1G28530	10031997	10035500	-	OB	37
						AT1G28540	10035664	10036956	+	A	118
371	9	Chr1	21919979	21920097	118	AT1G59650	21919745	21922923	+	DIE	*
371	17	Chr2	3883263	3883443	180	NO_GENE_DATA					
371	23	Chr3	8788662	8788731	69	AT3G24255	8788853	8793371	+	A	122
371	33	Chr4	5797218	5797350	132	AT4G09040	5794529	5797375	-	DIE	*
371	37	Chr4	11509807	11509887	80	AT4G21660	11505961	11509842	-	OB	35
377	24	Chr2	14267849	14267946	97	AT2G33730	14265450	14267972	-	DIE	*
377	42	Chr4	1228876	1228947	71	AT4G02760	1223959	1228941	-	OB	65
						AT4G02770	1229111	1229945	-	A	164
421	2	Chr1	13876141	13876312	171	NO_GENE_DATA					
421	35	Chr2	5598233	5598410	177	AT2G13440	5593060	5598296	-	OB	63
						AT2G13450	5598421	5599923	-	A	11
421	36	Chr2	10760215	10760303	88	AT2G25270	10759724	10762568	+	DIE	*
425	5	Chr1	3803879	3803951	72	AT1G11310	3800738	3804185	-	DIE	*
425	22	Chr2	6325160	6325237	77	NO_GENE_DATA					
425	57	Chr5	10934625	10934697	72	NO_GENE_DATA					
439	6	Chr1	28631078	28631250	172	AT1G76320	28631239	28634124	+	OA	11
439	12	Chr2	79850	80054	204	AT2G01090	79989	81710	+	OA	65
454	74	Chr4	8862711	8862828	117	AT4G15510	8860616	8862665	+	B	46
						AT4G15520	8862815	8864618	+	OA	13
458	15	Chr4	11302881	11303075	194	AT4G21200	11302685	11306601	+	DIE	*
458	17	Chr5	10192629	10192724	95	AT5G28220	10190357	10193119	-	DIE	*
466	1	Chr1	3774883	3774985	102	NO_GENE_DATA					
466	14	Chr2	15040636	15040717	81	AT2G35790	15040679	15042123	+	OA	38
466	15	Chr2	19125583	19125713	130	AT2G46560	19115371	19125856	-	DIE	*
466	16	Chr3	8115103	8115195	92	AT3G22890	8112723	8114992	+	B	111
						AT3G22900	8115177	8116253	+	OA	18
503	36	Chr2	3798899	3798964	65	NO_GENE_DATA					
503	73	Chr4	13577984	13578065	81	AT4G27050	13575800	13578051	-	OB	67
559	36	Chr3	20848404	20848562	158	AT3G56190	20845674	20848534	-	OB	130
559	36	Chr2	1797	1882	85	AT2G01008	1025	2810	+	DIE	*
559	105	Chr3	14195764	14195849	85	AT3G41762	14195741	14196603	-	DIE	*
						AT3G41761	14195917	14196078	+	A	68
559	145	Chr4	17512208	17512273	65	AT4G37210	17512259	17514625	+	OA	14
559	167	Chr5	12949683	12950233	550	NO_GENE_DATA					
559	176	Chr5	25775460	25775570	110	AT5G64460	25772849	25775568	-	OB	108

Annexe 6 : Candidats du crible sélectionnés pour l’avancement de la floraison.



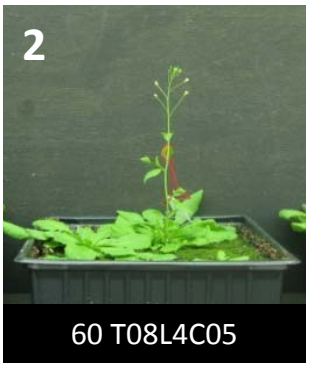
epiRIL 60 : candidats sélectionnés (partie 1)

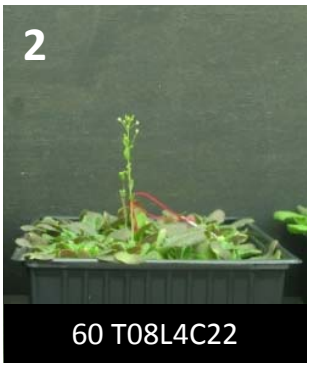


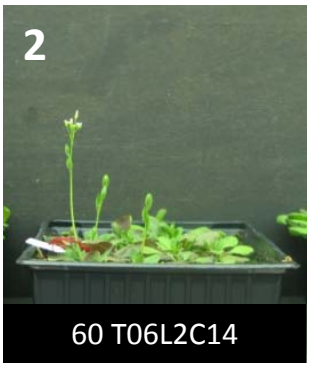


epiRIL 60 : candidats sélectionnés (partie 2)































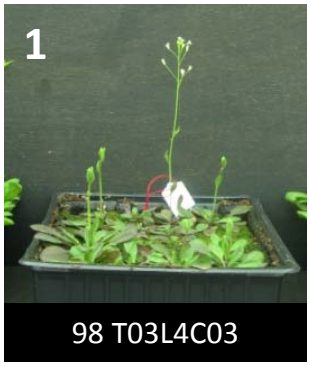




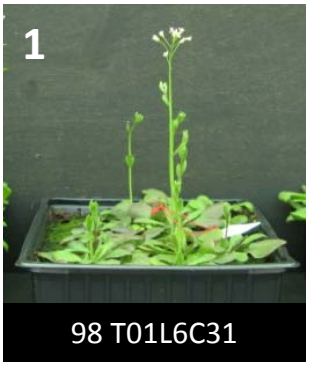




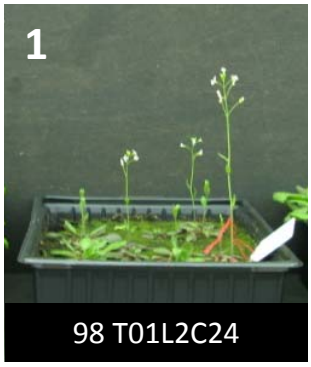
epiRIL 98 : candidats sélectionnés (partie 1)



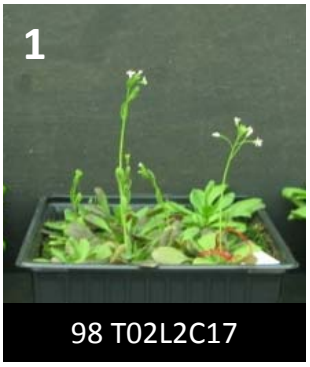
98 T03L4C03



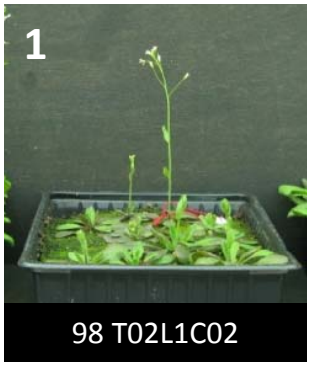
98 T01L6C31



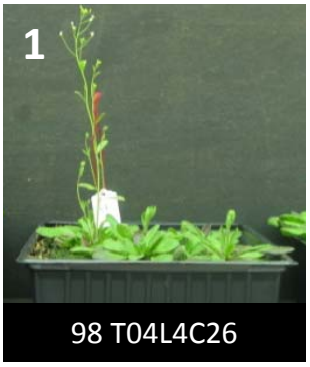
98 T01L2C24



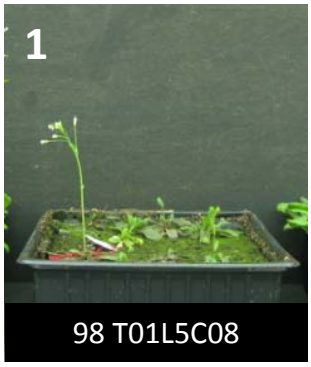
98 T02L2C17



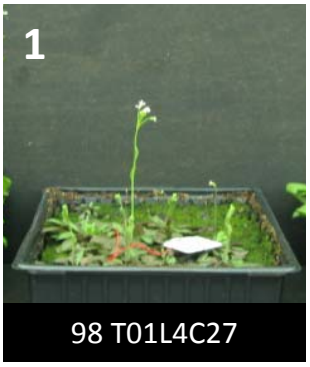
98 T02L1C02



98 T04L4C26



98 T01L5C08



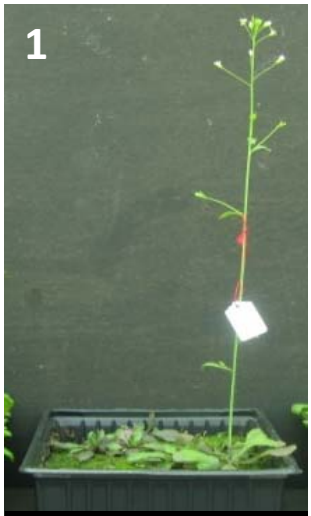
98 T01L4C27



98 T02L4C18



98 T01L7C06



98 T02L6C28



98 T02L4C21



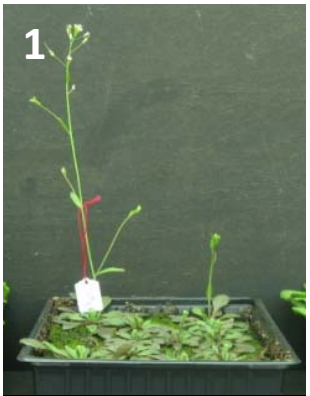
98 T03L7C17



98 T01L2C06

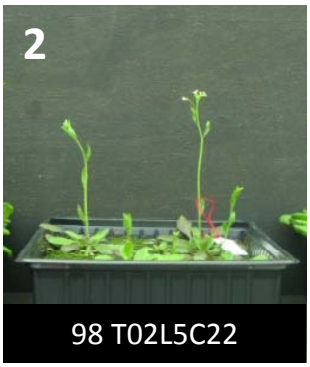


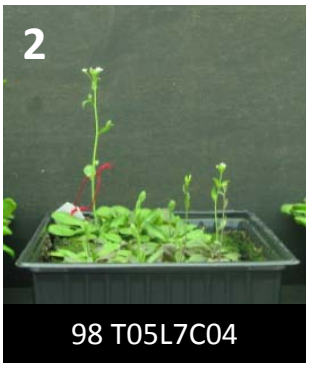
98 T05L2C17



98 T03L2C01

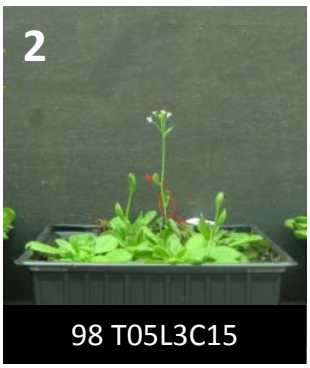
epiRIL 98 : candidats sélectionnés (partie 2)

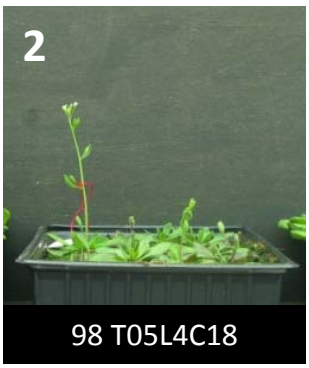


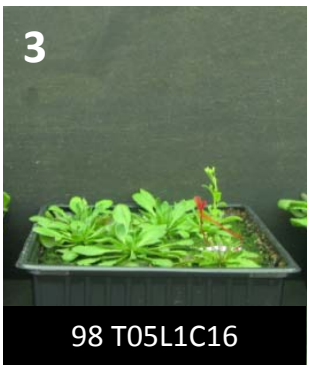


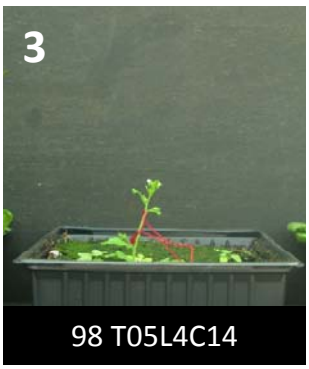






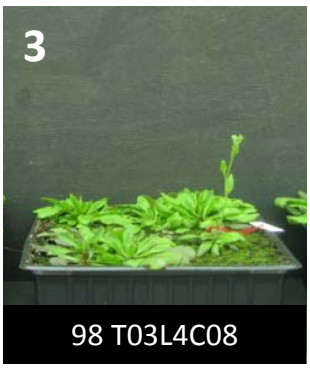












Abstract

Assessing the control of transposable elements by DNA methylation in *Arabidopsis thaliana*

Transposable elements (TEs) and their relics are major components of eukaryotic genomes. TEs are potentially highly mutagenic as their proliferation can cause chromosomal rearrangements, disrupt genes or affect gene expression through transcriptional interference. However, few TEs are usually mobile within genomes at any one time thanks to potent mechanisms that restrain their activity, such as DNA methylation in mammals and plants. Thus, in the flowering plant Arabidopsis, severe loss of DNA methylation caused by mutations in the chromatin remodeler gene *DDM1* triggers massive accumulation of transcripts corresponding to TEs. Yet, comparatively few TEs appear to be mobilized as a result. Here, we set out to determine (1) the extent to which DNA methylation prevents TE mobilization, (2) where do TEs insert following their reactivation and (3) what are the consequences of new TE insertions on the expression of neighboring genes. To this ends, we have sequenced the genome of over 50 epigenetic Recombinant Inbred Lines (epiRILs) that were derived from a cross between a wild type and an isogenic *ddm1* mutant line. After backcrossing of the F1 and selection of the progeny homozygous for wild-type *DDM1*, the epiRILs were propagated through six rounds of selfing. The epiRILs therefore permit a detailed assessment of transposition events soon after they have occurred. In order to identify TE mobilization in the epiRILs we developed TE-tracker, a pipeline based on Illumina sequencing of mate pairs libraries. Using this approach, we could show that although both retroelements and DNA transposons are mobilized in *ddm1* and the epiRILs, mobile TEs belong to only a dozen or so of the >300 TE families identified in the Arabidopsis genome. Furthermore the rate and dynamics of transposition vary dramatically between TE families, suggesting the existence of additional mechanisms controlling transposition. Molecular analysis performed on a subset of those mobile TEs belonging to distinct families show that these differences are greatly due to different modality of establishment of epigenetic control over newly inserted TEs. In addition, our analysis indicates that the distribution of new TE insertions differs dramatically from the one of resident copies. These findings provide compelling evidence that over accumulation of TE sequences in the pericentromeric regions of the Arabidopsis genome is not due to specific targeting of TEs, but rather to their elimination from chromosome arms. Finally, we assessed the extent to which repeat-associated DNA methylation impacts the expression of neighboring genes by studying mutants affected in different methylation pathways. Phenotypic and molecular analyses reveal that, even though most Arabidopsis genes are not detectably sensitive to the methylation status of neighboring repeats, two DNA methylation pathways act together to maintain the normal expression of only a very small number of genes near repeats and that these genes tend to have pleiotropic effects. Then, DNA methylation acts as a double-lock system to ensure the normal expression of a small number of key genes located near repeats.

Résumé

Etude du contrôle des éléments transposables par la méthylation de l'ADN chez *Arabidopsis thaliana*

Les éléments transposables (ET) et leur reliques sont des composants majeurs des génomes eucaryotes. Ils sont potentiellement hautement mutagéniques car leur prolifération peut engendrer des réarrangements chromosomiques, des interruptions de gènes ou affecter l'expression génique par interférence transcriptionnelle. Néanmoins, peu d'ET sont généralement mobiles dans les génomes grâce à l'action de mécanismes qui restreignent leur activité comme la méthylation de l'ADN chez les mammifères et les plantes. De fait, chez *Arabidopsis thaliana*, une perte sévère de méthylation de l'ADN causée par une mutation dans le gène codant la protéine remodeleuse de chromatine DDM1 (DECREASE IN DNA METHYLATION 1) engendre l'accumulation massive de transcrits correspondants à des séquences d'ET. En revanche, peu d'ET semblent être mobilisés suite à cette réactivation transcriptionnelle. Nous proposons ici de déterminer (1) l'étendue de la mobilisation des ET suite à la perte de méthylation de l'ADN, (2) la distribution des nouvelles insertions d'ET le long du génome d'Arabidopsis et (3) les conséquences des nouvelles insertions d'ET sur l'expression des gènes situés à proximité. Dans ce but, nous avons séquencé le génome d'une cinquantaine d'epiRIL (epigenetic Recombinant Inbred Lines) dérivées d'un croisement entre une plante sauvage et un mutant *ddm1.* Suite au croisement retour de la F1 avec une plante sauvage et sélection des individus F2 homozygotes pour l'allèle sauvage *DDM1*, les epiRIL ont été propagées au travers de 6 autofécondations successives. Les epiRIL permettent donc l'étude détaillée des évènements de transpositions juste après qu'ils aient eu lieu. Pour identifier les évènements de transpositions dans ces lignées nous avons mis au point TE-tracker, un programme basé sur les données issues du séquençage Illumina de banques maite-pair. Par cette approche, nous avons montré que les ET mobiles dans *ddm1* et les epiRIL appartiennent à seulement une quinzaine environ des >300 familles identifiées dans le génome d'Arabidopsis. Qui plus est, on observe des variations importantes de fréquences et dynamiques de transpositions entre les différentes familles d'ET ce qui suggère l'existence de mécanismes additionnels contrôlant la transposition. Les analyses moléculaires réalisées sur un sous-ensemble des ET mobilisés appartenant à différentes familles ont notamment montré que ces différences sont dues en grande partie aux différentes modalités d'établissement du contrôle épigénétique sur les ET nouvellement insérés. D'autre part, nos analyses indiquent que la distribution des nouvelles insertions d'ET diffère grandement de celle des copies résidentes. Ce résultat suggère donc que la suraccumulation des séquences d'ET dans les régions péricentromériques du génome d'Arabidopsis n'est pas due à un ciblage spécifique des insertions dans ces régions, mais est plutôt la conséquence de leur élimination des bras chromosomiques. Enfin, nous avons cherché à déterminer dans quelle mesure la méthylation de l'ADN associée aux séquences répétées a un impact sur l'expression des gènes situés à proximité en étudiant des mutants affectés dans les différentes voies de la méthylation de l'ADN. Par des analyses phénotypiques et moléculaires nous avons montré que, même si la plupart des gènes d'Arabidopsis n'est pas affectée par l'état de méthylation des séquences répétées situées à proximité, deux voies de la méthylation de l'ADN agissent ensemble pour maintenir l'expression normale d'un petit nombre de gènes ayant des effets pléiotropes situés proximité de séquences répétées. La méthylation de l'ADN agit donc comme un double système de contrôle pour assurer l'expression normale d'un petit nombre de gènes clefs localisés a proximité de séquences répétées.