

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 réparti 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.

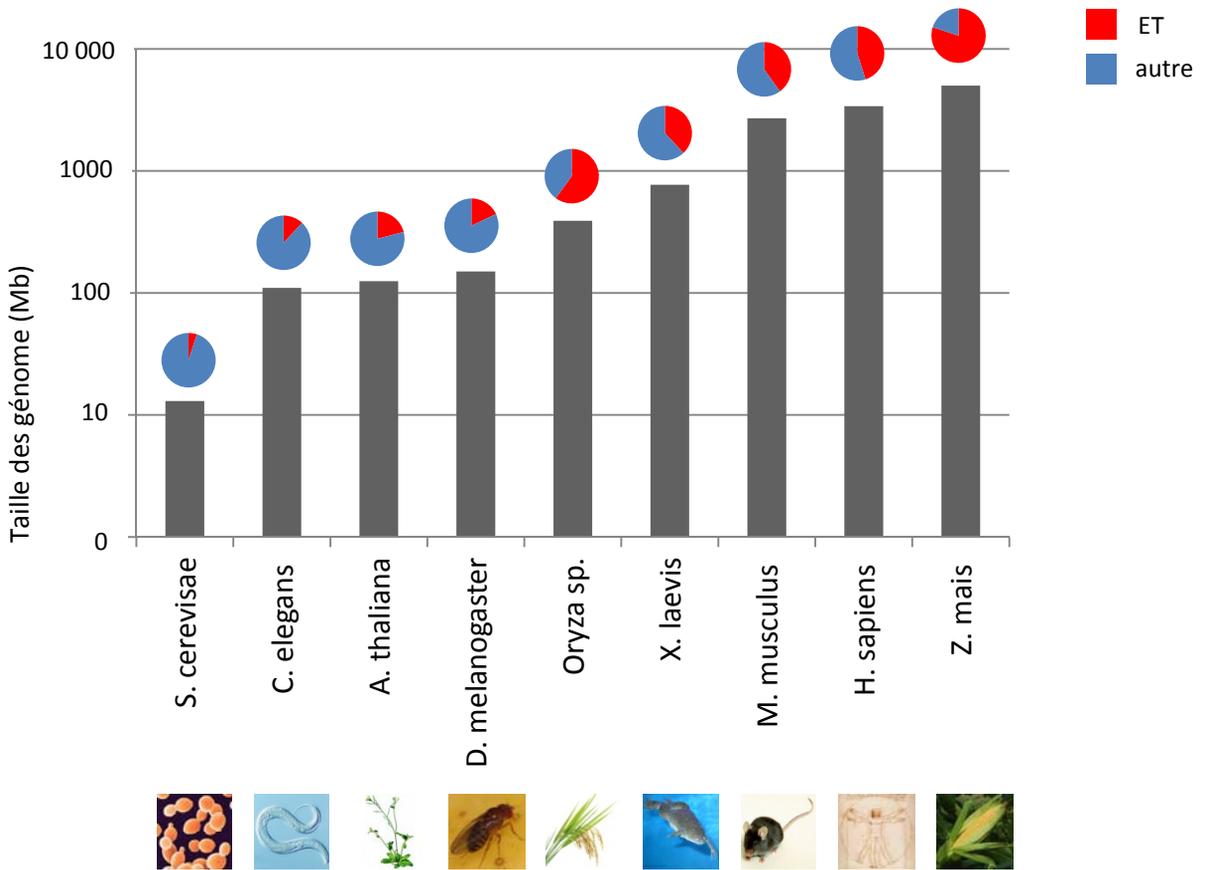


Figure 1.1 : Variation de la taille des génomes et de la fraction composée d'ET entre différentes espèces.

D'après Vierra et al. 2002, Tenailon et al. 2010 et Buisine et al. 2008.

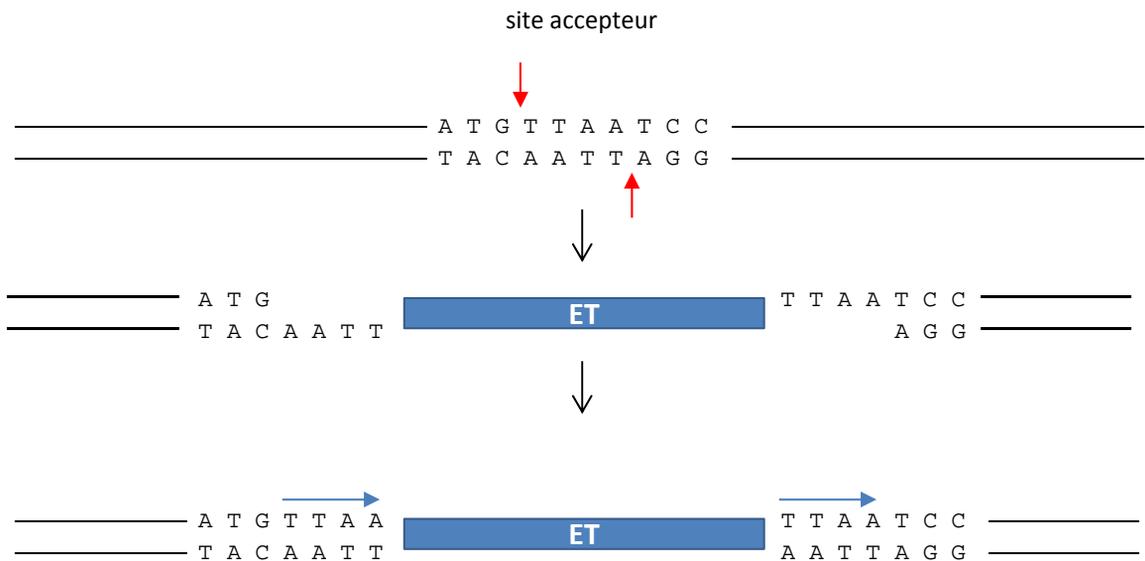


Figure 1.2 : Création de la Target Site Duplication lors de l'insertion d'un ET.

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 maïs. 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 été 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. cerevisiae* et pouvant atteindre jusqu'à 45% chez l'Homme et >80% chez certains 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.

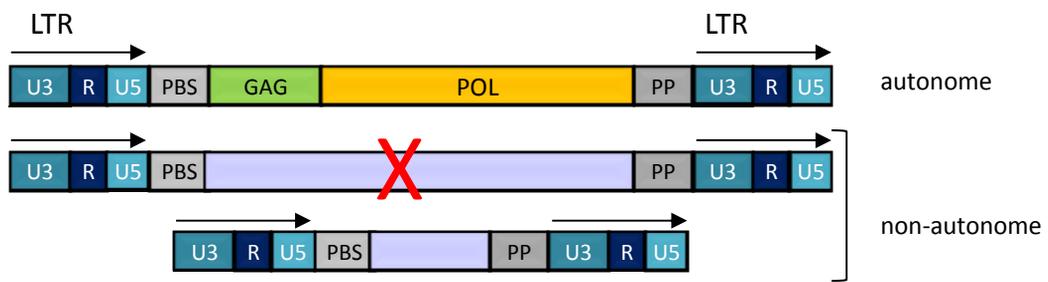


Figure 1.3: Structure générale des rétrotransposons à LTR.

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.

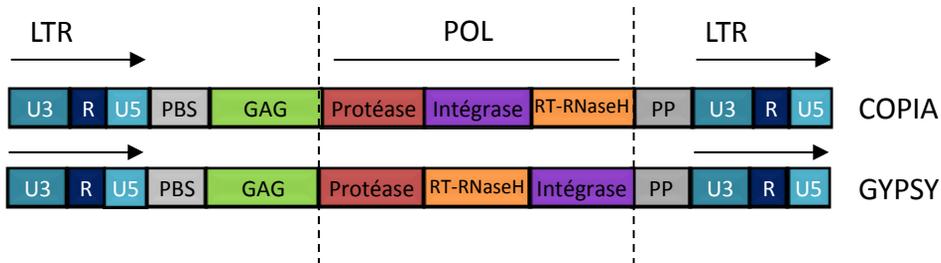


Figure 1.4 : Structure détaillée des rétrotransposons à LTR GYPSY et COPIA.

Adapté de Havecker, Gao and Voytas 2004 et Sabot and Schulman 2006.

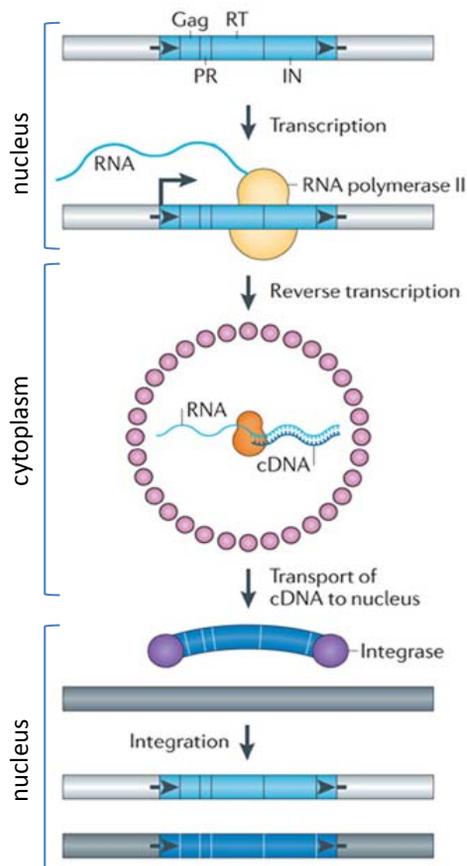


Figure 1.5 : Mécanisme de transposition des rétroéléments à LTR.

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 capsid, 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

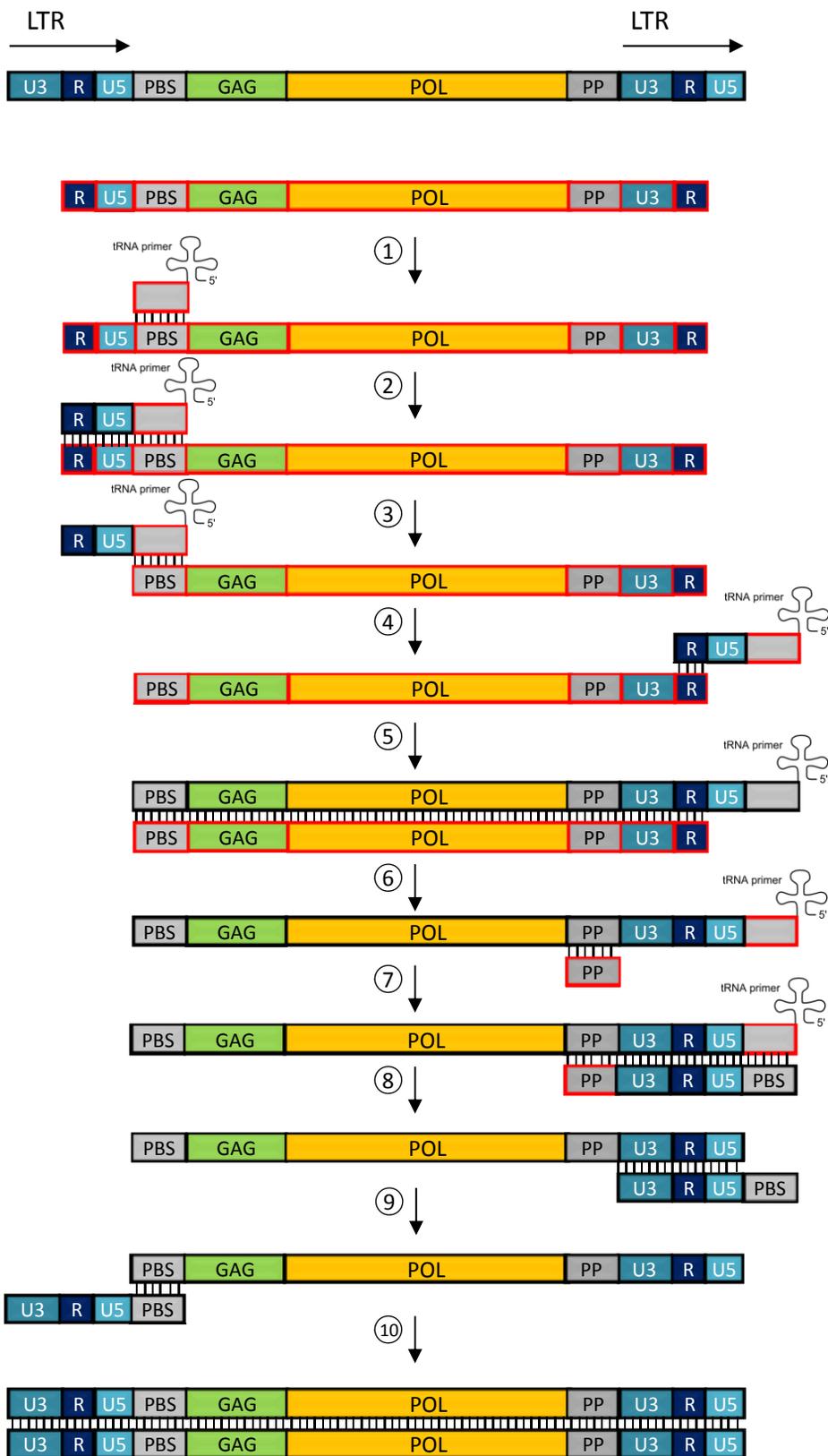


Figure 1.6 : Etapes de la transcription inverse des rétroéléments à LTR et des rétrovirus.

(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 capsid, dans le cytoplasme. Deux ARN sont en général présents au sein d'une même capsid 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 rétroéléments 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.

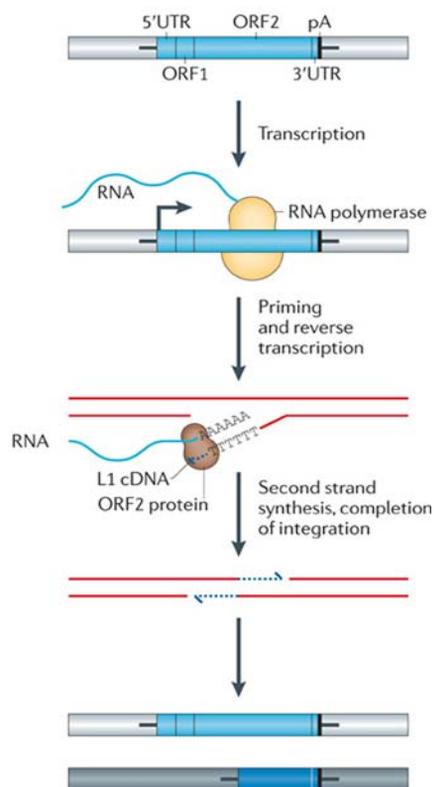


Figure 1.8 : Mécanisme de transposition des rétroéléments sans LTR.

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 rétroéléments, le mode couper/coller n'induit pas systématiquement une augmentation du nombre de copies. En

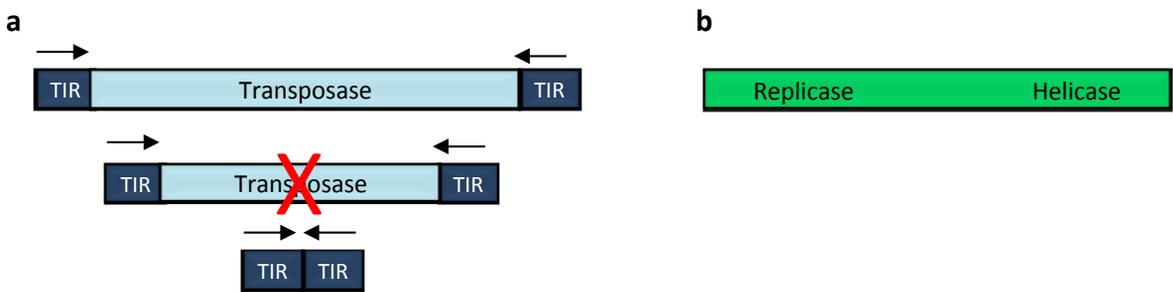


Figure 1.9 : Structure générale des transposons à ADN.

(a) Transposons à ADN de la sous-classe 1. (b) Helitrons (sous-classe 2). Adapté de Feschotte et al. 2002.

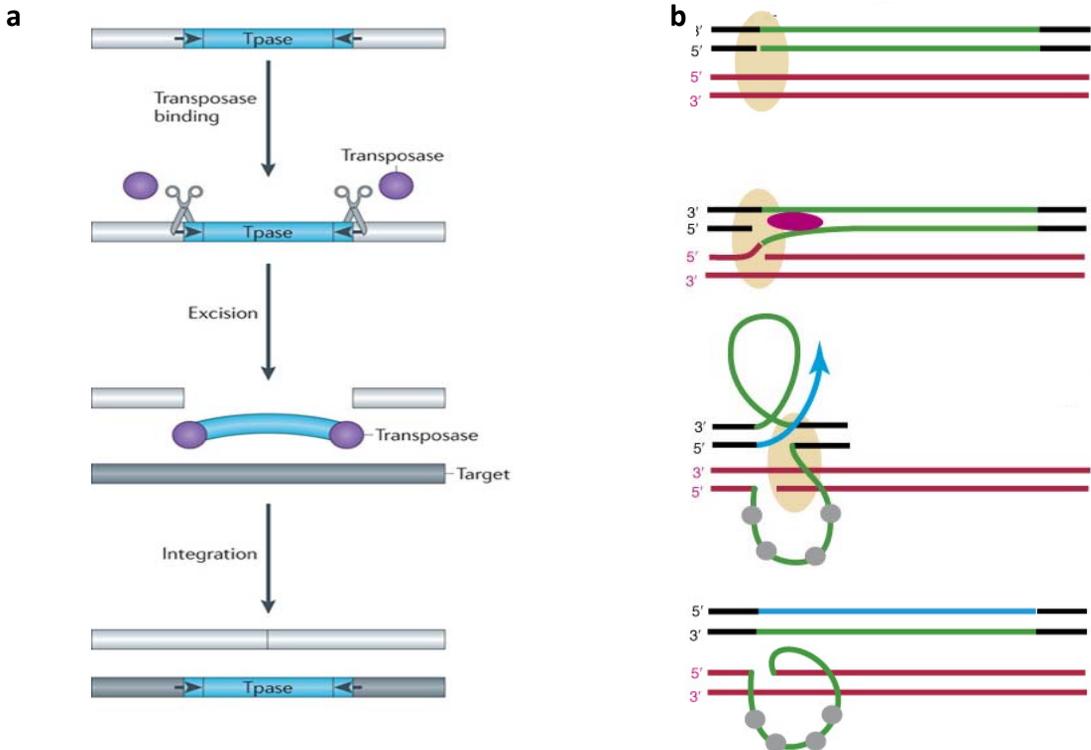


Figure 1.10 : Mécanisme de transposition des transposons à ADN.

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

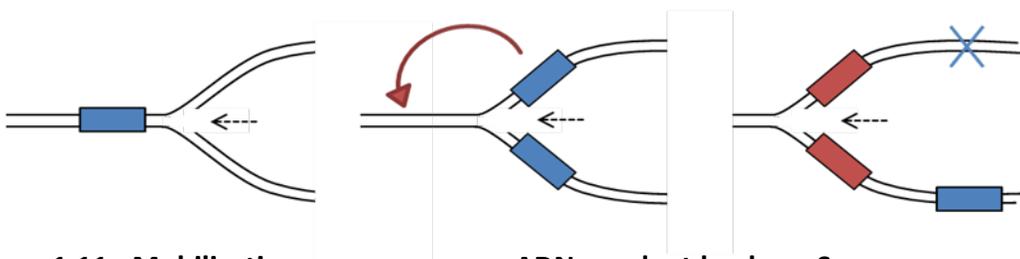


Figure 1.11 : Mobilisation des transposon ADN pendant la phase S.

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 reconnaître 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 réplcatif, encore peu connu, se ferait par un mécanisme de cercle roulant (rolling circle) de façon analogue aux transposons bactériens (fig. 1.10b). Il apparaît donc que, bien que classés dans le type II, les hélitrons ne présentent aucune similarité que ce soit

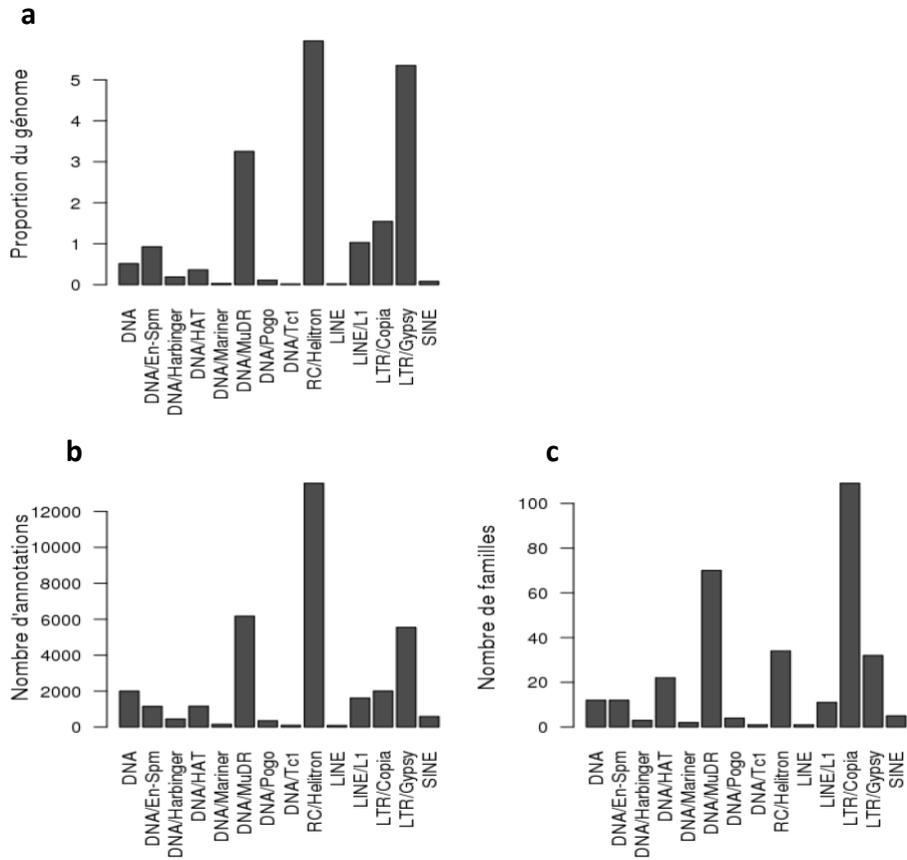


Figure 1.12 : Diversité des superfamilles d'ET d'Arabidopsis.

(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.

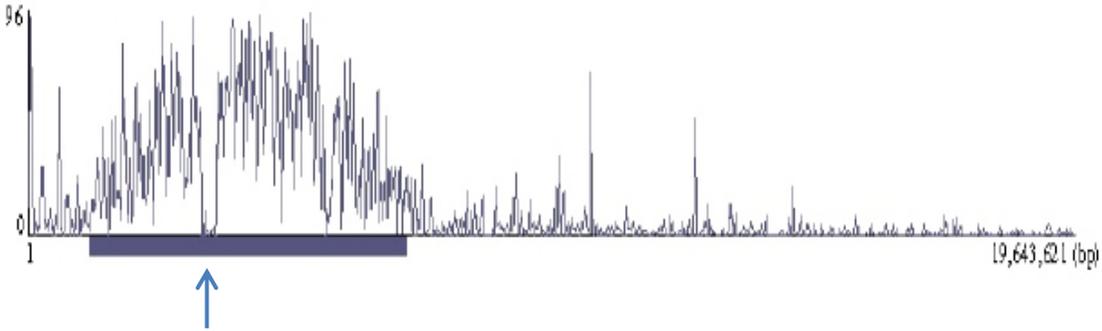
1.2.3 Les ET d'*Arabidopsis thaliana*

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 à 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 à 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

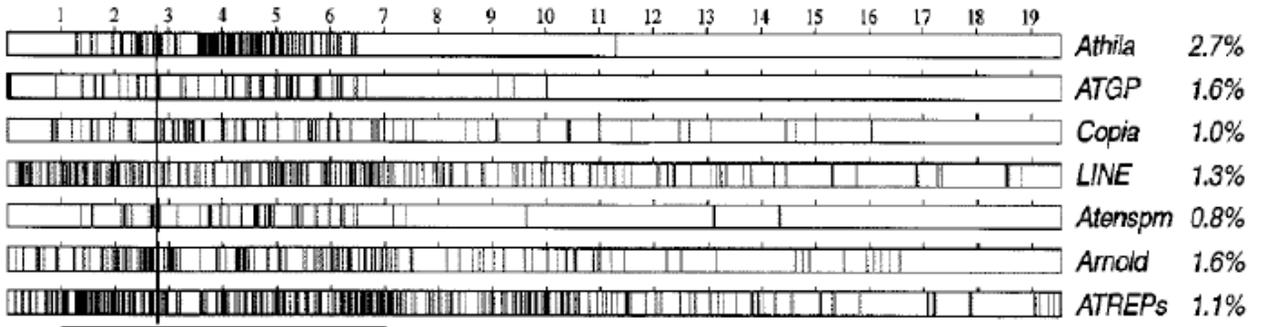


Figure 1.13 : Distribution des ET le long du chromosome 2 d'*Arabidopsis thaliana*.

(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 péri-centromériques 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 précentromériques 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 événements 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 îlots 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 maïs 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).

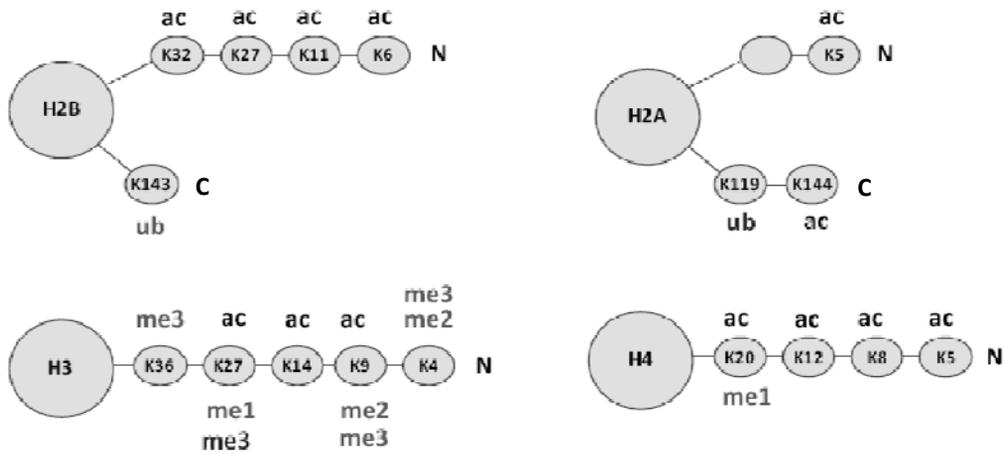


Figure 1.14 : Principales marques d'histones présentes chez Arabidopsis.

Représentation non exhaustive des résidus des histones H2A, H2B, H3 et H4 pouvant être acétylés (ac), méthylés (me), ou ubiquitinés (ub) chez Arabidopsis.

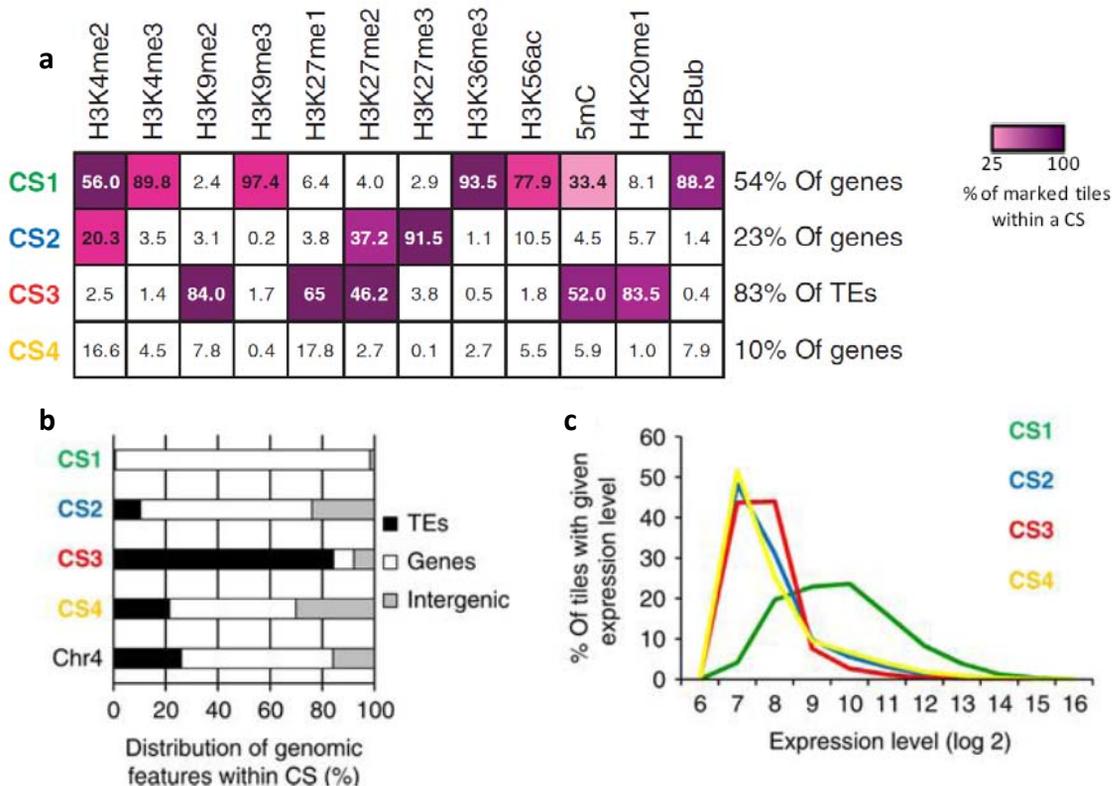


Figure 1.15 : Etats chromatinien chez Arabidopsis.

(a) Composition des 4 principaux états chromatinien (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 chromatinien 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 chromatinienne 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 transcriptionnellement actifs. Il est caractérisé par la di- et la triméthylation de la lysine 4 de l'histone H3 (H3K4me2

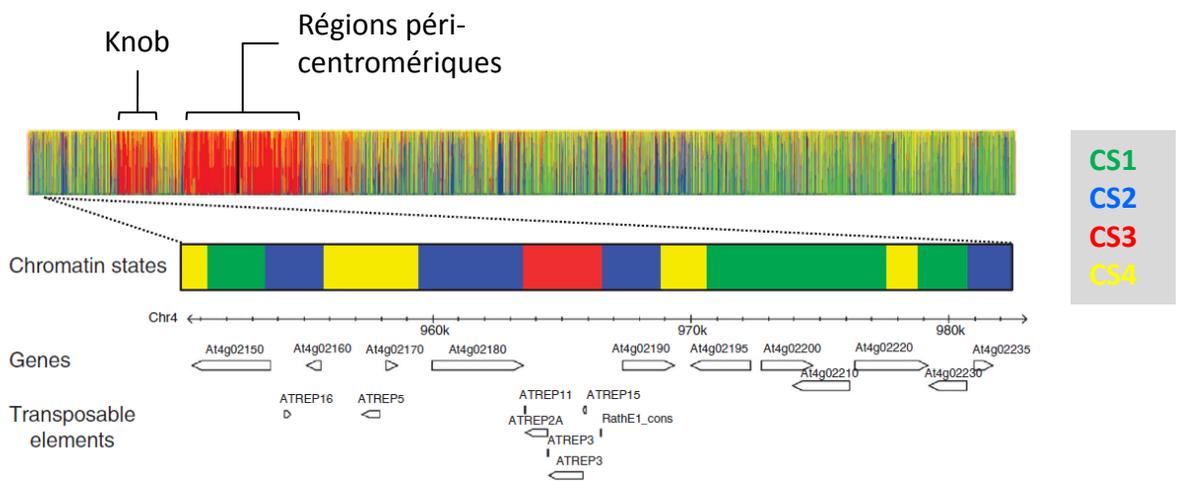


Figure 1.16 : Distribution des quatre états chromatinien le long du chromosome 4 d'Arabidopsis.

Les états chromatinien forment généralement des domaines de petite taille à l'exception du CS3 qui forme également de larges blocs au niveau des régions péri-centromériques 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 chromatinien 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 péri-centromériques 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 îlots d'hétérochromatine (fig. 1.16).

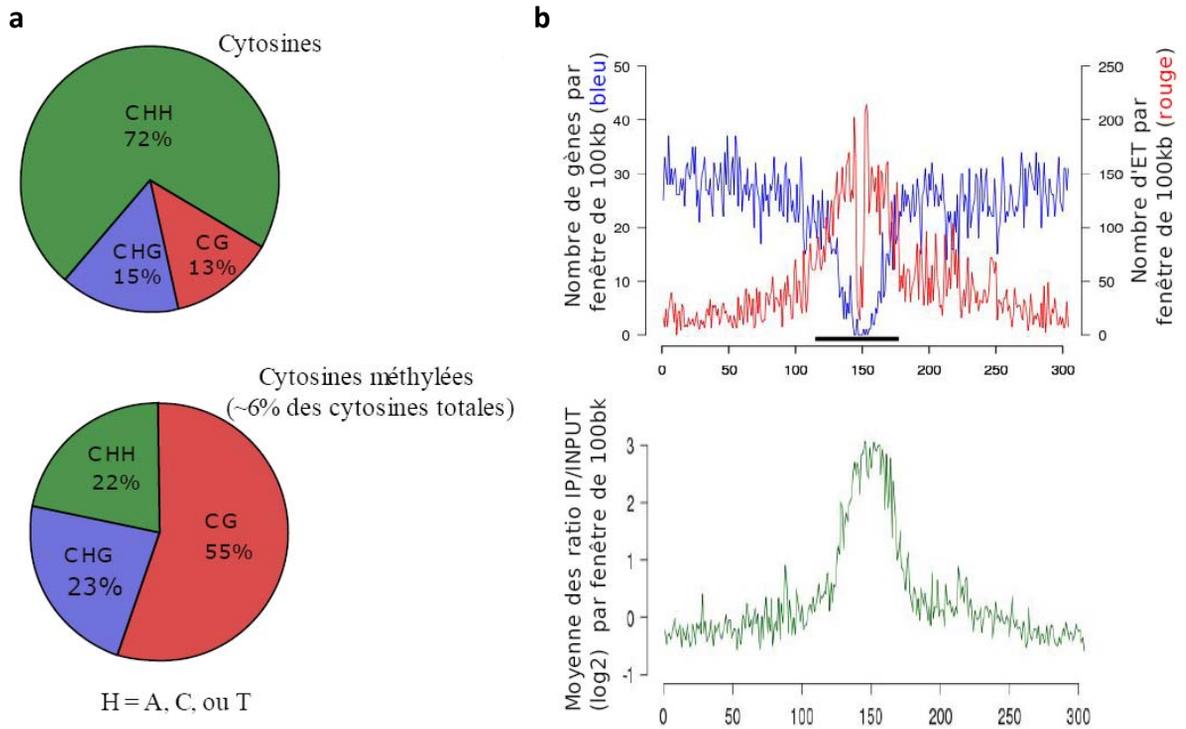


Figure 1.17 : Méthylation de l'ADN chez *Arabidopsis thaliana*.

(a) Répartition des cytosines du génome (en haut) et des cytosines méthylées (en bas) dans les trois contextes CG, CHG et CHH. D'après 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*.

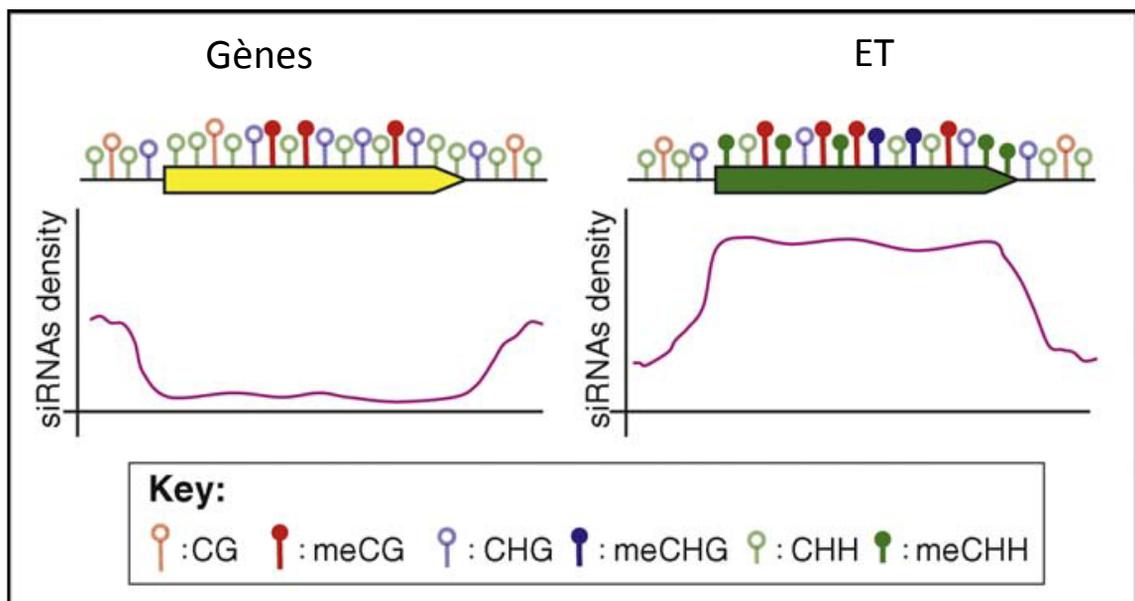


Figure 1.18 : Profils types de méthylation des cytosines le long des gènes et des ET chez *Arabidopsis thaliana*.

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 îlots 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 reconnaît 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 hémiméthylés (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

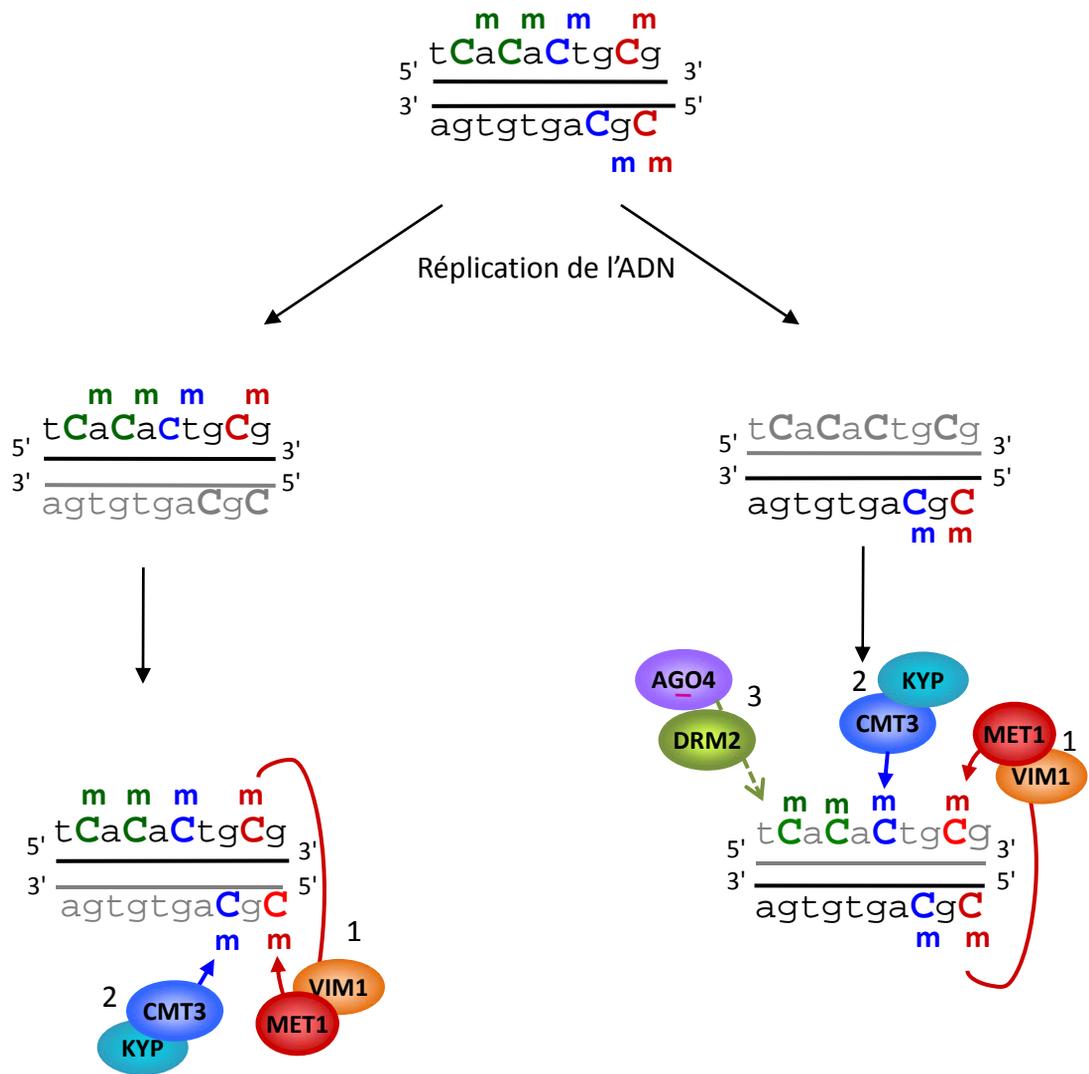


Figure 1.19 : Modèle du maintien de la méthylation de l'ADN sur les ET chez Arabidopsis.

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

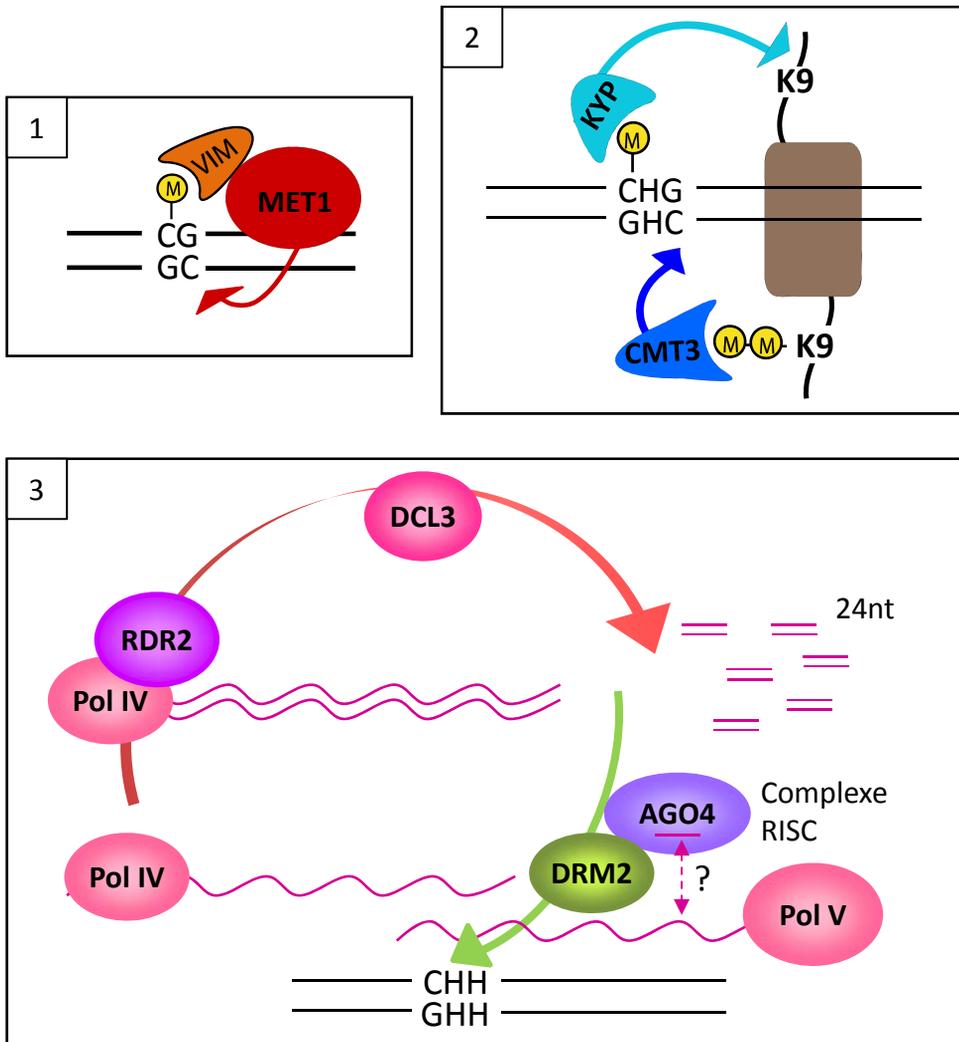


Figure 1.20 : Les trois voies de maintenance de la méthylation de l'ADN.

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'environ 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 remodelleuse 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, Jeddloh 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 reconnaît ses cibles. Ainsi, les données génétiques suggèrent l'existence d'une étape antérieure permettant à POLIV de reconnaître 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,

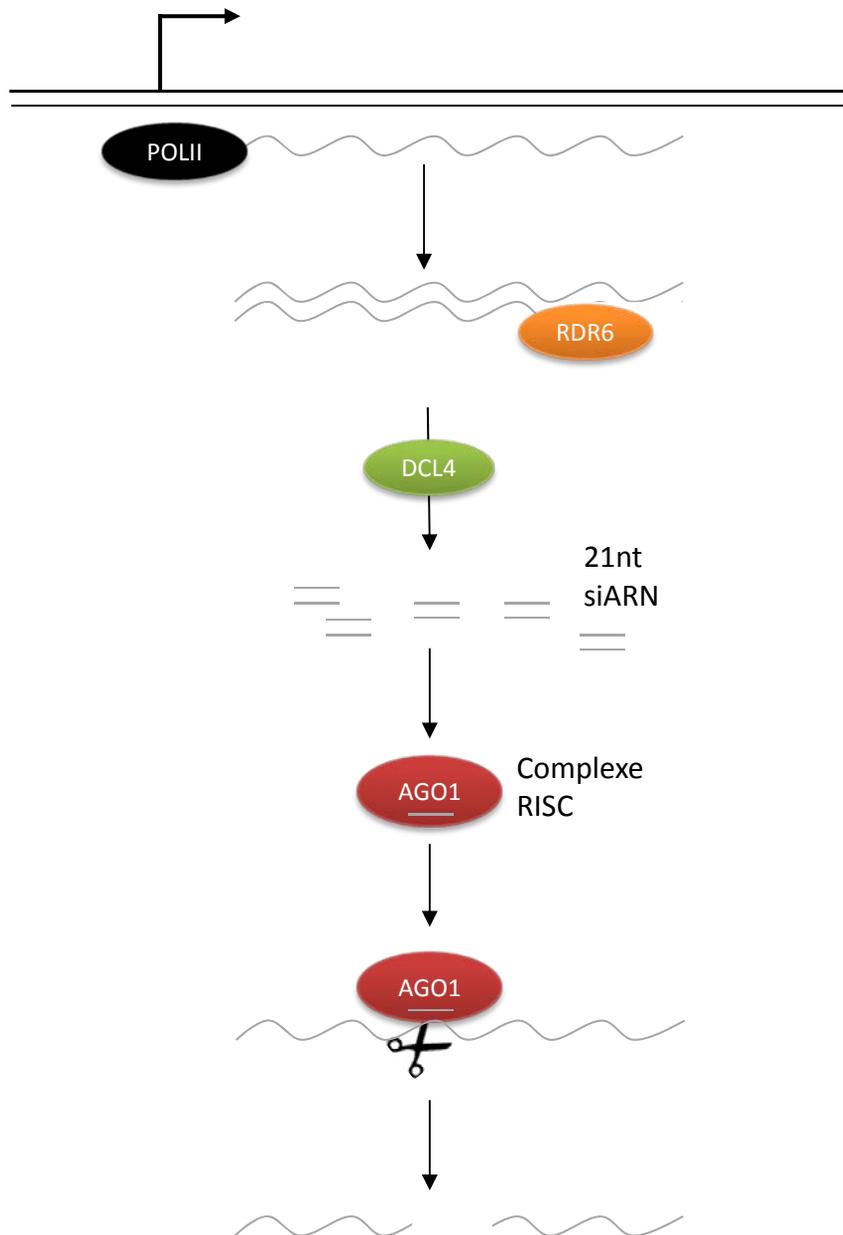


Figure 1.21 : Etapes principales de la voie du Sens-PTGS (SPTGS).
 Adapté de Brodersen et al. 2006.

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 réplifications. 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

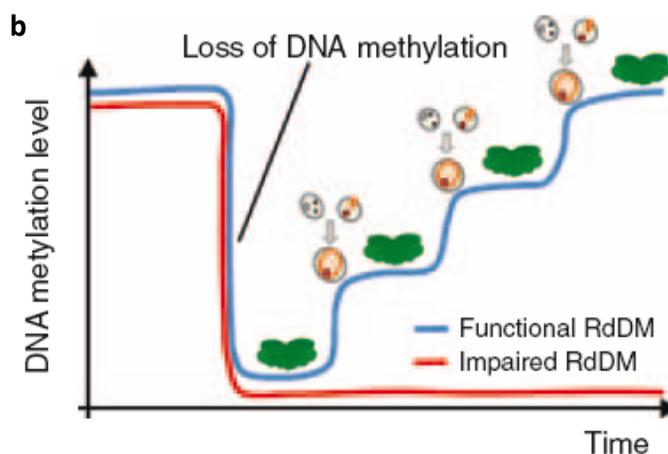
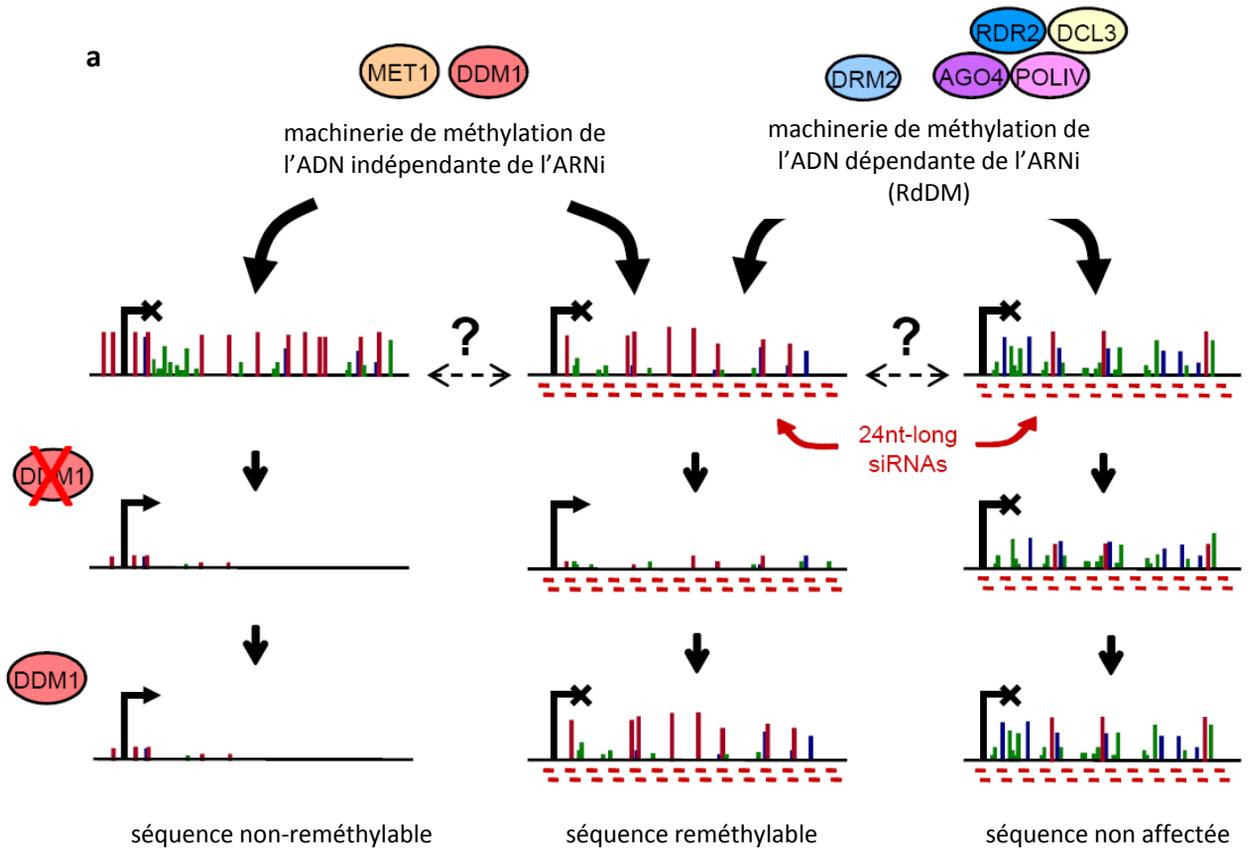


Figure 1.22 : Processus de restauration de la méthylation de l'ADN suite à une perte induite par une mutation dans le gène *DDM1*.

(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 chromatinienne 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 Microorchidia) 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

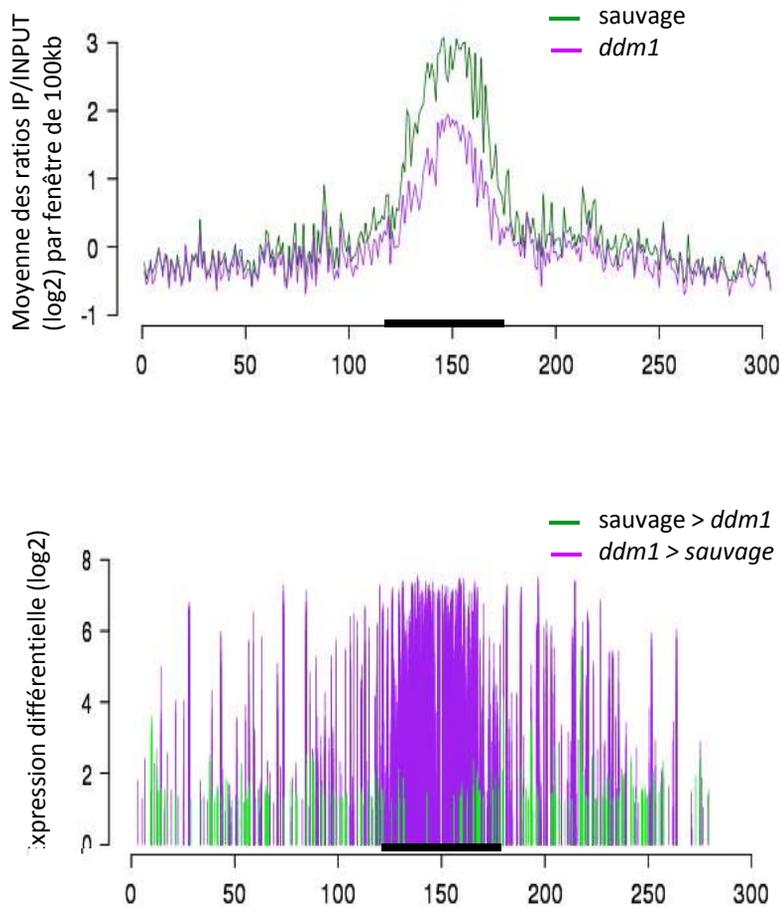


Figure 1.23 : La perte de méthylation induite par une mutation dans le gène *DDM1* cause la réactivation transcriptionnelle de très nombreux ET.

(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 pertes 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 apparaît 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 gaméophyte 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

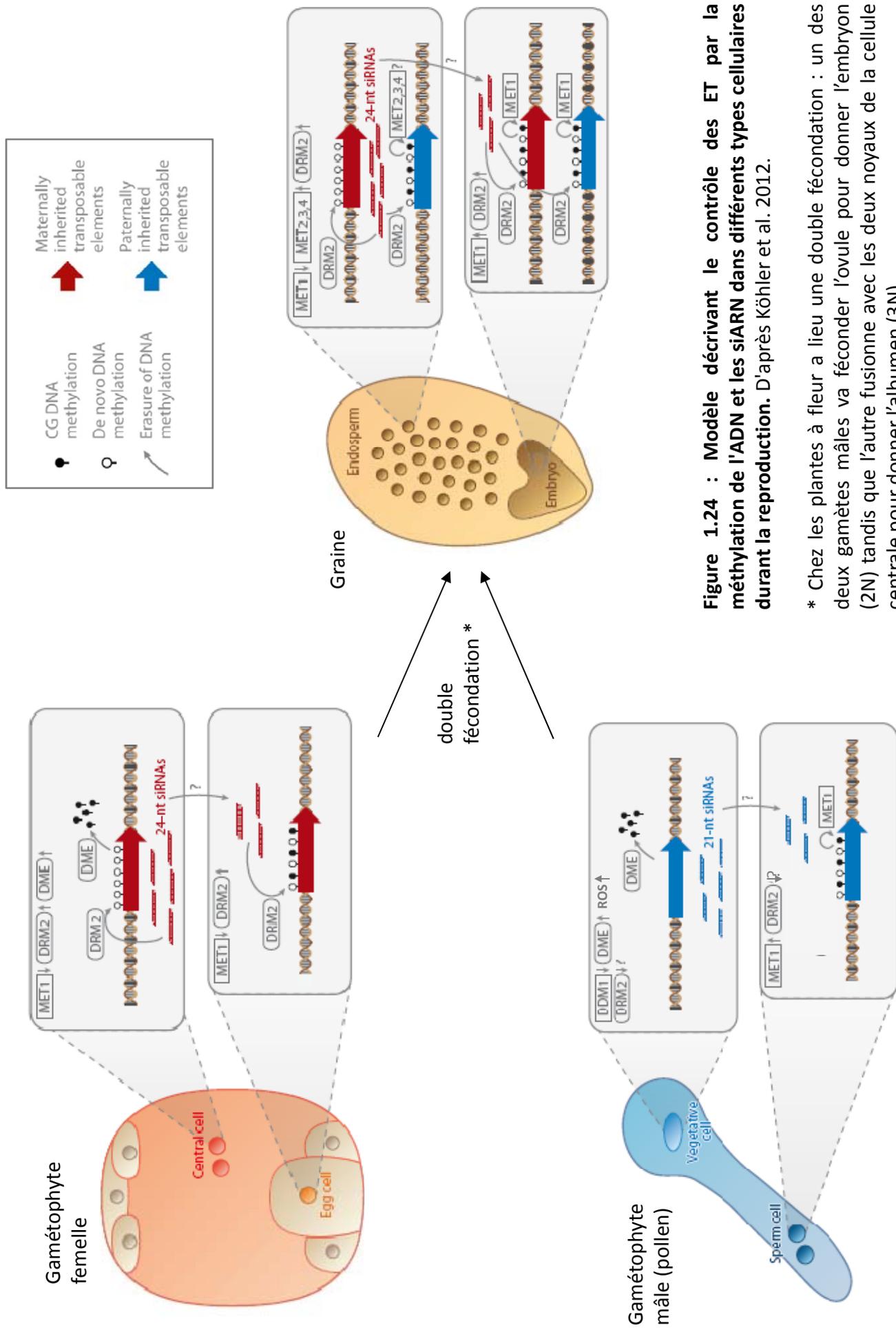


Figure 1.24 : Modèle décrivant le contrôle des ET par la méthylation de l'ADN et les siARN dans différents types cellulaires durant la reproduction. D'après Köhler et al. 2012.

* 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 côté 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 transcriptionnelle 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 péri-centromériques (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.

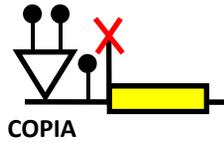
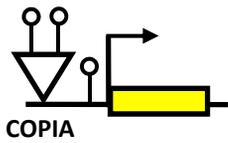
Des études récentes ont mis en évidence une corrélation négative chez *A. thaliana* entre le niveau d'expression des gènes et la densité d'ET méthylés (Hollister and Gaut 2009). De plus, des analyses basées sur des comparaisons entre *Arabidopsis thaliana* et *lyrata* à l'échelle du

génomique 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 (environ 200 bp) à partir des séquences ciblées par le RdDM (Ahmed et al. 2011).

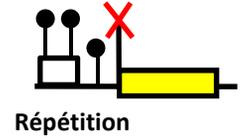
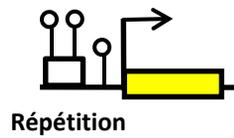
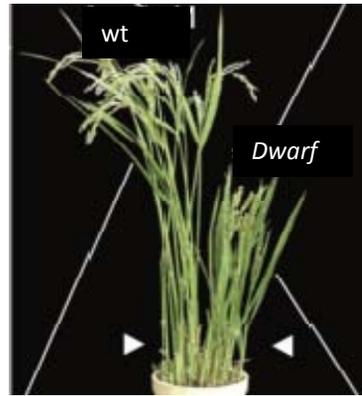
Il semble donc que l'état de méthylation des ET puisse jouer un rôle dans la régulation des gènes situés à proximité. Autrement dit, ces gènes vont potentiellement être induits/surexprimés dans les conditions développementales ou environnementales induisant l'hypométhylation des ET. Une fois encore, 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 engendrent 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).

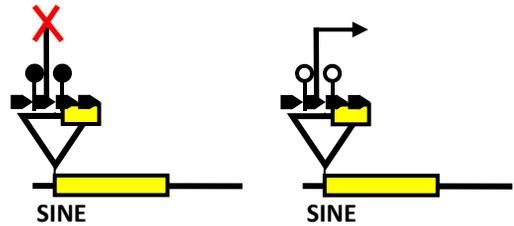
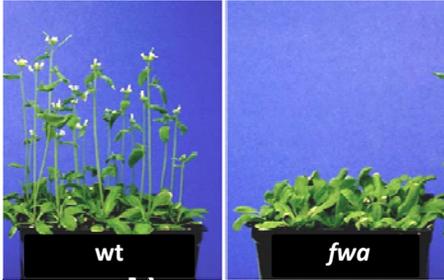
présence d'un transposon à ADN hAT en 3' du gène CmWIP1. La présence de l'ET méthylé entraîne la méthylation stable du promoteur de CmWIP1 et son extinction transcriptionnelle. Ce gène étant nécessaire à l'induction des étamines, son extinction a pour conséquence la production de fleurs femelles exclusivement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

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

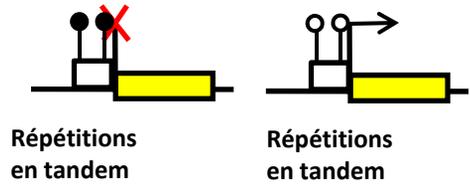
Une épimutation est une variation héritable de l'expression d'un gène qui n'est pas due à un changement de la séquence nucléotidique. De par leur rôle potentiel dans la régulation des gènes situés à proximité, les ET sont une source privilégiée d'épimutations. De fait, parmi les quelques épimutations 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)

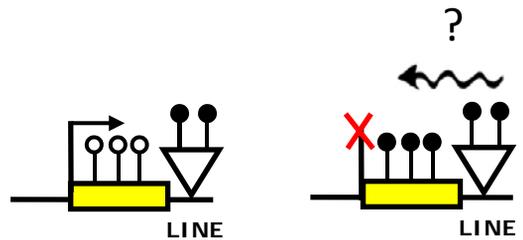


Figure 1.26 : Exemples de dérégulations géniques et leurs conséquences phénotypiques induites par des mutations affectant les voies de la méthylation de l'ADN chez Arabidopsis.

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'entraînant 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*. Étonnamment, 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 *SUPPRESSOR 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 été 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 hérissables 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 événements 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.