

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.

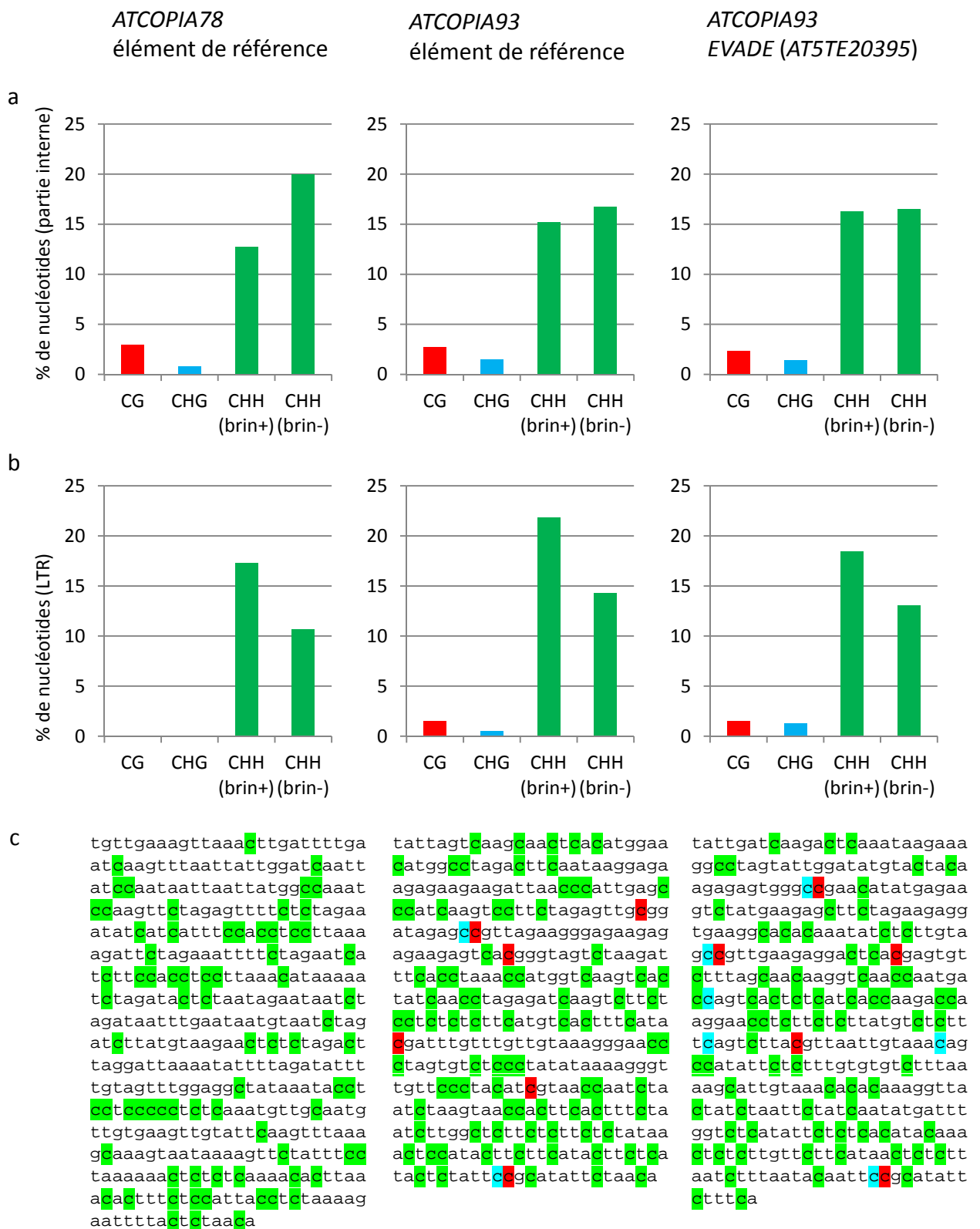


Figure 3.1 : Contexte nucléotidique des cytosines présentes dans *ATCOPIA78* et *ATCOPIA93*.

(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.

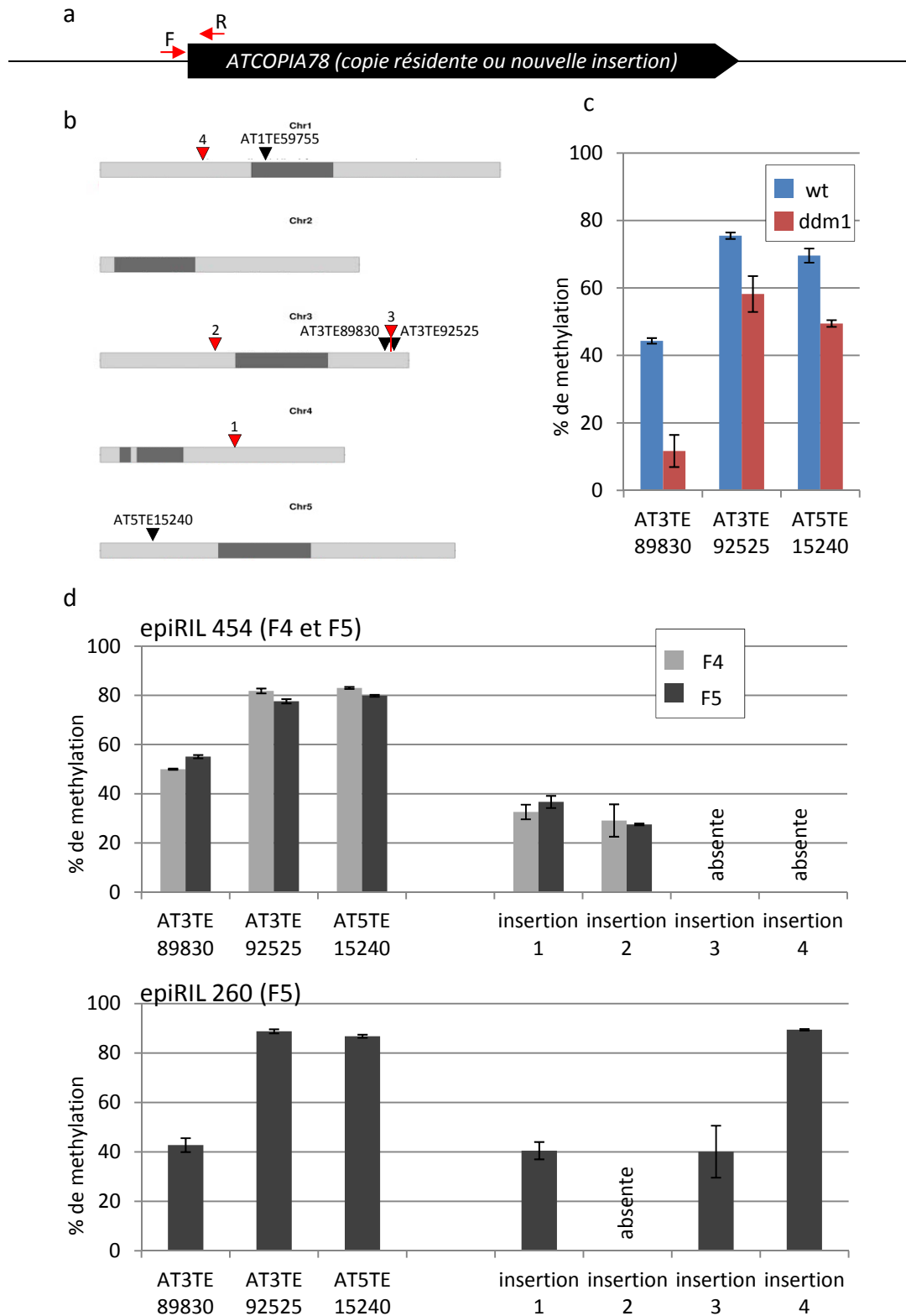


Figure 3.2 : Etablissement/Rétablissement du contrôle épigénétique sur les copies nouvellement insérées/résidentes de *ATCOPIA78*.

(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-transcriptionnel (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 apparaît 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 Colot³ & Olivier Voinnet^{1,2}

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 manner⁵, 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 strands⁶; *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 heterochromatin⁸.

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 transgenes¹⁵ 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 caveats¹⁶ 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*

¹Swiss Federal Institute of Technology (ETH-Z), Department of Biology, Zurich, Switzerland. ²Life Science Zurich Graduate School, Molecular Life Science Program, University of Zurich, Zurich, Switzerland. ³Institut de Biologie de l'École 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. ⁴Present 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. ⁵These 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 reactivation¹⁹.

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* mobilization¹⁸, 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 (F₈) 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 F₈ and F₁₁ (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.

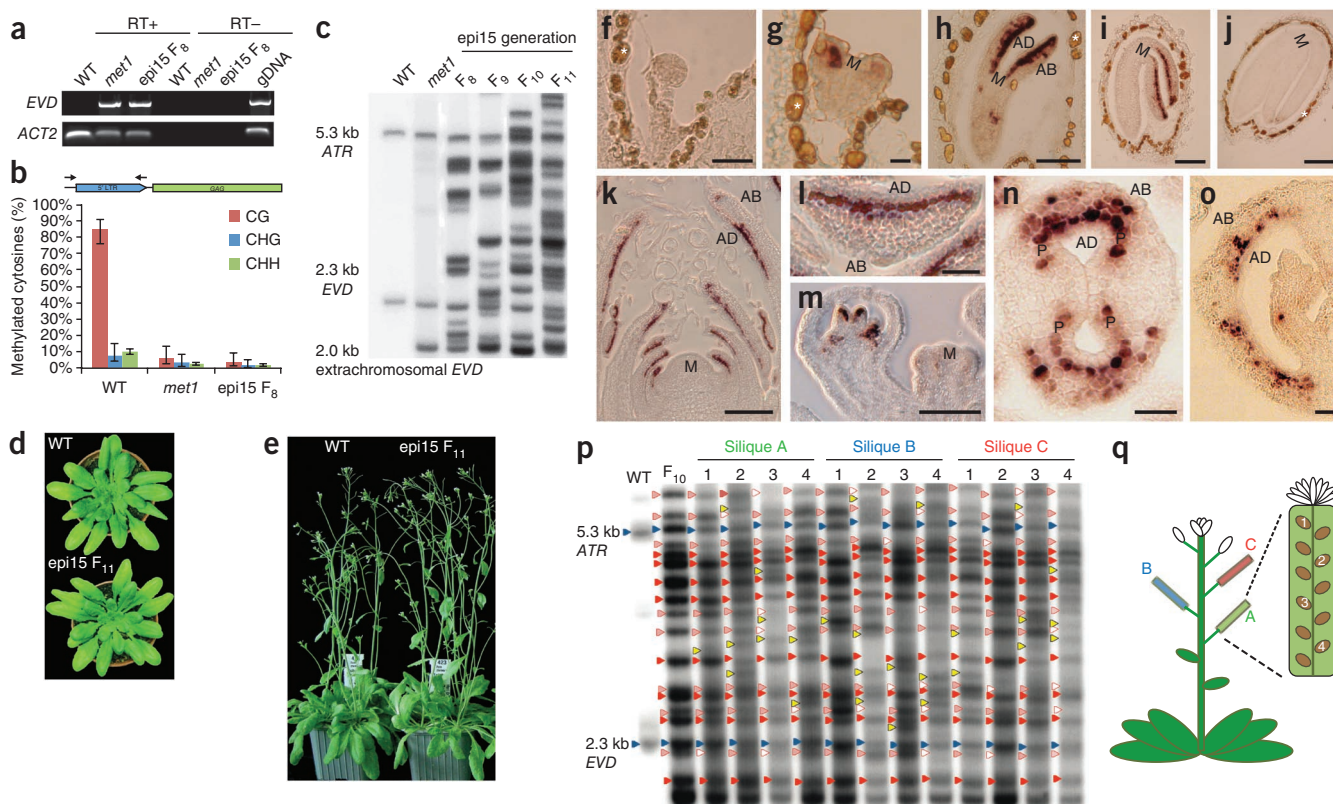


Figure 1 *EVD* transcript accumulation pattern. (a) RT-PCR analysis of *EVD* RNA levels in wild-type, *met1* and epi15 F₈ plants, normalized to *ACT2* (*ACT2*) levels. WT, wild type; RT, reverse transcriptase; gDNA, genomic DNA. (b) Bisulfite sequencing-based methylation analysis at CG, CHG and CHH sites of the *EVD* 5' LTR in wild-type, *met1* and epi15 F₈ 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 F₁₁ plants during vegetative (d) or reproductive (e) growth. (f–j) *In situ* hybridization of *EVD* RNA in an F₁₀ 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,i,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 F₁₀ 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).

To gain insight into the *EVD* life cycle in its host, we investigated *EVD* RNA accumulation patterns in epi15 F₁₀ 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 gynoceia—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 gametophyte²⁴ (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 F₁₀ 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 F₈ and F₁₁ 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* T₁ 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 F₈ and F₁₁ (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, respectively²⁷. Introducing 35S:*EVD* into the *rdp6-12* 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 F₈ onward (Fig. 2f), *EVD* expression and mobilization kept increasing from F₈ to F₁₁, 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 P19tm 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 RNA³². In *Saccharomyces cerevisiae*, VLP coating protects Ty3 LTR retrotransposon RNA from endonucleases via the conserved zinc-finger CCHC motif (CX₂CX₄HX₄C, 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*^{ARBD} (Fig. 2g). Having confirmed cognate *EVD*^{ARBD} polyprotein processing and *Gag*^{ARBD} stability (Supplementary Fig. 5b,c), 35S:*EVD*^{ARBD} was introduced into plants with or without expression of P19 or non-functional P19tm. Without P19, *EVD*^{ARBD} mRNA levels were strongly reduced compared to those of wild-type *EVD* that remained unaltered irrespective of P19 or P19tm (Fig. 2h, lanes 3 and 4 and lanes 5 and 6). In contrast, mRNA from *EVD*^{ARBD} reached wild-type levels with expression of P19 but not P19tm (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 F₁₁ and F₁₂, respectively, coinciding with the highest 3' *gag*-derived siRNA levels (Fig. 3a,b). *EVD* RNA levels declined beyond generation F₁₁, until their pronounced loss in generation F₁₄, a loss that was not attributable to reduced numbers of *EVD* insertions, as a plateau of ~40 copies per genome was attained from generation F₁₂ onward (Fig. 3a,b). F₁₄ 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 F₁₄; accordingly, F₁₄ coincided with the onset of LTR-derived 24-nt siRNAs (Fig. 3b,c). Initially undetectable

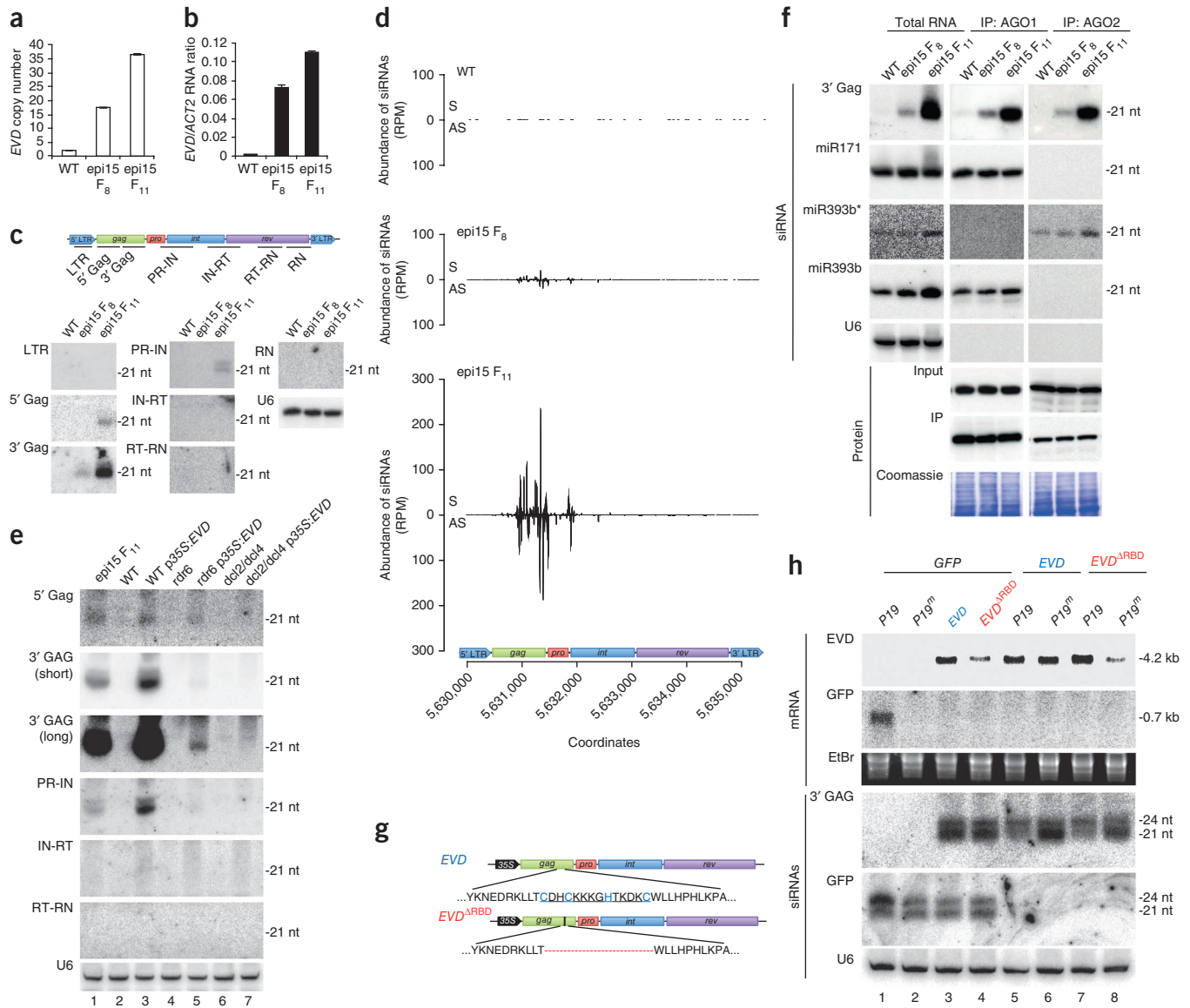


Figure 2 Antiviral PTGS is activated against and suppressed by *EVD*. **(a)** *EVD* copy number analysis by quantitative RT-PCR (qRT-PCR) in wild-type, epi15 F₈ and epi15 F₁₁ plants. **(b)** qRT-PCR analysis of *EVD* RNA levels in wild-type, epi15 F₈ and epi15 F₁₁ plants, normalized to *ACT2* levels. Error bars, s.d. in **a,b**. **(c)** RNA blot analysis of *EVD* siRNAs in wild-type, epi15 F₈ and epi15 F₁₁ 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 F₁₁ 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 siRNA 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^{ARBD}*. 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^{ARBD}* mRNA and siRNA levels with or without coexpression of *35S:P19* or *35S:P19^m*. EtBr, ethidium bromide staining of rRNA; U6, siRNA loading control.

between generations F₈ and F₁₁, these species also accumulated in F₁₂ and F₁₃, albeit less than in F₁₄, 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* background²¹ in which *EVD* mobilization was confirmed. In both lines epi454 and epi439, with 33 and 7 *EVD* copies

in generation F₈, 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.

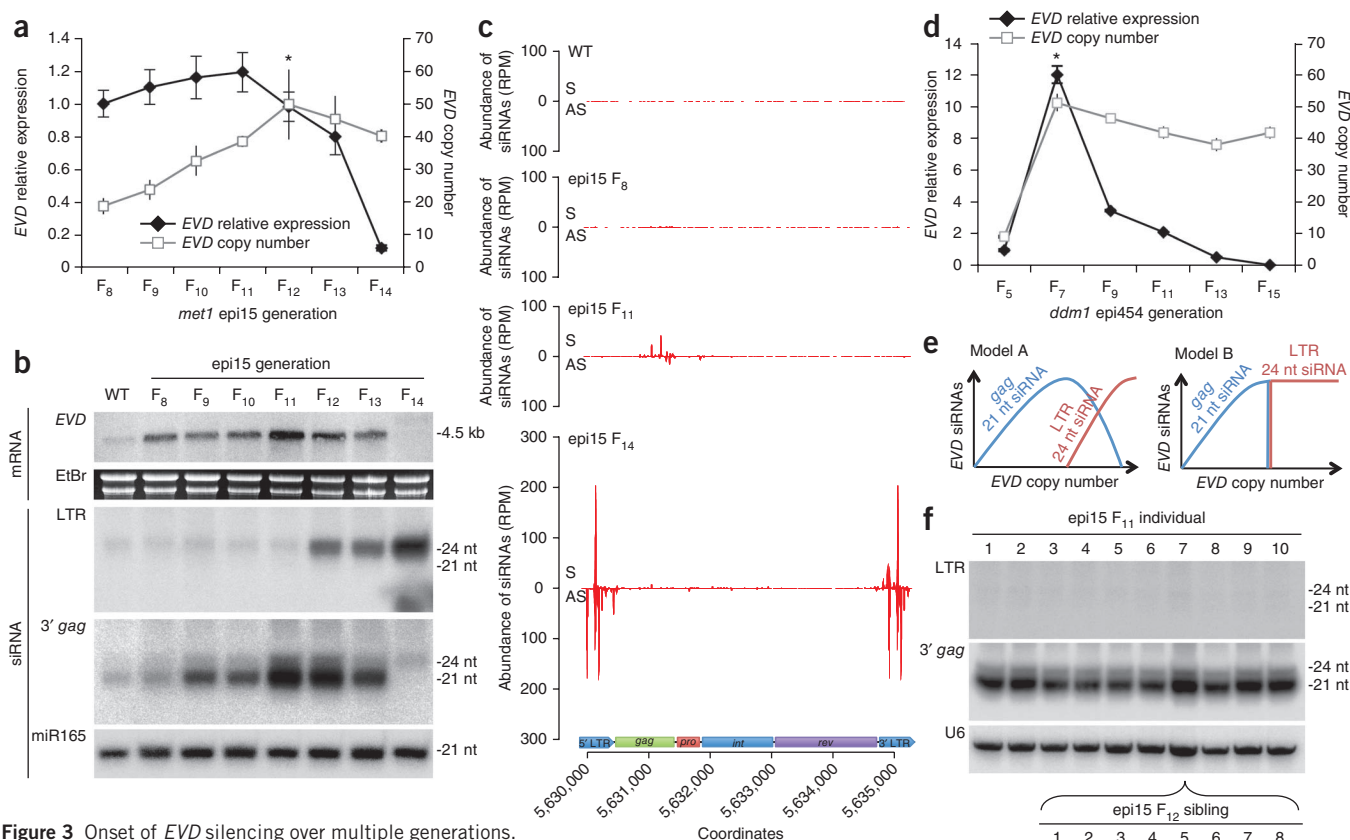


Figure 3 Onset of *EVD* silencing over multiple generations.

(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 F₈ levels, which were arbitrarily set to 1. The *EVD* copy number in the F₁₂ generation (*) take into account extrachromosomal DNA owing to active retrotransposition, stopped between generations F₁₃ and F₁₄. Error bars, s.d. (b) RNA blot analysis of LTR- and 3' *gag*-derived siRNA and *EVD* mRNA levels in wild-type and epi15 F₈ to F₁₄ plants. Hybridization to miR165 and ethidium bromide staining provide RNA loading controls. (c) Mapping of 24-nt deep sequencing siRNA reads to the *EVD* genome locus in wild-type, epi15 F₈, epi15 F₁₁ and epi15 F₁₄ generations (as in Fig. 2d). (d) qRT-PCR analyses of *EVD* expression and copy number in successive *ddm1* epi454 generations from F₅ to F₁₅. *EVD* mRNA levels were normalized to *ACT2* levels and then to epi454 F₅ levels, which were arbitrarily set to 1. The *EVD* copy number in the F₇ 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 F₁₁ plants and eight F₁₂ siblings from the single F₁₁ plant 7. U6, RNA loading control.

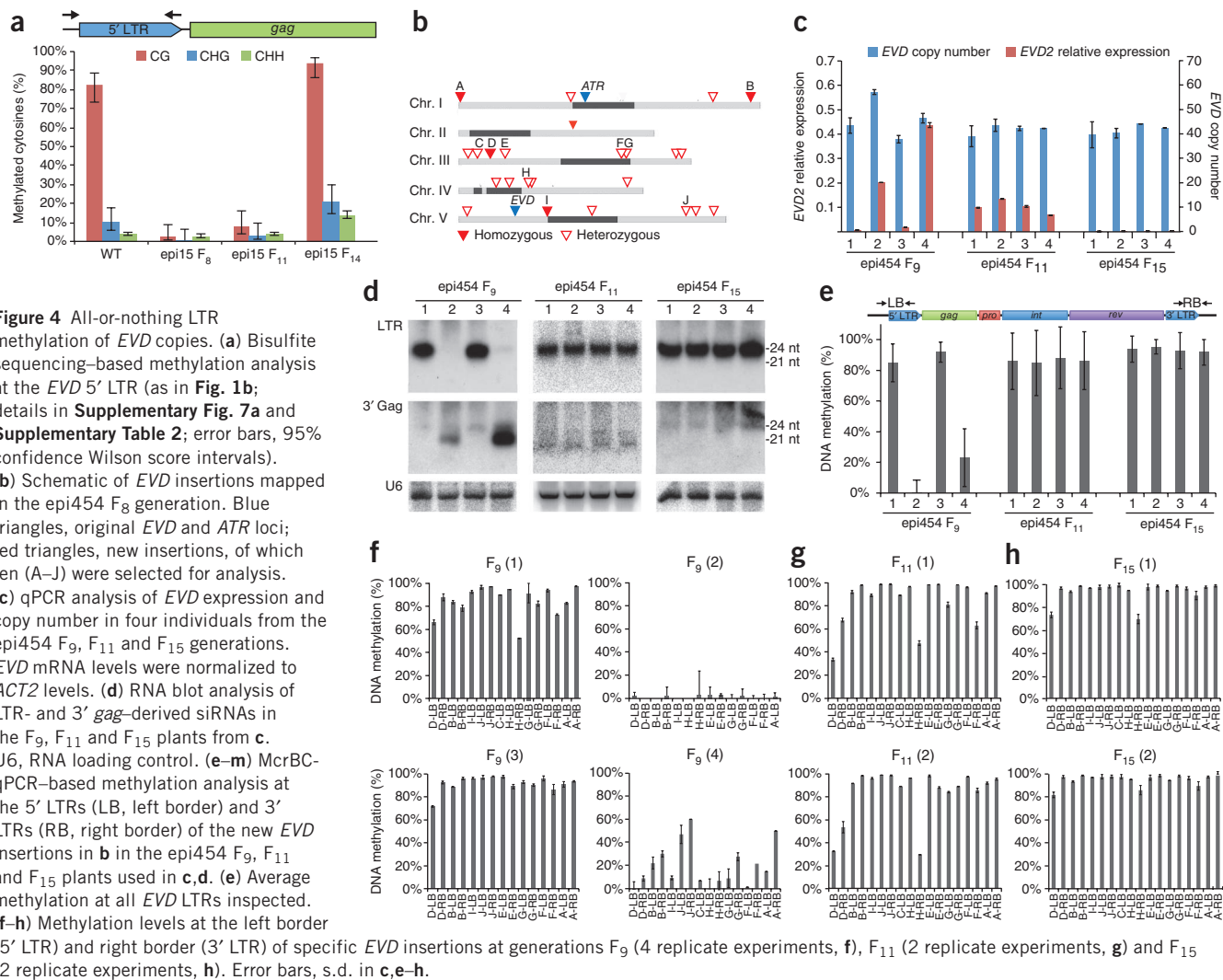
Two mutually exclusive *EVD* silencing states coexist in epiRILs

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 F₁₄ (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 F₁₁ to F₁₃, these distinct states would have been confounded by analysis of bulked individuals. This idea was confirmed by analyzing single F₁₂ siblings derived from a unique F₁₁ parent (Fig. 3f), validating model B. Therefore, the sharp onset of silencing at F₁₄ 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 F₁₄ but not F₈–F₁₁ 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 F₈ generations (ten and three insertions, respectively; Fig. 4b and Supplementary Fig. 7b).

In epi454 F₉ 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 F₉ plants (plants 2 and 4) accumulated only



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

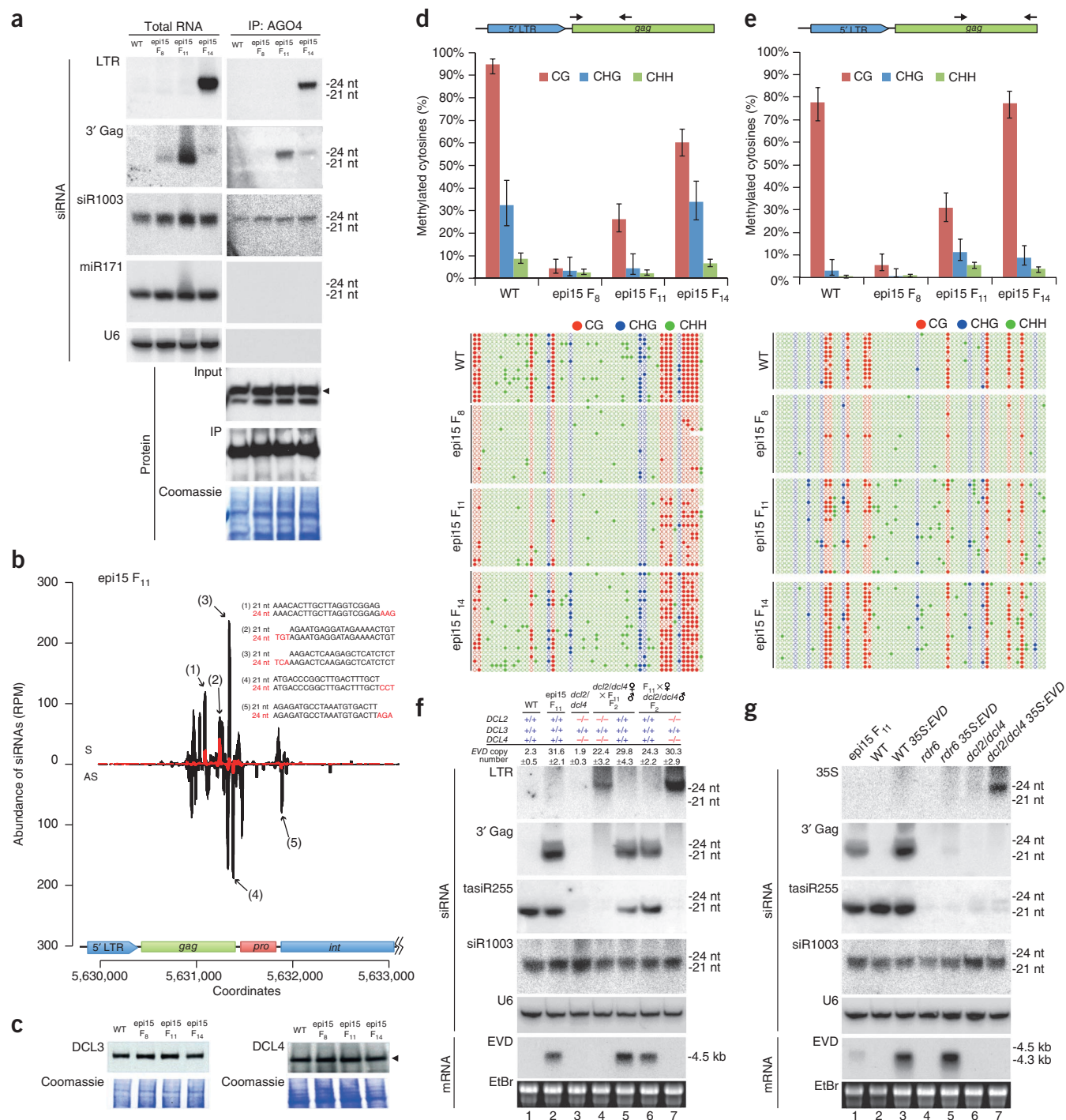
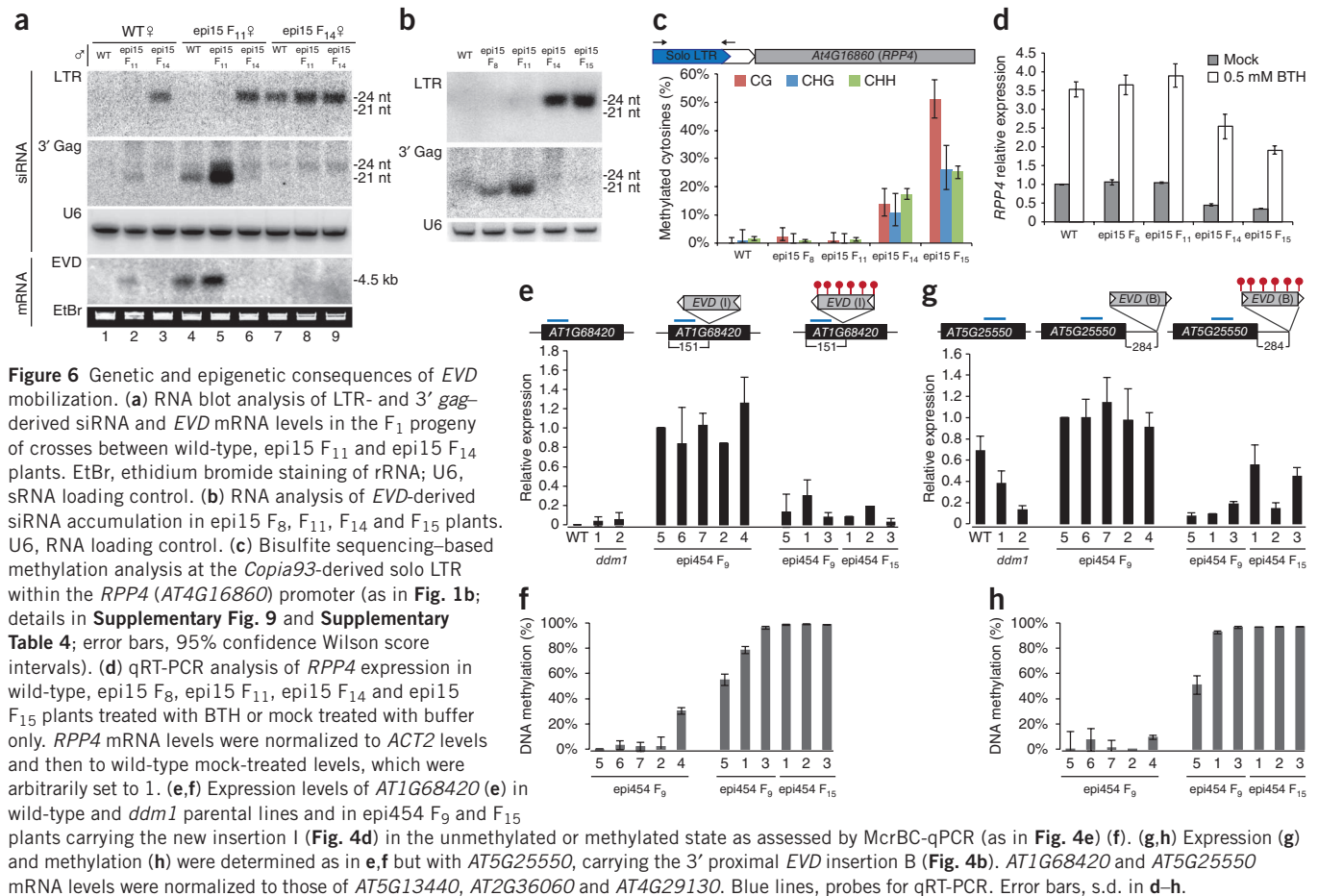


Figure 5 RDR6-DCL3-dependent RdDM triggers *EVD* TGS. **(a)** RNA analysis of LTR- and 3' *gag*-derived siRNAs in AGO4 immunoprecipitations. siR1003 (AGO4) and miR171 (AGO1) were probed to validate specific sRNA enrichment with AGO4. U6, control for RNA carryover from cleared immunoprecipitation extracts. 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 F₁₁ 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 F₈, epi15 F₁₁ and epi15 F₁₄ 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 F₁₁ 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.



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 F₈ and F₁₁ 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 F₁₁ plants; also as expected, this methylation was absent in F₈ 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 F₁₁ 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 F₁₁-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 F₁₄ generation but not in the F₁₁ 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 *rdm6* 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 F₁₁ and epi15 F₁₄ plants. When introduced into wild-type plants, F₁₄-derived copies remained silenced in F₁ 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, F₁₁-derived copies remained transcriptionally active,

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 F₁₁ and F₁₄ plants resulted in silencing of all F₁₁-derived *EVD* copies, as *EVD* RNA and 3' *gag*-derived siRNAs of 21–22 nt were below detection in F₁ 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 F₈ and epi15 F₁₁ plants. However, coinciding with the induction of LTR-derived 24-nt siRNAs, it became increasingly methylated from the F₁₄ generation onward (Fig. 6b,c and Supplementary Fig. 9), correlating with reduced *RPP4* expression, induced by treatments with the salicylate analog benzothiadiazole (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 F₈ generation (Fig. 4d) to *AT1G68420* and downstream of *AT5G25550*, respectively. Both full-length *EVD* insertions markedly increased transcription at both loci in F₉ individuals, displaying active *EVD* transcription and no LTR methylation (Figs. 4e and 6e–h). By contrast, in F₉ 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 F₁₅ 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 resiliency. 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.

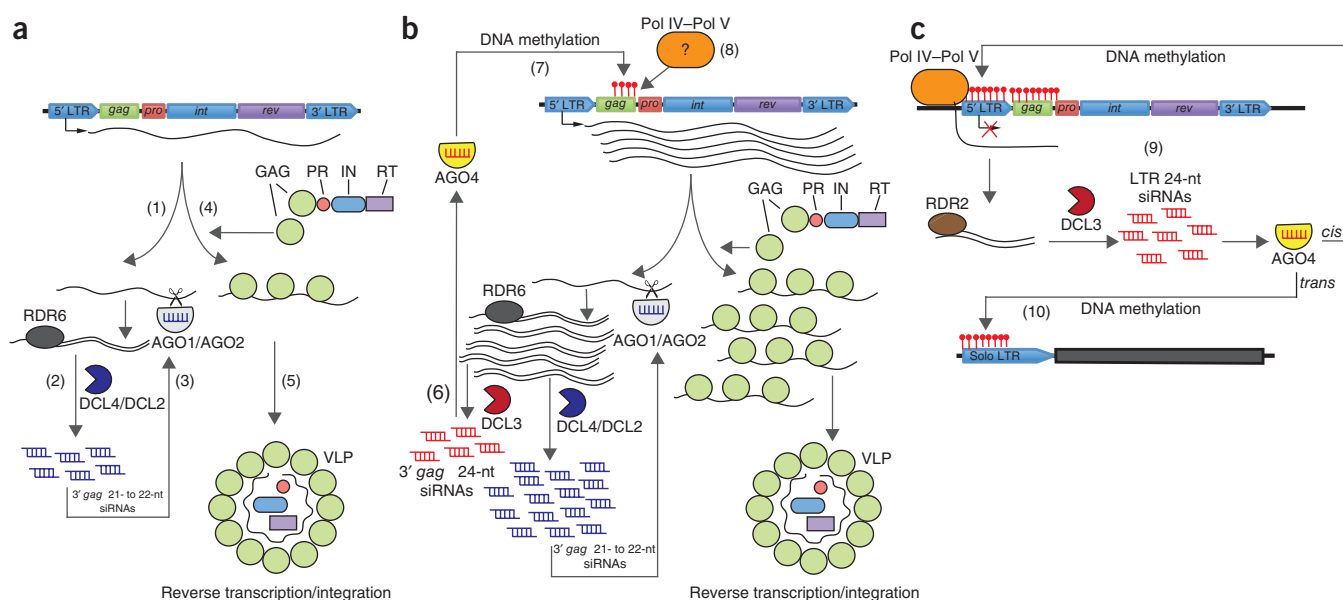


Figure 7 Rise and demise of *EVD*. (a) In early proliferation stages, a low *EVD* copy number accounts for moderate transcription and dsRNA synthesis by RDR6 (1). Processing of dsRNA by DCL4 or DCL2 produces 3' *gag*-derived siRNAs of 21–22 nt in length (2), which are loaded onto AGO1–AGO2 (3). This 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).

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 F₁₁ generation and markedly reduced in the F₁₄ 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 mutants⁴⁵ 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'École Normale 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>.

- McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216 (1956).
- Ito, H. *et al.* An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**, 115–119 (2011).
- 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).
- 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).
- Lisch, D. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* **60**, 43–66 (2009).
- Kankel, M.W. *et al.* *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122 (2003).
- 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).
- Zemach, A. *et al.* The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
- Yu, A. *et al.* Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc. Natl. Acad. Sci. USA* **110**, 2389–2394 (2013).
- Slotkin, R.K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461–472 (2009).
- Teixeira, F.K. *et al.* A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**, 1600–1604 (2009).
- Down, R.H. *et al.* Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. USA* **109**, E2183–E2191 (2012).
- Pereira, V. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol.* **5**, R79 (2004).

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. Gascioli, 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).

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*, *rdm6-12* and *ddm1* mutants and the *met1-1* or *ddm1-1*-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*^{ARBD} clones were recombined into pDONR1P2. 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 method⁶¹. 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 described³¹.

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 μ m 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. Immunological 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 recommendations⁶². 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 μ g) 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 indicated, 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 μ g 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 irradiation⁶⁴. 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 described⁶⁵, and 5–10 μ g was resolved on a denaturing 17.5% polyacrylamide-urea gel, transferred to membrane by electroblotting and chemically cross-linked⁶⁶. 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 -³²P]-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 -³²P]-dATP using T4 PNK (Thermo Scientific).

Southern blot analysis. Unless otherwise indicated, ten plants of the different epiRIL generations were bulked for genomic DNA extraction using the CTAB method⁶⁷, and 1 μ g 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 irradiation⁶⁴. [³²P]-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 μ M 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 (12,000g 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 (5,000g 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 Immobilon-P PVDF membrane (Millipore) and incubated with antibodies in PBS with 0.1% Tween-20 and 5% nonfat dried milk, following standard protein blotting procedures⁶⁴. 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 μ M 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 (12,000g at 4 °C). Protein was quantified in extracts to normalize total protein inputs. Extracts were precleared by incubation for 30 min with 10 μ l 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 (12,000g 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 μ l of Protein A-agarose beads. Beads were collected by centrifugation (4,000g at 4 °C), and immunoprecipitates

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 immunoprecipitation. 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 software⁷⁴ 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 software⁷⁵ (options: -e 50 -a -best -strata -m 5000 -nomaqround -y) together with the SAMtools⁷⁶ and BEDTools⁷⁷ 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. Grunthan, 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).



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 gynoeceum 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.

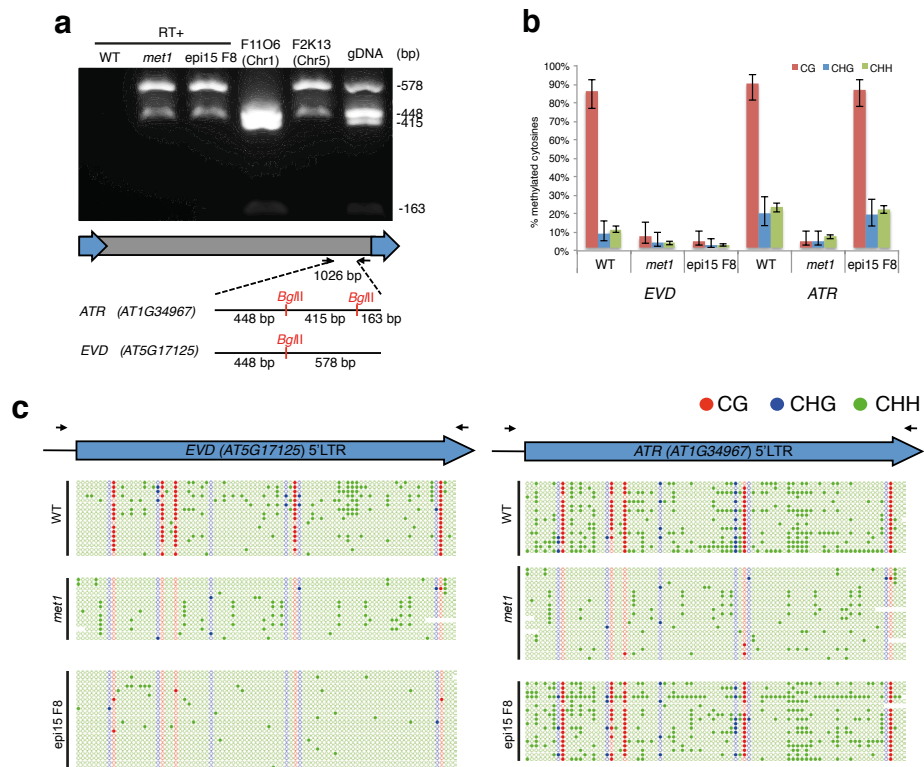
Supplementary Table 2: Figure 4 and Supplementary Figure 7 bisulfite sequencing raw data.

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.

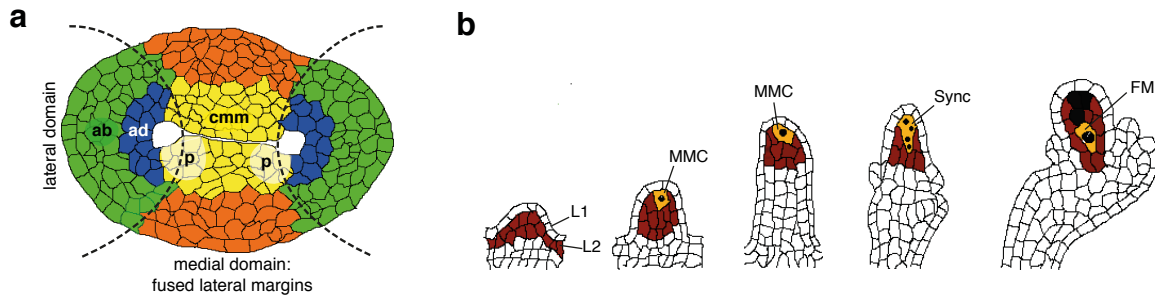
Supplementary Figure 1



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 *Bgl*III 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.

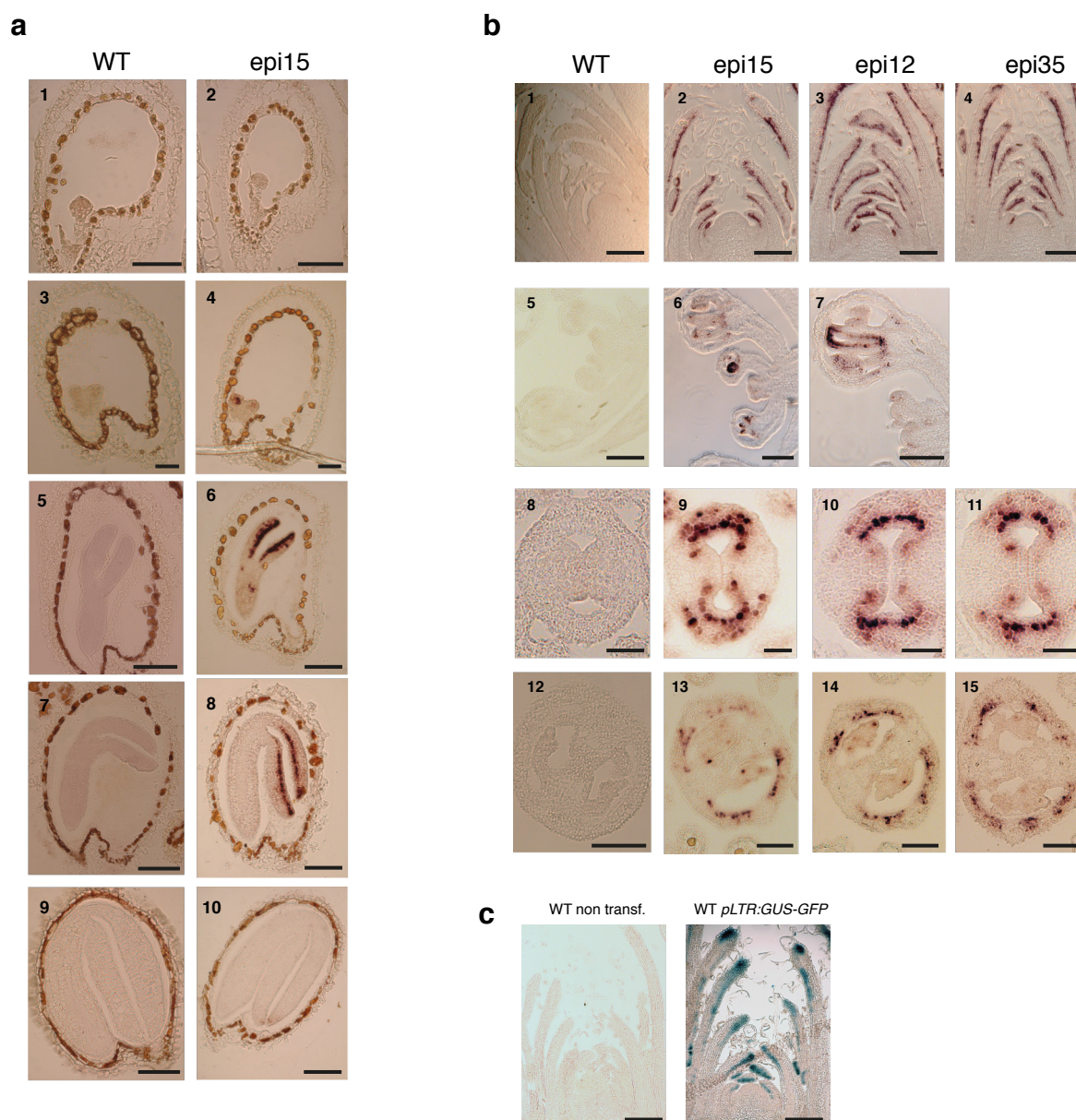
Supplementary Figure 2



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 domains²¹⁻²³. 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.

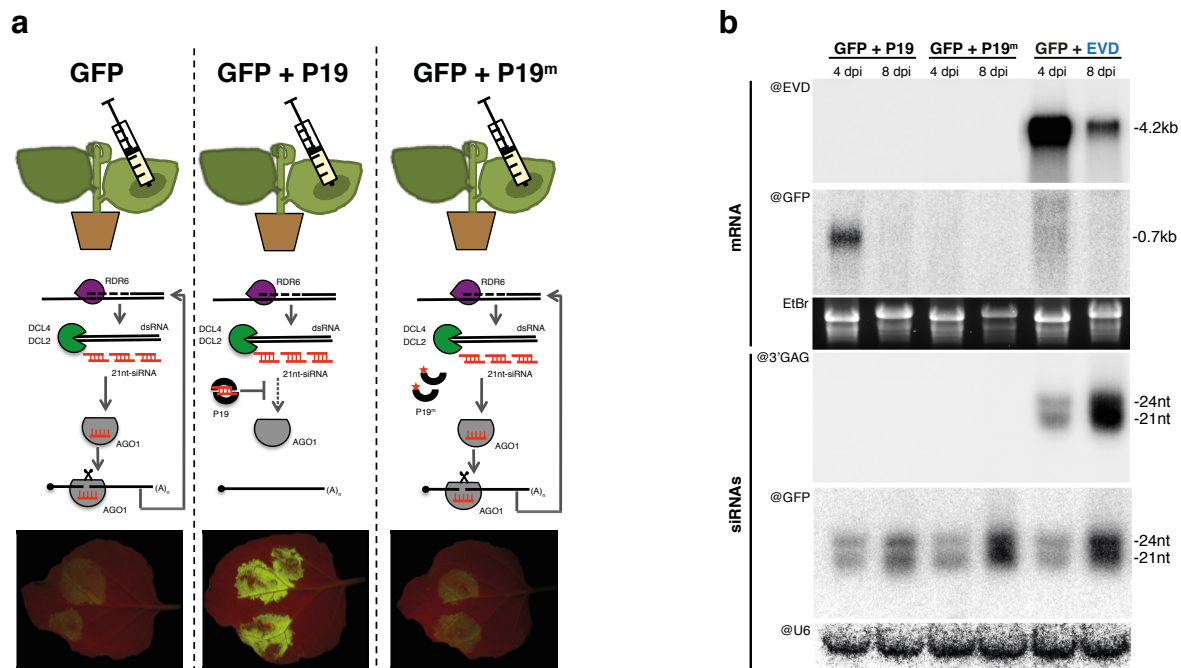
Supplementary Figure 3



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 μ m for **(a)** 1-4, **(b)** 12-15; 100 μ m for **(a)** 5-10, **(b)** 1-7 and **(c)**; 10 μ m for **(b)** 8-11.

Supplementary Figure 4

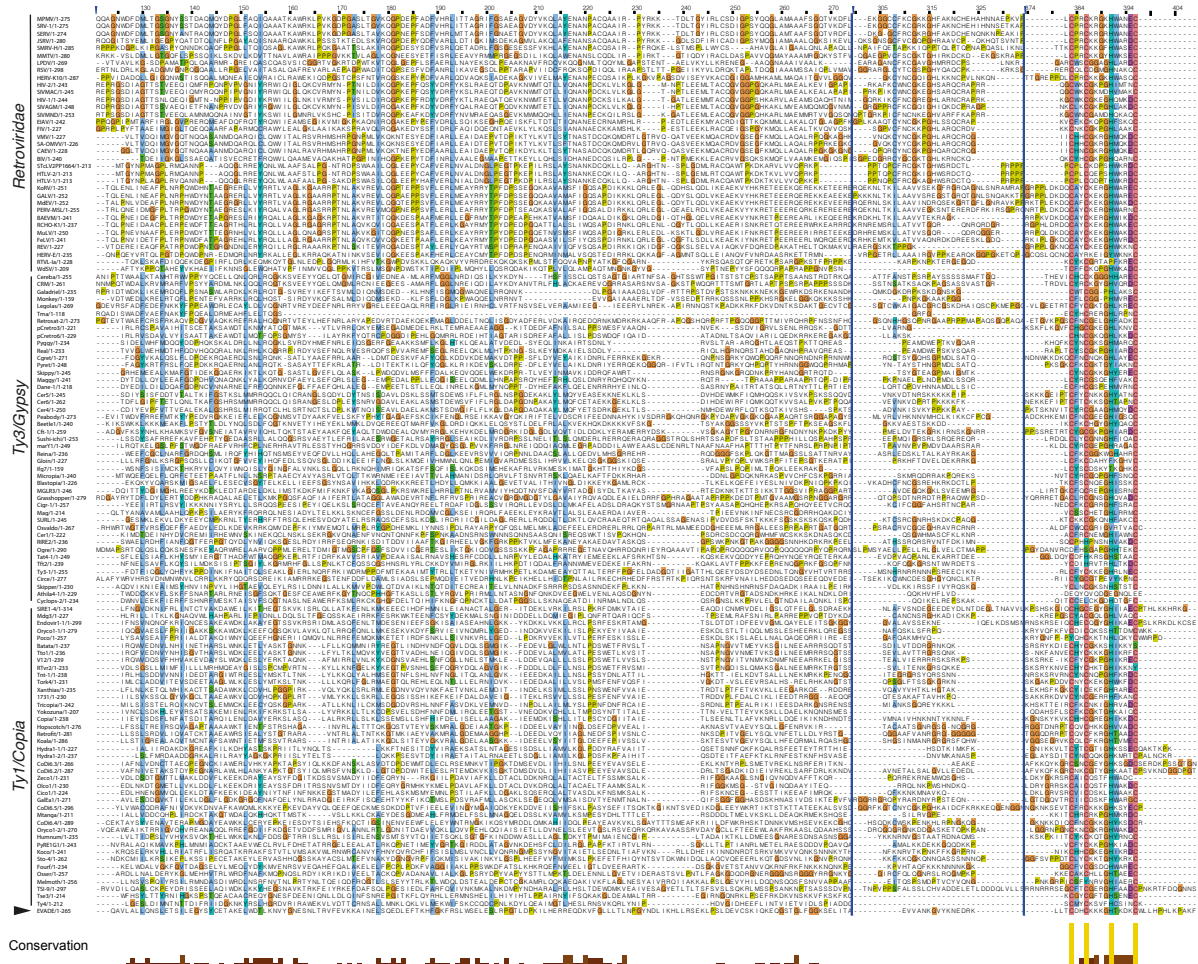


Supplementary Figure 4: Suppression of silencing assay in *N. benthamiana*.

(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 PTGS³⁰. 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 PTGS³⁰. 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:P19^m 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

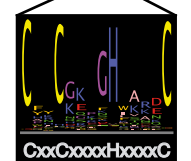
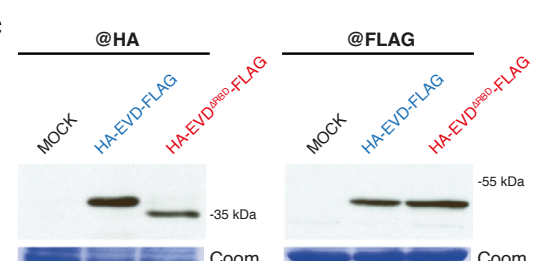
a



b



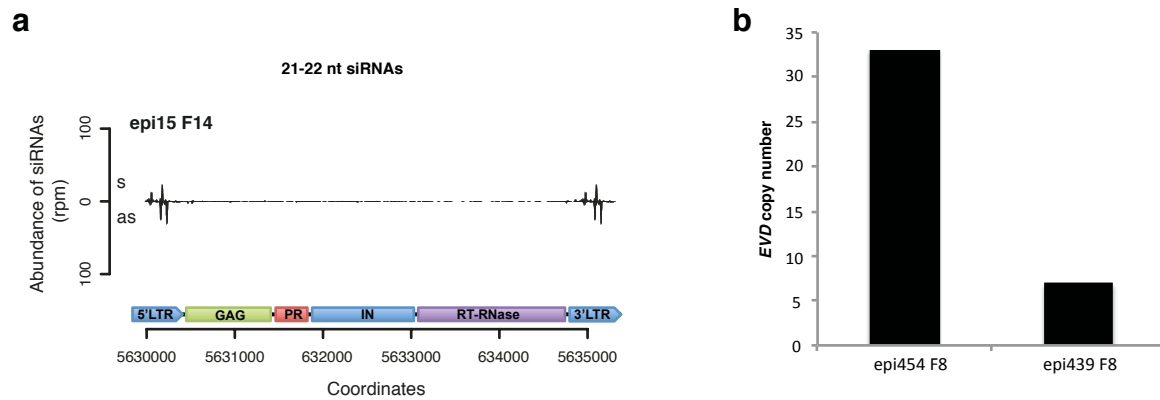
c



Supplementary Figure 5: GAG RNA binding domain conservation and EVD polyprotein 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 (Cx₂Cx₄Hx₄C). (c) Scheme of p35S:HA-EVD-FLAG and p35S:HA-EVD^{ARB}-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^{ARB}-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.

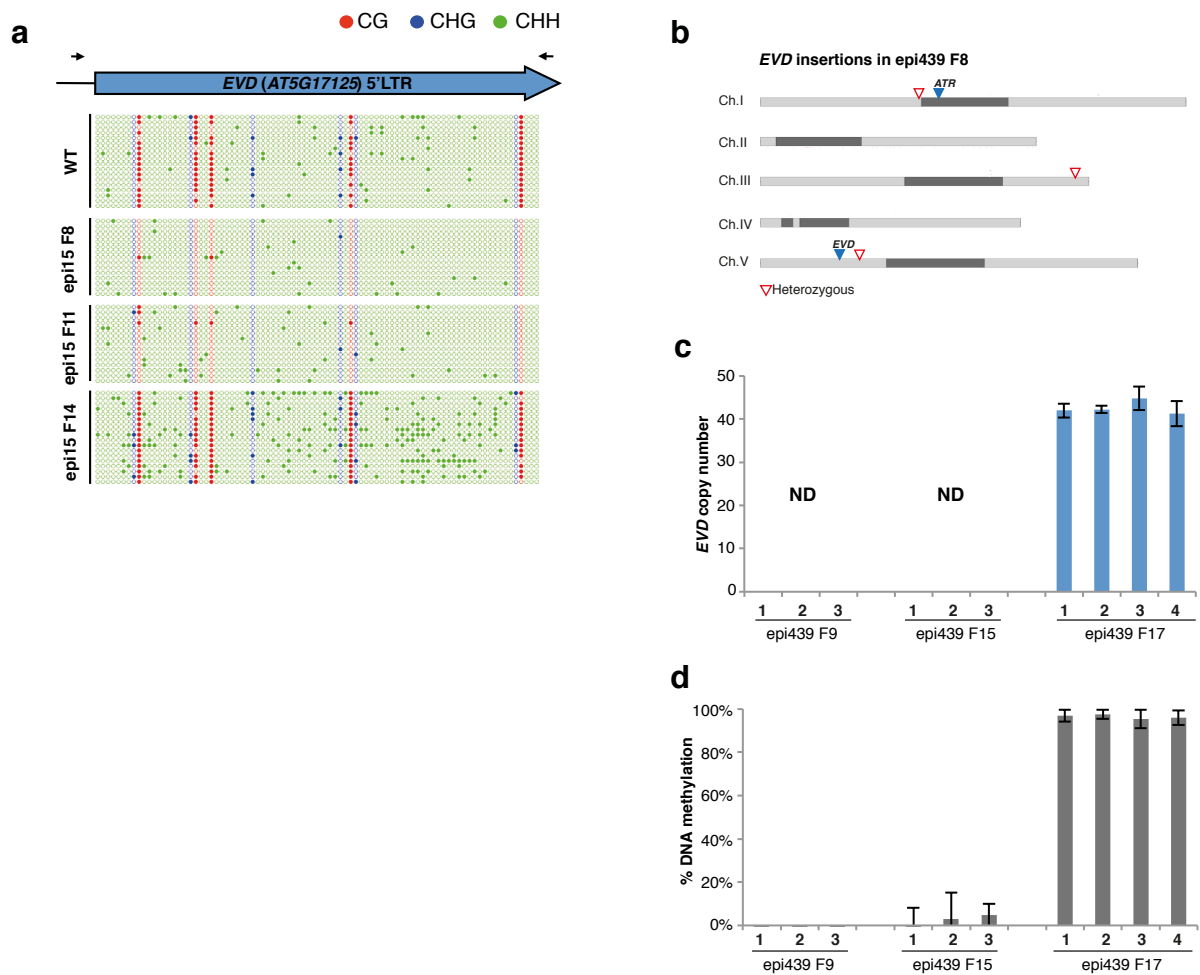
Supplementary Figure 6



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.

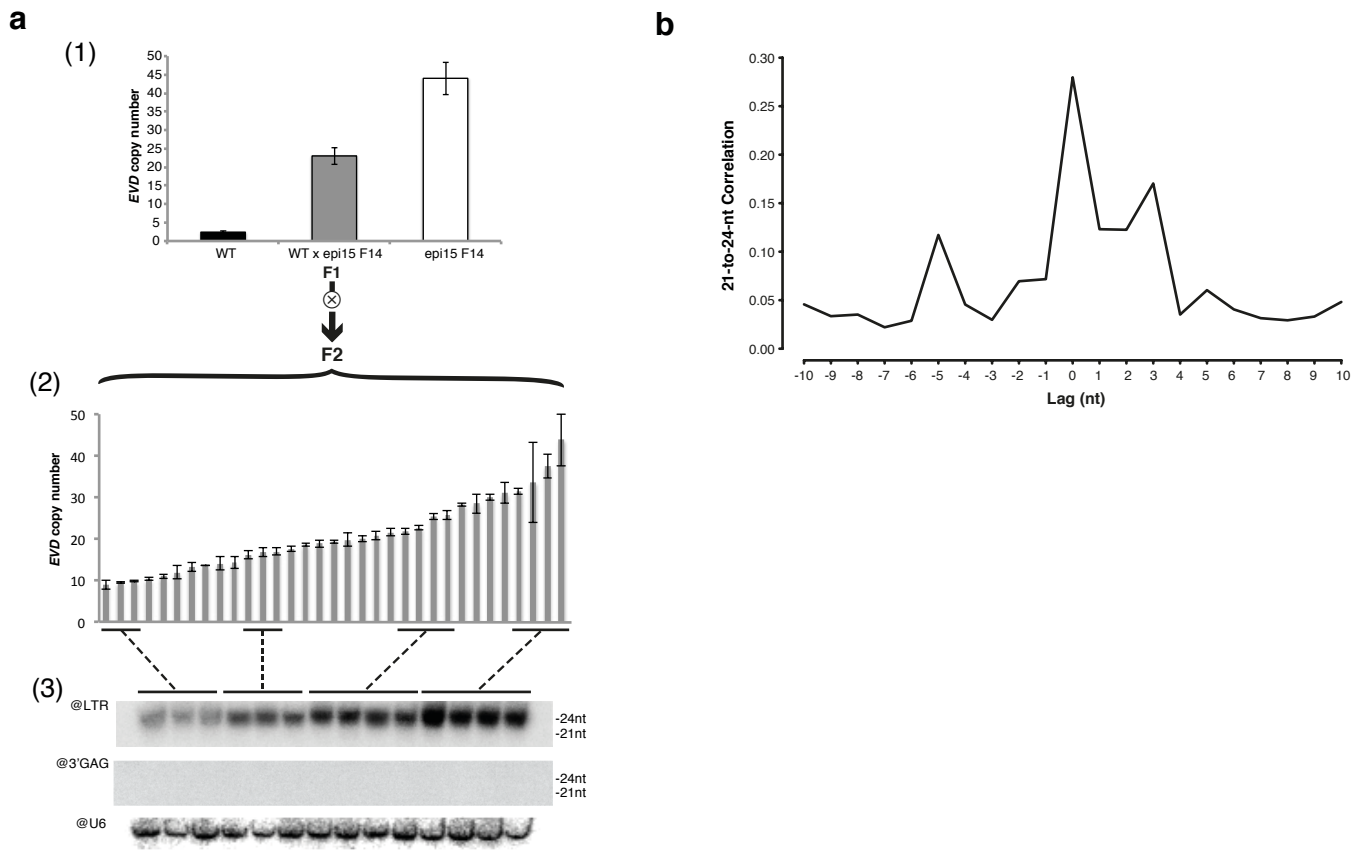
Supplementary Figure 7



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

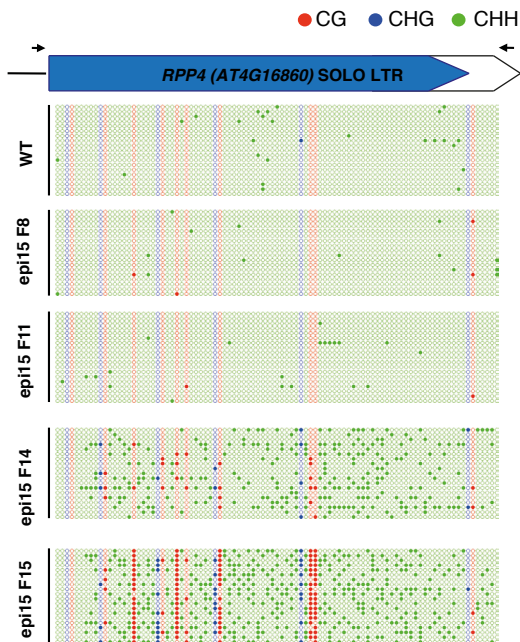


Supplementary Figure 8: Correlation of LTR 24-nt siRNAs abundance with EVD copy number and 21-24-nt siRNAs mapping correlation on EVD.

(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

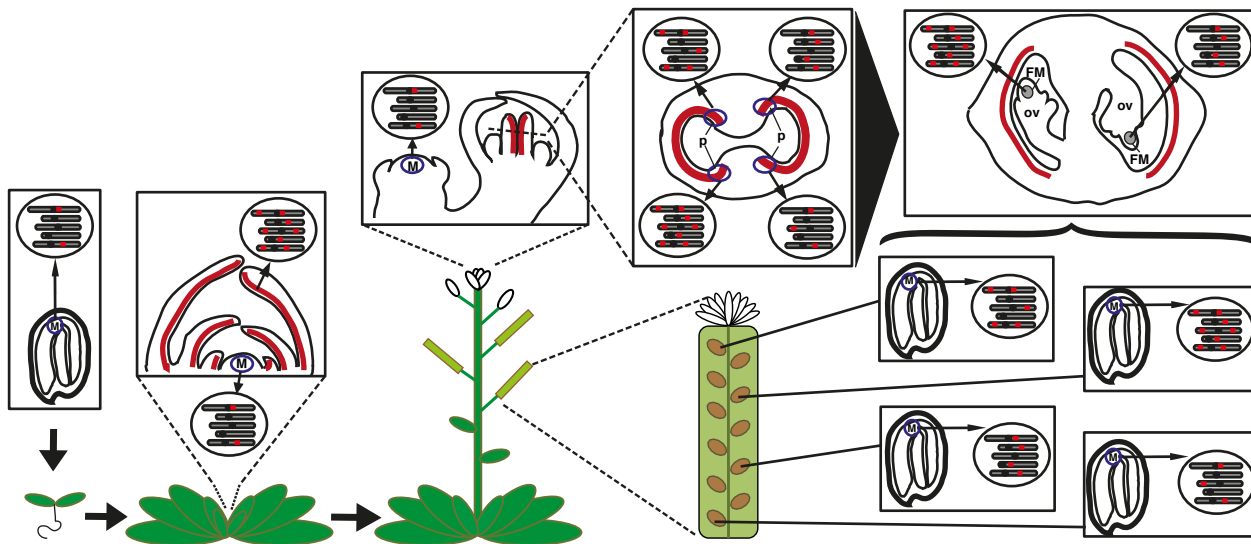


Supplementary Figure 9: Bisulfite sequencing dot-plots of *RPP4* Solo-LTR in epi15 generations.

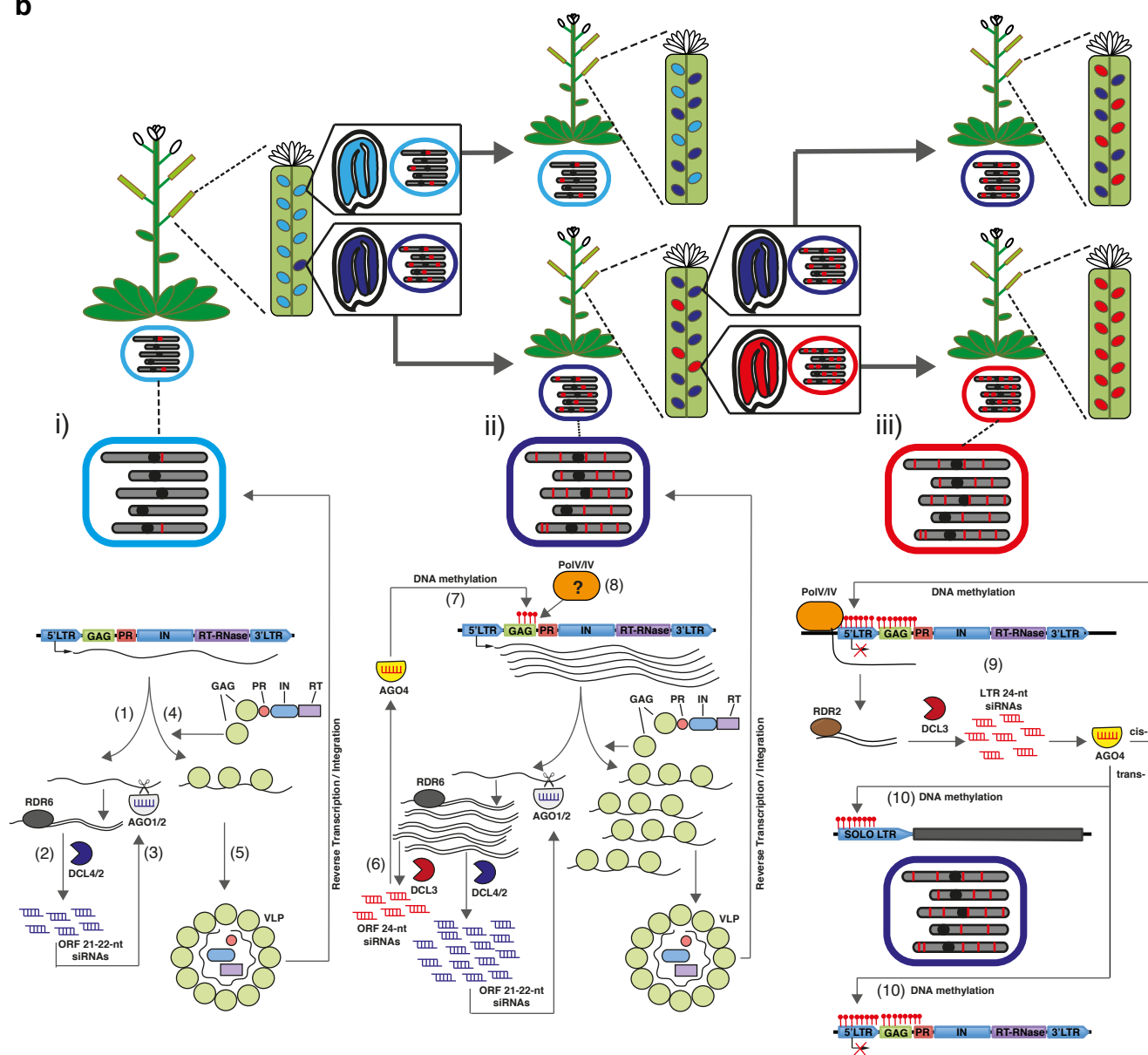
Dot-plots of bisulfite sequencing from individual clones for results shown in Figure 6b and Supplementary Table 4. Filled circles indicate methylated cytosine and empty circles unmethylated cytosine.

Supplementary Figure 10

a



b



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

(a) *EVD* expression pattern ensures host integrity through development 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) exclusion 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/PoIV, 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+
<i>EVD</i>	WT	85	102	1274	84,70%	7,84%	10,12%	17	2	75,57%	4,03%	8,58%	90,83%	14,71%	11,90%
	<i>met1-3</i>	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%
<i>ATR</i>	WT	80	96	1197	88,75%	18,75%	22,05%	16	2	79,98%	12,20%	19,79%	93,97%	27,70%	24,49%
	<i>met1-3</i>	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 Supplementary 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	GATCTTGGTGAACCTCAAATACTT
	EVD-RT R	TACTCCCCCTCAAGCGCAAGAC
EVD copy# qPCR/ qRT-PCR	EVD-qRT F	GATAGAGGAGATAGAAGATCTACAACCTGG
	EVD-qRT R	CTCTATACTCCGATTCTGCACTCGAACA
EVD ISH/Southern/ HMW-Northern probe	EVD-P F	TTTCATGGTATCAGAGCATAAGAT
	EVD-P R	CACCCGAGTCTATCACAAATTGAAC
LTR Bisulfite PCR	EVD-LTR BS F	TTTATAARCATAAAAACATAATCTTATRCTCTAATACCATA (TTTATAAGCATAAAGACATGATCTTATGCTCTGATACCATG)
	EVD-LTR BS R	GGAAAAAAYGAAGGAATTTAGAAATAYGATTTGGTAATTT (GGAAAAAACGAAGGAACCCCTAGAACACGATTTGGCAATTC)
	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	TTTTTAAGTGTATCCCAACTCTT
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	GAAGGATTGGGAGAACCTAAAGATCTTAC
	EVD-IR R	CTTATACTTGATTGTGAATACCCACCTTGA
RT-RN probe	EVD-RR F	CAAAGACGGTATAGACTCTACCAAGAC
	EVD-RR R	CTCTATACTCCGATTCTGCACTCGAACA
RN probe	EVD-RN F	GATAGAGGAGATAGAAGATCTACAACCTGG
	EVD-RN R	TACTCCCCCTCAAGCGCAAGAC
mGFP probe	mGFP-F	AGTAAAGGAGAAGAACTTTTCACT
	mGFP-R	TTCCGTCCTCCTTGAATCGA
EVD WT GAG-POL ORF Cloning	EVD ORF attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGACTTCTCAAAGATG
	EVD ORF attB2	GGGGACCACCTTTGTACAAGAAAGCTGGGTTTCAAGAGTGAGATAGATCC
EVD dRBD Site-directed Mut.	EVD dRBD F	AGGATAGAAAAGTGTGACTTGGTTGCTACATCCACACCT
	EVD dRBD R	AGGTGTGGATGTAGCAACCAAGTCAACAGTTTTCTATCCT
EVD HA-GAG-POL-FLAG Cloning	HA-EVD F	ATGTACCCATACGATGTTCCAGATTACGCTGGTGGAGGCGAGACTTCTCAAAGATGA
	EVD-FLAG R	TCACTTGTGCATCGTCATCCTTGTAAATCGCCACCTCCAGAGTGAGATAGATCCACAA
	HA-EVD attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTACCCATACGATGTTCC
	EVD-FLAG attB2	GGGGACCACCTTTGTACAAGAAAGCTGGGTTTCACTTGTGCATCGTCATCCTTGT
epi454/439 EVD insertions zygosity	454 EVD-A F	CTACTGGCGTGGATGCTTTT
	454 EVD-A R	GTAACACAGAGTTCAGATTTAAACCAA
	454 EVD-B F	TTCTTCCAAAATTTGATTACTTGC
	454 EVD-B R	TCCGAAAATGGCATAAGAGG
	454 EVD-C F	TTGGAATTCGATTCAACTAGAGG
	454 EVD-C R	CTCAAATTCATTGGAGACACATC
	454 EVD-D F	TGTGCGTAGCTTTTCTTGA
	454 EVD-D R	ATTGAGGAACCAGCATCACC
	454 EVD-E F	GCCTTAATGAATGGCCCAATA
	454 EVD-E R	ACTCCACGGGACTACCTCAC
	454 EVD-F F	GACCCCAACCAATCAATCAC
	454 EVD-F R	GACGTTATGGTGAATATGTGC
	454 EVD-G F	AAAATGTCACCTTATCATTGCAAGC
	454 EVD-G R	TTTTAGTGGTGCGGGCTTTA
	454 EVD-H F	CCAACGTGGGTTAGATCCAT
	454 EVD-H R	ACGGCTATTGGAGTTTTTGA
	454 EVD-I F	CGGTGTTTGTAGTGGGAACC
	454 EVD-I R	GCAAAAACATTGCTACATTTTCG
	454 EVD-J F	TCTCTTGTCTTGTGCGACTCA
	454 EVD-J R	CACACCATGGATAATGCTCTT
439 EVD-ChI F	ATCAACAAAATAGATATCTTTGCATTTT	
439 EVD-ChI R	TTCTTTTCGGATTCTGTGTGG	
439 EVD-ChIII F	ATGGTTCGATTGTGGAGGAA	

	439 EVD-ChIII R	CAACATTCTAGGGTTTTACCTTGC
	439 EVD-ChV F	TCTCCACCCAAAATCTCAG
	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	ATGAAAATATTTGGCGTGTCAT
	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	AGGACTGACGTACACCCTA
	454 EVD-F qF	CCTTCGAGTCGAGATCCAAT
	454 EVD-F qR	TCATATTCACATATGTTGGTGTGG
	454 EVD-G qF	CTGTACATTTCAATATTTATCCATCTTG
	454 EVD-G qR	AAATCACAACATATGTAATGTATCAAGT
	454 EVD-H qF	GAAGACTTTATGCCGTTTCATCA
	454 EVD-H qR	TTACATTTAGCGTTTCCCTGGT
	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	CCAGTTTGACGAAATTAACAA	
439 EVD-ChV qR	CGTAACCAAACATAAGGAGAACC	
5'GAG Bisulfite PCR	EVD-5GAG BS F	ATTTTAAGGGGAAATATTTAATTTGGTTAAGAATAAT (ACTTCAAGGGGAAACTACCTAACTTGGTCAAGAACAAC)
	EVD-5GAG BS F	ACCTAAACTTACCAAATATTTAATAAACTCTAAATCTTCTTA (ACCTGAACTTACCAAAGTGTTTAGTAAACTCTAAGTCTTCTTG)
3'GAG Bisulfite PCR	EVD-3GAG BS F	GTGTTTGAAGTGGAAGAAGGYGATTAAYGAATTAA (GTGTTTGAAGTGGAAGAAGGCGATTAACGAACATA)
	EVD-3GAG BS F	ATAACCCRACTTAACTTTTRCTCCTCATAAATTTCTTAA (ATGACCCGCTTGACTTTGCTCCTCATGAGTTTCTTGAG)
<i>RPP4</i> qRT-PCR	RPP4-qRT F	ACATGCGAGGACAACCAGAGG
	RPP4-qRT R	TTCAATTCGACGAAGTCACC
<i>RPP4</i> SOLO LTR Bisulfite PCR	RPP4-SLTR BS F	GGGAGGYGTTTAGGTAATTTATAYGTATATTATATTGATT (GGGAAGGCGTCCAGGTAATTCATACGTATATTATATTGATC)
	RPP4-SLTR BS R	TTCTCTACAAAATATTTTAAAAACACAAATTTTATAATTATCAA (TTCTCTACAAAGATATTTTAAAAACACAAATTTTGTGGTTATCAG)
<i>AT1G68420</i> qRT-PCR	AT1G68420 qF	GTCTCACACTTCAAGGCATCTTGC
	AT1G68420 qR	TTCATAGCTTCTCTACACCGGCA
<i>AT5G25550</i> qRT-PCR	AT5G25550 qF	GTTACAAGCGTGAAACGTGCCAT
	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

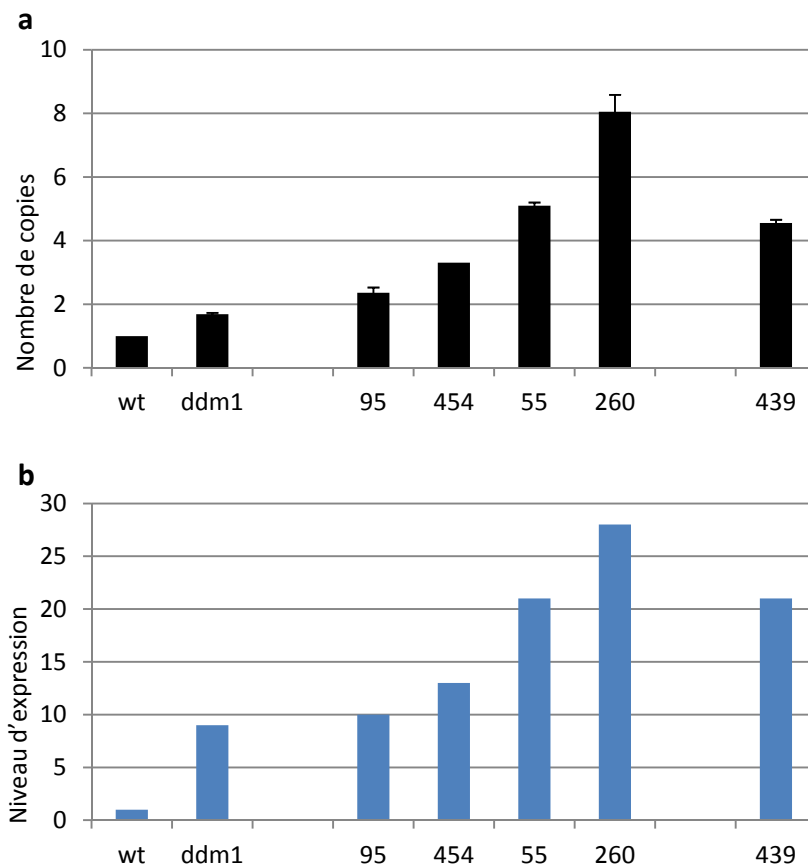


Figure 3.3 : Corrélation entre le nombre de copies d'AT2TE42810 (VANDAL21) et le niveau d'expression chez *ddm1* et les epiRILs.

(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 apparaît 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 peut 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.

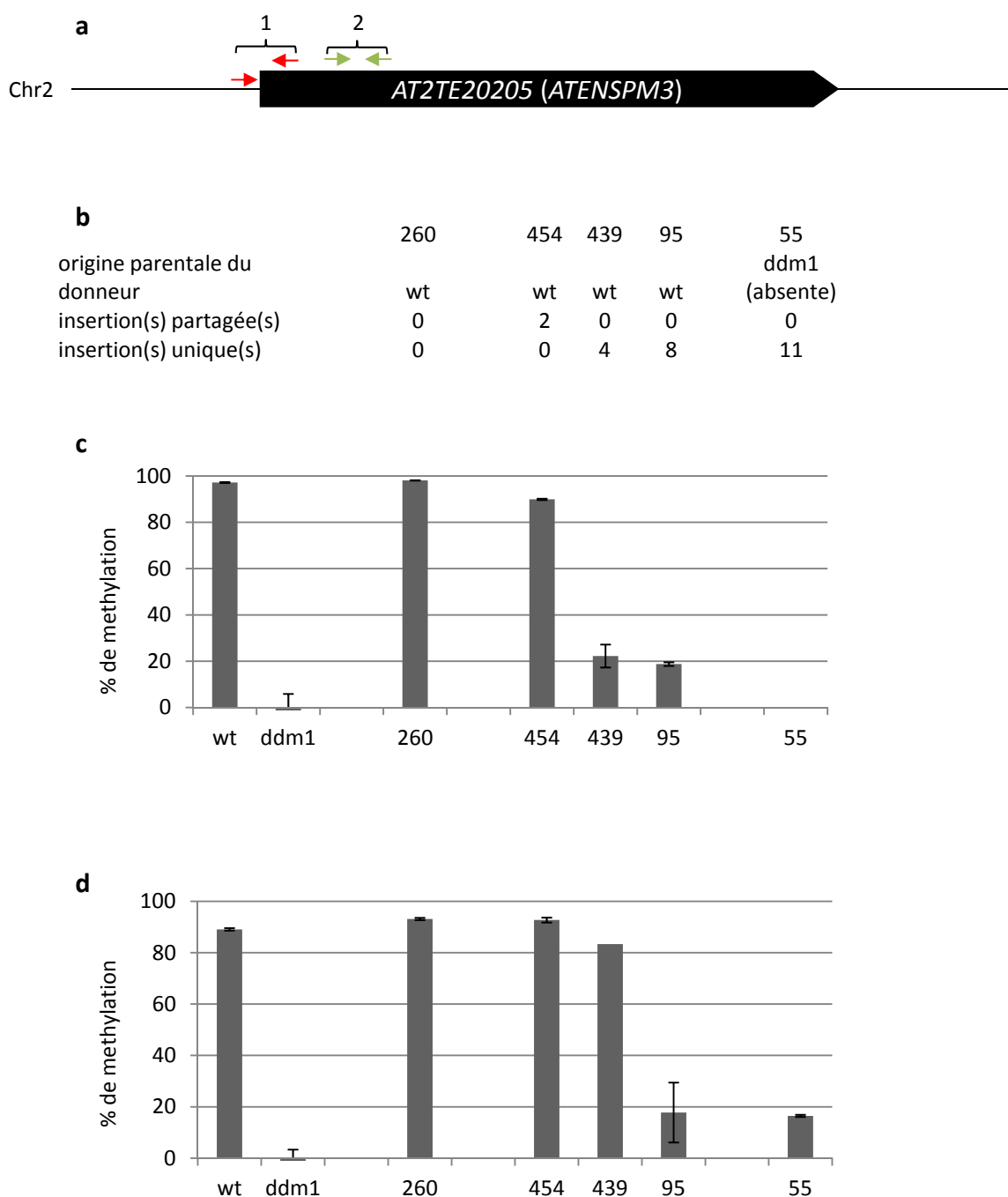


Figure 3.4 : Etude comparative des profils de méthylation et d'expression d'AT2TE20205 (ATENSPM3) chez le mutant *ddm1* et les épilIL.

(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

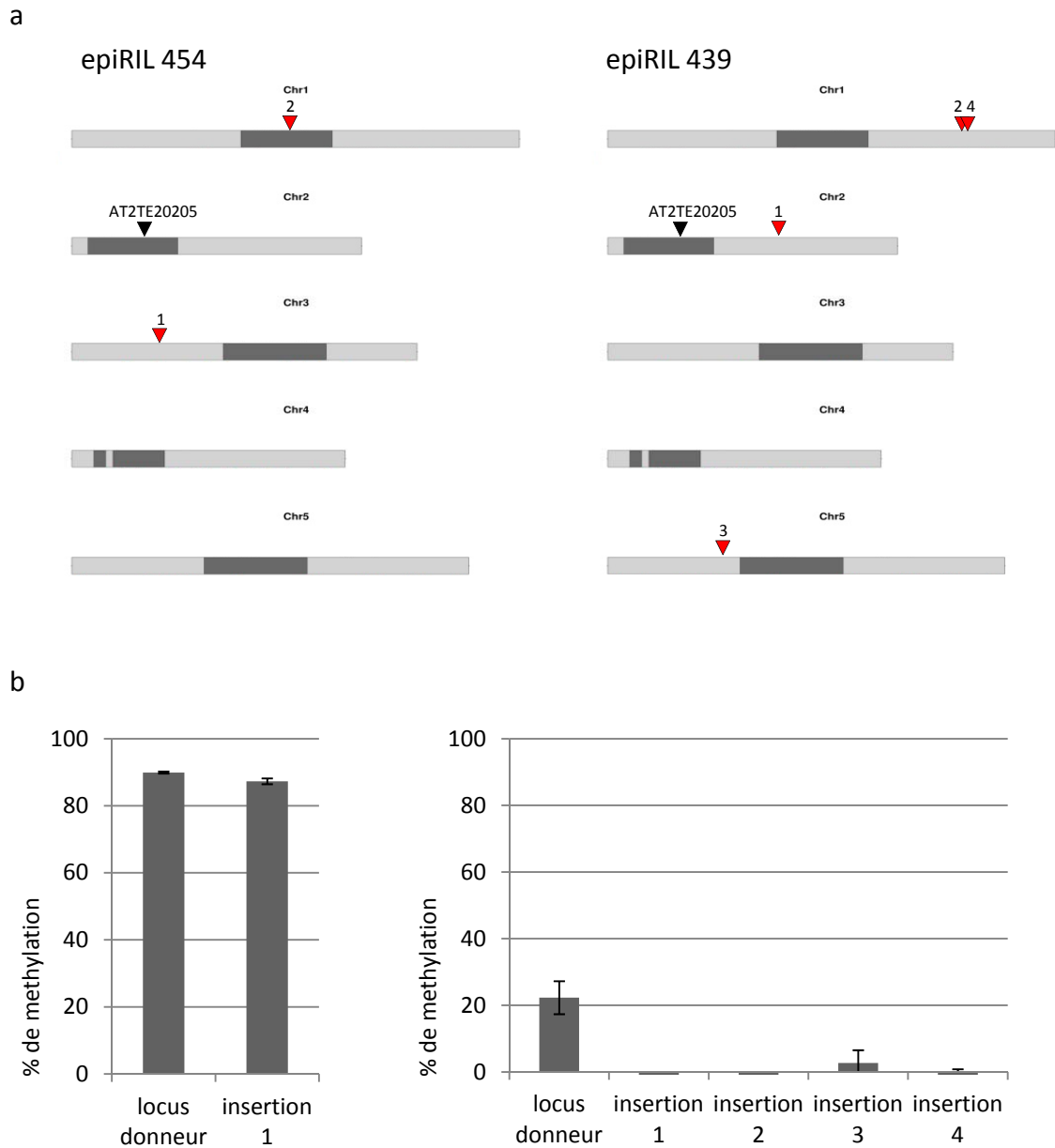


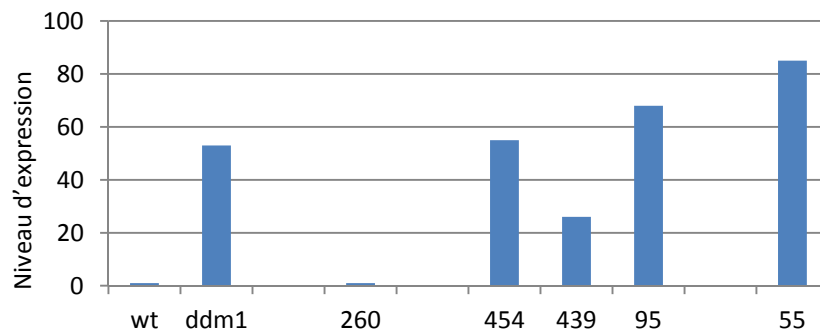
Figure 3.5 : Le % de méthylation au niveau de l'extrémité 5' des nouvelles insertions est identique à celui du locus donneur.

(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

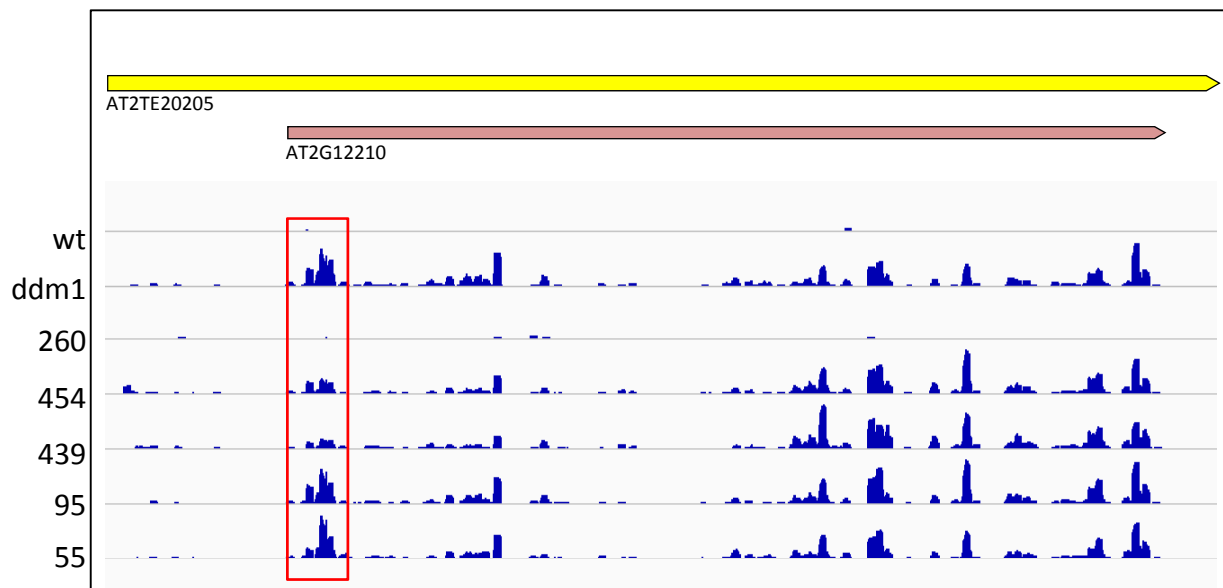


Figure 3.6 : Expression d'AT2TE20205 (ATENSPM3) chez le sauvage, *ddm1* et les epiRILs.

(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 apparaît 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 transcriptionnelle 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

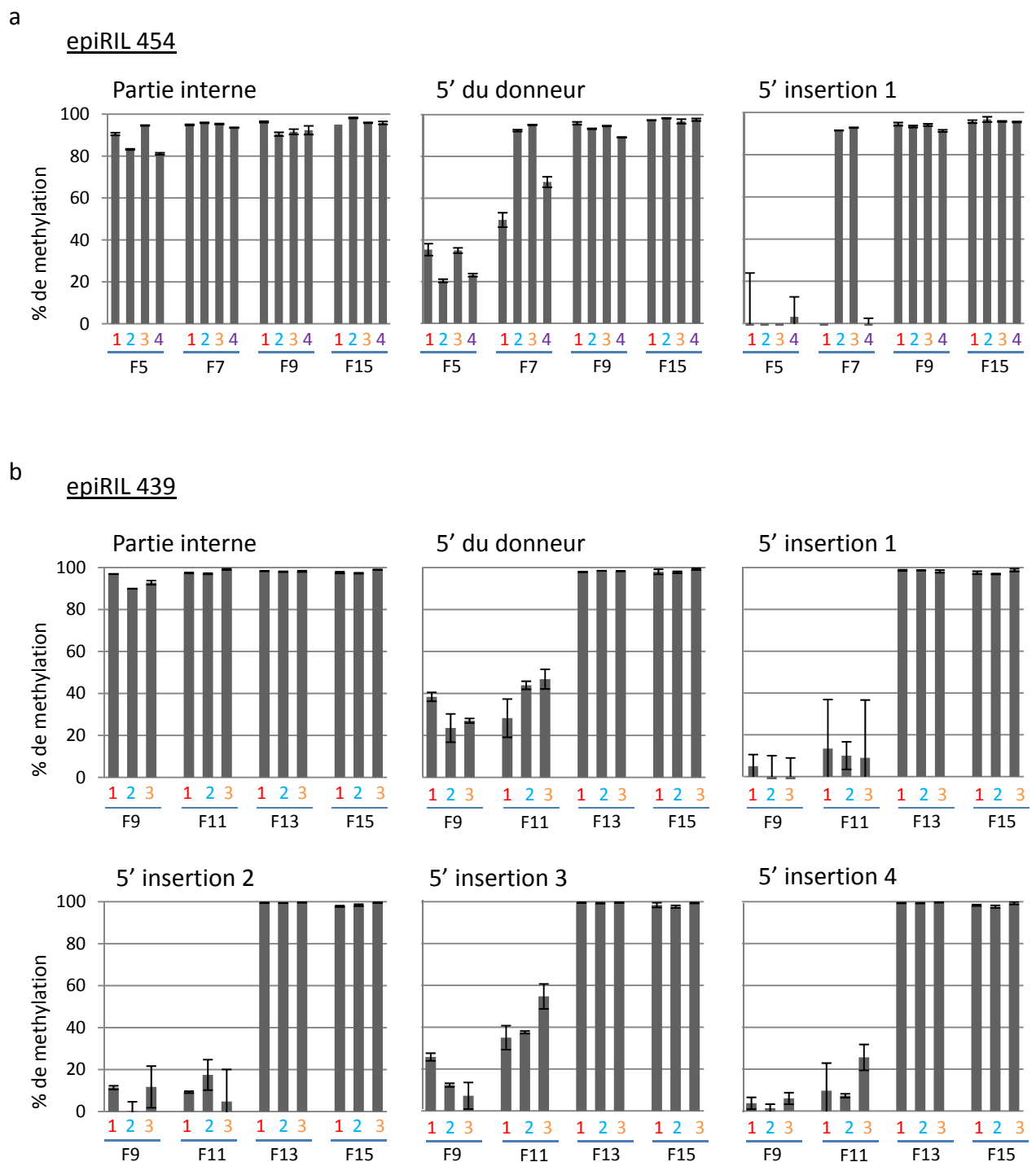


Figure 3.7 : Etude de l'établissement de la méthylation de l'ADN sur *AT2TE20205* (locus donneur et nouvelles insertions) au cours des générations dans les epiRIL 454 et 439 (mesures faites sur des plantes individuelles, 4 plantes par génération pour l'epiRIL 454 et 3 pour l'epiRIL 439). Les barres d'erreur 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 apparaît 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 *dmd1*, une approche génétique permettrait d'élucider ce point. Dans le cas où 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.

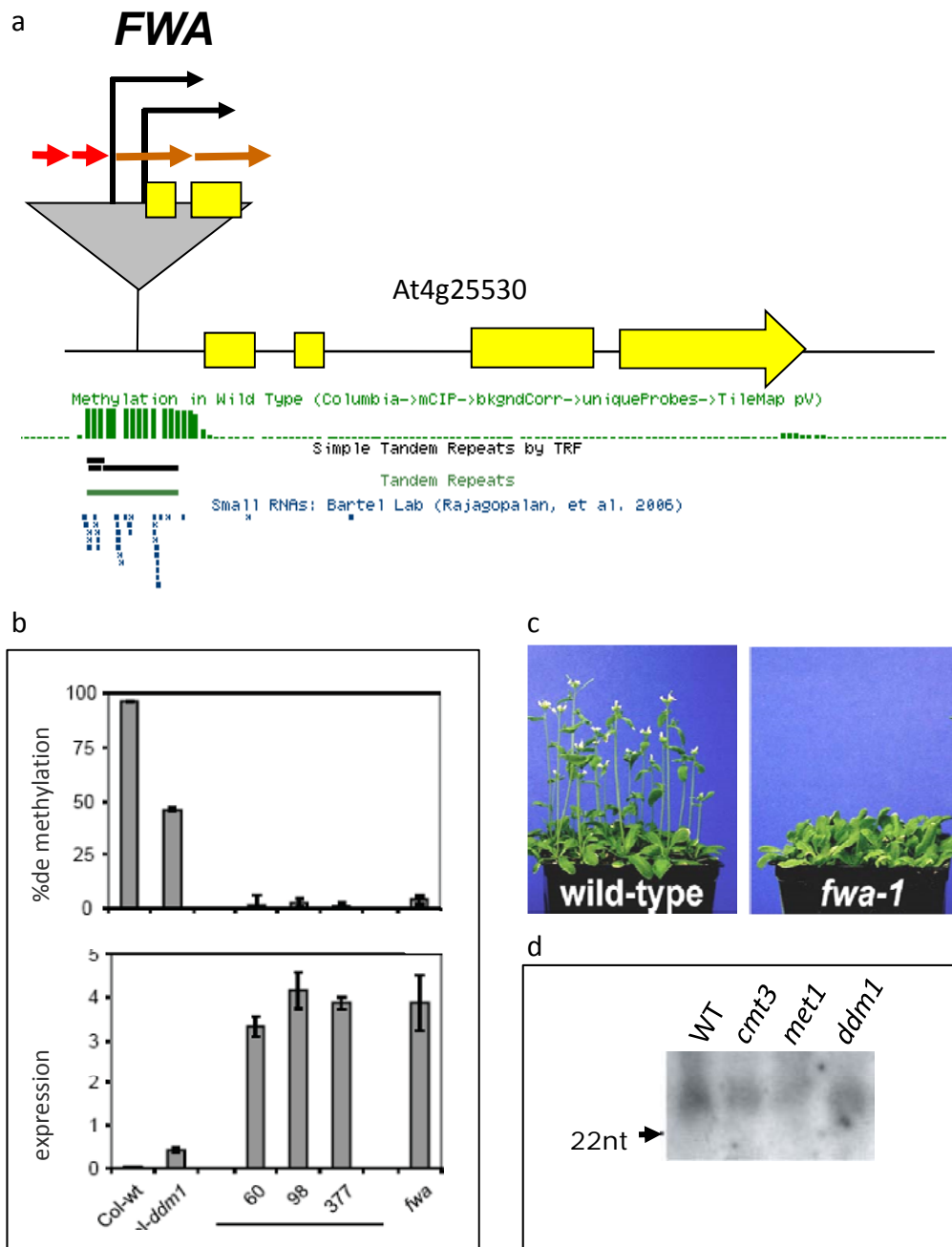


Figure 3.8 : caractéristiques et propriétés du locus *FWA*.

(a) représentation schématique du locus, les flèches rouges et ocres représentent 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

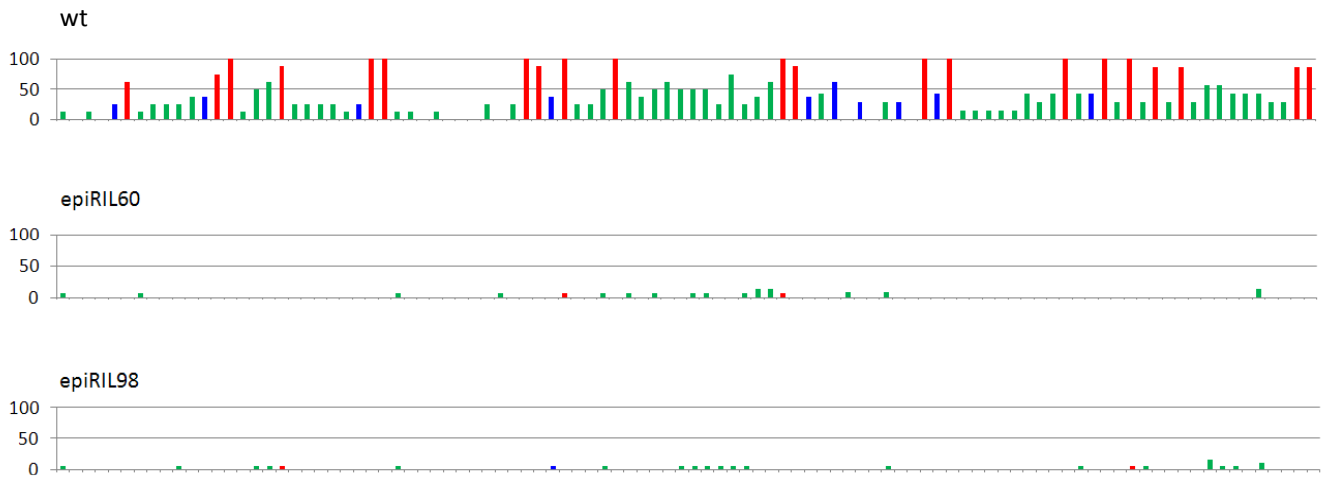


Figure 3.9 : Analyse de la méthylation de la séquence répétée en tandem qui compose le promoteur du gène *FWA* par séquençage bisulfite. Chaque barre représente une cytosine.

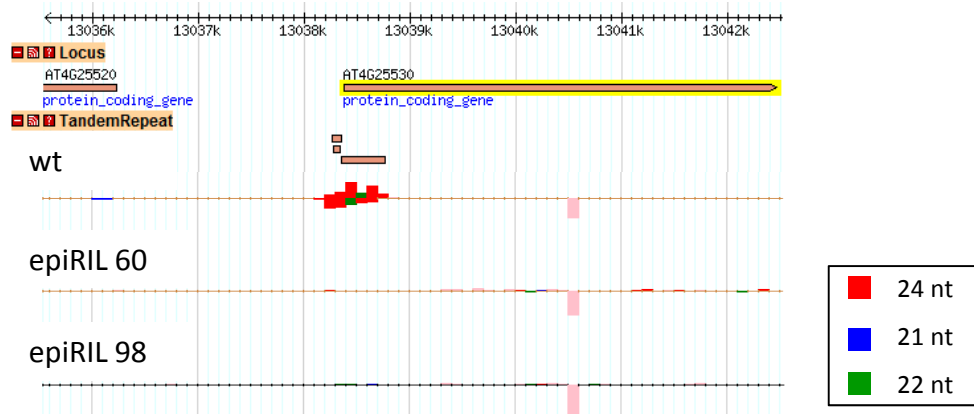


Figure 3.10 : Vue Gbrowse des données de séquençage profond des petit ARN au niveau du locus *FWA* pour le sauvage et les epiRIL 60 et 98.

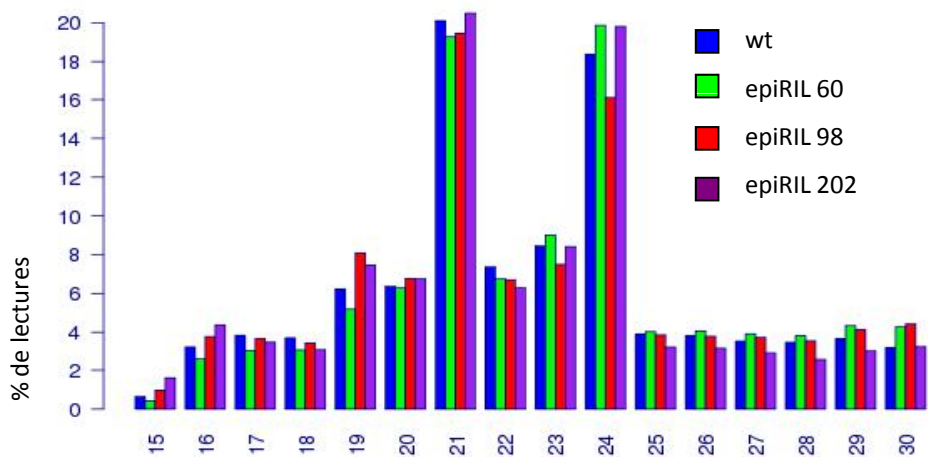


Figure 3.11 : Analyse des populations de siARN chez le sauvage et les épiRIL.

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

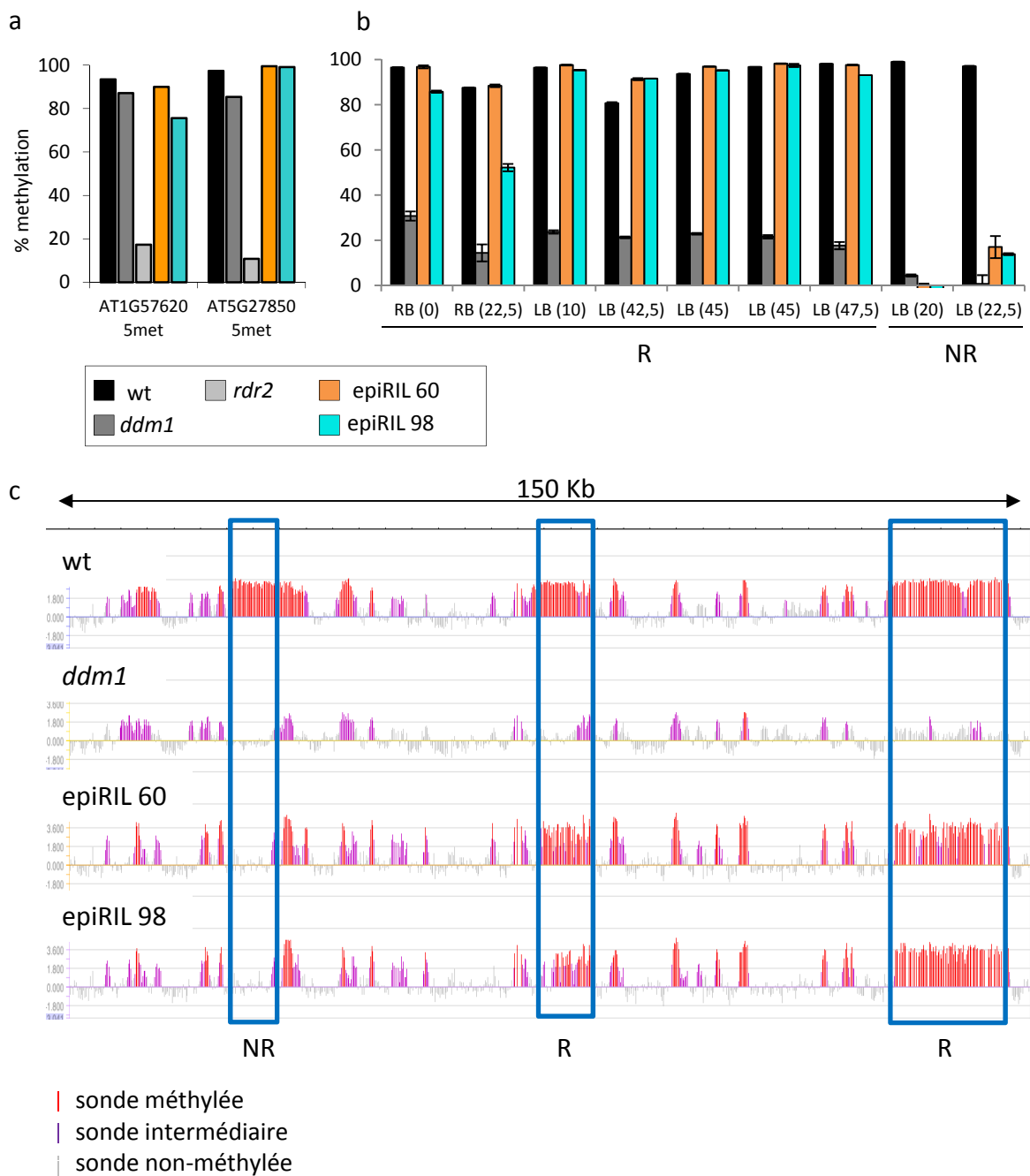


Figure 3.12 : Analyse de la méthylation de l'ADN dans les epiRIL 60 et 98.

(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 O⁶-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 par 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 événements distincts de mutagénèse. Dès lors, une mutation donnée ségrégera dans la moitié seulement

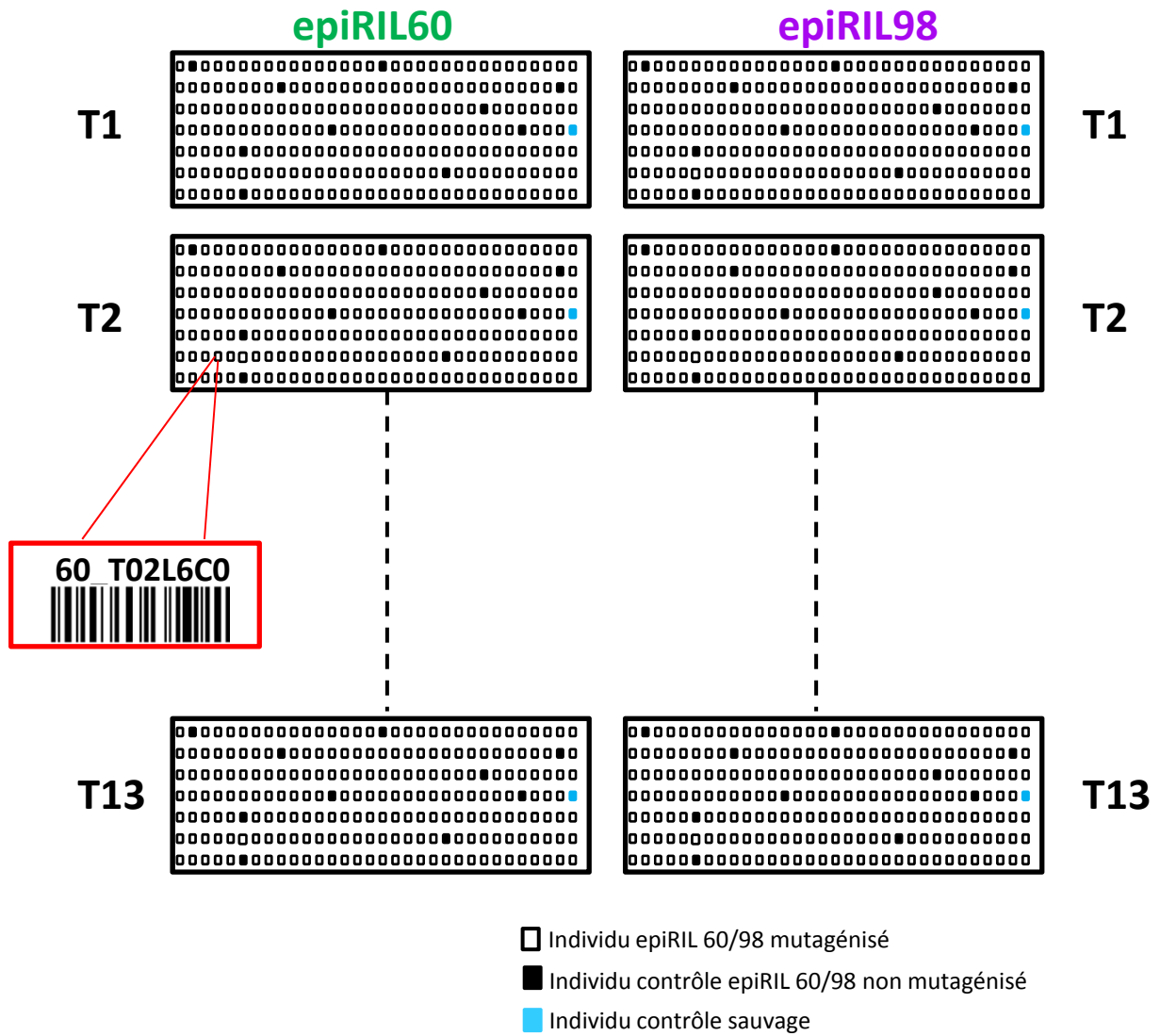


Figure 3.13 : Plan de la serre pour le semi des 6000 plantes M1.

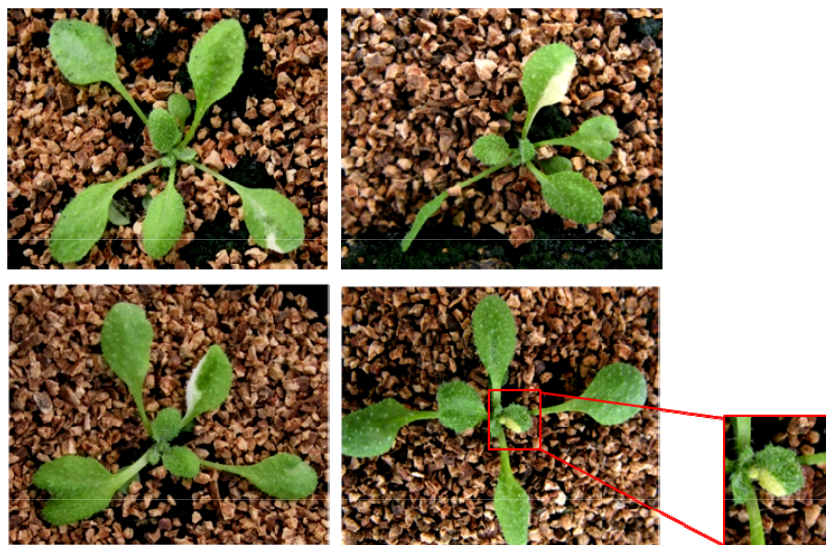


Figure 3.14 : Exemple d'individus présentant des secteurs albinos caractéristiques des plantes ayant subi une mutagenèse par traitement à l'EMS.

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 étudiés, 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.

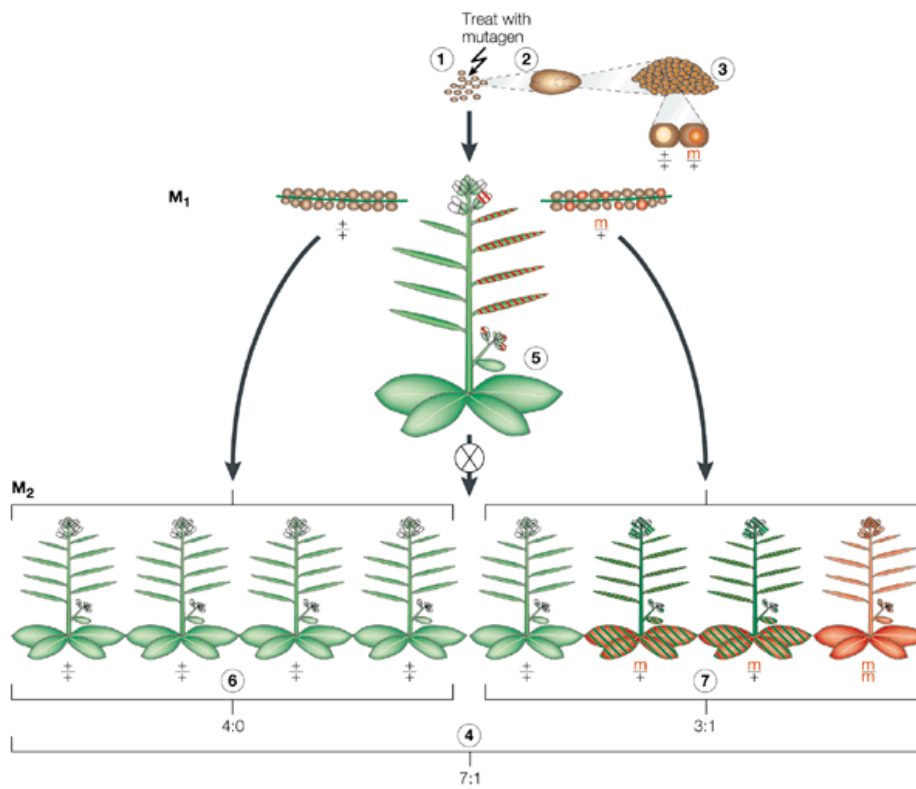


Figure 3.15 : Schéma explicatif illustrant la ségrégation des mutations suite à une mutagenèse par traitement à l'EMS. (Page et al., 2002)

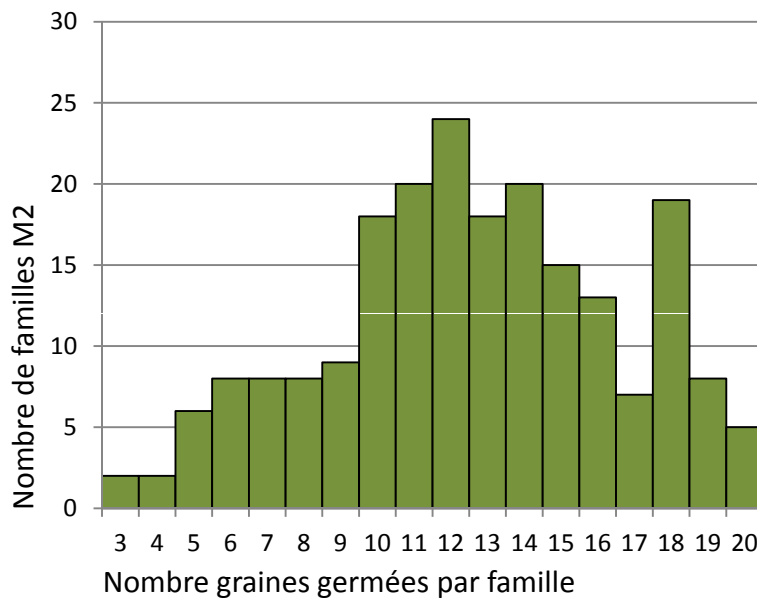


Figure 3.16 : Distribution du taux de germination des plantes mutagenisées.

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 mutagenèse) 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 première 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*.

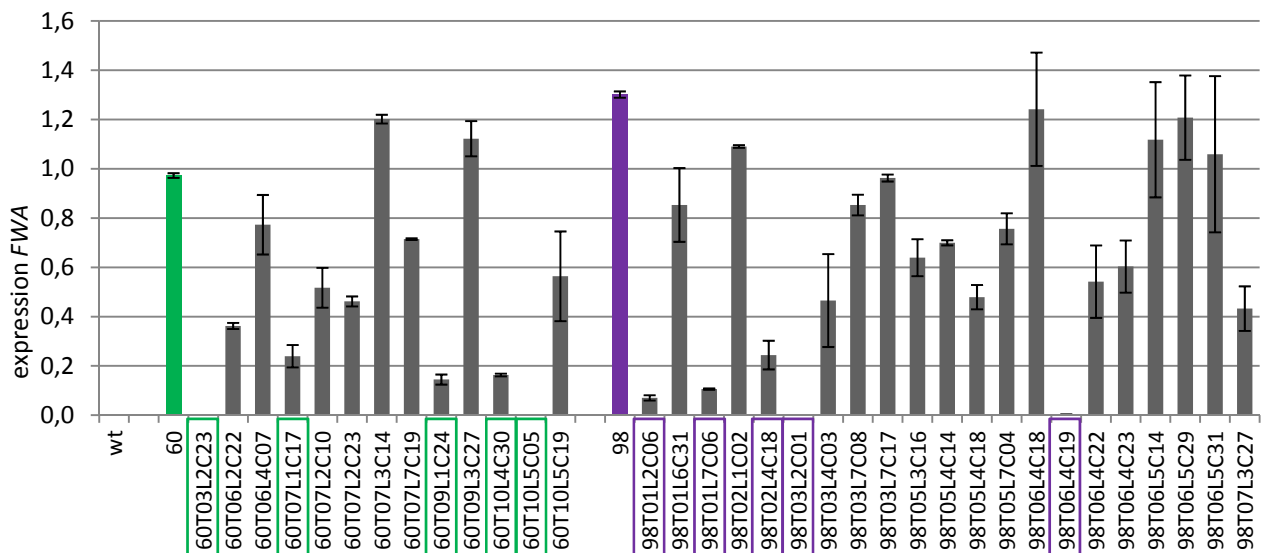


Figure 3.17 : Mesure de l'expression de *FWA* par qPCR exprimée en fonction du gène de l'actine chez les candidats isolés sur la base du critère phénotypique « avancement de la floraison ». Les candidats sélectionnés sont encadrés.

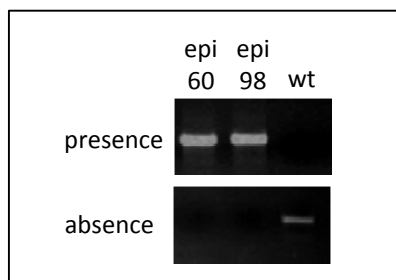


Figure 3.18 : Mise en évidence de la présence/absence par PCR d'une nouvelle insertion d'*ATCOPIA93* utilisée pour génotyper les candidats.

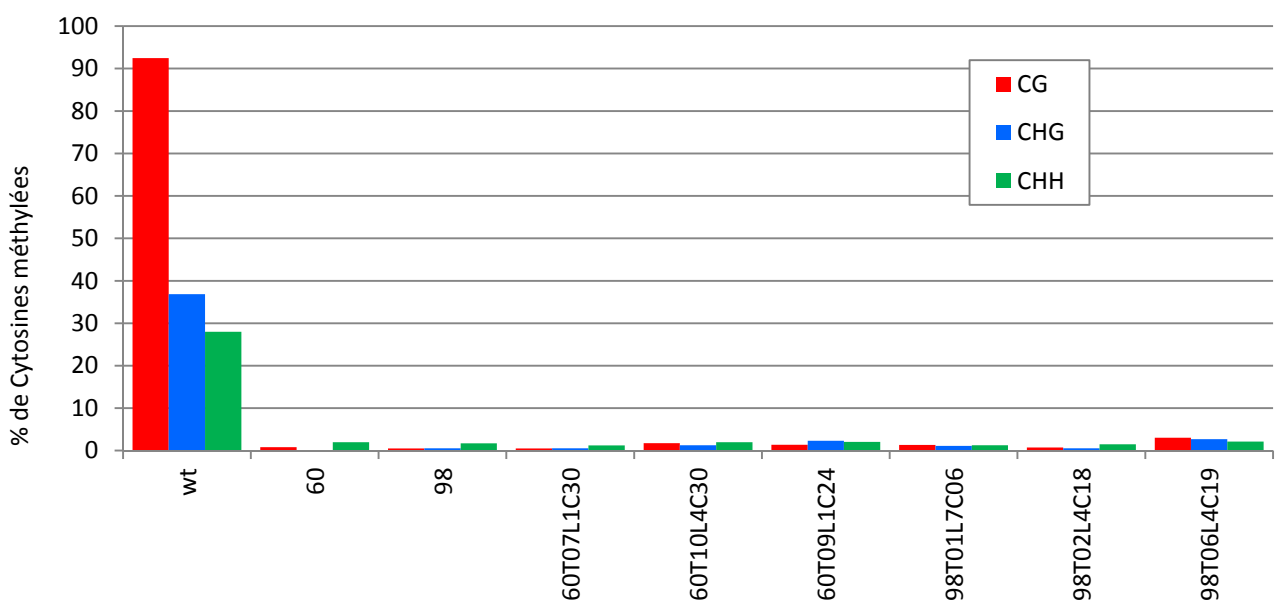


Figure 3.19 : Résultats du séquençage bisulfite de la séquence répétée en tandem de *FWA* chez les 6 candidats.

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 mutagenèse 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 mutagenèse 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.