Open questions about ongoing conflicts between transposable elements and their plant hosts

Alexandros Bousios[1] and Brandon S. Gaut[2]

[1] School of Life Sciences, University of Sussex, Brighton, UK

[2] Department of Ecology and Evolutionary Biology, UC Irvine, Irvine, CA 92697, USA

Author for Correspondence:

Brandon S. Gaut; Department of Ecology and Evolutionary Biology, UC Irvine, Irvine, CA 92697, USA; bgaut@uci.edu

Abstract

Transposable elements (TEs) comprise the majority of plant genomes, but most are epigenetically suppressed and therefore inactive. Research over the last decade has elucidated many of the components and mechanisms that contribute to TE silencing. In contrast, the evolutionary interactions between TEs and silencing pathways are less clear. Here, we discuss current information about this interaction both from mechanistic and evolutionary perspectives, by focusing on the possible states that TEs assume within a genome. Of special interest is the interphase between a TE's escape from silencing and the resumption of host control. We also discuss uncertainties about the host processes that reinitiate silencing, the regions of TEs that may be key targets for host recognition, the energetic costs for maintaining the silencing pathways and the long-term fate of TEs.

Introduction

Transposable elements (TEs) and their plant hosts engage in a continuous battle, whereby TEs seek to proliferate and hosts strive to control their proliferation. In evolutionary terms, it is difficult to declare a winner. On one hand, TEs have been successful by any measure; the majority of plant genomes are composed of TEs that vary from young intact insertions to old fragmented copies. This composition suggests that TEs often overcome host defenses, at least on evolutionary timescales. On the other hand, most (if not all) TEs appear to be epigenetically silenced under normal conditions, so that they do not proliferate. This near-universal suppression implies that plant hosts are largely in control of TE activity.

Epigenetic silencing relies heavily on a process known as RNA-directed DNA methylation (RdDM). RdDM has been studied intensively for many years, leading to the elucidation of its complex mechanisms and overlapping pathways. Here, we will not cover the interactions among RdDM components in much detail, as these have been presented in several recent reviews [*1-3]. Rather, we will first provide a brief overview of the potential states of a TE within a host genome, as well as the silencing pathways that influence those states, before highlighting questions of particular interest from mechanistic and evolutionary perspectives.

Overview of the TE epigenetic silencing pathways

RdDM employs two plant-specific RNA polymerases, *Pol IV* and *V*, both of which appear to be most important after a TE is initially silenced (Figure 1). [We discuss the initiation of silencing below.] *Pol IV* transcribes silenced TEs to produce 24 nucleotide (nt) siRNAs in concert with *RNA-directed RNA polymerase 2* (*RDR2*) and *Dicer-like 3* (*DCL3*). The 24nt siRNAs are then loaded onto *Argonaute 4* and *6* (*AGO4*/*AGO6*) proteins to direct them (by sequence complementarity) to scaffolding transcripts of the TEs, which are produced by *Pol V*. The association with *Pol V*-derived, chromatin-bound transcripts mediates cytosine methylation and the deposition of repressive histone marks, such as H3K9me, to the target TEs by chromatin modifying proteins [*1-3]. Crucially, recent work in Arabidopsis has shown that *Pol IV* and *Pol V* are recruited to TE loci that already contain H3K9me and cytosine methylation respectively [4,5]. These observations confirm that RdDM acts in a self-reinforcing loop that increases the deposition of silencing marks in a phase known as "establishment and spreading" of silencing [6],

eventually generating a dense heterochromatic environment that restricts access to the *Pol II* transcriptional machinery. The repressive chromatin modifications can be inherited to daughter DNA strands during cell replication independently of RdDM, while residual siRNA targeting may further increase heterochromatic formation to maximum levels; overall, this phase is known as "maintenance" of silencing [6]. Eventually, the process leads to deep silencing of TEs (Figure 1).

Nonetheless, the abundance of TEs indicates that they do escape suppression and proliferate. When this happens, a TE is likely transcribed by *Pol II* to start its life cycle towards producing new copies. At some point, however, the endogenous *RNAi* mechanism recognizes the TE mRNA and degrades it into 21-22nt siRNAs, aided by *RDR6*, *DCL2*/*DCL4* and *AGO1* [6,7]. This post-transcriptional step can lead to multiple cycles of *RNAi*, but, most importantly, it also triggers the "initiation" phase of transcriptional TE silencing (Figure 1). The mechanism of this transition was unknown until recently, when evidence from two independent groups led to two non-mutually exclusive models. The first suggests that some 21-22nt siRNAs are not loaded onto *AGO1*, which directs mRNA cleavage, but instead onto *AGO6* that guides chromatin modifications [*8]. The second proposes that the overproduction of TE mRNAs eventually overwhelms *DCL2*/*DCL4*, allowing *DCL3*-mediated biogenesis of 24nt siRNAs, which in turn direct heterochromatic formation through their loading onto *AGO4*/*AGO6* [*9]. Both models imply that *Pol V* is recruited to unmethylated loci to produce scaffolding RNAs, but this prediction has not yet been proven. In any case, "initiation" places the first heterochomatic marks on TEs, thereby triggering the "establishment and spreading" phase of silencing (Figure 1).

What happens between TE escape and the recovery of host control?

Escape from silencing

Given the prevalence of TEs within plant genomes, it is surprising that few have been observed to actively transpose and proliferate. Notable exceptions include the *Ac*/*Ds* and *Mutator* families of maize and the *mPing* element in rice [10-12]. However, genome-wide and/or biochemical analyses have shown that silenced elements can be activated during epigenetic loss or remodeling, as for example in mutants that establish or maintain silencing [13,14], in cell types such as the companion cell of the female gametophyte or the pollen vegetative nucleus [15,16], in developmental transitions from juvenile to adult [17], and during stress [18]. Furthermore, TEs may have themselves evolved mechanisms to escape, such as recombination between elements that generates new variants unrecognizable to host defenses [19,20] or integration near genes where epigenetic suppression may negatively affect gene expression [21]. In fact, some TE families tend to insert within genes [22]. Both the close proximity and physical overlap with genes suggest integration into regions where silencing cannot be robustly enforced, either due to its deleterious effects or because the chromatin environment must remain open for proper gene function [23-25].

Another way to escape silencing may be to enter a naïve genome through horizontal transfer. A growing body of data suggests that these events occur more often than initially thought between vertebrates and between invertebrates [26,27], but research on plants remain underrepresented [28]. Nonetheless, a study by El Baidouri et al. (2014) investigated the genomes of 40 angiosperms and found evidence for extensive horizontal

transfer of LTR retrotransposons [*29]. Actually, some elements underwent subsequent amplification bursts, suggesting both that they remained active after initial colonization and that this process may be an important factor in plant genome evolution.

Finally, in a fabulous review, Lisch suggested that TEs might counteract suppression by capturing gene fragments [30]. In theory, targeting of these elements by siRNAs could also affect the expression of the 'parent' genes from which the fragments were captured. Lisch's hypothesis was based primarily on the observation that *Pack-MULE* and *Helitron* TEs systematically acquire exons during transposition [31,32]. However, despite the extent of gene capture, most research has focused on the possible expression and on the evolutionary patterns of selection of these fragments [32-34] rather than on their potential effects to host targeting strategies. Surprisingly, the extent of gene capture by LTR retrotransposons, which make up the majority of plant genomes, is virtually unknown.

Reinitiating silencing: homology-dependent siRNA surveillance?

Once a TE is activated, host defenses appear to employ intricate ways to initiate their silencing. One possibility is that endogenous reactivated TEs can be suppressed by 24nt siRNAs produced by other members of the same family that have been silenced already. In this homology-dependent pathway, the *Pol IV*-derived 24nt siRNAs act as "immune memory" to identify highly similar TEs that are active [*1]. Although (as discussed previously) it remains unknown if and how *Pol V* is recruited to these unmethylated TEs, there is evidence that reactivated TEs in Arabidopsis and maize are silenced in *trans* by the *Pol IV*-RdDM pathway [18,35,36]. Perhaps, the presence or absence of immune

memory also explains whether a horizontal TE invasion will amplify in large numbers before it eventually becomes silenced. If, for instance, an endogenous silenced TE happens to share enough sequence homology with the invading TE, then its 24nt siRNAs may suffice to constrain the invader's proliferation.

There are, however, indications that the homology-dependent pathway is not always sufficient to silence active TEs. For example, the *Evade* element of the *ATCOPIA93* family in Arabidopsis remains active despite the presence of silenced relatives [37], and it is suppressed only when enough mRNA transcripts are produced to overwhelm *DCL2*/*DCL4* processing [*9]. Also, recent genome-wide studies of LTR retrotransposons [38,39] have shown that certain families, such as *Ji* and *Opie* in maize, have exclusive homology to large numbers of 21-22nt siRNAs, implying that (at least) some of their members are first expressed and then recognized by *RNAi* (Figure 1). Nonetheless, these families are also homologous to tens of thousands of 24nt siRNAs; if the homology-dependent pathway were able to quickly suppress active members, then one would not expect the presence of such large populations of smaller siRNA lengths.

Reinitiating silencing: identification of TE mRNA transcripts by RNAi

If the homology-dependent pathway fails, silencing of active TEs must be initiated by *RNAi*. As mentioned above, the first step likely entails the recognition of *Pol II*-derived TE mRNA. Importantly, *RNAi* needs to safely distinguish between TE and genic mRNAs, but this must be problematic given the high sequence and structural diversity among TE classes and, hence, the lack of a universal TE 'barcode'. *RDR6* is responsible for this crucial and poorly understood entry-point by converting the single-stranded

mRNA into double-stranded RNA [3]. It has been hypothesized that *RDR6* is assisted by 'primary' small RNAs that act either directly as primers or indirectly by their association with *AGO* proteins that cleave TE mRNAs to provide a starting point for double-stranded synthesis [40,41]. Indeed, some miRNAs play this role in Arabidopsis [**42]. However, miRNAs also participate heavily in genic mRNA cleavage, and so the selection of TE over genic mRNAs remains complicated. Furthermore, only a subset of Arabidopsis TE families are silenced via miRNAs [**42], implying that more mechanisms must be in place.

Intriguingly, one such mechanism could involve non-miRNA hairpin-derived small RNAs (hpRNAs) that are generated from hairpin structures within the TEs themselves. The first evidence for the importance of hpRNAs came from the *MuDR* family in maize. A spontaneous inverted repeat rearrangement within a single element (termed *Mu killer*) triggered *RNAi* and subsequent silencing of other *MuDR* members [35,43]. In fact, the *Mu killer* case prompted Lisch and Slotkin to suggest that the very low fidelity of TE replication may be the critical feature for attracting *RNAi*, because of the high chances for a 'killer' element with a hairpin structure to appear [22]. Axtell further discussed the potential importance of plant hpRNAs and noted that the numerous short hairpins in genomes (which also occur within TEs) may produce hpRNAs [44]. Although a systematic identification of hpRNAs has not yet been reported [45], they appear to be abundant in maize [46]. These arguments are supported by a study on Sirevirus LTR retrotransposons in maize [*39], which contain complex palindrome motifs in their LTRs that form stable secondary structures predicted to generate hpRNAs. Because *DCL* proteins can directly process hairpins into hpRNAs in Arabidopsis [47],

Sirevirus palindromes may act as the locus for hpRNA production, which then act as primers for *RDR6* synthesis of double-stranded RNA. Alternatively, the hairpins themselves could directly prime *RDR6* synthesis in a 'primary' small RNA-independent manner [48]. If true, palindromic regions may be crucial for reinitiating silencing within plant genomes, because, for example, the majority of Sireviruses from various plant hosts contain such structures [49]. Altogether, these studies suggest that recognition mechanisms may be specific to different TE families. It is thus likely that plants have evolved family-specific mechanisms to trigger silencing, based on their individual sequence characteristics.

The potential importance of *cis*-regulatory regions of TEs

From the selfish view of a TE, it is not entirely desirable to contain a sequence or region that provides a holdfast for host recognition. Why, then, might Sireviruses contain a palindromic region? The answer is simple: because it may be essential for an element's ability to proliferate. Indeed, the Sirevirus palindromes are found in the highly conserved *cis*-regulatory area of LTR retrotransposons, located in the upstream half of the LTRs [50-52]. Nonetheless, this region in Sireviruses is also a template for sequence evolution, because the palindromes differ along the sequence of an element, among family members, and among families [*39]. Furthermore, they appear to be epigenetic hotspots, as they are heavily targeted by siRNAs, especially of smaller lengths; however, each palindrome is targeted by different siRNAs, as a result of their differing sequence and positional characteristics. This example of strong interaction between *cis*-regulatory motifs and siRNAs raises the possibility that *cis*-elements may represent regions of

intense evolutionary arms race between TE and host defenses. Jacobs et al. beautifully describe an analogy of this interplay for the *L1* and *Alu* TE families in primates [**53]. Sequence and structural changes within their regulatory regions led to multiple amplification bursts during primate evolution, but these changes were always met by equivalent changes in their host's silencing components.

Detailed analyses of the epigenetic importance of the *cis*-regulatory regions of plant TEs are limited. However, studies in diverse organisms (mice, *Drosophila*, tomato, hominoids) and for various TE families have revealed that these regions are often arranged as arrays of repeats, with some also exhibiting symmetrical properties [54-57]. Therefore, the complex, possibly palindromic organization of regulatory motifs may be common across TE classes. In fact, for the TE family in tomato (termed *Tnt1*), the *cis*-regulatory region is the only highly diverse region among family members from several *Solanaceae* species, and this variability associates with different stress-related expression patterns [58,59]. This led to the hypothesis that the evolution of this region equipped *Tnt1* subgroups with diverse regulatory capacity to respond to different stimuli and colonize new hosts [59,60]. Similar patterns of diversity within the regulatory region was recently reported among *Tcs* elements in citrus [61]. Collectively, these observations denote the need for in-depth research on the evolutionary interplay between *cis*-regulatory TE motifs and epigenetic silencing.

The paradox of the time to regain control

There is increasing evidence that epigenetic defenses quickly respond to TE activation, achieving suppression within a few host generations. For example, the *ONSEN* LTR

retrotransposon was activated by heat-inducement of Arabidopsis seedlings, but it only managed to insert new chromosomal copies in progeny of a siRNA-compromised genetic background and not in the progeny of wild-type plants or in vegetative tissues [18]. Another study in Arabidopsis showed that, after severe loss of cytosine methylation, many TEs were remethylated over a few host generations [36]. Finally, when the *Evade* Arabidopsis element was awakened in epigenetic recombinant inbred lines that erased methylation, it was transcriptionally silenced after ~15 host generations or after ~40 new copies had been integrated in the genome [*9].

Based on these observations, we are faced with the following paradox: if TEs are rapidly resilenced, then how do amplification bursts occur that produce thousands of copies? Unfortunately, there is not yet a compelling answer to this question. In part, this may be because most analyses have thus far focused on plants with small genomes, where the number of identified full-length elements is relatively low. Additionally, the common categorization of TEs at the superfamily level (e.g. *Copia* and *Gypsy*) may obscure family and subfamily dynamics. However, comparative analysis of species from the *Oryza* genus [62], and work on the reference 1Gb sequence of chromosome 3B in wheat [63,64] and the fully sequenced 2.3Gb maize genome [65,66] have shown the accumulation of thousands of copies for several families, typically within an estimated timescale of thousands of years. It is likely that many TEs in plant hosts with moderate to large genomes have similar life histories. In such cases, do TE families repeatedly escape silencing for much longer lengths of time than have been observed experimentally thus far? Or, do they undergo repeated short bursts, producing a few new copies each time before they are resilenced? If so, how could they escape over and over again, given

homology-dependent silencing? We have very few insights into the timescale and conditions of bursts, although they presumably coincide with epigenetic losses due to recurring biotic or abiotic stresses.

What are the energetic costs of epigenetic silencing?

In evolutionary terms, it does not pay to maintain an energetically expensive process unless it contributes to fitness. The energetic costs of TE epigenetic mechanisms must be substantial to the host, but how costly are they and under what conditions are they maintained? Let us first consider the maintenance of TE repressive chromatin modifications (Figure 1) as a *de facto* necessity for cell integrity; we assume, therefore, that hosts cannot afford to diminish these processes regardless of energetic costs. Our focus then turns to siRNAs, which appear to be produced abundantly from both pericentromeric heterochromatin and from euchromatic chromosomal arms based on sequencing and mapping of *Pol IV*-derived transcripts and their associated siRNAs in Arabidopsis [**67]; in fact, 65% of the *Pol IV* transcripts mapped to TE loci, compared to 9% to genes [**67]. Surprisingly, few papers have directly calculated the proportion of siRNA libraries that has homology to TEs, despite abundant information about small RNAs from several plant species (for example, http://mpss.udel.edu/). However, limited evidence in Arabidopsis suggests that siRNAs that map to TEs constitute a substantial proportion of small RNAs [45], and up to 11.3% of siRNAs map to a set of ~6,500 maize Sireviruses that constitute only 2.8% of the genome [*39]. Given that Sireviruses totally occupy ~20% of the maize genome [66], it is likely that a much higher proportion of siRNAs corresponds to them. More detailed calculations across plant species and

tissues would be valuable, but the overarching impression is that most siRNAs are homologous to TEs.

This raises the question: why incur the energetic costs to produce siRNAs when most TEs are highly methylated [68-70] and, therefore, presumably deeply silenced? There are at least three potential answers. The first is that siRNA production is not costly in an evolutionary sense, so that siRNA production is subject to genetic drift rather than natural selection. If true, however, the system would be lost as (or more often than) retained under genetic drift. Second, siRNAs – and particularly 24nt siRNAs – may function as immune memory. Under this scenario, siRNAs act as a buffer against the possibility of TE activity, even though most methylation is maintained independently of RdDM in heterochromatic regions [**71]. The retention of siRNA surveillance would be similar to the evolution and retention of acquired immune memory in vertebrates, which is costly but maintained under frequent cycles of reinfection [72]. Consistent with this view, some observations suggest that the cycles of epigenetic emergencies may be frequent enough to be nearly constant. For example, many TE families are homologous to 21-22nt siRNAs [38,39], suggesting that TE expression, if not transposition, is common. Additionally, some TEs are expressed in mutant backgrounds with impeded 24nt siRNA synthesis [13,71], indicating that the mutant background releases them from the ongoing "establishment and spreading" phase of silencing (Figure 1).

Finally, another possibility is that siRNA production is necessary for a particular step of the host lifecycle, so that it cannot be lost. The fact that Arabidopsis TEs are epigenetically released in companion cells of the male and female gametophyte and then apparently reprogrammed in the sperm and egg cells is consistent with this assertion

[15,16]. However, if this were the only necessary function for siRNAs, evolution would be expected to dampen production in non-reproductive tissues, thereby conserving costs. In short, the evolutionary pressures that maintain this costly system of the host response are not yet clear, and they could be multifaceted in that they serve as immune memory, are often a response to ongoing emergencies, and may be essential for some aspects of host reproduction.

Do TEs age gracefully?

Much is known about siRNA targeting and methylation of TEs during different stages of silencing, but our understanding of the evolutionary dynamics of these epigenetic features as a function of TE age is limited. Two studies have investigated the relationship between siRNA targeting and insertion age to find that older TEs were targeted at lower levels by all siRNA lengths compared to their younger relatives [*39,73]. This implies that older TEs are likely to be deeply silenced, a state that can be maintained independently of siRNAs (Figure 1). If true, one also expects methylation to increase with age, as elements approach their maximal methylation levels. This expectation holds for maize Sireviruses [*39] (and apparently for mice TEs [74,75]), but not for rice LTR retrotransposons [76]. It is not yet clear if the differences between rice and maize are species-specific, or specific to the TE families examined.

The same study of maize Sireviruses revealed another additional intriguing pattern: while methylation levels increased with age, levels of both methylation and especially siRNA targeting were aberrant for a small subset of very old (>2.5 million years) elements [*39]. These old elements had epigenetic properties similar to those of

their youngest counterparts – i.e. lower methylation and high siRNA targeting. The reason for this pattern is mysterious, but one possibility is that the old elements constitute *in silico* evidence for the existence of 'zombie' TEs, which are defined as elements that are co-opted by the host because they can quickly trigger both their own and also the *trans*-silencing of active relatives [22,30]. The existence of zombies is, at this point, more hypothetical than proven, but they have an analogy to the loci that produce piwi-interacting RNAs in *Drosophila* [77].

There are also intriguing relationships between TE age and proximity to genes. This was first noted in Arabidopsis [21] and then rice [76], where it was suggested that elements were removed by natural selection as a consequence of their deleterious effects on genes when they were epigenetically silenced. In addition, genes tend to reside in regions of high recombination, where natural selection is more efficacious [78] and rates of TE removal higher [79]. To further investigate this relationship, we retrieved age information for ~6,500 maize Sireviruses from [80] and examined their distribution in relation to gene proximity (Figure 2). A positive (but weak) correlation was produced (Pearson $r = 0.1$; $P < 10^{-20}$), suggesting that elements are removed from the genome more quickly when they are near to genes. In other words, a TE has a better chance to reach a ripe old age if it inserts within, or becomes part of, a heterochromatic environment; yet, there may be exceptions to this rule, too. Notice that there is a discontinuity in Figure 2, which is caused by the old, potentially zombie elements residing closer to genes than expected based on their age alone. We do not know why they demonstrate this discontinuity, but if they were co-opted as zombies, they would then be expected both to

be retained (as a beneficial component of the host response) and also to lie in regions of open chromatin where *Pol IV* and *Pol V* transcription is ongoing.

Future directions

There is still a great deal of work required before the evolutionary and mechanistic features of the conflict between hosts and TEs are completely understood. While future research could take many directions, we emphasize on four:

Embrace Diversity: Much effort has been focused on elucidating RdDM processes in Arabidopsis, along with complementary studies in rice and maize [2]. Although there are many questions left to answer (some of which we have highlighted here), these efforts have yielded detailed insights into the pathways governing interactions between TEs and their hosts. It is worth noting, however, that these three species are not "normal", because they have much smaller genomes than the angiosperm average [81]. Moreover, Arabidopsis is particularly bizarre because its elements tend to be substantially older than those of its sister species *Arabidopsis lyrata* [82], suggesting a dearth of transposition events in the recent past. Is it possible that these model plants (and especially *A. thaliana*!) are atypical with respect to the steps involved in the host response? For this reason, we endorse continued epigenetic studies across a wide variety of hosts.

Comparative Epigenetics and Molecular Evolution of Host Response Genes: Elucidating the mechanisms of the host response will be difficult in non-model species. A helpful first-step will be to study the evolution of known genes that contribute to epigenetic mechanisms. For example, a recent paper has found that angiosperms are unique in containing some RdDM-related genes [83], indicating that TE:host interactions

may differ substantially among angiosperms and lower plants like gymnosperms, ferns and mosses.

Developmental Biology: There is an urgent need for more work on the host response as a function of development. In our opinion, one of the most exciting discoveries of the last decade was the finding that TEs were epigenetically released and then reprogrammed within the male and female gametophytes [15,16]. Is this reprogramming a necessary step in the Arabidopsis lifecycle? Is it shared with other plants? If so, how broadly?

Evolutionary Bioinformatics of TEs: We know very little about the focal regions of TE:siRNA interactions, because most studies have focused on mapping siRNA to consensus TE sequences [*8,23,42,46,68,71,84,85]. Unfortunately, the process of building a consensus is likely to omit the very regions of interest – i.e. those that evolve rapidly due to an ongoing arms race between TEs and hosts. The accurate identification and classification of large numbers of individual TEs remains, however, a challenging task [86,87]. Nonetheless, a recent review has stressed the need for fine-scale characterization of plant TEs to properly assess epigenetic dynamics [88], and here we amplify that sentiment.

Acknowledgements

We thank Nikos Darzentas for helpful discussions on the manuscript. AB is supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no: PIEF-GA-2012-329033. BSG is supported by NSF grant IOS-1542703.

Figure captions

Figure 1. States of TEs and their interactions with the epigenetic silencing pathways of the host. Active TEs may derive from either horizontal transfer or the escape of endogenous TEs from epigenetic suppression. The initiation of suppression of active TEs may depend on homology-dependent silencing or RNAi, which is facilitated by known or yet-to-be-identified triggers. RNAi cleaves TE mRNAs post-transcriptionally (PTS), but also places the first heterochromatic marks on TE insertions, hence initiating transcriptional silencing (TS). After initiation, RdDM strengthens suppression in a self-reinforcing loop termed establishment and spreading of silencing. These silenced TEs may slowly reach their maximum methylation levels while ageing, a process that may be largely maintained independent of RdDM. See text for additional details.

Figure 2. The proximity of Sirevirus elements to maize genes as a function of their insertion age. We retrieved information for 6,456 Sireviruses from MASiVEdb (http://databases.bat.infspire.org/masivedb/) [80] and allocated them into age groups as in [*39]. The number at the top of each boxplot indicates the number of elements within each group. Gene information was retrieved from the Filtered Gene Set of the maize B73 RefGen_V2 genome. my, million years.

References and recommended reading

*1. Fultz D, Choudury SG, Slotkin RK: **Silencing of active transposable elements in plants**. *Current Opinion in Plant Biology* 2015, **27**:67-76.

This review lucidly summarizes the complex pathways of epigenetic silencing, with an emphasis on the least understood step of initiation of silencing.

2. Matzke MA, Kanno T, Matzke AJM: **RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants**. In *Annual Review of Plant Biology, Vol 66*. Edited by Merchant SS; 2015:243-267. Annual Review of Plant Biology, vol 66.]

3. Matzke MA, Mosher RA: **RNA-directed DNA methylation: an epigenetic pathway of increasing complexity**. *Nature Reviews Genetics* 2014, **15**:394-408.

4. Johnson LM, Du JM, Hale CJ, Bischof S, Feng SH, Chodavarapu RK, Zhong XH, Marson G, Pellegrini M, Segal DJ, et al.: **SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation**. *Nature* 2014, **507**:124-+.

5. Law JA, Du JM, Hale CJ, Feng SH, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE: **Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1**. *Nature* 2013, **498**:385-+.

6. Panda K, Slotkin RK: **Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway**. *Plant signaling & behavior* 2013, **8**.

7. Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK: **The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs**. *Plant Physiology* 2013, **162**:116-131.

*8. McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK: **ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation**. *Embo Journal* 2015, **34**:20-35.

This paper shows how 21-22nt siRNAs, which are produced by cleavage of TE mRNA during RNAi, are loaded onto *AGO6* and direct chromatin modifications of active TEs.

*9. Mari-Ordonez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O: **Reconstructing de novo silencing of an active plant retrotransposon**. *Nature Genetics* 2013, **45**:1029-+.

This study follows the epigenetic re-silencing of an active TE in Arabidopsis. Initial protection from RNAi by its *Gag* capsid leads to its amplification. However, after ~40 copies or ~15 host generations *DCL2*/*DCL4* is unable to process all TE mRNA, allowing *DCL3* to produce 24nt siRNAs, guide *AGO4* to genomic copies and initiate silencing.

10. Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y, Tanisaka T, Wessler SR: **Dramatic amplification of a rice transposable element during recent domestication**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**:17620-17625.

11. Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, Conrad LJ, Kikuchi K, Kubinec TA, Hall BD, et al.: **Genome-Wide Distribution of Transposed Dissociation Elements in Maize**. *Plant Cell* 2010, **22**:1667-1685.

12. Lisch D: **Mutator and MULE Transposons**. *Microbiology spectrum* 2015, **3**:MDNA3-0032-2014.

13. Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D, Schnable PS: **Loss of RNA-Dependent RNA Polymerase 2 (RDR2) Function Causes Widespread and Unexpected Changes in the Expression of Transposons, Genes, and 24-nt Small RNAs**. *Plos Genetics* 2009, **5**.

14. Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T: **Bursts of retrotransposition reproduced in Arabidopsis**. *Nature* 2009, **461**:423-U125.

15. Ibarra CA, Feng XQ, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, et al.: **Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes**. *Science* 2012, **337**:1360-1364.

16. Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA: **Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen**. *Cell* 2009, **136**:461-472.

17. Li H, Freeling M, Lisch D: **Epigenetic reprogramming during vegetative phase change in maize**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:22184-22189.

18. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J: **An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress**. *Nature* 2011, **472**:115-U151.

19. Du J, Tian Z, Bowen NJ, Schmutz J, Shoemaker RC, Ma J: **Bifurcation and Enhancement of Autonomous-Nonautonomous Retrotransposon Partnership through LTR Swapping in Soybean**. *Plant Cell* 2010, **22**:48-61.

20. Sharma A, Schneider KL, Presting GG: **Sustained retrotransposition is mediated by nucleotide deletions and interelement recombinations**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:15470-15474.

21. Hollister JD, Gaut BS: **Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression**. *Genome Research* 2009, **19**:1419-1428.

22. Lisch D, Slotkin RK: **Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts**. In *International Review of Cell and Molecular Biology, Vol 292*. Edited by Jeon KW; 2011:119-152. International Review of Cell and Molecular Biology, vol 292.]

23. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK: **CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize**. *Genome Research* 2013, **23**:628-637.

24. Gent JI, Madzima TF, Bader R, Kent MR, Zhang X, Stam M, McGinnis KM, Dawe RK: **Accessible DNA and Relative Depletion of H3K9me2 at Maize Loci Undergoing RNA-Directed DNA Methylation**. *Plant Cell* 2014, **26**:4903-4917.

25. Sequeira-Mendes J, Aragueez I, Peiro R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C: **The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States**. *Plant Cell* 2014, **26**:2351-2366.

26. Bartolome C, Bello X, Maside X: **Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes**. *Genome Biology* 2009, **10**.

27. Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL: **Widespread horizontal transfer of retrotransposons**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**:1012-1016.

28. Wallau GL, Ortiz MF, Silva Loreto EL: **Horizontal Transposon Transfer in Eukarya: Detection, Bias, and Perspectives**. *Genome Biology and Evolution* 2012, **4**:801-811.

*29. El Baidouri M, Carpentier MC, Cooke R, Gao DY, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O: **Widespread and frequent horizontal transfers of transposable elements in plants**. *Genome Research* 2014, **24**:831-838.

This publication provides evidence for extensive horizontal transfer of TEs in several plant genomes.

30. Lisch D: **Epigenetic Regulation of Transposable Elements in Plants**. In *Annual Review of Plant Biology*. Edited by; 2009:43-66. Annual Review of Plant Biology, vol 60.]

31. Du C, Fefelova N, Caronna J, He L, Dooner HK: **The polychromatic Helitron landscape of the maize genome**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:19916-19921.

32. Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE: **The evolutionary fate of MULE-mediated duplications of host gene fragments in rice**. *Genome Research* 2005, **15**:1292-1297.

33. Barbaglia AM, Klusman KM, Higgins J, Shaw JR, Hannah LC, Lal SK: **Gene Capture by Helitron Transposons Reshuffles the Transcriptome of Maize**. *Genetics* 2012, **190**:965-975.

34. Yang LX, Bennetzen JL: **Distribution, diversity, evolution, and survival of Helitrons in the maize genome**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:19922-19927.

35. Slotkin RK, Freeling M, Lisch D: **Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication**. *Nature Genetics* 2005, **37**:641-644.

36. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, et al.: **A Role for RNAi in the Selective Correction of DNA Methylation Defects**. *Science* 2009, **323**:1600-1604.

37. Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O: **Selective epigenetic control of retrotransposition in Arabidopsis**. *Nature* 2009, **461**:427-U130.

38. Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP: **Repeat associated small RNAs vary among parents and following hybridization in maize**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:10444-10449.

* 39. Bousios A, Diez CM, Takuno S, Bystry V, Darzentas N, Gaut BS: **Palindromic structures in maize Sirevirus LTRs are central to the interplay between transposable element evolution and the host epigenetic response**. *Genome Research* 2015, **revision submitted**.

This study maps small RNAs to individual - i.e., not consensus - LTR retroelements and finds that the complex *cis*-regulatory regions are hotspots of epigenetic interactions and evolutionary change.

40. Bologna NG, Voinnet O: **The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis**. *Annual Review of Plant Biology, Vol 65* 2014, **65**:473-503.

41. Voinnet O: **Use, tolerance and avoidance of amplified RNA silencing by plants**. *Trends in Plant Science* 2008, **13**:317-328.

42. Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA: **miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* 2014, **508**:411-+.

This publication demonstrates that miRNAs can target mRNAs of some TE families both for post-transcriptional suppression and, more importantly, for initiation of transcriptional silencing.

43. Slotkin RK, Freeling M, Lisch D: **Mu killer causes the heritable inactivation of the Mutator family of transposable elements in Zea mays**. *Genetics* 2003, **165**:781-797.

44. Axtell MJ: **Classification and Comparison of Small RNAs from Plants**. *Annual Review of Plant Biology, Vol 64* 2013, **64**:137-159.

45. Coruh C, Shahid S, Axtell MJ: **Seeing the forest for the trees: annotating small RNA producing genes in plants**. *Current Opinion in Plant Biology* 2014, **18**:87-95.

46. Wang X, Elling AA, Li X, Li N, Peng Z, He G, Sun H, Qi Y, Liu XS, Deng XW: **Genome-Wide and Organ-Specific Landscapes of Epigenetic Modifications and Their Relationships to mRNA and Small RNA Transcriptomes in Maize**. *Plant Cell* 2009, **21**:1053-1069.

47. Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, Alioua A, Himber C, Voinnet O: **An endogenous, systemic RNAi pathway in plants**. *Embo Journal* 2010, **29**:1699-1712.

48. Devert A, Fabre N, Floris M, Canard B, Robaglia C, Crete P: **Primer-Dependent and Primer-Independent Initiation of Double Stranded RNA Synthesis by Purified Arabidopsis RNA-Dependent RNA Polymerases RDR2 and RDR6**. *Plos One* 2015, **10**.

49. Bousios A, Darzentas N, Tsaftaris A, Pearce SR: **Highly conserved motifs in non-coding regions of Sirevirus retrotransposons: the key for their pattern of distribution within and across plants?** *Bmc Genomics* 2010, **11**.

50. Grandbastien M-A: **LTR retrotransposons, handy hitchhikers of plant regulation and stress response**. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* 2015, **1849**:403-416.

51. Kumar A, Bennetzen JL: **Plant retrotransposons**. *Annual Review of Genetics* 1999, **33**:479-532.

52. Mergia A, Prattlowe E, Shaw KES, Renshawgegg LW, Luciw PA: **Cis-acting regulatory regions in the long terminal repeat of Simian Foamy virus type-1**. *Journal of Virology* 1992, **66**:251-257.

53. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D: **An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 2014, **516**:242-+.

This publication describes how the evolution of the *cis*-regulatory regions in the *L1* and *Alu* families during primate evolution was eventually met by analogous changes in zinc-fingers host proteins that are thought to bind to these TEs and mediate their silencing.

54. Araujo PG, Casacuberta JM, Costa APP, Hashimoto RY, Grandbastien MA, Van Sluys MA: **Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the Lycopersicon genus**. *Molecular Genetics and Genomics* 2001, **266**:35-41.

55. Fablet M, Rebollo R, Biemont C, Vieira C: **The evolution of retrotransposon regulatory regions and its consequences on the Drosophila melanogaster and Homo sapiens host genomes**. *Gene* 2007, **390**:84-91.

56. Ianc B, Ochis C, Persch R, Popescu O, Damert A: **Hominoid Composite Non-LTR Retrotransposons-Variety, Assembly, Evolution, and Structural Determinants of Mobilization**. *Molecular Biology and Evolution* 2014, **31**:2847-2864.

57. McDonald JF, Matyunina LV, Wilson S, Jordan IK, Bowen NJ, Miller WJ: **LTR retrotransposons and the evolution of eukaryotic enhancers**. *Genetica (Dordrecht)* 1997, **100**:3-13.

58. Beguiristain T, Grandbastien MA, Puigdomenech P, Casacuberta JM: **Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants**. *Plant Physiology* 2001, **127**:212-221.

59. Vernhettes S, Grandbastien MA, Casacuberta JM: **The evolutionary analysis of the Tnt1 retrotransposon in Nicotiana species reveals the high variability of its regulatory sequences**. *Molecular Biology and Evolution* 1998, **15**:827-836.

60. Manetti ME, Rossi M, Costa APP, Clausen AM, Van Sluys M-A: **Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera**. *Bmc Evolutionary Biology* 2007, **7**.

61. Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C: **Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges**. *Plant Cell* 2012, **24**:1242-1255.

62. Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, et al.: **Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice**. *Genome Research* 2006, **16**:1262-1269.

63. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, et al.: **Structural and functional partitioning of bread wheat chromosome 3B**. *Science* 2014, **345**.

64. Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, Barbe V, Mangenot S, Alberti A, Wincker P, et al.: **Organization and evolution of transposable elements along the bread wheat chromosome 3B**. *Genome Biology* 2014, **15**.

65. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, SanMiguel PJ, Bennetzen JL: **Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome**. *Plos Genetics* 2009, **5**.

66. Bousios A, Kourmpetis YAI, Pavlidis P, Minga E, Tsaftaris A, Darzentas N: **The turbulent life of Sirevirus retrotransposons and the evolution of the maize genome: more than ten thousand elements tell the story**. *Plant Journal* 2012, **69**:475-488.

67. Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD, Chen X: **Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. *Genome Research* 2015, **25**:235-245.

This paper identifies the genome-wide locations of *Pol IV*/*RDR2* transcripts and their associated siRNAs.

68. Feng SH, Cokus SJ, Zhang XY, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al.: **Conservation and divergence of methylation**

patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:8689-8694.

69. West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, Schmitz RJ, Springer NM: **Genomic Distribution of H3K9me2 and DNA Methylation in a Maize Genome**. *Plos One* 2014, **9**.

70. Zemach A, McDaniel IE, Silva P, Zilberman D: **Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation**. *Science* 2010, **328**:916-919.

71. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D: **The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* 2013, **153**:193-205.

This study reveals that maintenance of TE silencing in heterochromatic areas is based on methylation factor DDM1 and histone H1 independent of siRNAs, while RdDM is responsible for TE silencing near genic regions.

72. Best A, Hoyle A: **The evolution of costly acquired immune memory**. *Ecology and Evolution* 2013, **3**:2223-2232.

73. Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS: **Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:2322-2327.

74. Faulk C, Barks A, Dolinoy DC: **Phylogenetic and DNA methylation analysis reveal novel regions of variable methylation in the mouse IAP class of transposons**. *Bmc Genomics* 2013, **14**.

75. Reiss D, Zhang Y, Rouhi A, Reuter M, Mager DL: **Variable DNA methylation of transposable elements The case study of mouse Early Transposons**. *Epigenetics* 2010, **5**:68-79.

76. vonHoldt BM, Takuno S, Gaut BS: **Recent Retrotransposon Insertions Are Methylated and Phylogenetically Clustered in Japonica Rice (Oryza sativa spp. japonica)**. *Molecular Biology and Evolution* 2012, **29**:3193-3203.

77. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: **Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila**. *Cell* 2007, **128**:1089-1103.

78. Hill WG, A. R: **EFFECT OF LINKAGE ON LIMITS TO ARTIFICIAL SELECTION**. *Genetical Research* 1966, **8**:269:268.

79. Ma JX, Bennetzen JL: **Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**:383-388.

80. Bousios A, Minga E, Kalitsou N, Pantermali M, Tsaballa A, Darzentas N: **MASiVEdb: the Sirevirus Plant Retrotransposon Database**. *Bmc Genomics* 2012, **13**.

81. Tenaillon MI, Hollister JD, Gaut BS: **A triptych of the evolution of plant transposable elements**. *Trends in Plant Science* 2010, **15**:471-478.

82. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al.: **The Arabidopsis lyrata genome sequence and the basis of rapid genome size change**. *Nature Genetics* 2011, **43**:476-+.

83. Ma L, Hatlen A, Kelly LJ, Becher H, Wang W, Kovarik A, Leitch IJ, Leitch AR: **Angiosperms Are Unique among Land Plant Lineages in the Occurrence of Key Genes in the RNA-Directed DNA Methylation (RdDM) Pathway**. *Genome biology and evolution* 2015, **7**:2648-2662.

84. Diez CM, Meca E, Tenaillon MI, Gaut BS: **Three Groups of Transposable Elements with Contrasting Copy Number Dynamics and Host Responses in the Maize (Zea mays ssp mays) Genome**. *Plos Genetics* 2014, **10**.

85. Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR, et al.: **The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA**. *Genome Research* 2013, **23**:1651-1662.

86. Bousios A, Darzentas N: **Sirevirus LTR retrotransposons: phylogenetic misconceptions in the plant world**. *Mobile DNA* 2013, **4**.

87. Hoen DR, Hickey G, Bourque G, Casacuberta J, Cordaux R, Feschotte C, Fiston-Lavier AS, Hua-Van A, Hubley R, Kapusta A, et al.: **A call for benchmarking transposable element annotation methods**. *Mobile DNA* 2015, **6**.

88. Ragupathy R, You FM, Cloutier S: **Arguments for standardizing transposable element annotation in plant genomes**. *Trends in Plant Science* 2013, **18**:367-376.

Figure in original format
Click here to download high resolution image

Figure 1

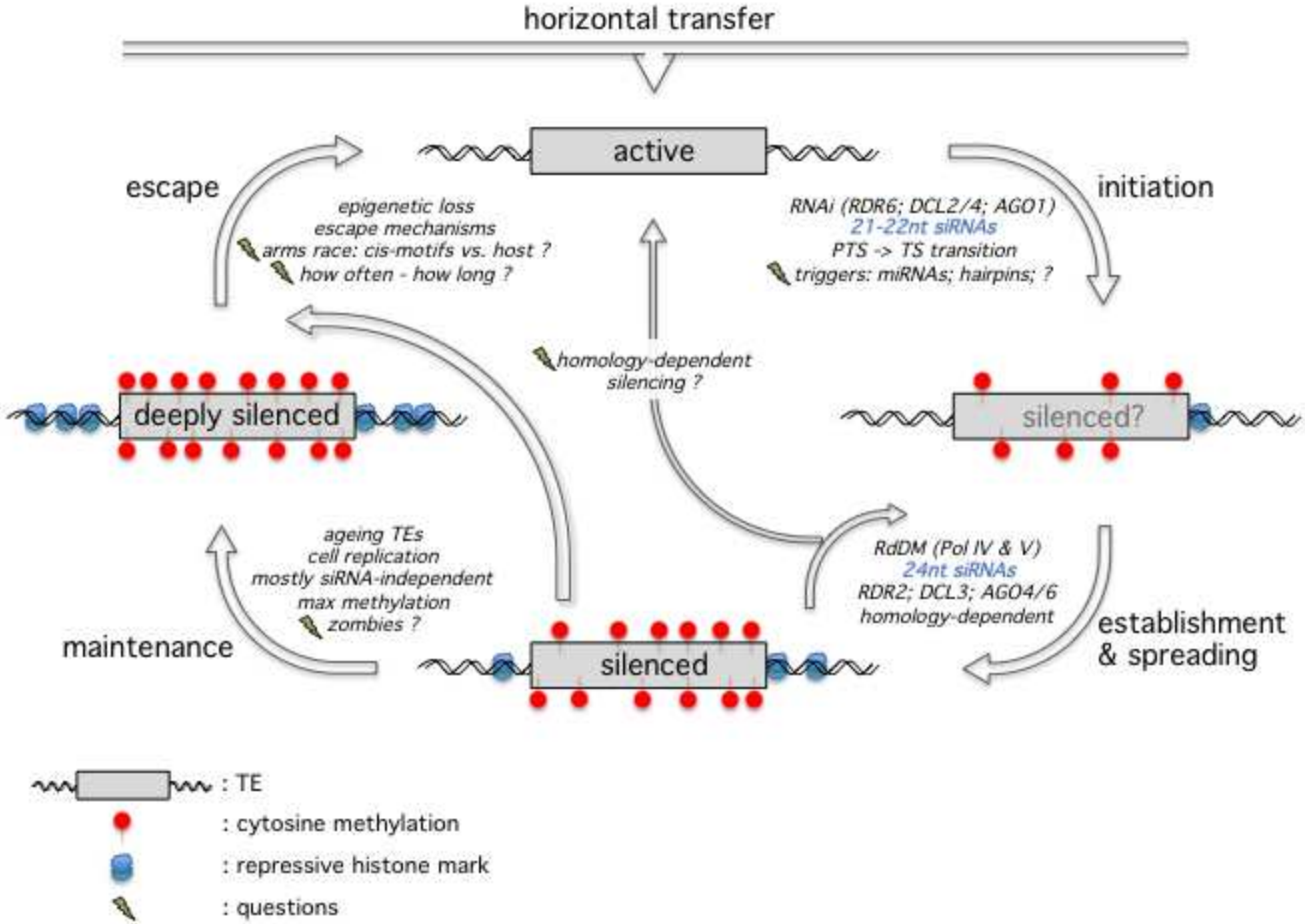

Figure in original format
Click here to download high resolution image

Figure 2