Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Epigenetic regulation and functional exaptation of transposable elements in higher plants

Xiekui Cui and Xiaofeng Cao

Transposable elements (TEs) are mobile genetic elements that can proliferate in their host genomes. Because of their robust amplification, TEs have long been considered 'selfish DNA', harmful insertions that can threaten host genome integrity. The idea of TEs as junk DNA comes from analysis of epigenetic silencing of their mobility in plants and animals. This idea contrasts with McClintock's characterization of TEs as 'controlling elements'. Emerging studies on the regulatory functions of TEs in plant genomes have updated McClintock's characterization, indicating exaptation of TEs for genetic regulation. In this review, we summarize recent progress in TE silencing, particularly in *Arabidopsis* and rice, and show that TEs provide an abundant, natural source of regulation for the host genome.

Addresses
State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Corresponding author: Cao, Xiaofeng (xfcao@genetics.ac.cn)

Current Opinion in Plant Biology 2014, **21**:83–88

This review comes from a themed issue on **Cell signalling and gene regulation**

Edited by **Xiangdong Fu** and **Junko Kyozuka**

For a complete overview see the Issue and the Editorial

Available online 25th July 2014

http://dx.doi.org/10.1016/j.pbi.2014.07.001

Introduction

Transposable elements (TEs) were discovered by Barbara McClintock over half a century ago in maize as mobile genetic entities that can alter gene expression patterns and induce chromosomal rearrangements [1,2]. McClintock referred to TEs as 'controlling elements' and proposed that they may serve two critically important functions: TEs may transpose 'in concert' to regulate gene expression during development, and they may function to restructure the genome under stress conditions [3,4]. TE studies in the following decades focused largely on the characterization of different types of TEs and their transposition mechanisms (for a review, see Ref. [5]). These studies quickly confirmed McClintock's descriptions of TE mobility, but they also showed that transposition generally occurs randomly, and that new TE insertions usually have deleterious effects on the host genome. For these reasons, TEs became viewed as selfish genetic parasites, and their persistence during evolution was simply attributed to their ability to out-replicate the host genome [6,7].

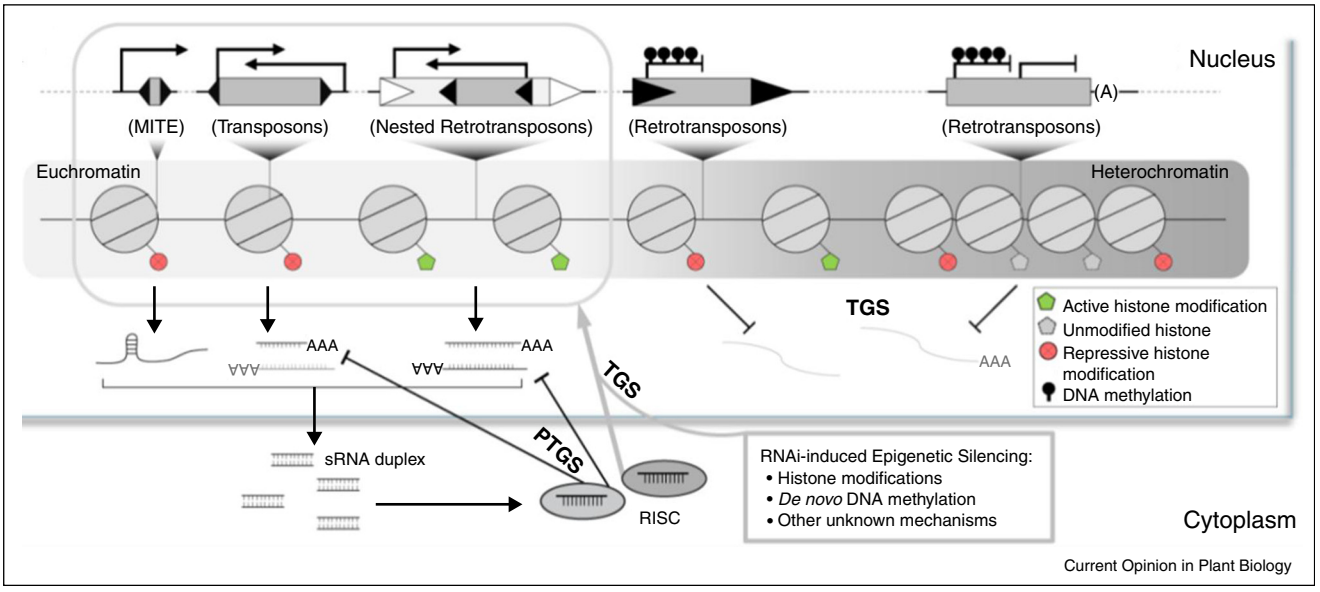
In recent years, questions regarding the effects of TEs on gene expression have reemerged with the rapid advances in plant genomics and epigenetics. Genome sequencing showed that TEs are major components of all plant genomes examined to date, and in many cases TEs constitute the most abundant class of nuclear DNA (e.g. ~85% of the genome in maize) [8,9]. In addition, in larger and more complex genomes (particularly of crop plants), TEs occur throughout the chromosomes and in close proximity to genes [9,10]. Importantly, comparisons among maize inbred lines showed that they largely share the same set of genes, but have drastically different assortments of TEs in the intergenic regions [11]. These results suggest that the differential accumulation of TEs may significantly affect intraspecific gene expression and phenotypic variation.

To minimize the deleterious effects of transposition, plants have evolved several epigenetic pathways that repress TE activities, including RNA interference (RNAi), DNA methylation and histone modifications [12]. These pathways are highly effective: despite the abundance of TEs, the vast majority of them are transpositionally inactive. Interestingly, mechanisms that silence TEs can also affect the expressions of endogenous genes in *cis* and in *trans* [1,2]. Recently, many reviews have examined the biology of TEs, including their structure, replication, function and evolution [1,2,5,8,12]. In this review, we will summarize recent findings on the epigenetic regulation of plant TE silencing and mobilization, and how TEs have been exapted to regulate gene expression.

Epigenetic mechanisms that silence TEs

Most host genomes employ multiple, interacting epigenetic controls to silence expression from TEs and thus prevent TEs from invading or proliferating; these epigenetic mechanisms include the small RNA pathway, DNA methylation and histone methylation/demethylation (Figure 1). In plants, 21/22-nt small-interfering RNAs (siRNAs) participate in post-transcriptional gene silencing (PTGS) and 24-nt siRNAs participate in transcriptional gene silencing (TGS). Produced by an RNase III-like protein (known as Dicer or Dicer-like, DCL), siRNAs

Figure 1



Epigenetic regulation of TEs in the host genome. Chromatin regions are divided into euchromatin and heterochromatin based on their accessibility to transcription. When TEs localize in euchromatic regions, the transcriptional machinery can produce aberrant transcripts. For instance, MITE insertion might cause production of a hairpin RNA similar to a miRNA precursor. Also, antisense transcripts can result from TEs inserted into transcribed portion of a gene in opposite orientation or from retrotransposons in a nested array. These aberrant double-stranded RNAs initiate post-transcriptional gene silencing (PTGS). Dicer-family proteins cleave dsRNAs into small interfering RNAs (siRNAs), which subsequently incorporate into the RISC complex, then cleaves transcripts that are complementary to the siRNA sequence. Alternatively, siRNAs can be loaded into a different RNA-induced transcriptional silencing complex to guide cleavage of nascent transcripts. Because of the attachment to RNA PolII and the DNA strand, these transcripts induced secondary modifications, including H3K9 methylation or possibly cytosine methylation, in the nearby chromatin. These processes induce heterochromatinization of the regions containing TEs (marked by the gray line box). Additionally, a TE (e.g. *Tos17*) located in a euchromatic region containing active genes (associated with H3K4me3, marked by green pentagons) can be silenced by addition of repressive marks like H3K9me2 (marked by red circles). A TE (e.g. *Karma*) located in heterochromatic region can be silenced by active H3K4me3 demethylation (with no H3K4me3, marked by gray pentagons). These TEs are also DNA hypermethylated, which maintains silencing states via a TGS pathway.

are often incorporated into distinct Argonaut-containing RNA-induced silencing complexes (RISC) to trigger degradation of target mRNAs. In higher plants, multiple paralogs of DCL/AGO/RDR genes have diversified to produce distinct small RNAs involving in either PTGS by the DCL4/AGO1/RDR6 pathway, or in TGS by the DCL3/AGO4/RDR2 pathway [13].

The RNA interference pathway can affect DNA methylation, as siRNAs can also recruit DRM2 (DOMAINS REARRANGEDMETHYLTRANSFERASE 2) to preferentially methylate one strand of DNA in the RNA-directed DNA methylation (RdDM) pathway [14,15] (Figure 1). DNA methylation can occur in both symmetric (CG, CHG) and asymmetric (CHH; H = A, T or C) sequence contexts. DNA METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLASE 3 (CMT3), and DRM2 and/or CMT2 methylases maintain distinct DNA methylation patterns in CG, CHG and CHH contexts, respectively [14,16]. Maintenance of the proper methylation status in all sequence contexts also requires the chromatin remodeling factor DDM1 (decrease in DNA methylation 1) [17]. Short TEs are prone to be targeted by CHH methylation, and CG and CHG methylation mainly targets longer TEs [16].

Furthermore, DNA methylation also affects histone modifications. For example, a recent structural analysis resolved the underlying mechanism in which H3K9me2 and CHG methylation act in a synergistic positive-feedback loop to maintain TE silencing [18–20]. In this scenario, the H3K9 methyltransferase Kryptonite (KYP) is recruited to methylated CHG DNA via its SET-ASSOCIATED and RING-ASSOCIATED (SRA) domain [20]. In addition, the bromo adjacent homology (BAH) domain and the chromo domain of CMT3 bind two adjacent H3K9me2 deposited by KYP (or its close homologs), recruiting CMT3 to methylate CHG sites [18,20].

Effective TE silencing is particularly important in germline cells, and plants have evolved several interesting pathways to reinforce this process. For example, in the pollen vegetative nucleus, the decrease of *DDM1* expression leads to a drastic loss of DNA methylation and the transcriptional reactivation of many TEs (e.g. the *Athila*