Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Transposable elements, genome evolution and transgenerational epigenetic variation

Aoi Hosaka[1] and Tetsuji Kakutani[1,2]

Although transposable elements (TEs) have been regarded as genomic parasites, accumulating evidence suggests that they can also have beneficial roles in evolution of diverse biological processes. In this review, we focus on epigenetic control of TEs as sources of selectable phenotypic variation, with an emphasis on their connections to defense responses.

Addresses
[1] Department of Integrated Genetics, National Institute of Genetics, Yata 1111, Shizuoka 411-8540, Japan
[2] Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Corresponding author: Kakutani, Tetsuji (tkak@bs.s.u-tokyo.ac.jp)

Current Opinion in Genetics & Development 2018, **49**:43–48

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Kathleen Burns** and **Damon Lisch**

https://doi.org/10.1016/j.gde.2018.02.012

Introduction

Transposable elements (TEs), mobile genetic elements found in virtually all organisms, were first identified in maize by Barbara McClintock. She described TEs as 'controlling elements' because they can affect activity of nearby genes [1]. In addition, McClintock found that TEs can shift between active and inactive states [2,3]. These changes can be heritably transmitted over multiple cell divisions or even multiple plant generations, although the activity can revert occasionally. These were pioneering observations which formed the foundation of our understanding of epigenetic regulation of TEs, which is heritable but reversible. After decades of research since then, we now know that the epigenetic control of TEs and the propensity of TEs to affect transcription of nearby genes provide powerful materials for gene control and evolution.

TEs can be regarded as genomic parasites. By increasing their copy number, TEs can propagate within genomes and populations even without benefit to the host [4,5]. Nonetheless, as insightfully summarized in recent reviews [6–8], the accumulation of genome and transcriptome information has revealed many examples of TE-derived sequences having been coopted to serve as regulatory sequences. An emerging view is that TEs survive over long periods of time within the genomes by avoiding damage to the host and can even benefit the host. Interactions between TEs and hosts lead to various layers of biological complexity, including epigenetic gene control and a diversity of defenses against other parasites.

In this review, we discuss the contribution of epigenetic control of TEs to selectable phenotypic variation. We also discuss examples of TEs that provide sources for evolution of defense functions, which paradoxically have evolved to regulate parasitic sequences and organisms.

Epigenetic controls of TEs and their evolution to *cis*-acting elements

Epigenetic mechanisms, which can be heritable but reversible, would account for the changes in TE activities found by McClintock. Indeed, molecular genetic analysis in maize has revealed that the changes in TE activities correlate with changes in DNA methylation; these TEs are cytosine methylated when they are inactive and unmethylated when they are active [9,10]. The connection between DNA methylation and TE activity has also been shown using mutants of *Arabidopsis* with reduced DNA methylation; when TEs lose DNA methylation by trans-acting mutations of genes encoding proteins involved in the DNA methylation machinery, many TEs are transcriptionally derepressed and mobilized [11–15]. *Arabidopsis* serves as a good model organism to investigate control of DNA methylation [16–18], as well as control of TEs [19,20]. In TEs of plants, DNA methylation is connected to other epigenetic marks, such as small RNA and histone modifications [16–18].

Among the histone modifications, methylation of lysine 9 of histone H3 (H3K9me) is an epigenetic mark of constitutively silent chromatin in microorganisms, animals and plants [21,22]. In these organisms, TEs can be de-repressed in mutants of H3K9 methyltransferases [21–24]. In addition to H3K9me, H3K4me also seems to affect TE activity. While H3K9me is a mark of constitutively silent chromatin, H3K4me is a mark of active chromatin. A rice protein, JMJ703, demethylates mono-methylated, di-methylated and tri-methylated H3K4, and a mutation of the JMJ703 gene induces transcriptional activation of multiple endogenous TEs, which is associated with accumulation of tri-methylated H3K4 (H3K4me3) [25].

H3K4me3 and H3K4me2 are found in promoters of transcribed sequences, while H3K4me1 is found in internal regions (bodies) of transcribed regions. H3K4me1 is also found in untranscribed regions and enhancers. In the fruit fly, RNAi screening for genes required for piRNA-mediated TE silencing identified the H3K4 demethylase gene LSD1 [26]. While jumonji-domain-containing proteins, such as JMJ703, can demethylate three types of H3K4me, me1/2/3, the LSD1 type of demethylases generally demethylate only di-methylated and mono-methylated lysine residues [27]. Therefore, at least in some of TEs in the fruit fly, TE activity is likely to be affected by H3K4me2 and/or H3K4me1.

In *Arabidopsis* mutants of H3K9 methylases, H3K4me1/2/3 accumulate in TEs that lose H3K9me, suggesting that H3K9me controls H3K4me directly or indirectly [28•]. H3K4me3 and H3K4me2 accumulate mainly in transcriptionally de-repressed TEs, while H3K4me1 is generally accumulated in TEs with reduced H3K9me2, irrespective of transcriptional activation. On the other hand, H3K4me1 accumulation was a prerequisite for transcriptional derepression of TEs. A simple interpretation of these observations is that H3K4me1 functions upstream of transcription as a response to H3K9me loss [28•]. In vertebrates, H3K4me1 is known as an epigenetic mark of enhancers. Interestingly, H3K4me1 accumulates in diverse human TEs when DNA methylation is lost during early development [29]. Similarly, H3K4me1 accumulates in some of mammalian SINEs (short interspersed elements) during their evolution, and these SINEs have been experimentally validated as *bona fide* active enhancers [30]. These observations suggest a general pathway whereby some of de-repressed TEs become sources for enhancers through changes in the epigenetic modifications.

As very nicely summarized in recent reviews, examples are accumulating that sequences derived from TEs constitute a significant portion of sequences controlling gene expression [6–8]. These TE-derived sequences are connected to diverse important phenomena, including chromosome territory formation [31], neural patterning [32], developmental reprogramming [33–35], and innate immunity [36•].

TE as sources for epigenetic variants

Epigenetic modifications and associated gene expression activity can be inherited transgenerationally (Figure 1). Examples of transgenerational epigenetic variation were initially found in maize [2,3,37–39], but examples are accumulating in other species including animals [39–43,44•]. Interestingly, the genes exhibiting transgenerational epigenetic variation often contain a TE insertion within or near them, and cytosine methylation of the inserted TE is associated with the change in the expression of that gene (Figure 1). A common pattern is that when the TE is methylated, the expression of the corresponding gene is normal or similar to expression of the allele without the TE insertion (Table 1); in other words, methylation of the TE masks its effect of the TE insertion on nearby gene expression. Unmethylated TEs can induce repression or overexpression, and in both cases, methylation of the TE masks the effect on the gene expression (Figure 1).

Thus a TE insertion can generate 'hidden' genetic variation without affecting gene expression. Still, such a TE potentially affects gene expression when the silent marks on the TE is lost. Such hidden genetic variation can be evolutionary tool, analogous to the idea of 'capacitor' of evolution [45,46].

Methylation of TEs can also mask the effect of intronic TEs on RNA processing. In the oil palm fruit, the 'mantled' abnormality is found as a somaclonal variant arising from tissue culture. Genome-wide DNA methylation analyses revealed that the abnormality correlates with loss of DNA methylation of a LINE element, called Karma, within an intron of the responsible gene (Figure 1d) [44•]. When Karma is not methylated, the gene terminates prematurely through alternative splicing at Karma. Similarly, in the *Arabidopsis* gene RPP7, a TE inserted within an intron affects the function of RPP7; RPP7 is functional when the intronic TE is enriched in H3K9me2, but when H3K9me2 is lost in this TE, alternative polyadenylation within the TE makes the transcript nonfunctional (Figure 1e) [47]. RPP7 is one of multiple copies of disease resistant genes (R genes) in the *Arabidopsis* genome, each of which encodes a protein recognizing a specific pathogen. Thus, the intronic TE might generate epigenetic variants for the RPP7 function (epigenetic variation in R gene is discussed later again). For the genes with intronic TEs, appropriate processing seems to depend on host factors, EDM2 and IBM2 [47–49] and these effects depend on silent marks in TEs [47,50,51]. Thus, these factors contribute to the masking of the potential deleterious effect of intronic TEs and allow the TEs to generate epigenetic variants of the genes harboring them.

Epigenetic variation as potential sources for robust defense response

Although it is becoming clear that TEs can provide diverse long-term benefits to the host, it is not clear how selection acts on individual new insertions. A powerful approach to understand short-term beneficial effects of specific TE insertions could be examination of polymorphic TE insertions within natural populations [52•,53•]. Examination of polymorphic TE insertions revealed that they are enriched in genes related to environmental responses. Especially interesting are preferential localization of TEs to R genes, which mediate defense responses by specifically responding to pathogens. As is the case for genes involved in animal immune