

REVIEW

Co-evolution of plant LTR-retrotransposons and their host genomes

Meixia Zhao¹, Jianxin Ma^{1,2}✉

¹ Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

² College of Life Sciences, Qingdao Agricultural University, Qingdao 266109, China

✉ Correspondence: maj@purdue.edu

Received May 4, 2013 Accepted May 22, 2013

ABSTRACT

Transposable elements (TEs), particularly, long terminal repeat retrotransposons (LTR-RTs), are the most abundant DNA components in all plant species that have been investigated, and are largely responsible for plant genome size variation. Although plant genomes have experienced periodic proliferation and/or recent burst of LTR-retrotransposons, the majority of LTR-RTs are inactivated by DNA methylation and small RNA-mediated silencing mechanisms, and/or were deleted/truncated by unequal homologous recombination and illegitimate recombination, as suppression mechanisms that counteract genome expansion caused by LTR-RT amplification. LTR-RT DNA is generally enriched in pericentromeric regions of the host genomes, which appears to be the outcomes of preferential insertions of LTR-RTs in these regions and low effectiveness of selection that purges LTR-RT DNA from these regions relative to chromosomal arms. Potential functions of various TEs in their host genomes remain blurry; nevertheless, LTR-RTs have been recognized to play important roles in maintaining chromatin structures and centromere functions and regulation of gene expressions in their host genomes.

KEYWORDS LTR-retrotransposons, recombination, evolution, epigenetic regulation, plants

INTRODUCTION

Transposable elements (TEs) are ubiquitous DNA components in all eukaryotes, and were often considered as “selfish” or “parasitic” DNA sequences because of their continuous reproduction and potential disruption of regular host genes (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Based on their

structural features and the mechanisms by which they transpose, TEs are generally grouped into two major classes, retrotransposons or class I transposons, and DNA transposons or class II (Wicker et al., 2007). Retrotransposons transpose via an RNA intermediate and through the “copy-and-paste” mechanism, while DNA transposons do not involve an RNA intermediate and employ the “cut-and-paste” transposition mechanism (Finnegan, 1989; Wicker et al., 2007). Retrotransposons are further divided into long terminal repeat retrotransposons (LTR-RTs) and non-LTR retrotransposons including long interspersed repetitive elements (LINEs) and short interspersed repetitive elements (SINEs). The former are most abundant in higher plants, while the latter make up the largest fraction of TEs in vertebrates (Grandbastien, 1992; Kumar and Bennetzen, 1999). In the past decade, an increasing number of plant genomes have been completely or nearly completely sequenced, providing unprecedented opportunities for identification and characterization, and structural, functional and evolutionary analyses of LTR-RTs in the plant kingdom. Here, we review the current knowledge gained from recent researches with respect to the abundance and distribution of LTR-RTs, the molecular mechanisms responsible for structural variation of LTR-RTs, genomic features and evolutionary forces shaping the patterns of distribution and structural variation of LTR-RTs, and potential interplays between LTR-RTs and genes.

STRUCTURES, CHARACTERISTICS, CLASSIFICATION, AND ORIGIN OF PLANT LTR-RTs

LTR-RTs are well characterized by their structural features and characteristics (Kumar and Bennetzen 1999). In general, intact LTR-RTs each contain two LTRs, which start with 5'-TG-3', end with 5'-CA-3', and harbor promoter and terminator sequences. Immediately downstream of the 5' LTR and upstream of the 3' LTR are primer binding site (PBS) and polypurine tract (PPT)

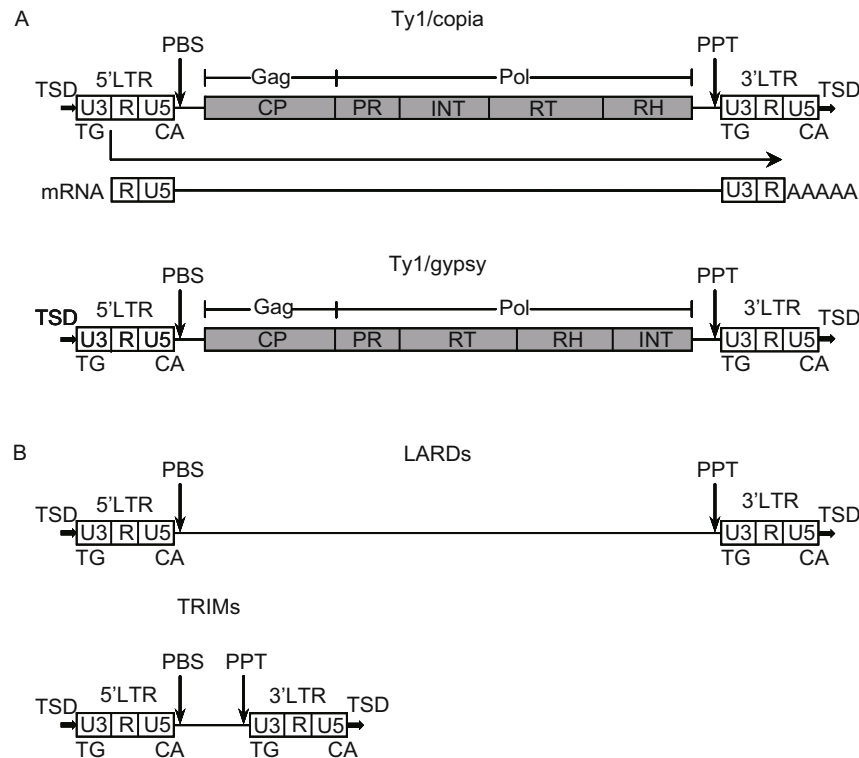


Figure 1. The structures and characteristics of LTR retrotransposons. (A) Autonomous elements (B) Nonautonomous elements. TSD, target site duplication; CP, capsid-like protein; PR, protease; INT, integrase; RT, reverse transcriptase; RH, RNAase-H; LARDs, large retrotransposon derivatives; TRIMs, terminal-repeat retrotransposons in miniature. The initial transcription starts from the R region within 5' LTR, and terminates at the R region of 3' LTR.

of ~20–25 base pairs, which are markedly conserved among widespread species (Fig. 1) (Wicker and Keller, 2007). These motifs, together with the promoter and terminator sequences, are *cis*-regulatory elements essential for the initiation of transcription and replication of an LTR-RT. Between the PBS and PTT sites in typical LTR-RTs are the coding regions that encode proteins, such as GAG, protease (PR), integrase (IN), reverse transcriptase (RT) and RNase H (RH), which are thought to be responsible for retrotransposition and integration of LTR-RTs into their host genomes as retroviruses behave (Fig. 1). Some LTR-RTs in plants (Du et al., 2010a) were found to carry an *env*-like gene (Song et al., 1994; Laten et al., 1998; Wright and Voytas, 1998; Wilhelm and Wilhelm, 2001; Wright and Voytas, 2002; Laten et al., 2003), which, commonly found in retrovirus, encodes a transmembrane coiled-coil protein that sponsors retroviral infection. These elements are often referred as putative plant endogenous retroviruses (Laten et al., 1998; Wright and Voytas, 1998, 2002), although the infection of a plant by a retrovirus has not been evidenced. Additional open reading frames (ORFs) were occasionally observed in some LTR-RTs (Ohtsubo et al., 1999; Steinbauerova et al., 2011), but why they exist and what function they serve, if any, are unknown.

LTR-RTs in plants are generally classified into different

superfamilies, lineages, and families. According to sequence similarity and the order of internal ORFs, LTR-RTs in plants are mainly divided into Ty1/*copia* (INT-RT-RH) and Ty3/*gypsy* (RT-RH-INT) superfamilies (Fig. 1). The former contains LTR-RTs that are generally defined into six evolutionary lineages (*Bianca*, *TAR*, *Angela*, *Ale*, *Ivana* and *Maximus*) (Wicker and Keller, 2007), while the latter contains LTR-RTs generally defined into five evolutionary lineages (*Tekay*, *Reina*, *CRM*, *Athila* and *Tat*) (Wang and Liu, 2008; Du et al., 2010a). These lineages are shared between monocots and eudicots, and thus assumingly to have been formed prior to the split of monocots and eudicots approximately 140–150 million years ago (Chaw et al., 2004). Based on the degrees of sequence similarity, LTR-RTs within individual lineages are generally classified into different families (Wicker et al., 2007). Hundreds or tens of LTR-RT families exist in a plant genome, although only a few proliferated to high copy numbers.

Despite the discovery of the *env*-like gene in some families belonging to two *Gypsy* lineages (*Tat* and *Athila*) and one *Copia* lineage (*Maximus*) in several plant species (Laten et al., 1998; Wright and Voytas, 1998, 2002; Laten et al., 2003; Du et al., 2010a), the origin of this gene remains to be a mystery. Ty3/*gypsy* elements were found to be more closely related to retroviruses than Ty1/*copia* (Wright and Voytas, 1998). A re-

cent study demonstrated that the *env*-like gene identified in the *Copia* lineage was probably derived from the *Gypsy* lineage through inter-element recombination, suggesting that the *env*-like gene in both the *Gypsy* and *Copia* lineages may share a single ancient origin (Du et al., 2010a).

In contrast to LTR-RTs with well-characterized internal ORFs (often defined as autonomous elements) that encodes essential proteins necessary for their retrotransposition activities, LTR-RT families with highly degraded protein coding sequences or without any coding sequences (often defined as nonautonomous elements) have been found in several plants. Different from many truncated LTR-RTs, which seem to have lost capability of amplification, these nonautonomous elements are thought to be able to utilize the retrotransposition machinery provided by their autonomous partners to proliferate, and in some cases, to very high copy numbers (Fig. 1) (Jiang et al., 2002a, b; Kalendar et al., 2004; Kejnovsky et al., 2006; Du et al., 2010b) (Fig. 1).

While no direct evidence for the autonomous and nonautonomous LTR-RT partnerships have been found, indirect evidence derived from sequence comparison suggests the existence of such partnerships, e.g., between autonomous element *RIRE2* and nonautonomous element *Dasheng* in rice (Jiang et al., 2002a, b); between autonomous element *Retand-1* and nonautonomous element *Retand-2* in *Silene latifolia* (Kalendar et al., 2004; Kejnovsky et al., 2006); and between autonomous element *SARE* and nonautonomous element *SNRE* in soybean (Du et al., 2010b). Some nonautonomous elements are large retrotransposon derivatives (LARDs) with several kilo-base pairs of internal sequences between two LTRs of individual elements (Kalendar et al., 2004), while others are terminal-repeat retrotransposons in miniature (TRIMs) with a few hundred base pairs of internal sequences (Fig. 1) (Witte et al., 2001; Yang et al., 2007; Gao et al., 2012). Despite the lack of direct evidence for the origin of non-autonomous elements, it is believed that they were derived from their respective autonomous partners (Jiang et al., 2002b; Kejnovsky et al., 2006; Du et al., 2010b). A recent study suggests that such partnerships can be enhanced and maintained by homogenization of LTR sequences between autonomous and nonautonomous partners by LTR swapping, perhaps through RNA template switch (Sabot and Schulman, 2007), during the reverse transcription process (Du et al., 2010b).

AMPLIFICATION, ELIMINATION, DISTRIBUTION AND ORGANIZATION OF PLANT LTR-RTs

LTR-RTs have been found to be the most abundant DNA components in all plant genomes that have been sequenced (The *Arabidopsis* Genome Initiative, 2000; International Rice Genome Sequencing Project, 2005; Paterson et al., 2009; Schnable et al., 2009; The International Brachypodium Initiative, 2010; Schmutz et al., 2010). However, the scales and timeframes of activity for amplification of LTR-RTs vary dramatically among lineages, species, and families (Du et al., 2010a).

This variation is largely responsible for genome size variation and genomic differentiation among related species. For example, *Arabidopsis* shows constantly low activities for LTR-RT proliferation from ancient to recent times relative to other plant species investigated, which, at least partially, explains why and how *Arabidopsis* maintains such a small genome (The *Arabidopsis* Genome Initiative, 2000; Du et al., 2010a). In contrast to *Arabidopsis*, many other plants have undergone extensive and periodic amplifications of LTR-RTs. In maize, the five largest LTR-RT families account for ~80% of maize retrotransposons, most of which are amplified in the past several million years (SanMiguel et al., 1996, 1998; Schnable et al., 2009). In *Oryza australiensis*, a wild relative of rice (*O. sativa*), bursts of three LTR-RT families within the last three million years have doubled its genome size (Piegu et al. 2006). A recent study revealed a high level of insertion polymorphisms of several LTR-RT families between the cultivated soybean (*Glycine max*) and its wild progenitor species *G. soja*, suggesting very recent amplification activities of these families (Tian et al., 2012). A substantial number of independent amplifications and insertions of LTR-RTs were also observed in two subspecies of rice, *japonica* and *indica*, and in different inbred lines of maize (Wang and Dooner, 2006; Dooner and He, 2008). These observations demonstrated the dynamics and plasticity of the plant genomes mediated by LTR-RT proliferation along evolutionary time.

Mechanisms that counteract plant genome expansions were also described (Bennetzen et al., 2005; Hawkins et al., 2009). One is thought to be unequal homologous recombination (UR), which often occurs between two LTRs of an intact element that removes its internal portion and one recombined LTR to form a solo LTR (Devos et al., 2002; Ma et al., 2004). Although UR can also occur between two linked LTR-RTs to remove their interval and one recombined LTR, or to remove their interval and one recombined element, the frequencies of such recombination were very low (Devos et al., 2002; Ma et al., 2004). Therefore, the ratios of solo LTRs to intact elements are usually used to evaluate and compare the relative rates of UR among different genomes, although the timeframes within which LTR-RTs amplified vary among species. Another mechanism for LTR-RT DNA removal is termed illegitimate recombination (IR), which eliminates TE sequences by generating and accumulating small deletions without requirement of homologous sequences (Devos et al., 2002; Ma et al., 2004; Vite and Bennetzen, 2006).

The extent of UR and IR in removing LTR-RT sequences seems to differ among species (Bennetzen et al., 2005; Vite and Bennetzen 2006). For example, UR was found to be a predominant mechanism, which removed about five times more LTR-RT DNA from the rice genome than IR did (Ma et al., 2004). By contrast, IR was extremely active in *Arabidopsis*, eliminating at least fivefold more LTR-RT DNA from the *Arabidopsis* genome than UR did (Devos et al., 2002). Regardless of relative extent of UR and IR, the processes of LTR-RT DNA removal appear to be remarkably effective and efficient, given the fact that the majority of intact LTR-RTs identified in most

plants were found to be amplified within the past several million years (Paterson et al., 2009; Schnable et al., 2009; The International Brachypodium Initiative, 2010; Du et al., 2010a). Because LTR-RTs were amplified over evolutionary time, the scarceness of aged intact elements in most plants would be interpreted as the outcome of rapid elimination of DNA sequences from aged elements. It was estimated, the average half-life of LTR-RTs in rice is approximately 3–4 million years (Ma et al., 2004; Bennetzen et al., 2005).

Despite the great variation in LTR-RT content and abundance across the plant kingdom, overall LTR-RTs tend to be organized in nested patterns (SanMiguel et al., 1996), and are preferentially accumulated in pericentromeric regions of most species investigated (International Rice Genome Sequencing Project, 2005; Paterson et al., 2009; Schnable et al., 2009; Schmutz et al., 2010). One group of LTR-RTs that show an extreme distribution pattern are centromeric-specific retrotransposons (CRs), such as CRRs in rice (Cheng et al., 2002) and CRMs in maize (Zhong et al., 2002). These elements were found to be preferentially enriched in the core centromeric regions of centromeres and appear to be the major components of functional centromeres (Cheng et al., 2002; Zhong et al., 2002; Nagaki et al., 2004). Nevertheless, LTR-RT families preferentially distributed in gene-rich chromosomal arms, but poorly represented around pericentromeric and/or centromeric regions were also observed (Meyers et al., 2001; Tian et al., 2009; Du et al., 2010a), indicating specific properties of individual families.

FACTORS SHAPING THE GENOMIC LANDSCAPE OF LTR-RTs

The biased accumulation of LTR-RT DNA in pericentromeric regions relative to chromosomal arms in plants was often interpreted as the outcome of preferential insertions of LTR-RTs in the former than in the latter regions (International Rice Genome Sequencing Project, 2005; Paterson et al., 2009; Schnable et al., 2009; Schmutz et al., 2010). However, such a pattern can also be arguably explained solely by selection against LTR-RT insertions in the latter regions than in the former regions. A recent analysis of LTR-RT insertion polymorphisms among 31 re-sequenced wild and cultivated soybean genomes revealed consistent patterns of the non-reference LTR-RT insertions (which were thought to have recently occurred during varietal diversification) and the accumulated LTR-RTs in the soybean reference genome. By contrast, the distribution patterns of the non-reference DNA TE insertions and the accumulated DNA TEs in the reference genome were significantly different, suggesting that the distribution patterns of TEs were shaped by both insertional preferences and purifying selection that purged TE sequences from the host genomes, although the intensities of selection against LTR-RT and DNA TE insertions appear to be different due to their distinct insertional preferences for two contrasting chromatin environments (Du et al., 2012; Tian et al., 2012).

Indeed, the biased accumulation of LTR-RTs in pericentromeric regions of the plant genomes is expected regardless of their insertional preferences. Firstly, pericentromeric regions are generally gene-poor regions, in which LTR-RT insertions would cause less frequent deleterious mutations than those inserted in chromosomal arms. Secondly, genetic recombination in pericentromeric regions are often severely or completely suppressed and thus purifying selection against LTR-RT insertions in these regions would be inefficient or less efficient than in chromosomal arms (Rizzon et al., 2002; Gaut et al., 2007; Tian et al., 2009, 2012). Even the recombination-suppressed pericentromeric regions were excluded, such associations of local genetic recombination (GR) rates with the abundance of LTR-RT DNA and gene densities along chromosomes were still detected in both rice and soybean (Tian et al., 2009; Du et al., 2010a; Tian et al., 2012). Furthermore, the ratios of solo LTRs to intact elements were found to positively correlate with both GR rates and gene densities across the entire genome of rice (Tian et al., 2009). These findings suggest that local genomic features such as genetic recombination and gene densities play important roles in shaping the genomic landscape of LTR-RTs in these two, and probably, many other plant genomes yet to be investigated. On the other hand, the biased accumulation and uneven distribution of LTR-RTs may also reshape some genomic and epigenomic features such as local GR, DNA methylation, and histone modification (Lippman et al., 2004).

EPIGENETIC SILENCING AND REACTIVATION OF LTR-RTs

In addition to UR and IR that have eliminated a large amount of LTR-RT DNA or generated numerous “dead” (non-functional) copies of LTR-RTs, epigenetic mechanisms, such as DNA methylation, histone modification and small RNA interfering, play important roles in LTR-RT silencing (Kasschau et al., 2007; Nobuta et al., 2007; Ziberman et al., 2007). Cytosine methylation is generally associated with histone modifications and heterochromatinization, leading to the suppression of transcription and inactivation of transposition (Lippman et al., 2004). It is particularly intriguing that this epigenetic feature appears to be heritable during DNA replication (Kinoshita et al., 2004; Slotkin et al., 2005; Downen et al., 2012) and thus important for maintenance of LTR-RT silencing.

Genome-wide profiling and comparison of DNA methylation and transcriptome between the wild type *Arabidopsis*, and *ddm1* (decrease in DNA methylation) and *met1* (cytosine DNA methyltransferase) mutants revealed a correlation between cytosine methylation and transcriptional activity of transposons (Vongs et al., 1993; Soppe et al., 2000; Miura et al., 2001; Kankel et al., 2003; Kinoshita et al., 2004, 2007). Despite the lack of direct evidence, it is reasonable to deduce that the increase of the transcriptional levels of LTR-RTs would facilitate the amplification activities of some elements with capacity of retrotransposition.

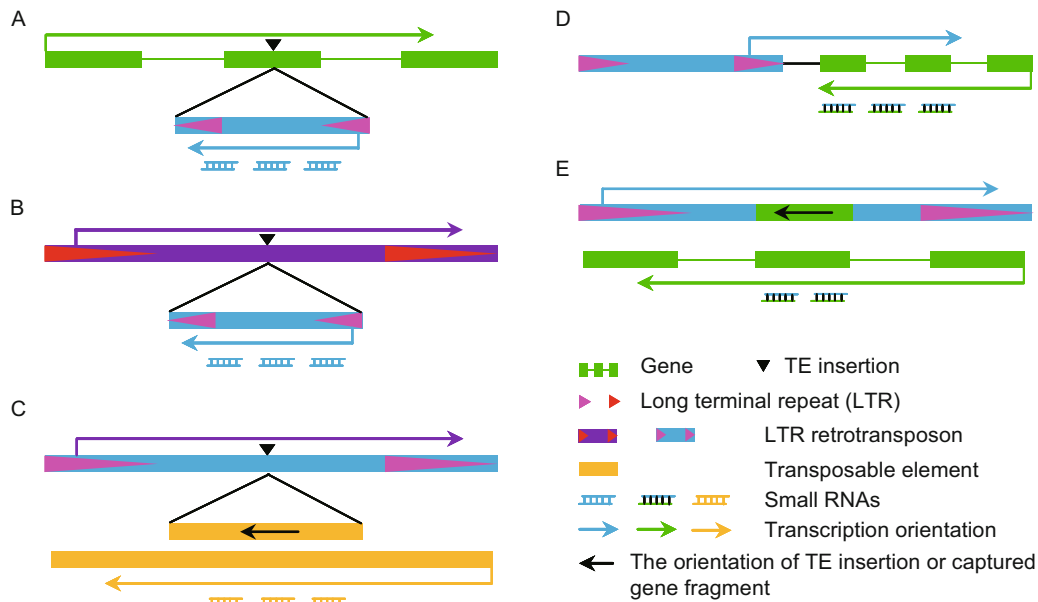


Figure 2. Potential sense-antisense transcripts and small RNAs generated by LTR retrotransposons. (A) An LTR retrotransposon inserts into a gene in reverse orientation. (B) An LTR retrotransposon inserts into another LTR retrotransposon in opposite orientation. (C) A nonautonomous element inserts into an LTR retrotransposon in opposite orientation. Transcript proceeds through this element, producing antisense transcript with its homologous elements outside of the same family. (D) The transcript readout from 3' LTR to the adjacent gene with opposite orientation. (E) An LTR retrotransposon captures gene fragment from reverse orientation. Transcript proceeds through this region, producing antisense transcript with its host gene outside.

Although little is known regarding how the *de novo* DNA methylation is initiated and maintained, more and more studies suggest that RNA-directed DNA methylation (RdDM) is involved in retrotransposon silencing (Cao and Jacobsen, 2002; Hamilton et al., 2002; Huettel et al., 2007; Matzke et al., 2007). It is likely that, during RdDM, double-stranded RNAs (dsRNAs) are processed to produce ~21–24 nt small interfering RNA (siRNAs), which direct *de-novo* methylation of original genomic locations and trigger epigenetic silencing (Aufsatz et al., 2002). Profiling of small RNAs in the plant genomes has demonstrated that LTR-RTs are one of the major sources to generate small RNAs (Lu et al., 2005; Nosaka et al., 2012). As illustrated in Fig. 2, antisense transcripts of LTR-RTs may be produced when they are *cis*-regulated by promoters from their genes or TEs arranged in reverse transcriptional orientations in their flanking regions, and subsequently, dsRNAs and siRNAs can be generated from the sense and antisense transcripts to induce silencing of the LTR-RTs generating the antisense transcripts and other elements sharing sequence similarities in siRNA target sites (Lisch 2009; Lisch and Bennetzen, 2011). Analysis of a tobacco LTR-RT element *Tnt1* transferred into *Arabidopsis*, suggests that silencing of this element was achieved in a copy number-dependent manner (Perez-Hormaeche et al., 2008). However, what a certain number of copies are reached and/or under what certain circumstances silencing would occur remains largely unknown.

While the majority of LTR-RTs are silenced, some elements can be reactivated under specific conditions including tissue

culture, low temperature, wounding, biotic and analogue signal molecular elicitors (Hirochika, 1993, 1996; Takeda et al., 1999; Hashida et al., 2003; Cheng et al., 2006). Such a kind of LTR-RTs include two well characterized elements—the tobacco *Tto1* and the rice *Tos17*, both of which are silenced in normal growth conditions, but can be reactivated by tissue culture (Hirochika, 1993; Hirochika et al., 1996). Further studies indicate that the reactivation of these elements is associated with demethylation of these elements. For example, in the homozygous *ddm1* background, the silent *Tto1* could regain transcriptional and transpositional activation (Hirochika et al., 2000). Functional analysis of LTR sequences indicated that the 13-bp motif is a *cis*-regulatory element involved in the reactivation process (Takeda et al., 1999; Sugimoto et al., 2000). It was reported that polyploidization events can also result in massive genome-wide methylation alternations at early generations of newly synthesized polyploids, which may be responsible for observed reactivation of TEs (Kraitshtein et al., 2010). The association of TE transcription and transposition activities with TE demethylation seems to be unarguable, but the trade-off between them remains unclear.

ALTERNATION AND REGULATION OF GENE EXPRESSION BY LTR-RTs

In general, insertions of LTR-RTs in untranslated regions (UTRs) or coding sequences (CDS) of the host genes can block the expression of these genes or produce premature

products or proteins that distinct from the regular forms. Indeed, most transcriptionally active LTR-RTs, such as *Tnt1* and *Tos17*, were discovered based on their phenotypic changes caused by their insertions into genes and subsequently used for generation of gene knockout mutant resources for functional genomics studies (Hirochika et al., 1996; Hirochika 2001; Miyao et al., 2003; Tadege et al., 2008). In addition to insertions within UTRs and CDS, some LTR-RTs, particularly transcriptionally or transpositionally active ones, have great potential to alter the expression of their flanking genes (Fig. 2D). This is because LTRs contain all requisite signals such as enhancers, promoters for regulation of gene expression. One example is the LTR-RT *Wis2-1A*, which transcribes at a very high level in newly synthesized wheat amphiploids and drives the readout synthesis of new transcripts from adjacent sequences including the antisense or sense strands of known genes (Kashkush et al., 2002, 2003; Kashkush and Khassan, 2007). These readouts could trigger the generation of double-stranded RNAs and eventually small RNAs to regulate corresponding genes (Fig. 2D). Furthermore, as illustrated in Fig. 2E, many plant LTR-RTs harbor genes or gene fragments (Jin and Bennetzen, 1994; Du et al., 2006; Wang et al., 2006). Upon the transcription of these fragments, sense or antisense transcripts can be generated and used to produce small RNAs to regulate corresponding genes (Jiang et al., 2004; Kawasaki and Nitasaka, 2004; Hanada et al., 2009).

Theoretically, LTR insertions within intronic sequences can be removed by splicing, but sometimes, in particular, insertions of large LTR-RT elements may lead to inaccurate recognition of the intron-exon boundaries by spliceosome and thus yield novel or non-sense products (Varagona et al., 1992; Marillonnet and Wessler, 1997). While the functional roles of LTR-RTs are still under debate, several studies have demonstrated that LTR-RT sequences may contribute to plant fitness by providing new regulatory components and properties to genes (White et al., 1994; Naito et al., 2009). Such elements are generally adjacent to the genes and provide *cis*-regulatory functions. For example, an LTR-RT *Hopscotch* in maize was found to act as the enhancer of a gene associated with a domestication trait (White et al., 1994; Bejerano et al., 2006). Nevertheless, a recent study in *Arabidopsis* reveals that heavily methylated TEs may have deleterious effects on their adjacent genes, perhaps by suppressing their expression (Hollister and Gaut, 2009).

CONCLUDING REMARKS AND PERSPECTIVES

LTR-RTs are the major constituents of the plant genomes and largely responsible for genome size variation, genome differentiation, and perhaps speciation. Despite the periodic amplification or recent burst of LTR-RTs as observed in a number of plant genomes, most elements are transcriptionally and transpositionally inactivated by DNA methylation and histone modification, and in all plants examined, a large proportion of LTR-RT DNA has been removed from the host genomes primarily through intra-element homologous recombination and

illegitimate recombination along evolutionary time. In addition to preferential integration, local genetic recombination appears to have played a major role, behind purifying selection, in shaping the removal, accumulation, and distribution of LTR-RT DNA.

While the contribution of LTR-RTs to genomic plasticity has been well understood, their origin, expression, insertional specificity, evolutionary fate, and potential impact on genetic and epigenetic gene regulation remain largely unclear. With the continuing accumulation of plant genome sequencing and resequencing data, it is expected that additional active LTR-RTs will be identified, which will facilitate future studies to address these questions.

ACKNOWLEDGEMENTS

This work was partially supported by Purdue Agricultural Research Program, Shandong Province Taishan Oversea Scholarship, and the United Soybean Board.

ABBREVIATIONS

CDS, coding sequence; CR, centromeric-specific retrotransposon; dsRNA, double-stranded RNA; GR, genetic recombination; IN, integrase; LINE, long interspersed repetitive element; LTR-RT, long terminal repeat retrotransposon; ORF, open reading frame; PBS, primer binding site; PPT, polypurine tract; PR, protease; RdDM, RNA-directed DNA methylation; RH, RNase H; RT, reverse transcriptase; SINE, short interspersed repetitive element; siRNA, small interfering RNA; TE, transposable element; UTR, untranslated region

COMPLIANCE WITH ETHICS GUIDELINES

Meixia Zhao and Jianxin Ma declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by the any of the authors.

REFERENCES

- Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J.M., and Matzke, M. (2002). RNA-directed DNA methylation in *Arabidopsis*. *Proc Natl Acad Sci U S A* 99, 16499–16506.
- Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441, 87–90.
- Bennetzen, J.L., Ma, J., and Devos, K. (2005). Mechanisms of recent genome size variation in flowering plants. *Ann Bot* 95, 127–132.
- Cao, X., and Jacobsen, S.E. (2002). Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* 12, 1138–1144.
- Chaw, S.M., Chang, C.C., Chen, H.L., and Li, W.H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58, 424–441.
- Cheng, C., Daigen, M., and Hirochika, H. (2006). Epigenetic regulation of the rice retrotransposon *Tos17*. *Mol Genet Genomics* 276, 378–390.

- Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., and Jiang, J. (2002). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14, 1691–1704.
- Devos, K.M., Brown, J.K.M., and Bennetzen, J.L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12, 1075–1079.
- Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603.
- Dooner, H.K., and He, L. (2008). Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. *Plant Cell* 20, 249–258.
- Downen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Downen, J.M., Nery, J.R., Dixon, J.E., and Ecker, J.R. (2012). Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* 109, E2183–2191.
- Du, C., Swigonova, Z., and Messing, J. (2006). Retrotranspositions in orthologous regions of closely related grass species. *BMC Evol Biol* 6, 62.
- Du, J., Tian, Z., Hans, C.S., Laten, H.M., Cannon, S.B., Jackson, S.A., Shoemaker, R.C., and Ma, J. (2010a). Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J* 63, 584–598.
- Du, J., Tian, Z., Bowen, N.J., Schmutz, J., Shoemaker, R.C., and Ma, J. (2010b). Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. *Plant Cell* 22, 48–61.
- Du, J., Tian, Z., Sui, Y., Zhao, M., Song, Q., Cannon, S.B., Cregan, P., and Ma, J. (2012). Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. *Plant Cell* 24, 21–32.
- Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet* 5, 103–107.
- Gao, D., Chen, J., Chen, M., Meyers, B.C., and Jackson, S. (2012). A highly conserved, small LTR retrotransposon that preferentially targets genes in grass genomes. *PLoS One* 7, e32010.
- Gaut, B.S., Wright, S.I., Rizzon, C., Dvorak, J., and Anderson, L.K. (2007). Opinion - Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet* 8, 77–84.
- Grandbastien, M.A. (1992). Retroelements in higher-plants. *Trends Genet* 8, 103–108.
- Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. *EMBO J* 21, 4671–4679.
- Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R.K., Lisch, D., Meyers, B.C., Shiu, S.H., and Jiang, N. (2009). The functional role of Pack-MULEs in rice inferred from purifying selection and expression profile. *Plant Cell* 21, 25–38.
- Hashida, S., Kitamura, K., Mikami, T., and Kishima, Y. (2003). Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in *Antirrhinum majus*. *Plant Physiol* 132, 1207–1216.
- Hawkins, J.S., Proulx, S.R., Rapp, R.A., and Wendel, J.F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc Natl Acad Sci U S A* 106, 17811–17816.
- Hirochika, H. (1993). Activation of tobacco retrotransposons during tissue-culture. *EMBO J* 12, 2521–2528.
- Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996). Retrotransposons of rice involved in mutations induced by tissue culture. *Proc Natl Acad Sci U S A* 93, 7783–7788.
- Hirochika, H., Okamoto, H., and Kakutani, T. (2000). Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* 12, 357–368.
- Hirochika, H. (2001). Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 4, 118–122.
- Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 19, 1419–1428.
- Huettel, B., Kanno, T., Daxinger, L., Bucher, E., van der Winden, J., Matzke, A.J., and Matzke, M. (2007). RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim Biophys Acta* 1769, 358–374.
- International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800.
- Jiang, N., Bao, Z., Temnykh, S., Cheng, Z., Jiang, J., Wing, R.A., McCouch, S.R., and Wessler, S.R. (2002a). Dasheng: A recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* 161, 1293–1305.
- Jiang, N., Jordan, I.K., and Wessler, S.R. (2002b). Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. *Plant Physiol* 130, 1697–1705.
- Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., and Wessler, S.R. (2004). Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431, 569–573.
- Jin, Y.K., and Bennetzen, J.L. (1994). Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bsl retroelement of maize. *Plant Cell* 6, 1177–1186.
- Kalendar, R., Vicent, C.M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A., and Schulman, A.H. (2004). Large retrotransposon derivatives: Abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics* 166, 1437–1450.
- Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddelloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003). *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* 163, 1109–1122.
- Kashkush, K., Feldman, M., and Levy, A.A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160, 1651–1659.
- Kashkush, K., Feldman, M., and Levy, A.A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* 33, 102–106.
- Kashkush, K., and Khasdan, V. (2007). Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics* 177, 1975–1985.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol* 5, 479–493.

- Kawasaki, S., and Nitasaka, E. (2004). Characterization of Tpn1 family in the Japanese morning glory: En/Spm-related transposable elements capturing host genes. *Plant Cell Physiol* 45, 933–944.
- Kejnovsky, E., Kubat, Z., Macas, J., Hobza, R., Mracek, J., and Vyskot, B. (2006). Retand: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat. *Mol Genet Genomics* 276, 254–263.
- Kinoshita, T., Miura, A., Choi, Y.H., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* 303, 521–523.
- Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T. (2007). Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J* 49, 38–45.
- Kraitshtein, Z., Yaakov, B., Khasdan, V., and Kashkush, K. (2010). Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics* 186, 801–812.
- Kumar, A., and Bennetzen, J.L. (1999). Plant retrotransposons. *Annu Rev Genet* 33, 479–532.
- Laten, H.M., Majumdar, A., and Gaucher, E.A. (1998). SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc Natl Acad Sci U S A* 95, 6897–6902.
- Laten, H.M., Havecker, E.R., Farmer, L.M. and Voytas, D.F. (2003) SIRE1, an endogenous retrovirus family from Glycine max, is highly homogeneous and evolutionarily young. *Mol Biol Evol* 20, 1222–1230.
- Lippman, Z., and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. *Nature* 431, 364–370.
- Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* 60, 43–66.
- Lisch, D., and Bennetzen, J.L. (2011). Transposable element origins of epigenetic gene regulation. *Curr Opin Plant Biol* 14, 156–161.
- Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J. (2005). Elucidation of the small RNA component of the transcriptome. *Science* 309, 1567–1569.
- Ma, J., Devos, K.M., and Bennetzen, J.L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14, 860–869.
- Marillonnet, S., and Wessler, S.R. (1997). Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing. *Plant Cell* 9, 967–978.
- Matzke, M.A., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A.J.M. (2007). Targets of RNA-directed DNA methylation. *Curr Opin Plant Biol* 10, 512–519.
- Meyers, B.C., Tingey, S.V., and Morgante, M. (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 10, 1660–1676.
- Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature* 411, 212–214.
- Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. (2003). Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* 15, 1771–1780.
- Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36, 138–145.
- Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T., and Wessler, S.R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461, 1130–1134.
- Nobuta, K., Venu, R.C., Lu, C., Belo, A., Vemaraju, K., Kulkarni, K., Wang, W., Pillay, M., Green, P.J., Wang, G.L., et al. (2007). An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol* 25, 473–477.
- Nosaka, M., Itoh, J., Nagato, Y., Ono, A., Ishiwata, A., and Sato, Y. (2012). Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice. *PLoS Genet* 8, e1002953.
- Ohtsubo, H., Kumekawa, N., and Ohtsubo, E. (1999). RIRE2, a novel gypsy-type retrotransposon from rice. *Genes Genet Syst* 74, 83–91.
- Orgel, L.E., and Crick, F.H.C. (1980). Selfish DNA: the ultimate parasite. *Nature* 284, 604–607.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556.
- Perez-Hormaeche, J., Potet, F., Beauclair, L., Le Masson, I., Courtial, B., Bouche, N., and Lucas, H. (2008). Invasion of the Arabidopsis genome by the tobacco retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing. *Plant Physiol* 147, 1264–1278.
- Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A., et al. (2006). Doubling genome size without polyploidization: Dynamics of retrotransposon-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16, 1262–1269.
- Rizzon, C., Marais, G., Gouy, M., and Biemont, C. (2002). Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res* 12, 400–407.
- Sabot, F., and Schulman, A.H. (2007). Template switching can create complex LTR retrotransposon insertions in Triticeae genomes. *BMC Genomics* 8, 247.
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., MelakeBerhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. *Nat Genet* 20, 43–45.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115.
- Slotkin, R.K., Freeling, M., and Lisch, D. (2005). Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat Genet* 37, 641–644.

- Song, S.U., Gerasimova, T., Kurkulos, M., Boeke, J.D., and Corces, V.G. (1994). An Env-like protein encoded by *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Gene Dev* 8, 2026–2057.
- Soppe, W.J.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J.M. (2000). The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* 6, 791–802.
- Steinbauerova, V., Neumann, P., Novak, P., and Macas, J. (2011). A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica* 139, 1543–1555.
- Sugimoto, K., Takeda, S., and Hirochika, H. (2000). MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. *Plant Cell* 12, 2511–2527.
- Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Gabriella, E., Zhao, P.X., Chabaud, M. et al. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant J* 54, 335–347.
- Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1999). A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J* 18, 383–393.
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- The International Brachypodium Initiative. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768.
- Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J.L., Jackson, S.A., Gaut, B.S., and Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res* 19, 2221–2230.
- Tian, Z., Zhao, M., She, M., Du, J., Cannon, S.B., Liu, X., Xu, X., Qi, X., Li, M.W., Lam, H.M., et al. (2012). Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. *Plant Cell* 24, 4422–4436.
- Varagona, M.J., Purugganan, M., and Wessler, S.R. (1992). Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* 4, 811–820.
- Vitte, C., and Bennetzen, J.L. (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc Natl Acad Sci U S A* 103, 17638–17643.
- Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). *Arabidopsis thaliana* DNA methylation mutants. *Science* 260, 1926–1928.
- Wang, H., and Liu, J.S. (2008). LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice. *BMC Genomics* 9, 382.
- Wang, Q., and Dooner, H.K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proc Natl Acad Sci U S A* 103, 17644–17649.
- Wang, W., Zheng, H., Fan, C., Li, J., Shi, J., Cai, Z., Zhang, G., Liu, D., Zhang, J., Vang, S., et al. (2006). High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18, 1791–1802.
- White, S.E., Habera, L.F., and Wessler, S.R. (1994). Retrotransposons in the flanking regions of normal plant genes - a role for copia-like elements in the evolution of gene structure and expression. *Proc Natl Acad Sci U S A* 91, 11792–11796.
- Wicker, T., and Keller, B. (2007). Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Res* 17, 1072–1081.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8, 973–982.
- Wilhelm, M., and Wilhelm, F.X. (2001). Reverse transcription of retroviruses and LTR retrotransposons. *Cell Mol Life Sci* 58, 1246–1262.
- Witte, C.P., Le, Q.H., Bureau, T., and Kumar, A. (2001). Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc Natl Acad Sci U S A* 98, 13778–13783.
- Wright, D.A., and Voytas, D.F. (1998). Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. *Genetics* 149, 703–715.
- Wright, D.A., and Voytas, D.F. (2002). Athila4 of *Arabidopsis* and Calypso of soybean define a lineage of endogenous plant retroviruses. *Genome Res* 12, 122–131.
- Yang, T.J., Kwon, S.J., Choi, B.S., Kim, J.S., Jin, M., Lim, K.B., Park, J.Y., Kim, J.A., Lim, M.H., Kim, H.I., et al. (2007). Characterization of terminal-repeat retrotransposon in miniature (TRIM) in Brassica relatives. *Theor Appl Genet* 114, 627–636.
- Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., and Dawe, R.K. (2002). Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14, 2825–2836.
- Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39, 61–69.