

# The genomes and transposable elements in plants: are they friends or foes?

Nam-Soo Kim<sup>1</sup>

Received: 23 January 2017 / Accepted: 25 January 2017 / Published online: 7 February 2017  
© The Genetics Society of Korea and Springer-Science and Media 2017

**Abstract** Transposable elements (TEs) are mobile genetic elements that are present in prokaryotes and eukaryotes. The ubiquity and abundance of these self-replicating entities, bereft of cellular function, had earned them the label of ‘genomic parasites’. However, the status of TEs has been revised, with ample genomic and biological evidence now portraying them as “genomic gold”. They are perceived as a major participant in the evolution of species. This review addresses the classification of TEs as well as their role and significance in the evolution of genomes, genetic diversity, gene regulation, and exaptation of contemporary species of the plant and animal kingdoms.

**Keywords** Transposable elements · Genome evolution · Gene regulation · Genetic diversity · TE exaptation

## Introduction

McClintock’s cytogenetic and genetic experiments in the 1940s with maize, *Zea mays*, enabled her to identify a line that was prone to breakage and instability of chromosome 9 at a specific locus, *Ds* (*Dissociator*). She attributed instability of the *Ds* locus in the presence of *Ac* (*Activator*) to genetic elements (*Ds* and *Ac*) capable of transposing their chromosomal locations (McClintock 1948). However, the genetic community was skeptical of McClintock’s concept of ‘moveable’ genes; contemporary geneticists were steadfast in their acceptance of the ‘Chromosome Hypothesis of

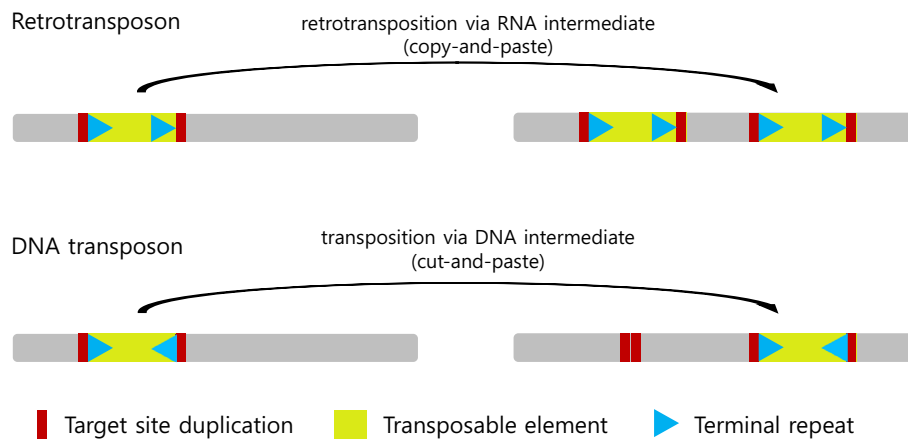
Heredity’, and Morgan’s legacy of genetic entities (genes) on chromosomes. Geneticists gradually became more accepting of mobile genetic elements after the discovery of similar elements in bacteria and other diverse eukaryotes. Initially transposable elements were labeled as ‘genomic parasites’ or ‘selfish DNAs’ because they were thought to propagate without obvious cellular functions in genomes (Doolittle and Sapienza 1980; Orgel and Crick 1980). Numerous findings have ascribed cellular functions to TEs (Volf 2006; Fedoroff 2012; Alzohairy et al. 2013) that have necessitated a revision of their moniker to ‘genomic gold’. However, the genomic role of TEs remains unclear, and is still being debated (Hua-Van et al. 2011; Abrusán et al. 2013; Brunet and Doolittle 2015). It is widely believed that TEs have been instrumental in evolution, and in shaping the genomes of many contemporary species of flora and fauna (Oliver and Greene 2012; Fedoroff 2012; Zhao et al. 2015). This review summarizes the system of classification of TEs, and evaluates their role as a major player in the evolution of plant genomes.

## Classifications of TEs

TEs can be grouped into two classes, based upon their manner of transposition: class 1 elements or retrotransposons move via a “copy-and-paste” manner while class 2 elements move via a ‘cut-and-paste’ manner (Finnegan 1989) (Fig. 1). They can also be classified as autonomous or non-autonomous elements, depending on their mobility. While some TEs can move by themselves because they are equipped with requisite molecular features for transposition, there are some that cannot move unless they are provided with proteins for their mobility *in trans*. Non-autonomous elements usually are derived from autonomous

✉ Nam-Soo Kim  
kimnamsu@kangwon.ac.kr

<sup>1</sup> Department of Molecular Bioscience, Kangwon National University, Chuncheon 200-701, South Korea



**Fig. 1** Class 1 retrotransposon (*top*) and class 2 DNA transposon (*bottom*). Class 1 retrotransposons transpose semiconservative copy-and-paste mechanism via RNA intermediate. Class 2 DNA transposon transpose cut-and-paste mechanism via DNA intermediate. For illus-

tration purpose, the retrotransposon inserts into the nearby chromosomal location for illustration purpose. However, the retrotransposons actually transpose more into different chromosomal sites. Whereas, the class 2 DNA transposons transpose into nearby chromosomal site

elements through mutations. Genome sequencing projects have revealed many thousands to millions of copies of fossilized transposable elements in diverse eukaryotic genomes (Kapitonov and Jurka 2003; Moutri et al. 2007; Smith et al. 2012).

A system of classification with appropriate annotations is necessary for the vast array of TEs from the newly emerging genome sequences in various species so that they can be classified into orders, super-families, and families on the basis of manner of transposition and sequence identities (Wicker et al. 2007). Class 1 retrotransposons have been classified into 5 orders: long terminal repeat elements (LTRs), *Dictyostelium* intermediate repeat sequence (DIRS), *Penelope*-like elements (PLEs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs). DIRS, PLEs, LINEs, and SINEs are referred to as non-LTR retrotransposons because they lack the LTR sequences. While non-LTR retrotransposons are highly abundant in animal genomes, LTR retrotransposons are predominant in plant genomes (Wicker et al. 2007). LTR-retrotransposons carry two genes, *gag* and *pol* (Voytas and Boeke 2002). The *gag* gene encodes capsid proteins that are responsible for packaging the retrotransposon RNA and proteins. The *pol* gene encodes three ORFs of reverse transcriptase (RT), RNaseH (RH), and integrase (INT), which are responsible for the retrotransposition of the element into new chromosomal locations. Depending on the order of the ORFs in the *pol* gene, the LTR-retrotransposons are further classified into *Gypsy* superfamily and *Copia* superfamily (Kumar and Bennetzen 1999). The order of the ORFs is RT-RH-INT in *Gypsy* and INT-RT-RH in *Copia* retrotransposons, respectively. Transcription starts from the LTR sequences on the left and proceeds until the right LTR; the transcript is exported from the nucleus to

cytoplasm where the RNAs serve as a template for translation for protein synthesis as well as cDNA synthesis. The first step is translation of the RNA to produce GAG and POL proteins. LTR-retrotransposon encoded aminopeptidase (AP) clips the POL protein into three components - RT, RNaseH, and INT. The GAG protein produces virus-like particles (VLPs) into which RNAs are packed together with the RT, RNaseH, and INT proteins. The RNAs are then copied into cDNA, starting from one LTR to another, and subsequently into double stranded DNAs within the VLPs. The double stranded DNAs re-enter the nucleus and then integrate into new chromosomal locations (Voytas and Boeke 2002; Levin 2002). Upon retrotransposition, LTR-retrotransposons create a 4–6 bp target site duplication (TSD). Non-LTR retrotransposons are mostly present in metazoans but are scarce in plants. These non-LTR retrotransposons lack both LTR and integrase; instead, they use the poly-A tail of the mRNA as primer to start DNA strand synthesis directly at the at the point of integration (Schulman and Wicker 2013). DIRS retrotransposons do not produce TSDs; PLEs, LINEs and SINEs produce TSDs of variable size on the insertion site (Wicker et al. 2007).

Class 2 TEs are often called DNA transposons, and can be differentiated into two subclasses: subclass 1 terminal invert repeat (TIRs) transposons and subclass 2 non-TIR transposons (Wicker et al. 2007). TIR transposons have transposase (TPase) and a variable size TIR at both ends. The TPase functions both as a DNA binding protein and an endonuclease (Reznikoff 2003). TPase in DNA transposons and INT in LTR-retrotransposons share the motif of aspartic acid (D)-aspartic acid (D)-glutamate (E) in the active site, implying that the divergence of DNA transposons and LTR-retrotransposons has been a very old evolutionary event that predates the divergence of prokaryotes

and eukaryotes because they are present in both prokaryotes and eukaryotes (Hickman et al. 2010; Schulman and Wicker 2013). Subclass 2 TIR transposons comprise 9 super-families. The size of TIRs is variable in each super-family, and all of them produce a few bases long TSD upon integration. The *Crypton* super-family also belongs to the subclass 1 transposon; but *Crypton* transposons lack TIRs and TPase (Goodwin et al. 2003; Kojima and Jurka 2011). *Crypton* transposons have 4- or 6-bp direct repeats, and move as circular DNA produced by recombination between direct repeats and catalyzed by tyrosine recombinase (YR) that is encoded by the *Crypton* transposons. The circular DNAs, then, reinsert into the targeting site by recombination where no TSD is produced (Goodwin et al. 2003).

Two transposons have been identified in subclass 2 - *Helitrons* (Du et al. 2009) and *Maverick* (Feschotte and Pritham 2005; Kapitonov and Jurka 2006). These two transposons are markedly different in structure and mode of replication from subclass 1 DNA transposons. *Helitrons* are widespread in plants, metazoans, and fungi. They were the first transposons to be discovered by computational analysis of whole genome sequences. *Helitrons* transpose by a rolling circle replication mechanism via a single-stranded DNA intermediate, and do not produce a TSD (Li and Dooner 2009; and Pritham 2014). *Helitrons* appeared frequently capturing gene fragments and moving in the genome, which has impacted greatly in genetic diversity in maize (Lai et al. 2005; Morgante et al. 2005; Yang and Bennetzen 2009). This will be discussed in detail in the session of the genetic diversity below. *Maverick* (also called *Politron*) transposons are unique in that a set of proteins is necessary for transposition, including DNA polymerase, retroviral integrase, cysteine protease, and ATPase. *Maverick* transposons have several hundred nucleotides long TIRs with 5'-AG and TC-3' termini, and produce a 6 bp TSD upon transposition (Kapitonov and Jurka 2006). They are excised from and reinserted into a genome by its encoded integrase.

## Genome size variations and TEs

C-value is a quantitative metric that is used to define genome size. It denotes the DNA content of a complete complement of chromosomes in an organism; thus, 1 C represents the amount of DNA in an unreplicated monoploid genome (Greilhuber et al. 2005). The parameter is species specific, and has been subject to strong selection during evolution (Lee and Kim 2014; Canapa et al. 2016). C-values generally correlate well with cellular and organismal genome complexities; the genome sizes of prokaryotes are smaller than those of eukaryotes while lower eukaryotes have smaller genomes than higher eukaryotes. Often

however, high genome size variations are present among closely related species and C-values correlate poorly with organismal complexity, suggesting a C-value paradox (Thomas 1971; Gregory 2001, 2002).

The C-value is often stated as picogram (pg) or base pairs (bp), where 1 pg represents 980 Mbp. In plant taxa, there is >2000-fold difference in the C-values of *Geng-lisea margaretae* (1C=63.4 Mbp) and *Paris japonica* (1C=148,880 Mbp), the smallest and largest members, respectively. In the genus *Elocharis*, comprising more than 250 species, the 20-fold difference in C-values between the smallest, *E. acicularis* (2n=20, 2C=0.5 pg) and the largest, *E. palustris* (2n=16, 2C=11.05 pg), is the highest genome size variation among flowering plant species analyzed for both ploidy and C-values (Zedek et al. 2010). C-values have been measured in >6000 plant species and are categorized in the Kew Plant DNA C-value Database (<http://data.kew.org/cvlaues>) (Bennett and Leitch 2011, 2012). Of the 6287 angiosperms (flowering plants) and 204 gymnosperms, the average 1 C value was 5.809 Gbp in angiosperms, and 18.157 Gbp in gymnosperms, respectively. The median values were 2.401 Gbp in angiosperms and 17.506 Gbp in gymnosperms; ~8.38% of the measured species have a C-value between 0.4 and 0.6 Gbp (Civán et al. 2011). Large genomes may have been derived from smaller genomes through one or more mechanisms of genome amplification. However, the rate of DNA loss is greater in small genomes than the rate of DNA gain by TE amplification, as evidenced in species of cotton and its relatives (Hawkins et al. 2006, 2009).

Whole genome duplication (WGD) or polyploidy is a plausible mechanism that can account for variations in genome size. The current eukaryotes, specifically angiosperms, have had multiple episodic WGDs (Bowers et al. 2003; Kellis et al. 2004; Smith et al. 2013); the small genome of *Arabidopsis thaliana* (130 Mbp) has experienced three WGDs since its divergence from gymnosperms (Bowers et al. 2003). Almost all flowering plants are paleopolyploids or polyploids (Wendel et al. 2016). Although WGD (polyploidization) can alter gene content and increase genome size in a single generation, it is not the only means for genomic variations. For example, the remarkable variations in genome size in *Elocharis* species have been attributed to the rapid amplification of copy numbers of a few families of transposable elements in these species (Vicent et al. 1999; Bennetzen et al. 2005; Hawkins et al. 2008; Civán et al. 2011). Table 1 shows the genome size and TE content of a few representative angiosperm species. Many reports are available on the linear correlation between genome size and content of transposable elements in the genome (Bennetzen et al. 2005; Vitte and Panaud 2005; Tenailon et al. 2010). Class I retrotransposons appear to have had the most impact on genome size

**Table 1** Genome sizes and TE contents in a few plant species

Species	Chromosome number (n)	Genome size (Mbp)	Total TE (%)	Class 1 TE (%)	Class 2 TE (%)	Unknown (%)
<b>Dicots</b>						
<i>Arabidopsis thaliana</i>	5	125	18.5	7.5	11	
<i>Fragaria vesca</i>	7	240	20.7	14.7	5.2	0.9
<i>Medicago trunculata</i>	8	375	18.3	16.9	1.4	
<i>Vitis vinifera</i>	19	487	21.5	19.4	1.4	0.7
<i>Musa acuminata</i>	11	523	43.7	42.4	1.3	
<i>Malus x domestica</i>	17	742	42.4	37.6	0.9	3.9
<i>Solanum tuberosum</i>	12	844	54.4	53.2	1.2	
<i>Solanum lycopersicum</i>	12	900	54.4	53.2	1.2	
<i>Glycine max</i>	20	1,115	58.7	42.2	16.5	
<b>Mono-cots</b>						
<i>Brachypodium distachyon</i>	5	272	28.1	23.3	4.8	
<i>Oryza sativa</i>	12	389	39.5	25.8	13.7	
<i>Setaria italica</i>	9	423	46.5	31.6	9.4	5.4
<i>Sorghum bicolor</i>	10	730	62	54.5	7.5	
<i>Zea mays</i>	10	2,300	84.2	75.6	8.6	
<i>Hordeum vulgare</i>	7	5,100	58.4	52.7	5.0	0.7
<i>Triticum aestivum</i>	21	17,000	79.8	63.7	14.9	1.2

The table was modified from the data of Oliver et al. (2013)

variations because of their semiconservative “copy-and-paste” manner of transposition. TE content in the small genome of *A. thaliana* is about 15%, of which only 4% are retrotransposons (Kaul et al. 2000). Large genome species on the other hand have a high content of TEs, of which a large proportion are class 1 TEs. For example, ~85% of the *Zea mays* genome contains TEs (1C=2.3 pg), of which 75% are class 1 L-retrotransposons and 8.6% are class 2 TEs (Schnable et al. 2009). Similarly, TEs occupy about 69% of the *Secale cereale* genome (1C=8.093 pg), in which class 1 L-retrotransposons and class 2 TEs are 64.3 and 5%, respectively (Bartoš et al. 2008). If amplification of LTR retrotransposons is the more plausible mechanism for ‘genome obesity’ (Hawkins et al. 2008; Civán et al. 2011), why are there large discrepancies in genome sizes in closely related species of a few angiosperms? Is a large genome adaptive to the host organism? The answer is no, because the large genome may impose problems in cell biochemistry and physiology for replication and metabolism as suggested by the ‘large genome constraint’ (Knight et al. 2005). In an analysis of over 6000 plant species, except for a few species with extremely large genomes, genome sizes of angiosperms were found to be skewed toward a smaller size (Civán et al. 2011). This implies that ‘genomic obesity’ through repeated sequences and TEs may have had a negative impact on the survival of lower, simple unicellular organisms. Genome size expansion is counterbalanced by illegitimate recombination between adjacent

LTR-retrotransposons (Devos et al. 2002; Hawkins et al. 2009). The illegitimate recombination between LTRs of different LTR-retrotransposons leads to the sequence deletion between LTRs and forms solo-LTR retrotransposon. That is, the abundance of solo-LTR retrotransposon might indicate the genome size reduction by illegitimate recombination between LTR-retrotransposon. Indeed, direct comparison between *Arabidopsis* (130 Mbp) and rice (430 Mbp) genomes revealed that the solo-LTR retrotransposons were more frequent in the small genome of *Arabidopsis* than that of rice (Bennetzen et al. 2005).

Polyploidy or WGD, and TEs might not act exclusively on genome size evolution; instead, they interact with each other. Hybridization can induce reactivation of the silent TEs in the parental genomes by genome shock (McClintock 1984; Levy 2013). During the episodic and cyclic polyploidization process, massive genome restructuring events occur, placing cells in a different environment from their diploid progenitors (Ågren and Wright 2011; Wendel et al. 2016). Since hyperactivities of TEs reduce host fitness, they are suppressed by host epigenetic silencing mechanisms such as methylation (Slotkin and Martienssen 2007). However, the new cellular environment in hybrids can erase methylation in TE sequences to reactivate the silenced TEs (Levy 2013). Genome shock (i.e., hybridization, WGD) is an internal stress to cells that induces TE activation; external stress (i.e., heat, cold, drought, pathogen) can also induce TE activation (Chadha and Sharma

2014; Makarevitch et al. 2015). Sunflower species in the genus *Helianthus* provide an example of retrotransposon proliferation from hybridization, and subsequent adaptation in a harsh environment (Ungerer et al. 2006). Three hybrid taxa, *H. deserticola*, *H. anomalus* and *H. paradoxus* are the products of ancient hybridization between diploid parental taxa, *H. annuus* and *H. petiolaris* (Rieseberg 1997). The hybrids have adapted to either a desert environment (*H. deserticola*, *H. anomalus*) or a salty marshland (*H. paradoxus*). Interestingly, the hybrids have nuclear genomes that are at least 50% larger than that of either parental species. The genome size differences in the hybrids were attributed to proliferation of the *Gypsy* retrotransposon that was caused by interspecific hybrid and abiotic stress acted (Ungerer et al. 2006; Staton et al. 2012). Here, not all TEs are activated; instead, a small number of TE families are activated to increase copy numbers, as evidenced by only a few lineages of TEs that are abundant in large genomes such as barley (Rostoks et al. 2002) and maize (Schnable et al. 2009). Estep et al. (2013) analyzed LTR retrotransposons in five panicoid grass genomes in which hyperactivity of a small family of LTR retrotransposons resulted in doubling the genome size of *Zea luxurians*, alleged to have happened in the last few million years. Moreover, in the genus *Zea*, one of the LTR retrotransposon amplification bursts was initiated by polyploidy, with most of the other TEs not being activated. If LTR retrotransposon bursts are responsible solely for genome expansion, the larger genomes should be teeming with younger LTR retrotransposons compared to the smaller genomes. Bennetzen et al. (2005) tested for this by analyzing the ages of LTR retrotransposons in six plant species; they found that the average age of LTR retrotransposon in rice was 2.5 million years old while that of barley was 2.8 million years old, suggesting that the age of LTR retrotransposons seems not to be related to genome size. The barley genome (4800 Mbp) is ten times larger than that of the rice genome (430 Mbp), implying that large genomes are due to amplification bursts of only a small number of families of LTR retrotransposons. This is consistent with the “Genome-Thrust” theory which states that TE bursts can cause extraordinary genetic changes in a short periods of time in a small population to drive genetic drift, and eventually lead to genome evolution (Oliver and Greene 2009, 2012; Oliver et al. 2013).

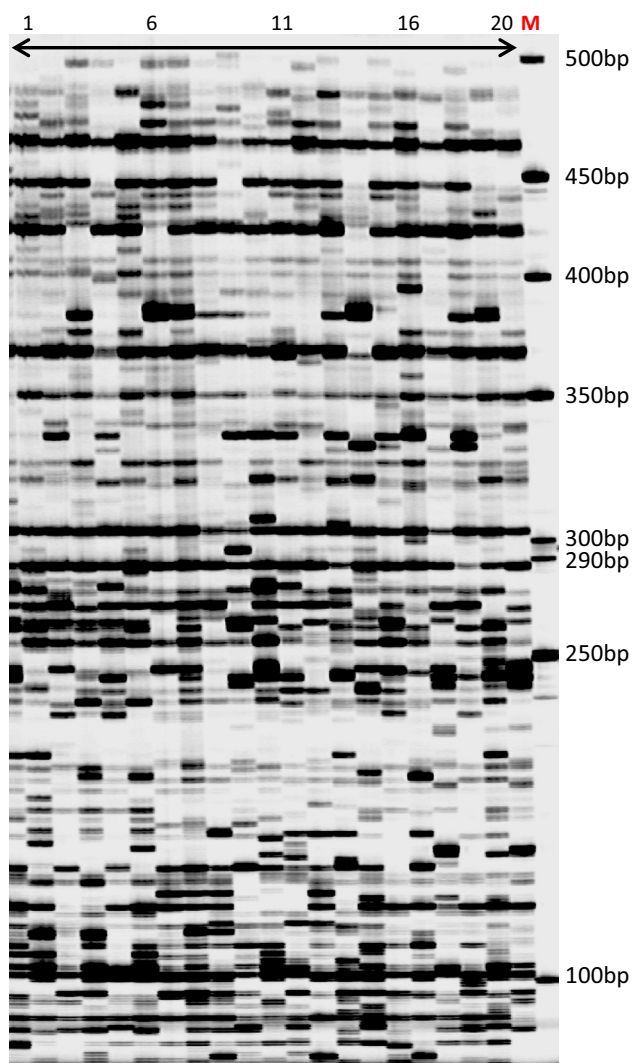
### Genetic diversity through TEs

Throughout evolution, waves of expansion or contraction of TE copies in the genome have resulted in the genomic variations that characterize eukaryotes. Genetic variations induced by TEs occur at both the genic and chromosomal levels. Genic level variation occurs either when TE

insertion disrupts coding function of a gene or when TE excision restores its coding function. A classic example of genic mutation through TE insertion is the wrinkled seed trait in Mendel’s peas. An *Ac/DS* family TE insertion into the starch synthesis gene resulted in an insufficient amount of starch in the developing seed to manifest as wrinkled, upon drying (Bhattacharyya et al. 1990). Another well-known example of TE insertion is the *hopscotch* TE insertion into the enhancer of *teosinte branched1* (*tb1*) to differentiate domesticated corn from its ancestral species teosinte (Studer et al. 2011). TEs cause chromosomal level variation by ectopic pairing between related TE families, which leads to chromosome rearrangements such as deletions, translocations, and duplications (Lönnig and Saedler 2002; Zhang et al. 2011). Chromosomal level disturbances affect many genes, often resulting in casualties that are fatal to the organism. TE-insertions into functional genes, however, undermine host adaptability so that the host organisms are sieved out from a population. Excluding a few families of TEs, such as *Helitrons* in maize as explained below, TEs in contemporary genomes are confined to the evolutionary neutral or ‘safe-haven’ chromosomal regions in a genome. Through their analysis of >500 eukaryotic species, Serra et al. (2013) postulated that an essentially neutral process governs the evolution of the abundance and diversity of TEs in all analyzed genomes.

Since the pioneer work of McClintock, TEs have been thoroughly characterized in maize. Considering the relatively short evolutionary timespan of modern maize lines, their diversity is extraordinary (Fig. 2). TE insertions and recombination in the common gene space have created high genetic variations among the maize lines. Wang and Dooner (2006) reported very high diversity of TE-insertions in the *bz* locus among 8 inbred maize lines that have been used in the breeding program of the USA. In addition to other TE insertions, two of their 8 inbred lines carried *Helitrons* in different sites, accounting for haplotype diversity in the *bz* locus (Wang and Dooner 2006). Of the various TEs, the rolling circle manner transposing *Helitron* best explains genetic diversity in maize (Morgante et al. 2005; Yang and Bennetzen 2009). *Helitrons* often capture different host gene fragments while they transpose, leading to intraspecific variations. Chimeric gene fragments by capturing different gene fragments were found in multiple chromosomal locations that are within inside of the individual *Helitron* elements (Lal et al. 2003; Gupta et al. 2005). Morgante et al. (2005) reported that as many as 10,000 genic content polymorphisms have occurred through *Helitron* insertions in the maize genome. The numbers were even higher in the report by Yang and Bennetzen (2009), who reported a total of 1,930 *Helitrons* from 8 families and >20,000 fragments which can account for approximately 2.2% of the maize genome (B73). The prominent *Helitron*





**Fig. 2** Genetic diversity in 20 different Korean maize landraces. The polymorphisms are detected by retrotransposon based SSAP (sequence specific amplified polymorphisms) technique

burst of amplification activity is dated as having occurred approximately 250,000 years ago, and elements in one of the eight families are still active in transposition.

Genetic diversity driven by TE burst can confer adaptability on the host in coping with stresses from a labile environment. This is more pronounced in the plant kingdom where, in contrast to the early separation of germ cells from soma cells in animals, the boundaries between germ cells and somatic cells are unclear. Variations from TE activities in plant somatic cells can therefore be inherited by the progenies, thereby conferring greater genetic diversity amongst populations. If the variation accords host adaptability, progenies with the variations may drift from the main population, form a subpopulation which has a bottleneck effect due to their size, and subsequently diverge from the population to give rise to subspecies (Belyayev

et al. 2010; Jurka et al. 2011; Levy 2013). This view is consistent with the hypothesis of punctuated evolution which attributes speciation to sudden events rather than a gradual accumulation of small mutations (Gould and Eldredge 1977). *Aegilops speltoides* is a diploid, cross-pollinated wild grass species, indigenous to the middle East. Belyayev et al. (2010) demonstrated that TE proliferations promote or intensify karyological and morphological changes in marginal populations, some of which may give potentially importance for the process of microevolution, which in turn all species with plastic genomes to survive new forms or species in this plant species during the period of rapid climatic change.

An experimental test of TE-mediated speciation and genomic plasticity is difficult; Jurka et al. (2011) have proposed a theoretical approach to assessing TE-mediated speciation by testing the potential impact of merging TE families on genetic diversification. Nevertheless, it is an interesting proposal that genome plasticity through TE diversity may have contributed to taxonation (Oliver and Greene 2009; Oliver et al. 2013). Apropos illustration for this can be found in the comparison of spermatophytes of angiosperms and gymnosperms. The number of extant angiosperm species is estimated as >350,000 (Soltis et al. 2008), making them the second largest group after the insects. The woody gymnosperms on the other hand, comprise less <1000 species, despite preceding the angiosperms evolutionarily (Megallon and Sanderson 2005). Both TE composition and diversity are different in the genomes of these two groups of eukaryotes (Kovach et al. 2010; Nystedt et al. 2013). While the genomes of gymnosperms have been hosts to a diverse set of TEs with low TE activity, those of angiosperms have accommodated TEs with much more activity. The TEs in gymnosperms are ancient, whereas in angiosperms, a few families of the TEs are younger, and have been repeatedly activated. Some lineages of TE families have been repeatedly amplified and others have been purged from the genomes in angiosperms, whereas those in gymnosperms have accumulated steadily without efficient removal. While differences in TEs may not be the sole determinant for the high fecundity and species richness of angiosperms, it is not unreasonable to conclude that they can account for the high species variability in plants.

## Gene regulation and TEs

We now know that TEs regulate gene expression in many ways, which was demonstrated by McClintock that TEs can cause gene expression changes in both qualitatively and quantitatively kernel color variegation as named TEs as “controlling elements” (McClintock 1956).

TE-insertions into coding genes generate null-alleles that often lead to critical casualties for the host. However, some null-alleles by TE-insertion can survive as exemplified by the wrinkled pea trait, caused by an *Ac/DS* family insertion into the gene for starch synthesis (Bhattacharyya et al. 1990). Another example is waxy foxtail millet (*Setaria italica*) that was derived from a TE-insertion into the *waxy* locus of *granule-bound starch synthase (GBSS1)* gene that has 14 exons (Kawase et al. 2005). Multiple insertions or excisions of several TEs at various locations in the *GBSS1* gene have resulted in waxy (no amylose) to low amylose types of foxtail millet, implying null-function to varying degrees of mild function of the encoded protein. The nature of TE-insertion induced null-phenotypes was also used for gene discovery by reverse genetics approaches (Piffanelli et al. 2007; Settles et al. 2007). Because *Tos17*, *Ty1-copia* retrotransposon in rice, is activated by tissue-culture, Piffanelli et al. (2007) constructed a large number of *Tos17* tagged mutant library after tissue-culture in rice, which was used for forward and reverse genetics strategies for novel gene discovery in rice.

Because most TE-insertion into exons are deleterious to fade out from the population, TE-insertions are often found near the genes and introns, often in the regulatory regions (Lisch 2013a, b; Zhao et al. 2015). While TE-insertion into the enhancer abolishes gene expression, its insertion in the repressor facilitates gene expression (Lisch 2013a). Gene regulation by TE-insertion into the regulatory region is often quantitative. Erucic acid is a non-edible plant oil found in seeds. There are four alleles, *E1*, *E2*, *E3*, and *e* of the *Fatty Acid Elongation1 (FAE1)* gene that determines the content of erucic acid in seeds of yellow mustard (Javidfar and Cheng 2013). *E1* is wild type and *e* is a mutant by *Sal-PIF*, *PIF/Harbinger*-like TE, inserted into the coding region of *FAE1* gene. In *E2* and *E3*, a *copia* type retrotransposon inserted between promoter and transcription start site, but they are different by the methylation at the promoter in *E3*. Erucic acid content in seeds in *E1* was 53%, the null mutant *e* showed no erucic acid. While *E2* showed 24% erucic acid, the promoter methylated *E3* revealed only 1.4% erucic acid content in seeds (Zheng and Cheng 2014). Another example is lignin content in maize. Cinnamyl alcohol dehydrogenase (CAD) is a key enzyme in lignin biosynthesis. *Brown midrib1 (bm1)* plants have reduced content of lignin, which is highly advantageous using in silage and biofuel. In maize, *bm1-cad1* mutant plant produced lignin as low as 24 to 30% compared to the wild type. The *bm1-cad1* mutant has *DS* element in the first intron (Chen et al. 2012). Thus, TE-insertion into gene or genic regions has an attenuating effect on gene regulation.

TE-insertions can also reprogram gene expression. A classic example of the reprogramming of gene expression by TE-insertion is anthocyanin pigment gene expression in

maize. Two regulatory genes, *myc* and *myb*, are required for pigment synthesis in maize. The *myc* gene has two alleles, *r1* and *b1*; *c1* and *pl* are alleles of the *myb* gene (Dooner and Robbins 1991). Alleles *r1* and *c1* are expressed in seeds while *b1* and *pl* are expressed in plant parts other than seeds. However, two variant alleles of *b1*, *B-Peru* and *B-Bolivia* are expressed in developing seeds instead of the plant body. Molecular analysis of these mutants reveals TE-insertions in the *b1* locus, with the independent TE-insertions causing ectopic expression of pigment synthesis genes (Selinger and Chandler 1999). Variation in grape skin color is also representative of TE-insertion into the regulatory region of a gene (Kobayashi et al. 2004). *Myb*-related genes, *VlmybA1-1*, *Vlmyb1-2*, and *Vlmyb1-3*, regulate anthocyanin accumulation in grape fruit skin. A *Ty3-gypsy* insertion into *Vlmyb1-1* gene at the 5'-regulatory end abolishes color accumulation, resulting in a color-less grape. LTR–LTR recombination resulted in solo-LTR and reverted the *Vlmyb1-1* gene function partially, resulting in pink grape (Lisch 2013a).

Methylated TE-insertions in genes often result in reduced expression of neighboring genes. Hollister and Gaut (2009), in their analysis of genomic, epigenomic and population genetic data of *A. thaliana*, observed three responses of neighboring gene expression to methylated TEs: (i) a negative correlation with TE methylation density, (ii) a purifying selection for methylated TEs and not for unmethylated TE, and (iii) presence of farther distances of older and methylated TEs from genes compared to younger TEs. The genome of *A. lyrata* (207 Mb) carries 2–3 times higher TE copy numbers than the genome of *A. thaliana* (130 Mb) (Hu et al. 2011). Genome-wide comparison of 24-nt siRNA of *A. thaliana* and *A. lyrata* revealed that siRNA-targeted TEs were associated with reduced gene expression in both species. However, the efficacy of RNA-directed DNA methylation silencing is lower in *A. lyrata*, suggesting differential TE proliferation between two congeners (Hollister et al. 2011). So, although the global cost of TE proliferation is unavoidable, there is a compensating trade-off between the benefit from TE silencing and the deleterious effects on the reduced gene expression nearby genes.

### Molecular exaptation of TE encoding genes

Exaptation is an evolutionary term that was coined to account for, “*adaptive features originally built by natural selection for one role, or even non-adaptive features, that have since been co-opted for a new role*” (Gould and Vrba 1982). TE encoded proteins mediate transposition of transposable elements, but sometimes they are exapted to perform different functions for host adaptability. These

exapted TEs have been called domesticated TEs (Miller et al. 1992), and numerous cases of TE domestication have been reported for the proteins encoded in both class 1 and class 2 TEs (Feschotte and Pritham 2007; Hoen and Bureau 2012; Alzohairy et al. 2013). Table 2 lists several domesticated TEs in plants.

TE activation can induce mutations that seriously hamper host fitness; in response, host genomes adopt epigenetic defense systems to control TE activities and mitigate their mutagenic potential (Blumenstiel 2011; Martienssen and Chandler 2013). TEs have two fates post-transposition into a new site; either they vanish through loaded mutations or they survive because the host acquires new function(s) (Rouzic et al. 2007; Hoen and Bureau 2012). TEs, upon integration into new chromosomal sites, are epigenetically silenced. The static TEs can either accumulate mutations (sequence changes or deletions), and eventually end up as no longer being recognized as the related ancestral elements. Alternately, they might acquire a function(s) that enhances host fitness (Miller et al. 1992; Muehlbauer et al. 2006). The latter situation is exemplified in domesticated TE elements which have several features that distinguish them from their ancestral autonomous counterparts. Domesticated TEs are usually present in single or low copy numbers, and are often detectable at orthologous loci in other organisms. In contrast, ancestral TEs are present in multiple copies, and often found at different chromosomal loci in divergent species. Domesticated TEs have lost the function for mobility through mutation while ancestral TEs often retain their mobility (Alzohairy et al. 2013). Although the comparison between non-synonymous changes ( $K_a$ ) and synonymous mutation ( $K_s$ ) with the ancestral TE protein provides indirect evidence of the domesticated TE proteins, unequivocally identifying the related ancestral TE protein(s) in the genome often posits challenges because the ancestral TE(s) is no longer present or, even if present,

they are fragmented or present only small fraction in the genome due to mutations (Donoghue et al. 2011; Joly-Lopez et al. 2016). Integrative computational analysis of whole genome sequences enabled to discover genome-wide novel plant genes derived from transposable elements (Hoen and Doug 2015), then reverse genetics approach can directly verify their phenotypic functions (Joly-Lopez et al. 2012, 2016). This genome-wide integrative approach revealed that TE exaptation events have occurred far more frequently to generate novel genes, coinciding with key evolutionary periods. With respect to this point, the evolution of angiosperms with FAR1 (Far-Red Impaired 1) and FHY3 (Far-Red Elongated Hypocotyl 3), *Mutator*-like transposase derived plant transcription factors, provide a good example (Lin et al. 2007). FAR1 and FHY3 regulate the expression of genes that are involved in the response of far-red light controlled by phytochrome A (Lin and Wang 2004). FAR1 and FHY3 also play a role in the diverse plant development and physiological processes including circadian rhythm (Li et al. 2011), chloroplast development (Gao et al. 2013), plant hormone signaling (Tang et al. 2013), shoot branching and stress response (Stirnberg et al. 2012), all of which are critical adaptive traits in eudicots. Another citable example is the *MUSTANG* (*MUG*) which is also *Mutator*-like element (*MULE*) derived domesticated plant gene (Cowan et al. 2005). In *Arabidopsis*, the *MUG* genes have a diverse role in flowering plant development such as chlorophyll production, flowering time, and seed yield (Joly-Lopez et al. 2012; 2015). In angiosperms, the *MUG* gene diverged into *MUGA* and *MUGB* that are present in basal eudicots, but absent in gymnosperms (Cowan et al. 2005; Joly-Lopez et al. 2012), implying that *MUG* genes might have occurred in early period in angiosperm evolution after gymnosperm-angiosperm divergence. Likewise, *Transib* transposase derived V(D)J recombinase in jawed vertebrates (Kapitonov and Jurka 2005) and *Ty3/Gypsy*

**Table 2** A few examples of domesticated transposable elements in plants

Gene Name	Function	Origin (TE)	Organisms	References
<i>FHY3/FAR1</i>	Light sensing Pleiotropic	<i>MULE</i>	<i>A. thaliana</i>	Lin et al. (2007)
<i>Daysleeper</i>	Plant development DNA repair	<i>hAT</i>	<i>A. thaliana</i>	Bundock and Hooykaas (2005)
<i>MUG1-8</i>	Plant development, pleiotropic	<i>MULE</i>	<i>A. thaliana</i>	Cowan et al. (2005), Joly-Lopez et al. (2016)
<i>GARY</i>	Unknown	<i>hAT</i>	Poacea (wheat, barley)	Muehlbauer et al. (2006)
<i>POGO</i>	Centromere associated protein CENP-B	<i>Pogo</i>	Drosophila, yeast, plants	Smit and Riggs (1996)
<i>ALP1</i>	Plant development pleiotropic	<i>PIF</i>	<i>A. thaliana</i>	Liang et al. (2016)
<i>COPIA-R7</i>	<i>Disease resistance</i>	<i>Copia retrotransposon</i>	<i>A. thaliana</i>	Tsuchiya and Eulgem (2016)
<i>ADOTE</i>	Plant development pleiotropic	<i>CACTA</i>	<i>A. thaliana</i>	Park and Kim (unpublished)



retrotransposon Gag protein derived *peg10* in placental mammals (Ono et al. 2006) provide examples of the roles of TE encoded protein exaptation in animal adaptation during evolution. In summation, it would not be unreasonable to contend that molecular exaptation of TE elements has been a significant determinant or turning point in accounting for the evolution of current forms of flora and fauna.

## Concluding remarks

Genomic investigations have enabled researchers to probe into the structural and organizational intricacies of an organism's holistic genetic make-up. It has been amply substantiated that genomes of organisms serve as a home for the panoply of motley genetic elements, of which TEs are the most prominent in terms of abundance and distribution. The notion of mobile elements was not readily accepted at the outset (Fedoroff 2012), and their “raison d'être” earned them such harsh epithets as “parasites”, and “junk” or “selfish” DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). However, the ubiquity of TEs in prokaryotes and eukaryotes necessitated a revision of their label, and has substantiated their status of an ancient and predated divergence of all life forms (Schulman and Wicker 2013). It is now evident that they have been the major drivers of a unique trajectory in evolutionary biology (Fedoroff 2012; Fedoroff and; Bennetzen 2013). TEs generate genetic variations that are the raw materials for evolution. Bereft of these TE-driven variations, the current taxonomical status of myriad species of flora and fauna would be completely different.

In plants, much of the work with TEs has been deliberately centered around domesticated and laboratory based model species rather than wild species in nature, primarily because of the plethora of relatively accurate and well-characterized data on traits in the former group of species. The citation of a temporal fluctuation of TE copy numbers in marginal populations of the wild diploid wheat, *Aegilops speltoides*, may facilitate an insight into the relatedness of the microevolution process and TE-driven plastic genomes in times of rapid climate change (Belyayev et al. 2010; Belyayev 2014). The industrial melanism mutation in the British peppered moth was recently attributed to a microevolutionary change caused by a TE-insertion in a gene involved in cell-cycle regulation (van't Hof et al. 2016). Investigations utilizing environmental epigenomics are the new focus for further elucidation of the role of TEs in the adaptability of wild unexplored species and untamed organisms (Bossdorf et al. 2008; Lira-Medeiros et al. 2010; Mirouze and Paszkowski 2011). Although complete comprehension of the ubiquity and abundance of self-replicating TEs remains in the offing, it is inexorably clear that

contemporary biology cannot be credibly and adequately explained without acknowledging the role of transposable elements in evolution and biology.

**Acknowledgements** Full appreciation and indebtedness is extended to Dr. Ramesh Bhambhani for his editorial assistance with the manuscript. This study has been worked with the support of a research grant of Kangwon National University in 2016–2017.

## Compliance with Ethical Standards

**Conflict of interest** Author (NSK) does not have a conflict of interest with the contents of the manuscript.

**Research involving human and animal rights** The manuscript did not utilize results generated from research with human and animal subjects.

## References

- Abrusán G, Sziláyi A, Zhang Y, Papp B (2013) Turning gold into ‘junk’: transposable elements utilize central proteins of cellular networks. *Nucl Acids Res* 41: 3190–3200
- Ågren J, Wright SI (2011) Co-evolution between transposable elements and their host: a major factor in genome evolution? *Chromosome Res* 19:777–786
- Alzohairy AM, Gyulai GG, Jansen RK, Bahieldin A (2013) Transposable elements domesticated and neofunctionalization by eukaryotic genomes. *Plasmid* 69:1–15
- Bartoš J, Paux E, Kofler R, Havránková M, Kopecky D, Suchánková P, Šimková H, Šafář J, Town CD, Lelley T et al (2008) A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. *BMC Plant Biol* 8:95
- Belyayev A, Kalendar R, Brodsky L, Nevo E, Schulman AH, Raskina O (2010) Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. *Mobile DNA* 1:6. doi:10.1186/1759-8753-1-6
- Bennett M, Leitch IJ (2011) Nuclear DNA amounts in angiosperms: targets, trends, and tomorrow. *Ann Bot* 107:467–590
- Bennett M, Leitch IJ (2012) Plant DNA C-value database (release 6.0. December 2012)
- Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. *Ann Bot* 95:127–132
- Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 60:115–122
- Blumenstiel JP (2011) Evolutionary dynamics of transposable elements in a small RNA world. *Trends Genet* 27:23–31
- Bossdorf P, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. *Ecol Lett* 11:106–115
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438
- Brunet TDP, Doolittle WF (2015) Multilevel selection theory and evolutionary functions of transposable elements. *Genome Biol Evol* 7:2445–2457
- Bundock P, Hooykaas P (2005) An *Arabidopsis* *hAT*-like transposase is essential for plant development. *Nature* 436:282–284

- Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E (2016) Transposons, genome size, and evolutionary insights in animals. *Cytogenet Genome Res* 147:217–239
- Chadha S, Sharma M (2014) Transposable elements as stress adaptive capacitors induce genomic instability in fungal pathogen *Magnaporthe oryzae*. *PLOS Genet* 9(4): e94415 doi:10.1371/journal.pone.0094415
- Chen W, VanOpdorp N, Fitzl D, Tewari J, Friedmann P, Greene T, Thomson S, Kumpatla S, Zheng P (2012) Transposon insertion in a cinnamyl alcohol dehydrogenase gene is responsible for a *brown midrib1* mutation in maize. *Plant Mol Biol* 80:289–297
- Civán P, Švec M, Hauptvogel P (2011) On the coevolution of transposable elements and plant genomes. *J Bot Article ID* 893546. doi:10.1155/2011/893546
- Cowan R, Hoen D, Scoen D, Bureau T (2005) MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. *Mol Biol Evol* 22:2084–2089
- Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination. *Genome Res* 12:1075–1079
- Donoghue MTA, Keshavaiah C, Swamidatta SH, Spillane C (2011) Evolutionary origins of *Brassicaceae* specific genes in *Arabidopsis thaliana*. *BMC Evol Biol* 11:47. doi:10.1186/1471-2148-11-47
- Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284:601–603
- Dooner HK, Robbins TP (1991) Genetic and developmental control of anthocyanin biosynthesis. *Annu Rev Genet* 25:173–199
- Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic *Helitron* landscape of the maize genome. *Proc Natl Acad Sci USA* 106:19916–19921
- Estep MC, DeBarry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panictoid grass evolution. *Heredity* 110:194–204
- Fedoroff NV (2012) Transposable elements, epigenetics, and genome evolution. *Science* 338:758–767
- Fedoroff NV, Bennetzen JL (2013) Transposons, genome shock, and genome evolution. In “Plant transposons and genome dynamics in evolution”. First Edition, NV Fedoroff ed. Wiley, Hoboken pp 181–201
- Feschotte C, Prtiham EJ (2005) Non-mammalian c-integrase are encoded by giant transposable elements. *Trends Genet* 21:551–552
- Feschotte C, Prtiham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Ann Rev Genet* 41:331–368
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5:103–107
- Gao Y, Liu H, An C, Shi Y, Liu X, Yuan W, Zhang B, Yang J, Yu C, Gao H (2013) *Arabidopsis* FRS4/CPD25 and FHY3/CPD45 work cooperatively to promote the expression of the chloroplast division gene ARC5 and chloroplast division. *Plant J* 75:795–807
- Goodwin T, Butler ML, Poulter RT (2003) *Cryptons*: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* 149:3099–3109
- Gould SJ, Eldredge N (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151
- Gould SJ, Vrba ES (1982) Exaptation—a missing term in the science of form. *Paleobiology* 8:4–15
- Gregory TR (2001) Coincidence, coevolution, or causation? DNA content, cell size, and C-value enigma. *Biol Rev* 76:65–101
- Gregory TR (2002) A bird’s-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evol Int J org Evol* 56:121–130
- Greilhuber J, Doležel J, Lysák, Bennett M (2005) The origin, evolution, and proposed stabilization of the term ‘genome size’ and ‘C-value’ to describe nuclear DNA contents. *Ann Bot* 95:255–260
- Gupta S, Gallavotti A, Stryker GA, Schmidt RJ, Lal SK (2005) A novel class of *Helitron*-related transposable elements in maize contains portions of multiple pseudogenes. *Plant Mol Biol* 57:115–127
- Hawkins JS, Kim H, Nason JD, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res* 16:1252–1261
- Hawkins JS, Grover CE, Wendel JF (2008) Repeated big bangs and the expanding universe: directionality in plant genome evolution. *Plant Sci* 174:557–562
- Hawkins JS, Proulx SR, Rapp RA, Wendel JF (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc Natl Acad Sci USA* 106:17811–17816
- Hickman AB, Chandler M, Dyda F (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol Biol* 45:50–69
- Hoen D, Bureau D (2012) Transposable element exaptation in plants. In “Plant transposable elements” Grandbastien MA, Casacuberta JM (eds). Springer Berlin pp 219–251
- Hoen D, Bureau D (2015) Discovery of novel genes derived from transposable elements using integrative genomic analysis. *Mol Biol Evol* 32:1487–1506
- Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 19:1419–1428
- Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci USA* 108:2322–2327
- Hu T, Patty P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genet* 43:476–481
- Hua-Van A, Rouzic AL, Boutin TS, Filée J, Capy P (2011) The struggle for life of the genome’s selfish architects. *Biol Direct* 6:19
- Javidfar F, Cheng B (2013) Construction of a genetic linkage map and QTL analysis of erucic acid content and glucosinolate components in yellow mustard (*Sinapis alba* L.). *BMC Plant Biol* 13:142
- Joly-Lopez Z, Forczek E, Hoen DR, Jurec N, Bureau TE (2012) A gene family derived from transposable elements during early angiosperm evolution has reproductive benefits in *Arabidopsis thaliana*. *PLoS Genet* 8:e1002931
- Joly-Lopez Z, Hoen DR, Blanchette M, Bureau TE (2016) Phylogenetic and genomic analyses resolve the origin of important plant genes derived from transposable elements. *Mol Biol Evol* 33:1937–1956
- Jurka J, Bao WD, Kojima KK (2011) Families of transposable elements, population structure and the origin of species. *Biol Direct* 6:44
- Kapitonov V, Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc Acad Natl Sci USA* 100: 6569–6574
- Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from *Transib* rtransposon. *PLoS Biol* 3:e181
- Kapitonov V, Jurka J (2006) Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci USA* 103:440–4545
- Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Town CD et al (2000) Analysis of

- the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Kawase M, Fukunaga K, Kato K (2005) Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol Gen Genomic* 275: 131–140.
- Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624
- Knight CA, Molinari NA, Petrov DA (2005) The large genome constraint hypothesis: evolution, ecology, and phenotype. *Ann Bot* 95:177–190
- Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. *Science* 304:982
- Kojima KK, Jurka J (2011) *Crypton* transposons: identification of new diverse families and ancient domesticated events. *Mobile DNA* 2: 12
- Kovach A, Wegrzyn J, Parra G, Holt C, Bruening GE, Loospra CA, Hartigan J, Yandell M, Langley CH, Korf I, Neale DB (2010) The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics* 11: 420 doi:10.1186/1471-2164-11-420
- Kumar A, Benntzen JL (1999) Plant retrotransposable elements. *Ann Rev Genet* 33:479–532
- Lai J, Li Y, Messing J, Dooner HK (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. *Proc Natl Acad Sci USA* 102:9068
- Lal SK, Giroux MJ, Brendel V, Vallejos CE, Hannah LC (2003) The maize genome contains a *Helitron* insertion. *Plant Cell* 15:381–391
- Le Rouzic A, Boutin TS, Capy P (2007) Long-term evolution of transposable elements. *Proc Natl Acad Sci USA* 49:19375–19380
- Lee SI, Kim NS (2014) Transposable elements and genome size variations in plants. *Genom Inform* 12: 87–97
- Levin HL (2002) Newly identified retrotransposons of Ty3/gypsy class in fungi, plants, and vertebrates. In: “Mobile DNA II” N. Craig et al (ed) edited. ASM Press, Washington DC, pp 684–701
- Levy AA (2013) Transposons in plant speciation. In “Plant transposons and genome dynamics in evolution”. First Edition, NV Fedoroff edited. Wiley, Hoboken, pp 165–179
- Li Y, Dooner HK (2009) Excision of *Helitron* transposons in maize. *Genetics* 182:399–402
- Li G, Siddiqui H, Teng Y, Lin R, Wan XY, Li J, Lau OS, Ouyang X, Dai M, Wan J et al (2011) Coordinated transcriptional regulation underlying the circadian clock in *Arabidopsis*. *Nat Cell Biol* 13:616–622
- Liang SS, Hartwig B, Perera P, Mora-Garcia S, de Leau E, Thorton H, de Alves FL, Rapsilber J, Yang S, James GV et al (2016) Kicking against the PRCs—A domesticated transposase silencing mediated by polycomb group proteins and is accessory component of polycomb repressive complex2. *PLoS Genet* 12: e1005812. doi:10.1371/journal.pgen.1005812
- Lin R, Wang H (2004) *Arabidopsis FHY3/FAR1* gene family and distinct roles of its members in light control of *Arabidopsis* development. *Plant Physiol* 136:4010–4022
- Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H (2007) Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. *Science* 318:1302–1305
- Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PCG (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE*. doi: 10.1371/journal.pone.0010326
- Lisch D (2013a) How important are transposons for plant evolution? *Nat Rev Genet* 14:49–61
- Lisch D (2013b) Transposons in plant gene regulation. In “Plant transposons and genome dynamics in evolution”. First Edition, NV Fedoroff edited. Wiley, Hoboken, pp 93–116
- Lönnig W-E, Saedler H (2002) Chromosome rearrangements and transposable elements. *Ann Rev Genet* 36:389–410
- Makarevitch I, Waters AJ, West PT, Stitzere M, Hirsh CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of genes I response to heat stress. *PLOS Genet* 11: e1005566. doi:10.1371/journal.pgen.1005566
- Martienssen R, Chandler V (2013) Molecular mechanisms of transposon epigenetic regulation. In “Plant transposons and genome dynamics in evolution”. First Edition, NV Fedoroff edited. Wiley, Hoboken, pp 71–92
- McClintock B (1948) Mutable loci in maize. *Carnegie Institute of Washington Year Book* 47: 155–169
- McClintock B (1956) Controlling elements and the gene. *Cold Spring Harb Quant Biol* 21: 197–216
- McClintock B (1984) The significance of responses of the genome to challenge. *Science* 226:792–801
- Megallon S, Sanderson M (2005) Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evol Int J org Evol* 59:1653–1670
- Miller WJ, Hagemann S, Reiter E, Pinsker W (1992) P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. *Proc Natl Acad Sci USA* 89:4018–4022
- Mirouze and Paszkowski J (2011) Epigenetic contribution to stress adaptation in plants. *Curr Opin Plant Biol* 14:267–274
- Morgante M, Brunner S, Pea G, Fengler K, Azucollo A, Rafalski A (2005) Gene duplication and exon shuffling by *helitron*-like transposons generate intraspecific diversity in maize. *Nature Genet* 37:997–1002
- Muehlhauber GJ, Bhat BS, Syed NH, Heinen S, Cho S, Marshall D, Pateyron S, Buisine N, Chalhoub B, Flavell AJ (2006) A *hAT* superfamily transposase recruited by the cereal grass genome. *Mol Genet Genom* 275: 553–563
- Moutri AR, Marchetto MCN, Coufal NG, Gage FH (2007) The necessary junk: new functions for transposable elements. *Hum Mol Genet* 16:R159–R167
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A et al (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 495:579–584
- Oliver KR, Greene WK (2009) Transposable elements: powerful facilitators of evolution. *Bioessays* 31:703–714
- Oliver KR, Greene WK (2012) Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TR-Thrust hypothesis. *Ecol Evol* 12:2912–2933
- Oliver KR, McComb JA, Greene W (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol* 5:1886–1901
- Ono R, Nakamura K, Inoue K, Naruse M, Usami T, Wakisaka-Saito N, Hino T, Suzuki-Migishima R, Ogonuki N, Miki H et al (2006) Deletion of *Peg10*, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat Genet* 38:101–106
- Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. *Nature* 284:604–607
- Piffanelli P, Droc G, Mieulet D, Nanau N, Bes M, Bourgeois E, Rouviere C, Gavory F, Cruaud C, Ghesquiere A, Giarderoni E (2007) Large-scale characterization of *Tos17* insertion sites in a rice T0DNA mutant library. *Plant Mol Biol* 65:587–601
- Reznikoff WS (2003) *Tn5* as a model or understanding DNA transposition. *Mol Microbiol* 47:1199–1206
- Rieseberg LH (1997) Hybrid origins of plant species. *Ann Rev Ecol Syst* 28:359–389

- Rostoks N, Park YJ, Ramakrishna W, Ma J, Druka A, Shiloff BA, SanMiguel PJ, Jiang Z, Brueggeman R et al (2002) Genome sequencing reveals gene content, genomic organization, and recombination relationships in barley. *Funct Integ. Genomics* 2:70–80
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
- Schulman AH, Wicker T (2013) A field guide to transposable elements. In *Plant Transposons and Genome Dynamics in Evolution*, First edition, HV Fedoroff ed. Wiley, Hoboken. pp 15–40
- Selinger DA, Chandler VL (1999) Major recent and independent changes in levels and patterns of expression of expression have occurred at the *b* gene, a regulatory locus in maize. *Proc Natl Acad Sci USA* 96:15007–15012
- Serra F, Becher V, Dopazo H (2013) Neutral theory predicts the relative abundance and diversity of genetic elements in a broad array of eukaryotic genomes. *PLoS ONE* 8(6): e63915. doi:10.1371/journal.pone.0063915
- Settles A, Holding DR, Tan BC, Latchaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Fajarado DS, Wroclawska E et al (2007) Sequence-indexed mutation in maize using the UniformMu transposon-tagging population. *BMC Genom* 8:116
- Slotkin RK, Martienssen R (2007) Transposable elements and epigenetic regulation of the genome. *Nat Rev Genet* 8:272–285
- Smit AF, Riggs AD (1996) Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci USA* 93:1443–1448
- Smith JJ, Suminaya K, Amemiya C (2012) A living fossil in the genome of a living fossil: Harbinger transposons in the Coelacanth genome. *Mol Biol Evol* 29:985–993
- Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Li W et al (2013) Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution". *Nat Genet* 45:415–421
- Soltis DE, Bell CD, Kim S, Soltis PS (2008) Origin and early evolution of angiosperms. *Ann NY Acad Sci* 1133:3–25
- Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, Ungerer MC, Knapp SJ, Rieseberg LH, Burker JM (2012) The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J* 72: 142–153
- Stirnberg P, Zhao S, Williamson L, Ward S, Leyser O (2012) FHY3 promotes shoot branching and stress tolerance in *Arabidopsis* in an AXR1-dependent manner. *Plant J* 71:907–920
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genet* 43:1160–1163
- Tang W, Ji Q, Huang Y, Jiang Z, Bao M, Wang H, Lin R (2013) FAR-RED ELONGATED HYPOCOTYL3 and FAR-RED IMPAIRED RESPONSE1 transcription factors integrate light and abscisic acid signaling in *Arabidopsis*. *Plant Physiol* 163:857–866
- Thomas CA Jr (1971) The genetic organization of chromosomes. *Ann Rev Genet* 5:2237–2256
- Thomas and Pritham (2014) *Helitrons*, the eukaryotic rolling-circle transposable elements. *Microbiol Spect* 3: 893–926.
- Tsuchiya T, Eulgem T (2016) An alternate polyadenylation mechanism coopted to *Arabidopsis RPP7* gene through intronic retrotransposon domestication. *Proc Natl Acad Sci USA* 110:E3535–E3543
- Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr Biol* 16:R872
- Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav V, Nevo E, Schulman AH (1999) Retrotransposon BARE-1 and the role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769–1784
- Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res* 110:91–107
- Volf JN (2006) Turning junk into gold: domestication of transposable elements and creation of new genes in eukaryotes. *Bioessays* 28:913–922
- Voytas DF, Boeke JD (2002) *Ty1* and *Ty5* of *Saccharomyces cerevisiae*. In: Craige et al (ed) edited. ASM Press, Washington, D.C, pp 631–662
- Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proc Natl Acad Sci USA* 103:17644–17649
- Wendel JF, Jackson SA, Meyers BC, Wing RA (2016) Evolution of plant genome architecture. *Genome Biol* 17:37. doi:10.1186/s13059-016-0908-1
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy O, Morgante M, Panaud O et al (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973–982
- Yang L, Bennetzen J (2009) Distribution, diversity, evolution, and survival of *Helitrons* in the maize genome. *Proc Natl Acad Sci USA* 106:19922–19927
- Zedek F, Šmerda J, Šmerda P, Bureš P (2010) Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*. *BMC Plant Biol* 10:265
- Zhang J, Yu C, Krishnaswamy L, Peterson T (2011) Transposable elements as catalysts for chromosome rearrangements. *Methods Mol Biol* 701:315–326
- Zhao D, Fergoson AA, Jiang N (2015) What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochim Biophys Acta* 1859: 366–380
- Zheng F, Cheng B (2014) Transposable element insertion and epigenetic modification cause the multiallelic variation in the expression of *FAE1* in *Sinapis alba*. *Plant Cell* 26:2648–2659