Chromosome Res (2014) 22:203–216
DOI 10.1007/s10577-014-9418-8

REVIEW

Diversity and evolution of transposable elements in *Arabidopsis*

Zoé Joly-Lopez · Thomas E. Bureau

Published online: 7 May 2014

Abstract Transposable elements are mobile genetic elements that have successfully populated eukaryotic genomes and show diversity in their structure and transposition mechanisms. Although first viewed solely as selfish, transposable elements are now known as important vectors to drive the adaptation and evolution of their host genome. Transposable elements can affect host gene structures, gene copy number, gene expression, and even as a source for novel genes. For example, a number of transposable element sequences have been co-opted to contribute to evolutionary innovation, such as the mammalian placenta and the vertebrate immune system. In plants, the need to adapt rapidly to changing environmental conditions is essential and is reflected, as will be discussed, by genome plasticity and an abundance of diverse, active transposon families. This review focuses on transposable elements in plants, particularly those that have beneficial effects on the host. We also emphasize the importance of having proper tools to annotate and classify transposons to better understand their biology.

Keywords transposons · mobile elements · *Arabidopsis lyrata* · *Arabidopsis thaliana* · exaptation · domestication · transduplication · transduction · transposase · retrotransposon · annotation · stress · regulatory networks

Responsible Editors: Martin Lysak and Paul Fransz.

Z. Joly-Lopez · T. E. Bureau (✉)
Department of Biology, McGill University,
Montreal, QC, Canada
e-mail: thomas.bureau@mcgill.ca

Abbreviations

Ac/Spm	*Activator/suppressor-mutator*
Bs1	*Barren sterile1*
cDNA	Complementary DNA
CNE	Conserved non-coding element
DBD	DNA-binding domain
dsRNA	Double-stranded RNA
DTE	Domesticated transposable element
FAR1	Far-red impaired response 1
FHL	FHY1-like
FHY1	Far-red elongated hypocotyl 1
FHY3	Far-red elongated hypocotyl 3
hAT	*hobo/Ac/Tam*
HY5	Elongated hypocotyl 5
IN	Integrase
KI	*KAONASHI*
LINE	Long interspersed element
LTR	Long terminal repeats
MITE	Miniature inverted transposable element
MULE	*Mutator*-like element
MUG	MUSTANG
ORF	Open reading frame
PB1	Phox and Bem1
RdDM	RNA-directed DNA methylation
RH	Ribonuclease H
RT	Reverse transcriptase
SINE	Short interspersed element
TE	Transposable element
TIR	Terminal inverted repeat
TRIM	Terminal-repeat retrotransposon in miniature
ULP	Ubiquitin-like protease

Introduction

A scientific milestone marked the start of the new millennium: the publication of the first plant genome sequence, *Arabidopsis thaliana* (Arabidopsis Genome 2000). Part of the Brassicaceae family, which encompasses valuable food (e.g., cabbage and mustard) and oil (e.g., canola), *A. thaliana* is a key model organism with a rapid life cycle and small size, making it ideal for laboratory studies. In the nearly 15 years since this milestone, 49 plant genomes have been published, 25 of them in the last 2 years (Michael and Jackson 2013).

These genomes have revealed that plant genomes are highly complex and subject to rearrangement events such as whole-genome duplication or triplication and chromosome fusion. While gene content oscillates between 20,000 and 50,000 genes (Wicker 2012), differences in genome size can be much greater, with published genomes as small as 82 megabase pairs (Mb) for the carnivorous bladderwort (*Utricularia gibba*) to as large as 19,600 Mb for the Norway Spruce (*Picea abies*) and a median of 480 Mb (Bennett 2011). Although they diverged only about 10 million years ago (MYA), *A. thaliana* (125 Mb) and *Arabidopsis lyrata* (207 Mb) have a nearly 50 % difference in genome size (Hu et al. 2011), both of which are relatively small genomes compared to other Brassicaceae species, which diverged up to 20 million years ago (MYA) and can be 5–10 times larger (Yang and Bennetzen 2009; Wicker 2012).

Differences in genome size are largely due to differences in the abundance of repetitive sequences, especially transposable elements (TEs) (El Baidouri and Panaud 2013). TEs are genetic elements that can transpose and persist in host genomes without conferring an apparent selective advantage to the host. In this review, we examine the TEs of *Arabidopsis*: their abundance, effects on the genome, and contributions to genome evolution. In particular, we investigate the possibility that TEs help to give rise to novel genes and regulatory networks.

TE types and survival mechanisms in *Arabidopsis*

Two main types of TEs in plants

TEs are ubiquitous in eukaryotic genomes and classified by the genes they contain and their mechanism of transposition (Wicker et al. 2007). Based on this, we find two main classes as follows: retrotransposons and DNA transposons. Retrotransposons transpose via "copy and paste" mechanisms, where the TE is transcribed into RNA, reverse transcribed, and the new complimentary DNA (cDNA) copy reinserted elsewhere into the genome (Fig. 1b). The retrotransposon class consists of different orders, among which we find those with long terminal repeats (LTRs), which in plants are the predominant order, and those without LTRs (Kazazian 2004; Wicker et al. 2007). Autonomous LTR retrotransposons resemble retroviruses and have similar open reading frames (ORFs), usually including *GAG*, which encodes the capsid protein, and *POL*, which contains reverse transcriptase (RT), ribonuclease H (RH), and integrase (IN) domains that catalyze cDNA production and mediate reinsertion into the genome (Kazazian 2004) (Fig. 1a). However, unlike retroviruses, LTR retrotransposons lack an apparent functional envelope (ENV) genes and additional sequences required for an infectious lifestyle. Non-LTR retrotransposons do not have terminal repeats but repeat sequences, often poly(A), at their 3′ terminus and variable sequence deletions at their 5′ terminus. They contain two open ORFs, encoding a GAG protein (ORF1) and an endonuclease and RT (ORF2) (Fig. 1a).

DNA transposons transpose via a "cut and paste" mechanism catalyzed by a transposase, which is encoded by the TE. Although different DNA transposon superfamilies encode widely diverged transposases, they have similar architectures. Transposases generally have an N-terminal functional DNA-binding domain (DBD) that binds to specific motifs in the terminal inverted repeats (TIRs) or subterminal sequences (Fig. 1a). They also have a C-terminal catalytic domain that mediates double-strand DNA cleavage and reinsertion into another genomic location (Feschotte and Pritham 2007; Sinzelle et al. 2009) (Fig. 1b). Some families, such as *CACTA* and *Mutator* in maize, encode additional products necessary for their transposition or integration (Yang and Bennetzen 2009; Raizada et al. 2001). Another type of DNA transposons, called *Helitron* was discovered relatively recently in *A. thaliana* and *Oryza sativa* (Kapitonov and Jurka 2001; Hollister and Gaut 2007). *Helitrons* may transpose by rolling-circle semi-replicative mechanism in which one strand of the TE is transferred to a new genomic site and serves as the template for DNA

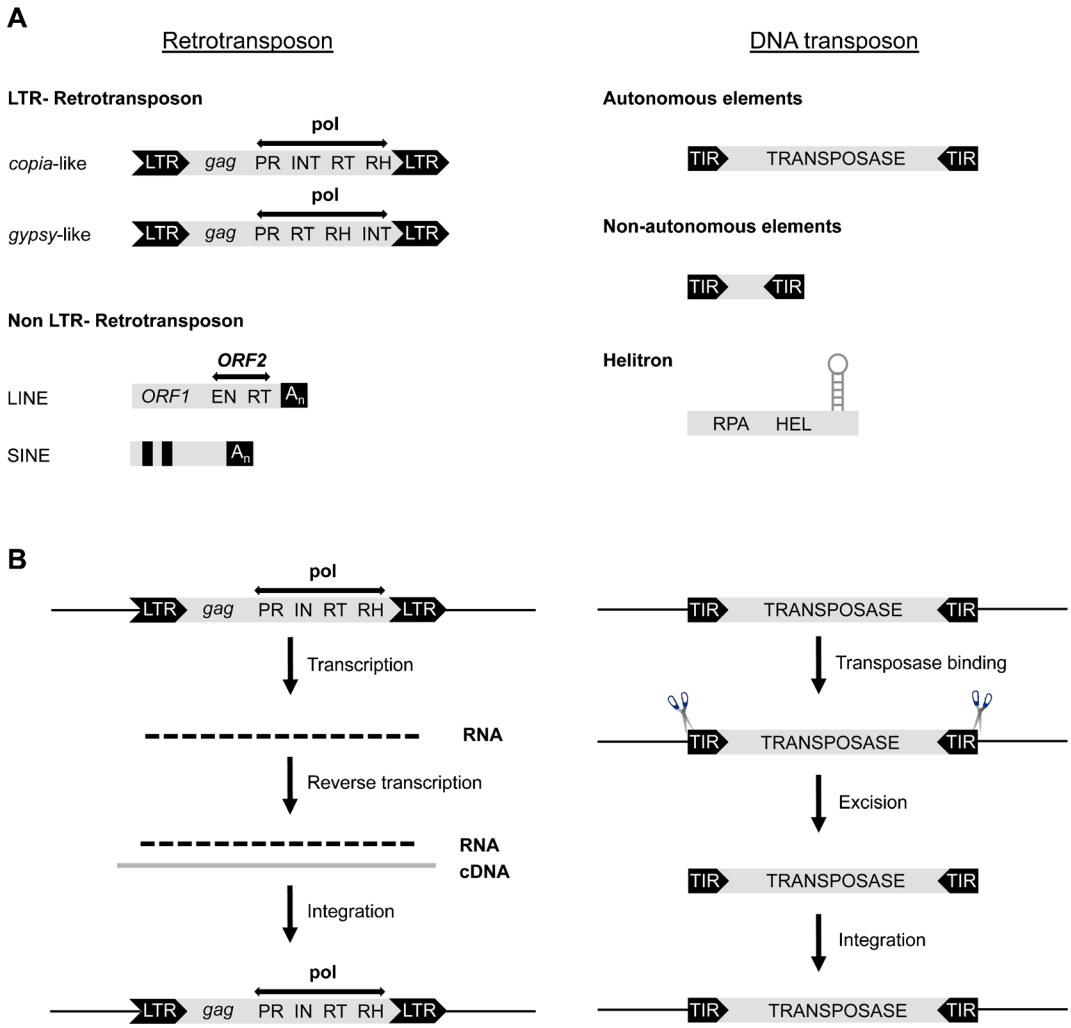

Fig. 1 Types of retrotransposons and DNA transposons in *Arabidopsis* and their mechanisms of mobilization. **a** Retrotransposons present in *Arabidopsis* are shown here by types arranged 5′ to 3′. LTR retrotransposons have long terminal repeats (*LTRs*) and contain two genes, *gag* and pol, that encode a protease (*PR*), an integrase (*INT*), a reverse transcriptase (*RT*), and an RnaseH (*RH*). Non-LTR retrotransposons terminate with a sequence repeat, often a poly A tail (A_n). Long interspersed nuclear elements (*LINEs*) contain two open reading frames (*ORFs*), where ORF2 encodes RT and endonuclease (*EN*), and short interspersed nuclear elements (*SINEs*) do not encode any proteins but contain *boxes A* and *B* of a pol III promoter (*thick black lines*). Autonomous DNA transposons have TIRs and encode a transposase. Non-autonomous elements do not encode for a transposase but have TIRs. Helitrons contain an ssDNA-binding replication protein A (*RPA*) and helicase (*HEL*) and have a stem-loop structure in the 3′ region. **b** On the *left*, example of retrotransposon mobilization. Transcription of a retrotransposon into RNA (*black dashed line*) is initiated from a promoter in the LTR. Then, the RNA strand is reversed-transcribed by the reverse transcriptase into cDNA (*gray continuous line*) and finally integrated into another genomic location. On the *right*, DNA transposons mobilize by a "cut and paste" mechanism. In most cases, the transposase is encoded and binds specific DNA motifs embedded within the TIRs. It then cleaves the double-stranded DNA (*scissors*) at these sites and mediates mobilization and ligation into another genomic location

synthesis catalyzed by the host repair machinery (Kapitonov and Jurka 2007).

TEs can sustain and propagate themselves by either encoding the genes necessary to promote their own transposition (autonomous elements) or by using the transposition machinery of closely related or even unrelated families to mediate their own transposition (non-autonomous elements). Examples of non-autonomous elements are short interspersed elements (SINEs) for retrotransposons and miniature inverted-repeat transposable elements (MITEs) for DNA transposons (Wessler et al. 1995).

TE abundance and activity in *Arabidopsis*

TE abundance varies considerably between *Arabidopsis* species, and estimates vary between studies. For *A. thaliana*, most estimates are near 15 % TE content (e.g., (Arabidopsis Genome 2000; de la Chaux et al. 2012)) but some are as high as 24 % (Hu et al. 2011), whereas in *A. lyrata*, estimates range from 25 (de la Chaux et al. 2012) to 30 % (Hu et al. 2011). Generally in plants, retrotransposons contribute the highest DNA coverage, due in part to their length, and *Arabidopsis* is no exception (de la Chaux et al. 2012; Schnable et al. 2009). Yet DNA transposons have approximately 50 % more copies in both *A. thaliana* and *A. lyrata* (de la Chaux et al. 2012). The most abundant DNA element families are Atrep3 and RS23, both Helitrons, followed closely by *Mutator*-like elements (MULEs). As in other plants, LTR retrotransposons are the most active, followed by MULEs. Most extant LTR retrotransposons have inserted within the past few million years, probably reflecting a high rate of turnover (i.e., insertion and deletion). For example, in *A. thaliana*, more than 90 % of *copia* superfamily insertions postdate the divergence between *A. thaliana* and *A. lyrata*, 5–10 MYA (Pereira 2004; Ziolkowski et al. 2009). *A. lyrata* has greater TE activity, with the average age of LTR retrotransposons being approximately 1.1 MY compared to 3.1 MY in *A. thaliana* (Hu et al. 2011). There is also a wide variation in TEs even between *A. thaliana* populations. Cao et al. (Cao et al. 2011) performed whole-genome sequencing of 80 strains from eight populations in different geographical regions and found that 80 % of TEs in the reference genome are partially or completely absent in at least one of the strains. Moreover, a comparison of the genome sequences of two *A. thaliana* lines, Columbia (Col) and Landsberg *erecta* (L*er*), detected that since these ecotype diverged about 200,000 years ago, nearly 200 TEs have been active, including retrotransposons, MULEs, and Helitrons (Ziolkowski et al. 2009).

However, because TE detection and annotation remain difficult, these numbers ought to be treated with caution. To illustrate, a detailed manual curation of selected *A. lyrata* genomic reveals that TEs are often misannotated as protein-coding genes (T.E. Bureau, unpublished data). We will elaborate on issues related to TE annotation in a subsequent section, "The importance of TE annotation". Furthermore, several genes in *A. thaliana* annotated as TEs are likely domesticated TE (DTE) genes (e.g., Fig. 2b), which we discuss in the section entitled "Exaptation (molecular domestication)".

TE survival

TEs, which populate virtually all extant genomes, have been characterized as selfish and parasitic because they can persist without producing immediate beneficial phenotypes (Doolittle and Sapienza 1980; Orgel and Crick 1980). Through transposition, TEs generate multiple copies that sustain mutations independently, thus a TE family can escape disabling mutations and deletions so long as its transposition rate is sufficiently high relative to the mutation rate. By replicating sufficiently frequently in the gametes, at least one autonomous copy can be passed to the next generation. Because they are selected not to directly improve the fitness of their host genome but to maintain self-replicating copies, TEs are not primarily under phenotypic selection, like host genes, but are instead under self-replicative selection.

Although they must self-replicate to survive, large-scale TE activation could result in an overwhelmingly high level of mutations that could jeopardize the survival of the host (and its TEs). To prevent this, mechanisms have evolved to recognize and silence TEs, both transpositionally and transcriptionally, and maintain this silencing memory through multiple rounds of cell division. TEs are primarily silenced via post-transcriptional gene silencing and RNA-directed DNA methylation (RdDM), both of which are initiated by the production of double-stranded RNAs (dsRNAs) from TE sequences. dsRNAs are processed into small interfering RNAs, which can target TE-related mRNAs for degradation and trigger DNA and histone modifications via RdDM (Lisch 2009; Castel and Martienssen 2013). To maximize their self-replicative success despite these silencing pathways, TEs have evolved self-regulatory mechanisms like tissue-specific promoters. For example, promoters in autonomous maize MULEs contain nested sets of pollen-specific motifs, which increase reporter gene expression in more than 20-fold compared to leaves (Rudenko and Walbot 2001). For additional details about TE silencing, see "Control of transposable elements in *Arabidopsis thaliana*" by Ito and Kakutani, also in this issue.

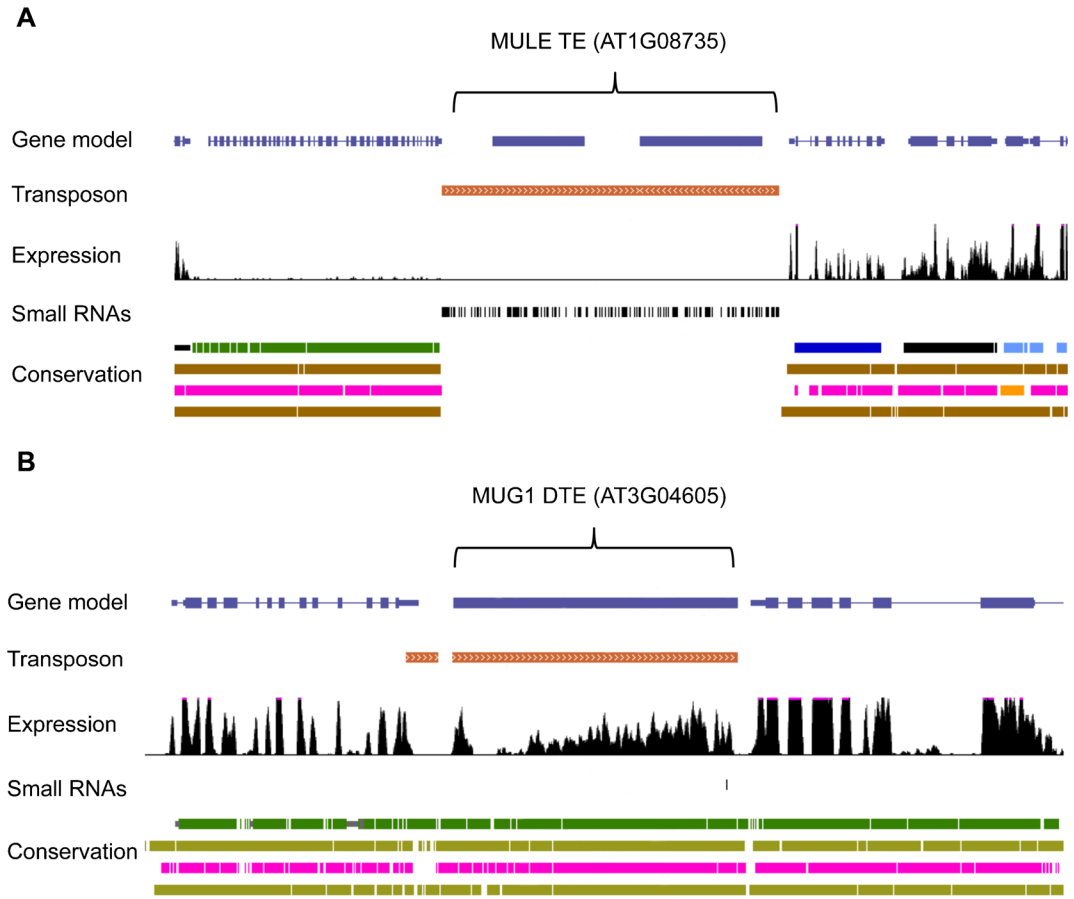

Fig. 2 Illustration of signature differences between transposable elements (*s*) and domesticated transposable elements (*DTEs*). The tracks are taken from the UCSC Genome Browser (http://mustang.biol.mcgill.ca:8885/) (Haudry et al. 2013). In each example, the tracks represent, from top to bottom: (1) Gene model (Gene Prediction track from NCBI), (2) Transposon (Transposon coordinates from TAIR; http://www.arabidopsis.org/), (3) Expression (Sequence Read Archive NCBI), (4) Small RNAs (TAIR sncRNAs), and (5) Conservation in Brassicaceae (Orthologous tracks for *Aradopsis lyrata*, *Capsella rubella*, *Sisymbrium irio*, and *Laevenworthia alabamica* where *colors* indicate chromosome number in the browser. **a** The MULE is detected by the transposon track (*orange rectangle*), has no gene expression because it is epigenetically silenced by the presence of several small RNAs (*small black lines*), and finally is likely active and not conserved as it is only present at this genomic location in *A. thaliana* and not in the other Brassicaceae as shown by the gap in the genomic chains. **b** *MUG1* DTE has an imprecise gene model (known to have two exons) and is detected by the transposon track despite being a confirmed DTE, which is an example of misannotation, is expressed and not epigenetically silenced (*absence of black lines*), and finally (5) is conserved in the Brassicaceae species and has maintained synteny with its neighboring genes, as shown by the absence of gap in the tracks

The importance of TE annotation

TE annotation challenges

Advances in DNA extraction techniques combined with next-generation sequencing technologies continue to make genome sequencing ever faster and more affordable. Despite and indeed because of this progress, genome annotation remains a great challenge. This is particularly true for TE annotation, where determining TE content, location, abundance, and classification are persistent difficulties. Although detecting high copy number and previously characterized families is relatively easy, detecting new and low copy-number families remain a daunting challenge. Also, immobilized TE fossils accumulated point mutations, indels, and rearrangements, eventually making it difficult to precisely determine their boundaries or indeed to find them at all. This is exacerbated by the propensity of TEs to insert into other TE copies to form nested configurations, as well as into host genes (Lerat 2010).

Programs to detect TEs

Saha et al. (2008) and Lerat (2010) have compared the most commonly used tools available to detect and annotate TEs in sequenced genomes. Here, we emphasize those specifically used with *Arabidopsis*. The most well-known and popular tool is RepeatMasker (http://www.repeatmasker.org/), which detects TEs similar to sequences in a consensus library. Originally designed to mask repeats in sequences in order to facilitate gene assembly and annotation, RepeatMasker was commonly the only analysis performed in early genome-sequencing projects. Although RepeatMasker can detect low copy-number families, it cannot detect previously unknown TEs or even TEs that are sufficiently diverged from known consensus sequences. Because of these key limitations, purely homology-based approaches using programs like RepeatMaster and similar ones like Greedier should serve only as one of several aspects of TE annotation (Li et al. 2008).

To overcome the limitation of using libraries of known repeats, various types of ab initio repeat identification tools have been developed. The most apparent characteristic of many TE families is repetitiveness. RECON (http://selab.janelia.org/recon.html) and RepeatScout (http://repeatscout.bioprojects.org/) use all-vs-all similarity searches to identify sequences with more than a specified number of copies and construct consensus sequences. This approach does not distinguish TE from non-TE, so non-TE sequences, such as high copy-number gene families, must be filtered from the results. Finally, occurrences of the consensus sequences are identified in the assembled genome sequence using the same approach as with libraries of known TE sequences. In addition to being unable to distinguish between TEs and other high copy-number sequences, repetitiveness-based approaches also have the limitation that they cannot detect low copy-number families. Another type of ab initio approach is signature-based, which identifies sequences with specified patterns of structures and motifs that are characteristic of a given type of TE. This approach can be used to find new elements but not new classes of elements since basic knowledge of class characteristics is not always available. Different programs have been developed and used in *Arabidopsis* to detect certain types of TEs, such as LTR retrotransposons (e.g., LTR_ FINDER http://tlife.fudan.edu.cn/ltr_finder/) and non-autonomous DNA MITEs (e.g., TRANSPO, http://alggen.lsi.upc.es/recerca/search/transpo/transpo.html). A drawback is that some of these programs have not been previously tested on real sequences and therefore may be unusable on biological data.

Although many tools are available, there is considerable variation between them and a need for improvement at several levels. Many programs are difficult to use and require substantial expertise to interpret and compare results. Indeed, most major genome projects use the same basic approach for TE annotation, as detailed above: create a library by combining known TEs (e.g., RepBase http://www.girinst.org/repbase/) with ab initio sequences discovered based on repetitiveness, as well as motif- or structure-based programs, then scan the genome for similarity. However, this standard approach disregards key structural evidence (e.g., terminal repeats and target site duplications), comparative evidence (e.g., interspecies alignments and genome re-sequencing), and functional evidence (e.g., small RNAs and methylation). The approach also provides no way to assess the confidence of its predictions, which is critical for certain applications. For instance, studies of TE-derived regulatory elements ought to focus on high-confidence TEs, whereas some studies of genome evolution may want to include low-confidence TEs. To address these limitations, research is underway to develop a system to integrate the predictions of various existing TE discovery programs with diverse genomic evidence using a probabilistic framework similar to state-of-the-art gene prediction procedures to assign probabilities to its predictions (D. Hoen, G Hickey, T. Bureau and M. Blanchette, *pers comm*.). If successful, such a platform could greatly improve the quality of TE annotation in many existing and newly sequenced genomes. As we will see in the next section, the growing interests in TEs and the need to properly identify and annotate them comes from numerous studies that have reported various effects of TEs on the genomes of all kingdoms.

Effects of TEs on plant genomes

TEs are most immediately under self-replicative selection, but their effects on the genome are subject to phenotypic natural selection and can therefore be selectively retained or removed. Here, we present a few examples of the effects of TEs on plant genomes.

Genome structure

Genome organization

Allopolyploidization (i.e., the hybridization of divergent genomes) has been coupled with rapid structural and functional alterations in genomes, most drastically in the repetitive component (Leitch and Leitch 2008), where it has been thought to induce bursts of transposition. No evidence for transposition was reported in synthetic allopolyploids resulting from a cross between *A. thaliana* and *A. lyrata* subsp. *patraea* as well as in the recent allopolyploid *Arabidopsis suecica* (Parisod et al. 2010). However, it has been suggested that in polyploids of *A. thaliana* and *Arabidopsis arenosa*, there are transposition events involving the DNA transposon *Sunfish* (Madlung et al. 2005). It appears that TE effects on genome organization after polyploidy are not necessarily transposition bursts, but rather substantial rearrangements in the wide epigenetic changes induced by the TE fraction (Parisod et al. 2010).

Genome size

Variation in genome size among closely related species has often been shown to be in large part due to differential accumulation of TEs (Agren and Wright 2011). For example, *A. thaliana* has a relatively low TE content compared to maize, which has more than 85 % TE content (Schnable et al. 2009). Important differences in genome size between *A. thaliana* and *A. lyrata* can be explained in part by major rearrangements, including three chromosome fusions, that led to a karyotype of five chromosomes for *A. thaliana* compared to eight for *A. lyrata* and other Brassicaceae and the loss of an entire *A. lyrata* centromere (chr4); but to a greater extent, by a reduction in TE activity and potentially more efficient TE elimination in *A. thaliana* (Hu et al. 2011). Other reports have also shown that genome size reductions in *A. thaliana* occur predominantly in selectively unconstrained, TE-rich regions (Haudry et al. 2013; de la Chaux et al. 2012).

Genome maintenance

Centromere and telomere maintenance

In most eukaryotic organisms, centromeres and telomeres contain large numbers of retrotransposons, raising the possibility that retrotransposons have acquired structural or functional roles in these heterochromatic regions. Indeed, *Drosophila melanogaster* has lost the telomerase gene common to most eukaryotes, and its telomeres are instead maintained by non-LTR retrotransposons that target the ends of its chromosomes (Pardue and DeBaryshe 2011). In plants, it remains to be determined whether retrotransposon sequences have direct roles in centromere or telomere function, but certain superfamilies of retrotransposons are highly enriched in telomeres and in centromeric or pericentromeric regions. For example, *copia*- and *gypsy*-like LTR retrotransposons and LINEs are enriched in the pericentromeric regions of *Arabidopsis* (Kumar and Bennetzen 1999; Miller et al. 1998).

Generation of variation

Chromosomal rearrangement

DNA transposons have been implicated in the induction of large-scale chromosomal rearrangements in plants and animals, mainly because their transposition mechanisms involve multiple double-strand breaks and repair events (Feschotte and Pritham 2007).

Somatic variation

Somatic variation was the first activity of TEs to be reported, when Barbara McClintock discovered that *Ac/Spm* activity in somatic tissues could lead to variegated corn color phenotypes (McClintock 1950). Additional somatic effects of TEs have since been reported in plants, such as variation in branching patterns and fruit cluster size in the grapevine cultivar Carnigan caused by the transposition of a DNA TE (*Hatvine1-rrm*) into the promoter region of the gene *VvTFL1A* (Levin and Moran 2011).

Allelic recombination

Polyploidy events are unlikely to account for all the structural variation that exists among plant genomes. For example, the most recent polyploidy event in the *Arabidopsis* lineage occurred 25–100 MYA, too long ago to contribute to differences in genome size and chromosome number between *A. thaliana* and *A. lyrata*. Therefore, recombination is thought to be a strong driving aspect in plant genome variability (Gaut

et al. 2007). This is based on observations that recombination is mutagenic and that plant genomes seem to be organized along recombinatorial gradients. For example, gene density and recombination, which are associated with euchromatic regions, are positively correlated in rice and *A. thaliana*. Interestingly, there seems to be a selective pressure to limit large TEs to gene-poor, low-recombination regions, whereas the TEs in gene-rich, high-recombination regions are predominantly small (Gaut et al. 2007).

Evolutionary innovation

Gene duplication and rearrangement

Transposons have the potential to mediate gene duplication, as observed in alignments of orthologous genomic sequences from *A. thaliana* with *Carica papaya* and *Vitis vinifera*, two outgroup species (Flagel and Wendel 2009). They found that about 11 % of *A. thaliana* genes have "transposed" either into or out of syntenic regions shared with its common ancestor, i.e., papaya, and many of these events are likely due to transposons. Transposon release has also been shown to coincide with polyploidy, potentially leading to an episodic expansion of transposon-mediated duplication shortly after polyploidy.

Epigenetic regulation

TE populations are silenced epigenetically by host mechanism like methylation and small RNAs; however, these silencing effects not only target TEs but also neighboring genes, especially if the TEs are inserted in gene-rich regions or near promoters. Since it is not thought that plants undergo global epigenetic resetting in each generation as it is observed in metazoans (Lisch 2009; Castel and Martienssen 2013), it is more likely that the epigenetic effects will be propagated to the next generation and possibly cause short- or long-term silencing of host genes. Interestingly, a MULE TE named Hiun (Hi) has been recently shown in *Arabidopsis* to encode a protein different from a transposase that acts as an anti-silencing factor to mediate transposition (Fu et al. 2013). For additional details about Hi, see "Control of transposable elements in *Arabidopsis thaliana*" by Ito and Kakutani, also in this issue.

Response to stress

Stress conditions have also been associated with sometimes-large increases in TE activities in plants to induce transcription or to redirect TE integration to alternative target sites (Levin and Moran 2011). Whether silencing mechanisms are insufficient to prevent TE reactivation under these conditions or whether they are down-regulated under stress is not yet known, but TE activation under environmental challenge supports the hypothesis of Barbara McClintock who suggested that this effect may create the genetic diversity needed to overcome threats to host survival (McClintock 1984).

Novel regulatory networks

TE movement and accumulation may participate in rewiring existing regulatory networks and creating new networks. TE insertions can disrupt and effectively eliminate *cis*-regulatory elements and thus remove genes from existing networks. TEs can also modify an existing gene network through TE-mediated shuffling and duplication of *cis*-regulatory elements. The preferential insertion and accumulation of TEs, especially certain DNA transposons, into the upstream region of genes might further enhance the reorganization of *cis*-regulatory elements. In addition, DNA transposon excisions are often imprecise, leaving small stretches of sequences or rearranging flanking host sequences, which can induce regulatory mutations resulting in phenotype variations. One example of TEs playing a crucial role in regulatory networks can be observed in mammalian genomes, where hundreds to thousands of binding sites for key mammalian transcription factors are found in distinctive repeat families (Bourque et al. 2008). In addition, mammalian genome comparisons have revealed the occurrence of numerous conserved non-coding elements (CNEs) in the human genome, many of which appear to be derived from ancient TEs (Cordaux and Batzer 2009; Xie et al. 2006). A recent study looking for CNEs conserved in nine sequenced Brassicaceae species located more than 90,000 CNEs under selection, providing evidence of transcriptional and post-transcriptional regulation (Haudry et al. 2013). Based on the results in the human genome, it is conceivable that TEs could also contribute to CNEs in plants; however, no evident cases were found among

these 90,000 CNEs (M. Blanchette, personal communication).

Transduplication

TEs can acquire non-TE sequences, including duplicated fragments of genes. In retrotransposons, a phenomenon known as transduction involves read-through transcription from a retrotransposon promoter into adjacent host gene sequences and subsequent incorporation of the host gene into the transposon sequence during reverse transcription (Bureau et al. 1994; Kazazian 2004). The few reported cases of TE with transduced host cellular sequences in plants are the LTR-retrotransposon *Bs1* in maize that contains a transduced ATPase domain (Elrouby and Bureau 2010) and *Katydid-At1*, an element of the TRIM family (terminal-repeat retrotransposons in miniature) of non-autonomous LTR retrotransposons that contains an ORF corresponding to a putative non-sense-mediated mRNA decay *trans*-acting factor (Witte et al. 2001). In DNA transposons, the insertion of gene fragments is referred to as transduplication; and although the mechanism remains unknown, the presence of intronic sequences in transduplicated host fragments argues that it occurs differently than transduction. In plants, transduplication is most common in MULEs and *Helitrons* in maize, rice and *Arabidopsis* and CACTA in the Japanese morning glory (Dooner and Weil 2007). Although an intriguing possibility is that transduplicated host genes could have the potential to create novel protein-coding genes (Jiang et al. 2004), it appears that the vast majority are actually truncated and pseudogenic (Hoen et al. 2006; Kawasaki and Nitasaka 2004) or under epigenetic suppression (Kapitonov and Jurka 2007; Yang and Bennetzen 2009). One exception was found in *Arabidopsis*, a family of non-TIR MULEs called *KAONASHI* (KI) (Hoen et al. 2006). KI-MULEs contain a second ORF in addition to transposase with a conserved sequence similar to the peptidase C48 domain usually found in ubiquitin-like protein-specific proteases (ULP). KI-MULEs are targeted by small RNAs and have the capacity to mobilize, with generally intact termini necessary, and at least one KI-MULE is mobile in mutants deficient in histone and DNA. This suggests that KI could be conserved for a selfish function rather than beneficial to the host. Recently, the DTE gene family *MUSTANG* in *A. thaliana* (see the section entitled "Exaptation (molecular domestication)") was shown to contain the host domain Phox and Bem1p (PB1) known to be involved in protein signaling (Joly-Lopez et al. 2012). It is not clear whether this domain was acquired through transduplication while the *MUSTANG* TE ancestor was still active or after domestication.

Speciation

Many of the aforementioned effects of TEs, such as TE abundance, TE-derived genomic features, and chromosomal rearrangements involving TE sequences, are lineage-specific. This suggests that TEs are related to the process of speciation, either as a cause or as an effect. In the following section, we examine the phenomenon of exaptation, in which TEs create new genes and produce innovations that can also contribute to speciation.

Exaptation (molecular domestication)

The term exaptation was coined to refer to adaptations that radically alter the function of a feature from that for which it was originally selected (Gould and Lloyd 1999). In the context of this section, we refer to TE exaptation, or molecular domestication (Miller et al. 1992), as the direct exaptation of TE genes and binding sites that will contribute directly to the phenotypic function of the host organism and therefore come under phenotypic rather than self-replicative selection (Hoen and Bureau 2012). Although the exact process of exaptation is not known, certain events need to take place to lead to a different evolutionary signature between TEs and domesticated transposable elements (DTEs). For example, TEs must be immobilized by mutations in the terminal repeats or in the transposition protein (transposase or reverse transcriptase) (Fig. 2). An immobilized TE that becomes a DTE will have fewer copies and, if conserved, will be syntenic between different species. Also, because they produce beneficial phenotypes, domesticated genes will tend not to be targeted by small RNAs and to be expressed rather than silenced (Fig. 2). For more details on the exaptation process, we recommend the following reviews: (Hoen and Bureau 2012; Feschotte and Pritham 2007; Sinzelle et al. 2009).

Because of their potential to generate major evolutionary innovations, interest over DTEs has grown rapidly in recent years. Known innovations include the

RAG genes and their binding sites in the vertebrate adaptive immune system (Fugmann 2010; Agrawal et al. 1998), SETMAR in the mammalian placenta (Kaneko-Ishino and Ishino 2012; Ono et al. 2006), and telomeres (Pardue and DeBaryshe 2011). In *Arabidopsis*, three families of DTEs have been identified.

The *FHY3* family

FAR1 (FAR-RED IMPAIRED RESPONSE 1) was the first DTE discovered in plants although its finding was made fortuitously. During forward screens looking for *A. thaliana* mutants that undergo normal de-etiolation under white light but fail to do so under far-red light, *fhy3* (*far-red elongated hypocotyl 3*) (Whitelam et al. 1993) and *far1* (Hudson et al. 1999) were identified and shown to disrupt the high irradiance phyA response. The connection to TEs was then made in a search for *Mutator* transposases in grasses, where the authors surprisingly found that *FAR1* had sequence similarity with *Mutator* transposase genes, but yet lacked the TE termini, suggesting that *FAR1* had been co-opted to perform a host function (Lisch et al. 2001). *FAR1* and *FHY3* are paralogs that arose from a duplication event subsequent to molecular domestication and diversified which explains why they have partially overlapping functions. Functional dissection of their functional domains revealed that the three domains of the ancestral MULE have been co-opted, where an N-terminal C2H2 zinc finger domain mediates DNA binding, and a central MULE domain and an N-terminal SWIM domain are both required for homo- and heterodimerization and to activate expression (Lin et al. 2007, 2008).

In addition to de-etiolation, various studies have reported additional roles for *FAR1* and *FHY3* in *Arabidopsis* such as interaction with the transcription factor HY5 (ELONGATED HYPOCOTYL 5) involved in promoting photomorphogenesis, in order to repress *FHY1* and *FHL* transcription, as well as interaction with central oscillator transcription factors to generate rhythmic expression of *ELF4* (Li et al. 2011). *FHY3* seems to have a more pleiotropic effect as shown in a chromatin immunoprecipitation-based sequencing (ChIP-Seq) analysis to identify FHY3 binding sites in the *Arabidopsis* genome, which revealed that FHY3 directly recognizes binding sites associated with more than 1,700 genes (Ouyang et al. 2011).

Taken together, these findings suggest that FHY3/FAR1 act as transcription factors and play various roles in *Arabidopsis*, some of which have yet to be uncovered. FHY3/FAR1 are members of a family (the *FHY3* family), which also includes 12 additional genes in *A. thaliana*, referred to as *FAR1-related sequences* (FRSs) (Lin and Wang 2004). The *FHY3* family consists of five phylogenetic clades, all with at least two members in *A. thaliana* and all but one with members in both eudicots and monocots, suggesting that they were domesticated prior to the monocot-dicot split. Interestingly, the *FHY3* phylogram published by Lin et al. 2007 includes two internal branches containing active TEs, which suggests that the family may have been formed by at least three independent domestication events. If independent domestication of related TEs did in fact produce host genes with similar functions, it raises important questions regarding the nature of exaptation and its role in genome evolution.

The *SLEEPER* family

Like FHY3/FAR1, *DAYSLEEPER* was discovered fortuitously during a yeast one-hybrid screen for proteins binding the upstream of a specific DNA repair gene (*Ku70*), in which *DAYSLEEPER* was found to bind a Kubox1 motif present upstream of several other developmental genes (Bundock and Hooykaas 2005). *DAYSLEEPER* shares sequence similarities with a *hAT* DNA transposon, where it retained its dimerization domain but not its termini or key residues necessary for transposition. *DAYSLEEPER* is the only plant DTE identified so far that is essential to development. *Arabidopsis* homozygous *daysleeper* knockout mutants exhibit severe phenotypes, such as failure to expand cotyledon growth and develop normal leaves or floral organs (Bundock and Hooykaas 2005). Also, *Arabidopsis* lines overexpressing DAYSLEEPER have multiple abnormalities, including slow growth, altered flower morphology, and sterility. Transcriptome analysis of overexpressed seedlings has shown that suites of genes have altered expression, although none contain the Kubox1 DNA motif to which *DAYSLEEEPER* binds. Recently, genes homologous to *Arabidopsis DAYSLEEPER* have been identified in angiosperms and together they form the *SLEEPER* family (Knip et al. 2012).

The *MUSTANG* family

The *MUSTANG* (*MUG*) DTE family was discovered not through systematic screens, but through an in silico search for signatures of DTE sequences derived from MULE transposons (Fig. 2; (Cowan et al. 2005)). The family consists of eight genes in *A. thaliana* divided into two subfamilies that differ both phylogenetically and structurally: *MUGA* (genes *MUG1-4*) and *MUGB* (genes *MUG5-8*). Most clades in the *MUGA* subfamily include genes orthologs in both monocots and dicots, whereas *MUGB* has separate clades for monocots and dicots, indicating that the *MUGA* family diversified prior to the monocot-eudicot split, whereas *MUGB* diversified after the split. Moreover, all *MUG* genes contain the three conserved domains of the ancestral TE, namely an N-terminal MuDR DNA-binding domain, a core MULE transposase domain, and a C-terminal SWIM zinc finger domain that can either interact with DNA or proteins. However, *MUGB* genes have an additional domain, PB1, a host domain not usually found in TEs (Joly-Lopez et al. 2012). Certain *MUG* homozygous mutants have pleiotropic phenotypes throughout the lifespan of *A. thaliana*, from delayed development and delayed flowering to aberrant flower morphology and reduced seed set (Joly-Lopez et al. 2012). *mug* mutants are also responsive to a number of stress conditions and appear to cause changes in the expression of a large number of genes (Z. Joly-Lopez, E. Forczek, unpublished data). Subsequent work on this family could determine whether the *MUSTANG* family functions in transcription regulation like the other MULE-derived family, *FHY3*.

Identification of novel DTEs in *A. thaliana*

Until recently, these three were the only families of DTEs known in *Arabidopsis*. However, a new study using a comprehensive approach similar to but more general than that used to identify the *MUG* family, which searched for DTEs derived not just from MULEs but any TE superfamily and which included more lines of evidence (Hoen and Bureau 2014). This search represents the first report to measure the extent of TE exaptation at a genome-wide scale in non-mammalian lineages. The search identified 39 novel DTEs in 24 families. In addition, six genes were identified that had been previously reported in the literature but described as DTEs, illustrating why a more comprehensive understanding of TEs and DTEs is important. Of the 67 total DTEs now known in *A. thaliana*, five times more have originated from DNA transposons than retrotransposons, among which most were derived from *MULEs* (like *FHY3* and *MUG*) and many from *hAT* (like *DAYSLEEPER*).

Functions of exaptation

In plants, the DTEs' function that stands out is transcription regulation, with all characterized DTEs derived from DNA transposons. This propensity to function as transcription factors (TFs) is not necessarily surprising for DNA transposons, since they have innate characteristics that predispose them to exaptation as de novo regulatory networks (Hoen and Bureau 2012). For example, in all cases of DTEs derived from DNA transposons, the entire transposase or at least its DNA-binding domain has been co-opted, which suggests a role in recognizing motifs and binding DNA. TFs are central to the regulation of gene expression; and in plants, they are involved in all phases of development, with a particularly high number in defense and stress responses. Recently, it has been suggested that most of the DBDs of plant-specific TFs have originated from endonucleases associated with TEs in unicellular eukaryotes and have then experienced plant-specific expansion by acquiring new functions (Yamasaki et al. 2013). Since several DTEs are already known to function as TFs, we could hypothesize that the novel DTEs may also be involved, at in least in part, in transcription regulation, and if it were to be the case, it would be an important contribution of about 50 new TFs. If only a handful of these genes have similar effects than the other known DTEs, we may have been missing key components of the big picture, which should receive considerable attention in the future.

We have mentioned that a significantly higher proportion of DTEs are derived from DNA transposons than retrotransposons in *Arabidopsis*. To explain this gap, it is possible that DNA transposons may be more amenable sequences to be domesticated. They may also be producing beneficial phenotypes that have been preferentially selected or favored functions during the context of the domestication events. For example, the *FHY3* and *MUSTANG* families are likely to have been domesticated at the onset of angiosperms (Lin et al. 2007; Joly-Lopez et al. 2012), which was an evolutionary climate of profound change where novel transcription factors

linking previously unconnected pathways may have been highly beneficial. Retrotransposons have likely been domesticated to fulfill other functions. Plant DTEs derived from retrotransposons have retained either the Retrotrans_gag domain or the RnaseH domain from the ancestral LTR retrotransposons. *gag* encodes proteins involved in the maturation and packaging of retrotransposon RNA and proteins into a form suitable for integration in the genome whereas Rnase H is an enzyme with endonuclease activity that catalyzes the cleavage of RNA and is required for the replication/transposition of the retrotransposons (Kumar and Bennetzen 1999). The function of these DTEs is yet to be uncovered; however in humans, *gag* has been domesticated repeatedly (and convergently) in the placenta to play roles in fusing layers of epidermis and other tissues (Cornelis et al. 2012) It could be that such DTEs may have a similar function, like in fusing layers of tissues in plants or in vesicle trafficking to membranes.

Concluding remarks

TEs have come a long way since being described as genomic parasites, selfish, and junk DNA. The complex relations between plant transposons and their hosts have contributed to plant genomes in a number of ways, from genome/gene reorganization to gene expression and gene co-option. Even in a genome as well studied as *Arabidopsis*, certain aspects remain elusive such as the full extent to which key host genes are derived from TEs. Further work in TE annotation and characterization of novel DTEs is important in elucidating the roles that TEs play in plant genome evolution.

References

Agrawal A, Eastman QM, Schatz DG (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394:744–751

Agren JA, Wright SI (2011) Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? Chromosom Res 19:777–786

Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815

Bennett MDAIJL (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Ann Bot 107:467–590

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18:1752–1762

Bundock P, Hooykaas P (2005) An *Arabidopsis* hAT-like transposase is essential for plant development. Nature 436: 282–284

Bureau TE, White SE, Wessler SR (1994) Transduction of a cellular gene by a plant retroelement. Cell 77:479–480

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956–963

Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet 14:100–112

Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691–703

Cornelis G, Heidmann O, Bernard-Stoecklin S, Reynaud K, Veron G, Mulot B, Dupressoir A, Heidmann T (2012) Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. Proc Natl Acad Sci U S A 109:E432–E441

Cowan RK, Hoen DR, Schoen DJ, Bureau TE (2005) MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Mol Biol Evol 22:2084–2089

de la Chaux N, Tsuchimatsu T, Shimizu KK, Wagner A (2012) The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. Mob DNA 3:2

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Dooner HK, Weil CF (2007) Give-and-take: interactions between DNA transposons and their host plant genomes. Curr Opin Genet Dev 17:486–492

El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5:954–965

Elrouby N, Bureau TE (2010) Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. Plant Physiol 153:1413–1424

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183:557–564

Fu Y, Kawabe A, Etcheverry M, Ito T, Toyoda A, Fujiyama A, Colot V, Tarutani Y, Kakutani T (2013) Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. EMBO J 32:2407–2417

Fugmann SD (2010) The origins of the Rag genes–from transposition to V(D)J recombination. Semin Immunol 22:10–16

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007) Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet 8:77–84

Gould SJ, Lloyd EA (1999) Individuality and adaptation across levels of selection: how shall we name and generalize the unit of Darwinism? Proc Natl Acad Sci U S A 96:11904–11909

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM,

Dewar K, Stinchcombe JR, Schoen DJ, Wang X, Schmutz J, Town CD, Edger PP, Pires JC, Schumaker KS, Jarvis DE, Mandakova T, Lysak MA, van den Bergh E, Schranz ME, Harrison PM, Moses AM, Bureau TE, Wright SI, Blanchette M (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet 45:891–898

Hoen DR, Bureau T (2012) Transposable element exaptation in plants. In: Grandbastien MA, Casacuberta JM (eds) Plant transposable elements. Springer Berlin Heidelberg, Heidelberg

Hoen DR, Bureau TE (2014) Discovery of novel genes derived from transposable elements using integrative genomic analysis. In preparation

Hoen DR, Park KC, Elrouby N, Yu Z, Mohabir N, Cowan RK, Bureau TE (2006) Transposon-mediated expansion and diversification of a family of ULP-like genes. Mol Biol Evol 23:1254–1268

Hollister JD, Gaut BS (2007) Population and evolutionary dynamics of Helitron transposable elements in *Arabidopsis thaliana*. Mol Biol Evol 24:2515–2524

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KF, van de Peer Y, Grigoriev IV, Nordborg M, Weigel D, Guo YL (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481

Hudson M, Ringli C, Boylan MT, Quail PH (1999) The FAR1 locus encodes a novel nuclear protein specific to phytochrome A signaling. Genes Dev 13:2017–2027

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Joly-Lopez Z, Forczek E, Hoen DR, Juretic N, Bureau TE (2012) A gene family derived from transposable elements during early angiosperm evolution has reproductive fitness benefits in *Arabidopsis thaliana*. PLoS Genet 8: e1002931

Kaneko-Ishino T, Ishino F (2012) The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. Front Microbiol 3:262

Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A 98:8714–8719

Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23:521–529

Kawasaki S, 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

Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632

Knip M, de Pater S, Hooykaas PJ (2012) The SLEEPER genes: a transposase-derived angiosperm-specific gene family. BMC Plant Biol 12:192

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science 320:481–483

Lerat E (2010) Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity (Edinb) 104:520–533

Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615–627

Li X, Kahveci T, Settles AM (2008) A novel genome-scale repeat finder geared towards transposons. Bioinformatics 24:468–476

Li G, Siddiqui H, Teng Y, Lin R, Wan XY, Li J, Lau OS, Ouyang X, Dai M, Wan J, Devlin PF, Deng XW, Wang H (2011) Coordinated transcriptional regulation underlying the circadian clock in *Arabidopsis*. Nat Cell Biol 13:616–622

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

Lin R, Teng Y, Park HJ, Ding L, Black C, Fang P, Wang H (2008) Discrete and essential roles of the multiple domains of *Arabidopsis* FHY3 in mediating phytochrome A signal transduction. Plant Physiol 148:981–992

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 60:43–66

Lisch DR, Freeling M, Langham RJ, Choy MY (2001) Mutator transposase is widespread in the grasses. Plant Physiol 125: 1293–1303

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L (2005) Genomic changes in synthetic *Arabidopsis* polyploids. Plant J 41:221–230

Mcclintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A 36:344–355

Mcclintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome 6:0

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics 150:1615–1623

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 U S A 89(9): 4018–4022

Ono R, Nakamura K, Inoue K, Naruse M, Usami T, Wakisaka-Saito N, Hino T, Suzuki-Migishima R, Ogonuki N, Miki H, Kohda T, Ogura A, Yokoyama M, Kaneko-Ishino T, Ishino F (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

Ouyang X, Li J, Li G, Li B, Chen B, Shen H, Huang X, Mo X, Wan X, Lin R, Li S, Wang H, Deng XW (2011) Genome-wide binding site analysis of FAR-RED ELONGATED HYPOCOTYL3 reveals its novel function in *Arabidopsis* development. Plant Cell 23:2514–2535

Pardue ML, Debaryshe PG (2011) Retrotransposons that maintain chromosome ends. Proc Natl Acad Sci U S A 108:20317–20324

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186:37–45

Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biol 5:R79

Raizada M, Benito M, Walbot V (2001) The MuDR transposon terminal inverted repeat contains a complex plant promoter directing distinct somatic and germinal programs. Plant J 25: 79–91

Rudenko GN, Walbot V (2001) Expression and post-transcriptional regulation of maize transposable element MuDR and its derivatives. Plant Cell 13:553–570

Saha S, Bridges S, Magbanua ZV, Peterson DG (2008) Computational approaches and tools used in identification of dispersed repetitive DNA sequences. Trop Plant Biol 1:85–96

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, Mcmahan L, van Buren P, Vaughn MW et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Sinzelle L, Izsvak Z, Ivics Z (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci 66:1073–1093

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Whitelam GC, Johnson E, Peng J, Carol P, Anderson ML, Cowl JS, Harberd NP (1993) Phytochrome A null mutants of *Arabidopsis* display a wild-type phenotype in white light. Plant Cell 5:757–768

Wicker T (2012) So many repeats and so little time: how to classify transposable elements. In: Grandbastien MA, Casacuberta JM (eds) Plant transposable elements. Springer Berlin Heidelberg, Heidelberg

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, Sanmiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Witte CP, Le QH, Bureau T, 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

Xie X, Kamal M, Lander ES (2006) A family of conserved noncoding elements derived from an ancient transposable element. Proc Natl Acad Sci U S A 103:11659–11664

Yamasaki K, Kigawa T, Seki M, Shinozaki K, Yokoyama S (2013) DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. Trends Plant Sci 18:267–276

Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc Natl Acad Sci U S A 106:19922–19927

Ziolkowski PA, Koczyk G, Galganski L, Sadowski J (2009) Genome sequence comparison of Col and Ler lines reveals the dynamic nature of *Arabidopsis* chromosomes. Nucleic Acids Res 37:3189–3201