Biochimica et Biophysica Acta 1859 (2016) 366–380

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

What makes up plant genomes: The vanishing line between transposable elements and genes

Dongyan Zhao, Ann A. Ferguson, Ning Jiang *

Department of Horticulture, Michigan State University, 1066 Bogue Street, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 17 August 2015
Received in revised form 9 December 2015
Accepted 11 December 2015
Available online 17 December 2015

Keywords:
Transposon
Retrotransposon
Gene expression
Gene duplication
Domestication

ABSTRACT

The ultimate source of evolution is mutation. As the largest component in plant genomes, transposable elements (TEs) create numerous types of mutations that cannot be mimicked by other genetic mechanisms. When TEs insert into genomic sequences, they influence the expression of nearby genes as well as genes unlinked to the insertion. TEs can duplicate, mobilize, and recombine normal genes or gene fragments, with the potential to generate new genes or modify the structure of existing genes. TEs also donate their transposase coding regions for cellular functions in a process called TE domestication. Despite the host defense against TE activity, a subset of TEs survived and thrived through discreet selection of transposition activity, target site, element size, and the internal sequence. Finally, TEs have established strategies to reduce the efficacy of host defense system by increasing the cost of silencing TEs. This review discusses the recent progress in the area of plant TEs with a focus on the interaction between TEs and genes.

1. Introduction

Transposable elements (TEs) were first discovered by Barbara McClintock in the 1940s using maize (*Zea mays*) as a model system. Maize, a member of the grass (Poaceae) family, is an excellent model for genetic studies. Its large flowers with separate male (the tassels) and female (the ears) reproductive organs allow easy genetic crosses. In addition, the large size of maize chromosomes makes them readily visible under a light microscope. In a normal maize plant, chromosome breakage is rarely observed. Nevertheless, McClintock noticed very frequent breakage on a particular locus of chromosome 9 in one special maize line [1]. Subsequently, she discovered that two loci were required for the breakage. One locus found at the site of breakage and was called *Ds* (*Dissociation*). The other, which was required to "activate" the breakage, was therefore called *Ac* (*Activator*). Since the location of *Ac* and *Ds* appears to be variable between generations, McClintock proposed that they were actually genetic elements capable of transposition [1]. Thus, the first TEs were discovered through a forward genetics approach.

In 1983, more than 30 years later, McClintock won a Nobel Prize for her discovery of TEs. This coincided with the cloning and sequencing of the *Ac* and *Ds* elements by multiple research groups [2,3,4,5]. It was discovered that *Ac* encodes a transposase protein responsible for the transposition of itself as well as *Ds*. Meanwhile, more TEs were identified from maize and other organisms including other plants, animals, fungi, algae, and bacteria. These TEs form distinct superfamilies, families, and subfamilies (defined below)—*Ac*/*Ds* represent only one of numerous TE families. Moreover, not all TEs at present are capable of transposition; the ones with current transposition activity are called active TEs. When an active TE creates a new copy, the two copies are often identical at the time of transposition. Over time, the two copies diverge and become less similar.

In 2002, approximately 20 years after the cloning of *Ac*/*Ds*, the entire genome of rice (*Oryza sativa*), a relative of maize, was sequenced [6,7,8]. Rice was the first crop genome sequenced and the quality of its sequence remains the best among all crops. In 2003, the first active DNA element in rice, called *mPing*, was discovered independently by three research groups through two distinct approaches [9,10,11]. One group identified *mPing* through a mutation called *slender glume*, caused by the insertion of this element [11], resembling the classical pathway for TE discovery. A reverse genetics approach was employed by the two other groups [9,10]. With this method, candidate active elements were identified through computational search of available genomic sequences. The rationale for this search is that if an element is active, or was active in the recent past, identical or highly similar copies should be identified in the genome. Among the 52 copies of *mPing* in Nipponbare, the sequenced rice cultivar, 40 are identical to each other, strongly suggesting a recent activity. Since *mPing* appeared to be a good candidate for an active TE, its transposition activity was tested in tissue culture, where excisions and new insertions of *mPing* were observed [9,10]. The advances of technology, especially the availability of genomic and transcriptomic sequences, have revolutionized our approaches to study TEs and brought about a burst of information about TE biology. In this review, we will discuss some of the latest progress in plant TEs.

---

* Corresponding author.
E-mail address: jiangn@msu.edu (N. Jiang).

http://dx.doi.org/10.1016/j.bbagrm.2015.12.005
1874-9399/© 2015 Elsevier B.V. All rights reserved.

2. Classification of plant transposable elements

Based on the transposition mechanism, TEs fall into two classes. Class I, the retrotransposons, use a "copy and paste" mechanism and utilize an RNA intermediate for transposition [12]. Class II, the DNA transposons, transpose via a DNA intermediate through a "cut and paste" mechanism [13]. Based on their coding capacity, both classes of TEs can be divided into autonomous and non-autonomous elements. Autonomous elements (such as *Ac*) encode the protein products required for their transposition. Non-autonomous elements (such as *Ds*) do not encode the relevant products and rely on their cognate autonomous elements for transposition. When an element inserts into a genomic locus, a small piece of flanking sequence in the insertion site is duplicated and this sequence is called target site duplication (TSD).

For both RNA and DNA TEs, they are often further categorized into superfamilies. A superfamily is classified by elements sharing transposases with significant similarity at the protein level and having similar length of TSDs. A family of elements typically shares a sequence that is critical for transposition, such as the element ends. A subfamily is composed of members wherein the entire element sequence or at least the major portion of the elements is conserved (Fig. 1). However, during genome-wide annotation, a family is defined by a practical method using sequence similarity among different elements due to the difficulty in defining elements sharing transposition machinery or *cis*-sequences. A common practice is that if two elements share 80% or higher identity at the nucleotide level for over 80% of the length of the shorter element, the two elements are considered to belong to the same family [14]. It is helpful to follow this community standard whenever possible, but certainly, these parameters may be modified to suit the objectives of each individual study.

2.1. Class I elements (retrotransposons)

Class I elements can be further divided into several groups, including the long terminal repeat (LTR) elements, long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs) (Table 1). Due to the lack of LTRs, LINEs and SINEs are called non-LTR retrotransposons. Non-LTR retrotransposons are often associated with a poly-A tail at the 3′ end of the elements as a consequence of transcription. LTR elements are further classified into *Gypsy* superfamily and *Copia* superfamily, depending on the gene order in the internal region of the elements [12]. The internal regions of LTR elements encode *gag*, *pol*, and *int* genes that are synthesized as a polyprotein. The *gag* gene encodes structure proteins of virus-like particles which are responsible for packaging of retrotransposon RNA and proteins. The *pol* gene encodes reverse transcriptase and RNase H activities that are required for replication/transposition of the retrotransposon, and *int* encodes the integrase that allows the DNA form of the retrotransposon to insert at a new chromosomal location [15]. During the transposition of class I elements, the element mRNAs are converted into cDNA through the action of reverse transcriptase (from the *pol* gene) which is encoded within the element, and the TE cDNAs are then inserted into a target site in the genome. LTR elements are associated with a TSD of 5 bp whereas non-LTR retrotransposons generate TSDs with variable length and in some cases even deletion of the flanking sequence [16].

Because of their replicative transposition mechanism, class I elements can amplify very rapidly and constitute the largest portion of DNA content in most plant genomes. Particularly, LTR elements are often responsible for the expansion of plant genomes, which is different from that in mammals where non-LTR retrotransposons are very abundant [17]. For instance, 75% of the maize genome is occupied by LTR retrotransposons, including both *Gypsy* and *Copia* like elements [18]. In contrast, non-LTR retrotransposons are relatively scarce in plants, usually accounting for less than 5% of the genome with some exceptions, *e.g.*, apple (*Malus domestica*, 7.95%), sacred lotus (*Nelumbo nucifera*, 6.4%), sugar beet (*Beta vulgaris*, 5.67%), and banana (*Musa acuminata*, 5.41%) [19,20,21,22].

2.2. Class II elements (DNA transposons)

DNA transposons are usually associated with terminal inverted repeats (TIRs) and transpose via a DNA intermediate [13]. In general, DNA transposons excise from one site (donor site) and reinsert elsewhere (target site) in the genome. The departure of element from donor site is called "excision" whereas the integration of element into target site is called "insertion/reinsertion" or "forward transposition".

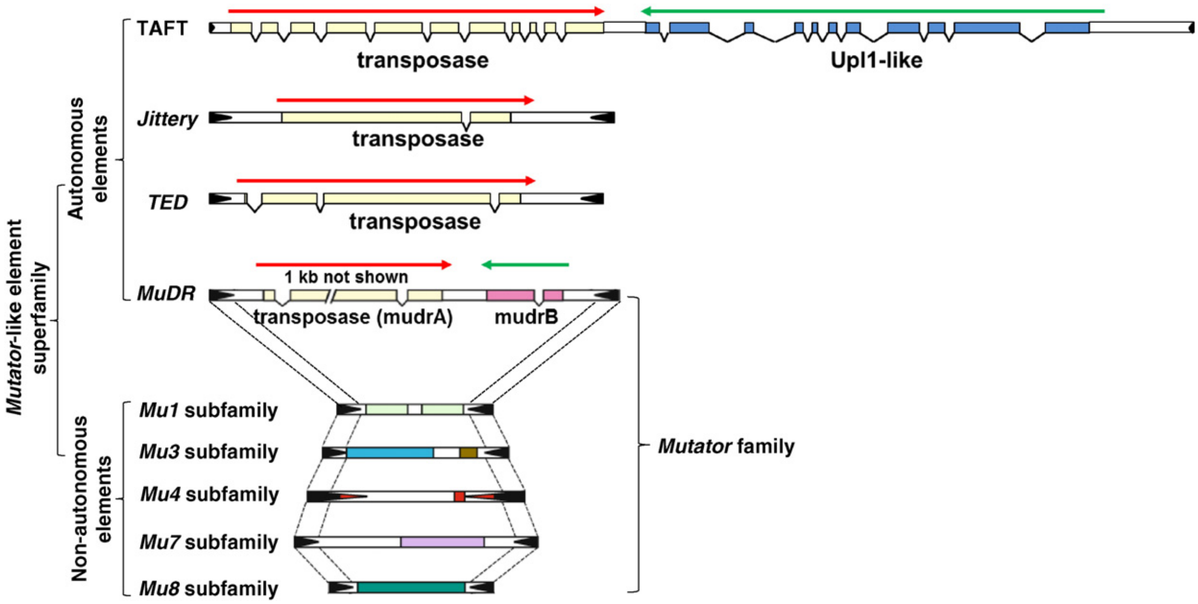

Fig. 1. Structural organization of a superfamily of transposable elements. *Mutator*-like superfamily (DNA transposons) in maize is used as an example. Terminal inverted repeats (TIRs) are depicted as black triangles. Colored boxes represent open reading frames in autonomous elements or acquired gene fragments in non-autonomous elements, and other sequences are depicted as white boxes. The homologous sequences between *MuDR* and its non-autonomous counterparts are connected with dashed lines. Introns are shown as lines connecting exons. In *Mu4*, the red triangles indicate that a portion of the acquired gene fragment was duplicated, inverted, and became part of the TIR. Colored arrows above the elements indicate the transcribed regions as well as the orientation of transcription.

Table 1
Structural features and classification of transposons in plants.

Class	Superfamily		TSD	Element size		TIR or LTR size	Terminal sequence (5′…3′)
	Traditional name	Systematic classification		Autonomous	Non-autonomous		
Class I	LTR-*Copia*	RLC	5 bp	4–11 kb	292 bp to 4 kb	85 bp to 3 kb	TG…C/G/TA
	LTR-*Gypsy*	RLG	5 bp	5–20 kb	2–5 kb	120 bp to 6 kb	TG…CA
	LINE	RIL/RII	Variable	Up to 9 kb	NA	NA	…AAAA
	SINE	RST/RSL	Variable	NA	80–500 bp	NA	…AAAA
Class II	*Ac*/*Ds*/*hAT*	DTA	8 bp	3–6 kb	110 bp–3 kb	5–22 bp	C/TA…TA/G
	En/*Spm*/*dSpm*/CACTA	DTC	3 bp	6–21 kb	200 bp–6 kb	12–28 bp	CACTA/G…C/TAGTG
	MuDR/*Mutator*/*Mu*/MULE	DTM	7–11 bp, mostly 9	4–16 kb	120 bp–3 kb	0–800 bp	G/C…G/C
	PIF/*Harbinger*/*Tourist*	DTH	TNA	3–7 kb	80 bp–3 kb	14–60 bp	GGG/CC…GG/CCC GG/AGCA…TGCT/CC
	Tc1/*Mariner*/*Stowaway*	DTT	TA	3–7 kb	80 bp–3 kb	11–120 bp	CTCCCTCC…GGAGGGAG
	Helitron	DHH	None	5–17 kb	150 bp–20 kb	None	TC…CTRR

NA: Not available.

For most DNA elements, excision is not necessarily accompanied by re-insertion in the genome. In plants, there are several superfamilies of DNA transposons that are briefly introduced below (Table 1).

2.2.1. *Ac/Ds/hAT or DTA elements*

The founder elements of this superfamily are *hobo* from fruit fly (*Drosophila melanogaster*), *Ac*/*Ds* from maize, and *Tam3* from snapdragon (*Antirrhinum majus*) [23,24,25]. The first letter of the three elements contributes to the name "*hAT*". "DTA" is the systematic nomenclature [14]. Autonomous *hAT* elements are a few kilobases (kb) in length whereas non-autonomous elements can be as short as less than 200 bp. Most *hAT* elements have short TIRs that are less than 30 bp, starting with "CA/TA" and ending with "TG/TA". The TSD of *hAT* elements is usually 8 bp in length. Since *Ac*/*Ds* were the first identified transposons, they are the best-characterized DNA transposons in plants and have been widely used in gene tagging and functional genomics studies [26,27,28].

2.2.2. *CACTA or DTC elements*

The founder elements of this superfamily are *Spm*/*dspm* in maize [29]. They are named CACTA elements because their terminal sequences are "CACTA/G…C/TAGTG". In general, autonomous CACTA elements are large in size. The *Tam1* element in snapdragon is 17 kb [30], and the *Tgmt* element in soybean (*Glycine max*) is over 20 kb [31]. Like *hAT* elements, CACTA elements are associated with short TIRs, and frequently contain many subterminal repeats. CACTA elements generate 3-bp TSD upon insertion. Unlike other DNA transposons, which are enriched in genic regions, CACTA elements do not have a significant preference for genic regions [32]. Due to the large size of CACTA elements [30,31,32], their amplification can result in significant genome mass. For example, CACTA elements represent 13% of the wheat genome, accounting for 87% of the DNA transposons in wheat.

2.2.3. *Mutator/MULE or DTM elements*

Mutator/*Mu*/*MuDR* elements in maize are the founder elements [33]. Similar elements in other organisms are called *Mutator*-like elements (MULEs). MULEs are the most complex superfamily of DNA transposons in plants in terms of sequence, structure, size, and transposition behavior. Most MULEs have extended long TIRs (100-800 bp) compared to other DNA transposons. Yet there are a group of MULEs called "non-TIR MULEs," which have short (usually <50 bp) TIR with low similarity between the two termini [34]. Some non-TIR MULEs, such as the TAFT element in maize, specifically target "TATATA…" satellite sequences [35]. In addition, some MULEs have two or more tandem TIRs at the termini of the elements [36]. Autonomous MULEs can be rather large in size. For example, the CUMULE element is 12 kb in length [37]. The terminal sequences of MULEs vary among elements but most of them start or end with G or C. Most MULEs create a 9-bp TSD upon insertion. Non-autonomous MULEs usually are not deletion derivatives of the autonomous elements; instead, they often carry non-TE genomic sequences including genes. In rice, about ¼ (2812) of the TIR MULEs carry gene fragments [38], which are referred to as "Pack-MULEs".

2.2.4. *PIF/Harbinger/Tourist or DTH elements*

For this superfamily, the non-autonomous elements (*Tourist*) were described nearly a decade prior to the discovery of their autonomous elements (*PIF*/*Harbinger*) [9,39,40,41,42]. Like CACTA elements, *PIF*/*Harbinger*/*Tourist* elements are associated with a 3-bp TSD. Nevertheless, the TSD of CACTA elements can be any combination of the four nucleotides, whereas that for *PIF*/*Harbinger*/*Tourist* elements is mainly "TAA" or "TTA", suggesting a more stringent target selection during transposition. Compared with other autonomous DNA elements, *PIF*/*Harbinger* elements are relatively small in size, usually a few kb in length. *Tourist* elements are usually less than 600 bp. The terminal sequences for *PIF*/*Harbinger*/*Tourist* elements are "GGG/CCC…GGC/GCC" or "GA/GGCA…TGCC/TC" [41].

2.2.5. *Tc1/Mariner/Stowaway or DTT elements*

Like *PIF*/*Harbinger*/*Tourist* elements, the non-autonomous *Stowaway* elements were reported prior to their autonomous elements (*Tc1*/*Mariner*) in plants [43,44,45]. *Tc1*/*Mariner* elements specifically insert into "TA" sequence and the "TA" sequence is duplicated as a TSD. *Tc1*/*Mariner* elements in animals, such as those in *C. elegans*, are very compact, usually less than 2.5 kb in size [46]. In contrast, it is not unusual for *Tc1*/*Mariner* elements in plants to be over 5 kb. The terminal sequences for *Tc1*/*Mariner*/*Stowaway* elements are "CTCCCTCC…GGAGGGAG". Like *Tourist* elements, *Stowaway* elements are small in size (<600 bp). *Tc1*/*Mariner* elements are detected in many plant genomes but absent from banana, grape (*Vitis vinifera*), sacred lotus, *Amborella*, and *Selaginella* [22,47,48,49].

When the non-autonomous *Tourist* and *Stowaway* elements were initially identified from plants, it was unclear which types of autonomous elements were associated with these elements [39,43]. Hence, it was impossible to assign them to a specific family or superfamily. As a result, they were classified as "miniature inverted repeat transposable elements" (MITEs) based on their structural features. About a decade later, it was revealed that *Tourist* elements are related to *PIF*/*Harbinger* elements [40,41,42], and the *Stowaway* elements are related to *Tc1*/*Mariner*-like elements [44,45,50]. Technically, MITEs can refer to any small TEs (<600 bp) with TIRs, although it is the term most commonly used for *Tourist* and *Stowaway* elements due to historical reasons. In many plant genomes, MITEs are numerically the most abundant elements although they account for a limited genome fraction due to their small size.

2.2.6. *Helitron or DHH elements*

All of the above autonomous DNA transposons encode transposases with a DDE motif for "cut and paste" transposition and are associated with TIR [51]. In contrast, *Helitrons* replicate through a rolling-circle mechanism and do not generate a TSD upon insertion [52]. Autonomous *Helitrons* encode a Y2-type tyrosine recombinase resembling that in bacterial rolling-circle transposons, with a helicase domain and replication initiator activity. *Helitrons* specifically insert into "AT" sequence, *i.e.*, flanked by "A" at the 5′ end and "T" at the 3′ end. *Helitrons* have no TIRs, and the terminal sequence is "TC…CTRR", where R stands for "G" or "A". In addition, a short GC-rich stem loop (a 16–20 bp stem separated by 10–12 bp loop) is often found near the 3′ end of the element. Similar to Pack-MULEs, non-autonomous *Helitron* elements frequently carry genes or gene fragments [53,54,55,56].

To date, there are a total of 12 superfamilies of DNA transposons described but only 6 of them (see above) are reported in plants. The six superfamilies that are absent from plants include *Molin*, *P*-element, *PiggyBac*, *Transib*, *Crypton*, and *Maverick* [14]. Since most of the TE studies are focused on flowering plants, it is likely that novel superfamilies of DNA transposons will be reported when the genomes of non-flowering plants, such as mosses and ferns, are thoroughly studied.

3. Factors that contribute to the proliferation of TEs in plants

TEs are often referred to as "repetitive elements". It is true that some elements can amplify to hundreds, even thousands of copies. Nevertheless, most TEs families in a genome are only found either low copy or moderately repetitive. For example, in maize, there are over 400 families of LTR retrotransposons which occupy 75% of the genome. Nevertheless, the top 20 families contribute to ~70% of the genome, whereas the remainder (380 or more) only contributes to 5% of the genome [57]. Moreover, more than 250 families only have one or two intact members. Low copy numbers are common for DNA transposons as well. Among the 3000 Pack-MULEs in rice, over half have only one copy, with an average copy number of 3 [58,59]. It suggests that the majority of TEs are not significantly amplified but the few families that are amplified contribute to the critical mass of the genome. The high amplification activity of a few families results in the false perception that most TEs are highly repetitive. This is analogous to the fact that some non-TE genes belong to large gene families while others are single copy genes or belong to small gene families. The only difference is that most non-TE gene families achieve copy numbers in a rather long time frame. In contrast, TEs could increase their copy numbers very rapidly and may become unrecognizable in a similar path (see below). For example, the amplification of LTR elements has doubled the maize genome in the last few million years [60], which is very recent in an evolutionary time scale. Below we will briefly discuss some of the factors that contribute to the variation of abundance of TEs within and among genomes.

3.1. Transposase activity

Given the dramatic difference in copy numbers of different TEs in a genome, it is conceivable that the element itself is responsible for the degree of amplification. For a transposition event to occur, the autonomous elements must be transcribed, translated, and a functional transposase (or reverse transcriptase, integrase, *etc.*) be generated. An astonishing fact about transposase is that those from naturally occurring TEs are often not the most competent form, with various types of defects, suggesting selection against transposases with high activity. For example, the truncated form of *Ac* transposase demonstrated higher activity than the wild-type form [61], which aggregates in cytoplasm at high concentrations [62]. Similarly, truncation of the transposase from a MULE in rice significantly enhances its activity [63]. It is likely that the truncation alters the conformation of the transposase so that it influences the activity. For the *Ping*/*Pong* elements from rice, the mutation of a putative nuclear export signal in the transposase leads to a 6-fold increase of excision activity, which is likely due to the elevated concentration of transposase in the nuclei [64].

3.2. Element size

Besides the transposase, element size is critical for the transposition activity [65,66,67]. For DNA transposons, it is obvious that non-autonomous elements, which are small in size, significantly outnumber their autonomous counterparts [32]. In a recent experiment, we directly compared the transposition efficiency of two elements which differ by 1 kb (469 bp vs. 1485 bp) [63], with the smaller element constructed as a deletion derivative of the larger one. It was shown that the smaller element is more active in both excision and reinsertion, and its overall transposition activity is 36-fold that of the larger element [63]. The small element size may confer a mechanistic advantage during transposition. For example, some transposons form a synaptic complex prior to excision, as two transposon ends on the same molecule are brought together via transposase-mediated oligomerization [68,69]. The shorter distance between the two ends in small elements may be more conducive to form the synaptic complex. In addition, small elements are usually less deleterious than larger elements when they insert around the genes, so their insertions are more likely to be retained. Nevertheless, the success of small elements could be due to factors beyond size. In rice, a *Stowaway* MITE is able to utilize the transposase encoded by a distantly related autonomous *Osmar* element [70]. The transposition activity of the deletion derivatives of the *Osmar* element is extremely low while that for the MITE is very high (approximately 1 vs. 50). This is due to the presence of a suppressive motif inside the autonomous element and a region that enhances transposition in the *Stowaway* element. It was proposed that the presence of a suppressive motif in the autonomous elements would minimize their deleterious effect of large insertions and facilitate the long-term persistence of the entire family in the genome [70].

Similar to the presence of MITEs among DNA transposons, there are terminal-repeat retrotransposons in miniature (TRIM) among LTR retrotransposons [71]. TRIMs are non-autonomous LTR elements, usually less than 1 kb in length. Their size is in stark contrast to the autonomous LTR elements with sizes ranging from a few kb to over 10 kb. It appears that TRIMs are widespread in plants [72,73,74,75]; however, none of them has amplified to the degree that is comparable to that of MITEs. This is possibly due to the different transposition mechanism of DNA transposons and retrotransposons. Unlike DNA transposons that could act on distantly related elements (see below), the transposition machinery of retrotransposons demonstrates *cis*-preference for its encoding mRNAs [76]. Thus, non-autonomous retrotransposons are relatively unfavorable in the competition compared to non-autonomous DNA transposons. Although the non-autonomous SINEs have successfully amplified in human [17], human SINE/LINE copy number ratio (~2) is still not comparable to that of MITE/autonomous elements in plants, which is at least in the order of hundreds [41,44].

3.3. Cross-mobilization and sequence swapping

As stated above, non-autonomous elements may utilize transposases from distantly related autonomous elements (such as the *Stowaway* and *Osmar* element), and this process is called cross-mobilization. In another study, the efficacy of the cognate autonomous element was directly compared with the distant relative. The *Ping*/*Pong* elements in rice can both mobilize *mPing* which is the deletion derivative of *Ping*, and *Pong* is a distant relative which share only 70% sequence similarity [9]. Interestingly, *mPing* seems to be equally or more active with *Pong* as the source of transposase than with *Ping* [64]. It is unclear why *Pong* is more effective; however, cross-mobilization relationship maximizes the opportunity for TEs to persist in the genome if the cognate autonomous element is lost or non-functional. On the other hand, if the relevant TE is associated with self-preference, the divergence between autonomous and non-autonomous elements may reduce or abolish the activity of the non-autonomous element. To enhance the partnership with the autonomous element, a non-autonomous LTR element in soybean has been continuously recombining with the LTR sequences from the autonomous elements in order to maintain its activity [77].

3.4. Target selection

The majority of TEs have target site specificity which means that insertion into the genome does not occur randomly [78]. Target selection is critical for the success and persistence of TEs. In general, DNA transposons are located in genic regions especially the 5′ ends of genes [79,80,81], while LTR retrotransposons are mostly located in intergenic, repetitive sequence, or heterochromatic regions [12], with some exceptions. CACTA elements, as mentioned above, do not demonstrate a strong preference for genic regions like other DNA transposons [18,32]. Some LTR elements, like *Tos17* element in rice, preferentially target genic regions [82]. If we consider that the impact of insertion is proportional to the element size, we would expect only small elements to be successful in genic regions. This premise is consistent with the high copy number of MITEs within or around genes. It also explains why *Tos17* never amplifies to high copy number in natural populations [83] and most of the low copy number LTR elements in maize are located in genic regions [57].

3.5. Interaction with other TEs

The genome could be considered as an ecosystem where TEs interact with genes and other TEs. First, since most TE insertions are deleterious, the overall transposition frequency has to be controlled at a certain level to ensure the long-term fitness of the organism. This is in accordance with the fact that only a very small subset of the TEs in the genome is currently active and most existing transposases are not the most competent form. Second, if TEs target genic regions or low copy sequences, they must compete for target sites because low copy number sequences that could tolerate insertions are limited in a genome. When a low copy region is occupied by a TE, the relevant site becomes repetitive and no longer available for elements with the same target preference, which represents a negative feedback. Third, if TEs target repetitive sequences, they may insert into other TEs to form nested insertions. This creates a positive feedback because one insertion provides targets for further insertion. From this point of view, LTR retrotransposons are more competitive than DNA transposons because the former can insert into each other as well as into DNA TEs. The consequence of such preference is that LTR elements may abolish the transposition activity of DNA TEs by inserting into them. Meanwhile, DNA TEs are incapable of inflicting the same damage towards LTR elements since they rarely insert into other repetitive sequences.

The interaction between DNA TEs and retrotransposons was demonstrated in a comparative study using MULEs in maize and rice. Maize has a much larger genome than rice, but the number of MULEs in maize is only one third of that in rice. One of the limiting factors is that many putative autonomous MULEs in maize are targeted by LTR retrotransposons which interrupt the open reading frame (ORF) of MULEs [84]. Due to the selective target choice, LTR elements can amplify indefinitely while the amplification of DNA TEs (those that target genic regions) would plateau at a certain point as the target sites diminish. Moreover, the presence of active LTR elements is a threat to DNA TEs but not vice versa. This may explain why in some plant genomes, the amount of DNA transposons is negligible. Those include papaya (*Carica papaya*, 0.2%), *Utricularia gibba* (0.4%), tomato (*Lycopersicum esculentum*, 0.9%), cucumber (*Cucumis sativus*, 1.2%), saltwater cress (*Thellungiella parvula*, *1.2%*), and banana (*Musa acuminate*, 1.2%) [20,85,86,87,88,89]. Although in some cases, the low abundance of DNA transposons may reflect incomplete annotation, the fact that not a single plant genome (if it does contain transposons) has such low abundance of retrotransposons suggests that extinction of DNA TEs occurs more often than that of retrotransposons.

3.6. TE silencing

Due to their mutagenic role, high activity of TEs is always a threat to the host, and silencing mechanisms have evolved to suppress TE activity [90,91]. At the center of the silencing mechanism is the generation of double strand RNAs (dsRNAs) and small interfering RNAs (siRNAs). The dsRNA could be produced from the transcription of elements with extended TIRs, the transcription of different copies of an element from opposite orientations, or bi-directional transcription from a single element. Since the chance to have transcripts from different copies increases with copy numbers, high copy number TEs are more likely to generate dsRNA and be silenced. The dsRNA is then cleaved by members of the Dicer family of proteins into short 21–30-nucleotide siRNAs that guide RNA-degrading complexes to a complementary transcript, which causes the degradation of the relevant RNAs. Alternatively, the siRNA can feed into a different protein complex and cause chromatin modification, including DNA methylation, modifications of histone, and alterations in chromatin packing and condensation [90,92]. Such modification makes the chromatin inaccessible so that the transcription of TEs is suppressed.

A comparative genomic analysis indicates that the efficacy of such silencing mechanism determines the degree of amplification of TEs in a genome-wide fashion [93]. *A. lyrata* is a relative of *A. thaliana* and they share a common ancestor about 10 million years ago [94]. The genome of *A. lyrata* is 50% larger than that of *A. thaliana*, and a significant part of the increase is attributed to the amplification of TEs [93]. *A. lyrata* has two- to three-fold higher copy numbers of every major TE family compared with *A. thaliana*, indicating little family-specific expansions or reductions since the two species diverged. This suggests a genome-wide suppression of TE activity in *A. thaliana*. Consistent with this notion, TEs in *A. lyrata* are targeted by a lower fraction of uniquely matching siRNAs. Since uniquely matching siRNAs are associated with more effective silencing of TE expression, TEs in *A. lyrata* are subject to less effective silencing [93]. The differential amplification of TEs in *A. thaliana* and *A. lyrata* indicates that TE activity is tightly regulated by their host and the efficacy of regulation varies dramatically among different genomes. Nevertheless, as discussed in Section 4, a fitness cost might be associated with the host defense system, and the silencing mechanism becomes a double edged sword.

3.7. Elimination of TE sequences from the genome

There are a variety of mechanisms that mitigate the genome expansion caused by TE amplification. It is known that unequal recombination occurs between the two LTRs of an LTR retrotransposon, and this leads to the formation of solo LTRs. In this case, one LTR and the internal sequences between the two LTRs are eliminated. TE sequences may be deleted by other mechanisms such as illegitimate recombination, which removes more DNA sequences than unequal homologous recombination in Arabidopsis [95]. In rice, it is estimated that 190 Mb of retrotransposon sequences, accounting for nearly half of the rice genome, was eliminated in the last 8 million years by various deletion mechanisms [96]. A subsequent comparison of LTR elements between *japonica* (Nipponbare) and *indica* (93–11) rice cultivars demonstrated that both local genetic recombination rate and gene density were negatively correlated with the amount of LTR elements, suggesting that high recombination rate promotes elimination of TE sequences [97].

The presence of various types of deletion processes combined with relaxed constraint explains why TEs are only recognizable for a few million years and why they tend to go extinct upon the loss of transposition activity. Like the silencing mechanism, the efficiency of sequence elimination varies among different species. Through examination of the genomes of cotton and its relatives, Hawkins et al. demonstrated that the rate of DNA loss is significantly higher in the species with smaller genomes and it is sufficient to reverse the genome expansion caused by TE amplification [98,99,100,101]. The physical removal of TEs minimizes the opportunity for TEs to be reactivated. Taken together, the abundance of TEs depends on the balance between amplification and sequence elimination, and it appears that plants with small genomes are associated with efficient silencing machinery as well as rapid sequence loss.

4. The consequence of TE activity

When TEs amplify in a genome, it alters genome structure genetically and epigenetically. One obvious consequence of TE amplification, especially that of LTR retrotransposons, is the expansion of genome size [60,102,103,104]. This has been well established and will no longer be discussed in this review. Beside the addition of genomic sequences, TEs may influence gene expression, duplicate normal genes, and become domesticated genes. All of these activities have significant impact on allelic and phenotypic diversity.

4.1. The influence of TEs on gene expression

4.1.1. Patterns of alteration

In yellow mustard (*Sinapis alba*), the *FATTY ACID ELONGATION1* (*FAE1*) gene determines the erucic acid content in seeds, an important agronomic trait. Erucic acid is an important raw material for many industrial applications but is undesirable in edible oil [105]. *FAE1* has four different alleles, designated as E1, E2, E3, and e, with progressively declining erucic acid content [106]. The E1 allele is the full functional version of the gene with 53% of erucic acid in seeds [107]. The null functional allele e is caused by an insertion of a *PIF*/*Harbinger* element in coding region which interrupts the gene ORF resulting in almost no erucic acid in the seeds. The E2 allele contains a *Copia* type LTR retrotransposon between the transcription start site and the promoter. This leads to the reduction of *FAE1* expression and 24% of erucic acid in seeds. E3 has the exact same sequence as E2; however, its promoter region is methylated, which results in further reduction of *FAE1* expression and only 1.4% erucic acid (Fig. 2A). It is clear that the presence of the LTR element is essential, but not sufficient, to cause methylation of the promoter. Thus, TEs control the erucic acid content in yellow mustard by creating genetic and epigenetic alleles to manipulate the function of *FAE1* [107].

In addition to knocking out genes or influencing the expression level, TEs may modify the structure of gene transcripts. In maize, the *brown midrib1* (*bm1*) mutant is caused by a *Ds* insertion in the first intron of cinnamyl alcohol dehydrogenase 2 (CAD) gene [108]. CAD catalyzes the conversion of hydroxycinnamyl aldehydes to monolignols, a key step in lignin biosynthesis. The *Ds* insertion provides an alternative splicing site for transcription, which leads to the formation of a premature stop codon in *cad2* mutant mRNA [108] (Fig. 2B). The loss of CAD2 function causes a reduction of total lignin content by 24%, which favors the utilization of maize as a biofuel crop. In tetraploid sour cherry (*Prunus cerasus*), the self-incompatibility is determined by a pair of linked genes: the stylar *S-RNase* and pollen *S F-box* (*SFB*) genes. The breakdown of self-incompatibility in one of the sour cherry cultivars is due to the insertion of a small *Helitron* element downstream of the stop codon. The presence of the *Helitron* element interferes with polyadenylation so that the majority of the mRNAs are not associated with a poly(A) tail [109] (Fig. 2C). The lack of polyadenylation leads to a non-functional allele, making the mutant self-compatible.

TEs may not only influence the genes nearby, but also affect the expression of genes in *trans*. In rice, Pack-MULEs are associated with the generation of small RNAs. If the small RNAs are derived from regions homologous to the parental genes and are produced at sufficient levels, the expression of the parental genes is suppressed [58]. Such suppression in *trans* is not limited to gene-carrying TEs and their parental genes. In Arabidopsis, many genes that do not contain a TE are regulated through partial and fortuitous complementarity to the siRNAs generated from TEs. Interestingly, the siRNAs generated by a LTR retrotransposon *Athila* targets many genes including *OLIGOURIDYLATE binding PROTEIN1b* (*UBP1b*), which suppresses the translation of TE proteins [110,111] (Fig. 2D). As a result, the siRNAs derived from *Athila* may suppress the expression of normal genes but enhance the expression of TEs. Thus there is a trade-off between defense against TE activity and host gene expression, and TEs have been utilizing the silencing system for their survival in the arms race.

4.1.2. The influence of TEs at genome-wide level

The above examples demonstrate how transposons could cause dramatic changes in gene expression, which leads to detectable phenotypic changes. In reality, however, not all transposition events are associated with a detectable phenotype. On one hand, many TEs do not insert into genic regions so the chance for them to influence a gene or genes is relatively low. On the other hand, the effect of TE insertions in genic regions could be subtle, especially when the element is very small. In rice, the 430 bp *mPing* element is highly active and its copy number ranges from 1 to 1000 in different rice cultivars [112]. It was estimated that in certain rice cultivars *mPing* was increasing its copy number by ~40 copies per plant per generation [112]. Most of these newly amplified *mPing* elements insert near genes, and some of them insert into exons. However, most of the insertions in exons disappear from the subsequent generation [112], either through excision or segregation. Surprisingly, the remainder of the *mPing* insertions either up-regulates or has no detectable impact on the expression of nearby genes [81]. It appears that *mPing* carries regulatory motifs which can be activated by stress [81]. The rapid elimination of potential deleterious insertions and the positive or neutral effect on gene expression enables *mPing* to attain high copy numbers in a relatively short period [112].

If a TE, such as *mPing*, contains a *cis*-regulatory motif and amplifies to high copy numbers, it may rewire new genes into the existing transcriptional network, when it inserts at the 5′ end of genes and becomes part of the regulatory region of the downstream genes. Even the TE is far away from a gene, it may influence the network by diluting the effect of the relevant transcription factors. A recent study in *Brassica* species demonstrates that TEs may modulate certain transcription networks by amplification of *cis*-regulatory sequences. The E2F transcription factor family plays a critical role in regulation of the cell cycle, DNA replication, and development [113]. Interestingly, in some of the *Brassica* species, up to 85% of E2F binding sites are located inside TEs, with detectable binding by E2F proteins [114]. This suggests that TEs may have played important roles in the evolution of E2F-related transcriptional networks in *Brassica*.

In plants, polyploidization is very common. *Brassica rapa* is a hexaploid, and its genome consists of three subgenomes compared with its fellow crucifer, Arabidopsis. One of the subgenomes is dominant over the other two subgenomes. In general, the recessive genomes are more fractionated with lower level of gene expression. This is because the recessive genomes harbor more TEs in the promoter regions of genes resulting in very intense silencing by siRNAs [115]. The attack of siRNAs on TEs causes suppression of nearby genes, followed by degeneration of the subgenomes. Thus, the amount of TEs in the parental genome, or the accumulation of TE insertions in each subgenome after polyploidy events, may define which subgenome becomes dominant.

If a subset of TEs impacts the expression of genes, the total genes affected by TEs in a genome should be largely proportional to the total number of TEs in the genome, or the number of TEs in genic regions. In *A. thaliana*, 16% of the genes have a TE insertion within 500 bp of the genes [116]. This is in contrast to *A. lyrata*, which has more TEs, and 24% of the genes harbor a TE within 500 bp. Moreover, TEs that are targeted by siRNAs are correlated with reduced expression of adjacent genes in both *A. thaliana* and *A. lyrata* [93]. In maize, 66% of the genes have a TE insertion within 1 kb to the gene. Up to 25% of the maize transcriptome is responding to the expression of the *Mutator* family [117]. More importantly, TEs influenced up to 20% of the genes up-regulated in response to abiotic stress, and as many as 1/3 of the genes that are solely responsive to stress [118]. In most cases, TE

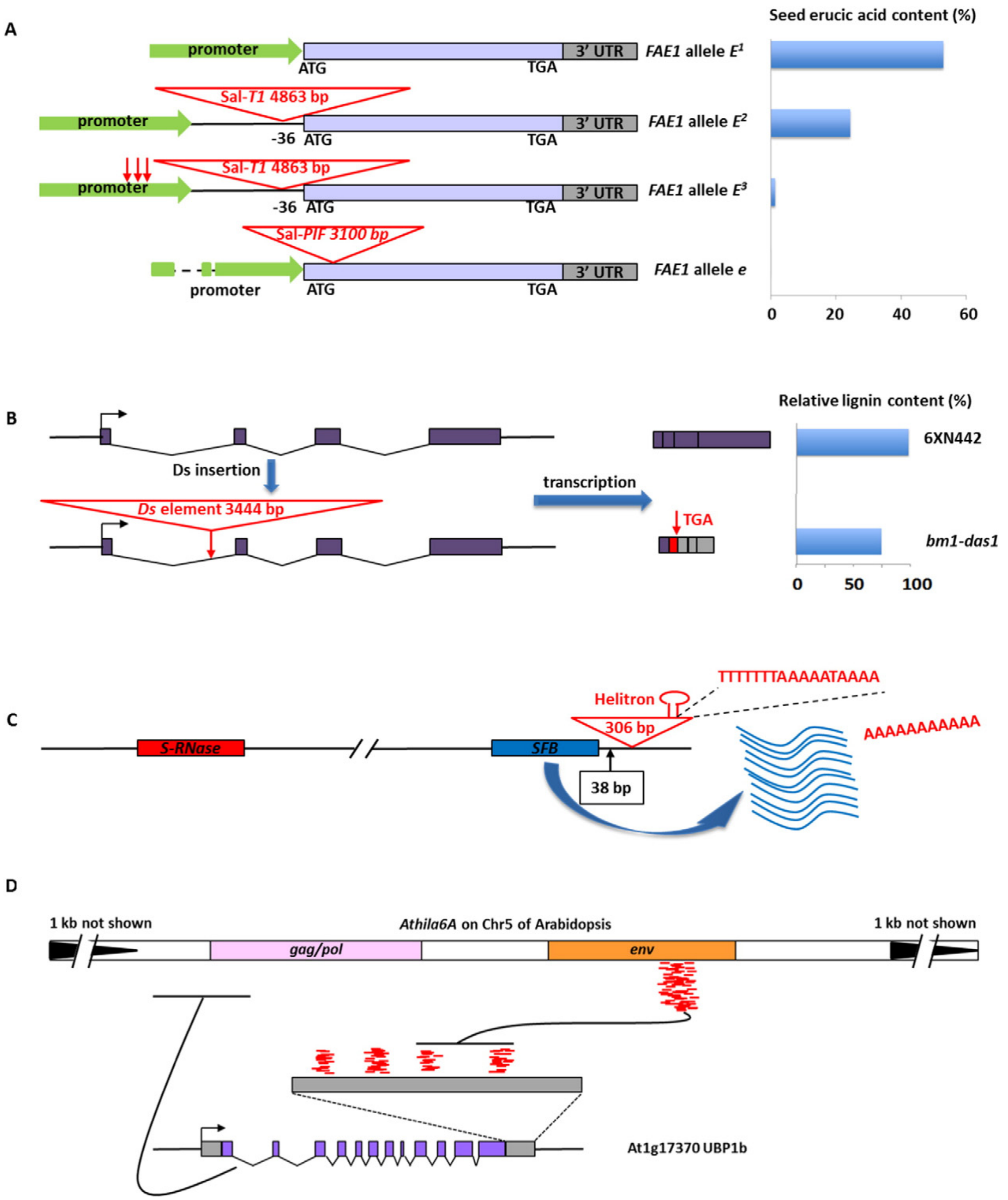

Fig. 2. The impact of transposable elements (TEs) on gene expression. *A*. The structure of the four FAE1 alleles, E1, E2, E3, and e, in yellow mustard. Green arrows indicate promoter regions and red arrow points to the three methylation sites. The bar graph on the right panel in A and B indicates erucic acid or lignin content linked to the corresponding alleles. *B*. The structure of maize cinnamyl alcohol dehydrogenase 2 (CAD2) gene in wild-type and *brown midrib1* (*bm1*) mutant plants. *C*. The self-incompatibility locus in sour cherry cultivar MSU 27e 9. RNAs are shown as waved lines. *D*. Arabidopsis *Athila* LTR retrotransposon and *OLIGOURIDYLATE binding PROTEIN1b* (*UBP1b*) gene. Small RNAs are shown as short red lines and black triangles indicate LTRs. TE insertion in genes is shown as open triangles. Colored boxes represent coding regions and untranslated regions are shown as grey boxes. Introns are shown as lines connecting exons. Other sequences are shown as horizontal lines or open boxes.

insertions act predominantly as local enhancers of expression rather than as novel promoters, which may contribute to heritable variation for stress response [118]. Taken together, TEs may activate or suppress the expression of adjacent genes, and a substantial portion of the genes in a genome are influenced. Such regulation may confer adaptive advantages for host plants.

4.2. Duplication of non-TE sequences including genes

4.2.1. The scale of gene duplication by plant TEs

The first incident of gene duplication/acquisition by TEs was reported 28 years ago [119], where the *Mu1*/*Mu2* subfamily of *Mutator* elements was found to carry a piece of sequence from a host gene, which will be referred as parental gene hereafter. Subsequently, a few additional acquisition events were reported in maize, morning glory (*Ipomoea nil*), Arabidopsis, and rice [34,120,121,122,123]. However, it was unclear whether those acquisition events represented a rare curiosity, or whether acquisition of genomic sequence was a general mechanism.

Now it is clear that all superfamilies of DNA transposons as well as LTR retrotransposons in plants are capable of duplicating genomic sequences including genes [18]. Duplication by non-LTR retrotransposons (LINEs and SINEs) has not been unambiguously reported; however, given the fact that L1 element (a LINE element) duplicated 1% of the human genome, it is likely that they are involved in gene duplication. Different TEs vary dramatically in their ability of duplicating genomic sequences. Among plant DNA transposons, three superfamilies are associated with gene duplication in large scale. In morning glory, the CACTA element *Tpn1* has 500–1000 members, and most of them carry different types of parental gene fragments [124]. In rice, there are ~3000 Pack-MULEs which carry sequences from over 1500 parental genes [38,59], accounting for 4% of all non-TE genes. Pack-MULEs are also prevalent in *Lotus japonicas* and sacred lotus [22,125]. In maize, there are ~30,000 copies of *Helitron*s and over 1000 intact elements carry gene fragments [55,56]. Given the fact that transposons are only recognizable for a few million years, most of the ancient duplication events by TEs cannot be recovered or revealed. As a result, it is likely that a large portion of plant genes have been duplicated at certain point of the evolutionary history [126].

4.2.2. Mechanism of sequence duplication

Genomic sequence duplication by retrotransposons is likely to occur during transposition since the duplicated sequences are processed by the reverse transcription complex. For non-LTR elements such as L1 in humans, it is known that the transcription of elements is often extended to flanking sequences during the transcription process, and the flanking sequences are transposed to new genomic loci together with the element [127]. For LTR elements, this process is initiated with the accidental packaging of genomic mRNAs in the virus-like particles. Subsequent reverse transcription and integration into the genome result in an independent retrogene. If template switching occurs between the LTR element and the gene mRNA, the gene fragment could be located inside the LTR element [128].

The mechanism of sequence duplication by DNA transposons is less clear. Gene fragments inside DNA elements sometimes contain introns, so it is likely that the duplication occurs at the DNA level, not RNA or cDNA levels. Three models have been proposed for the acquisition of genomic sequences inside Pack-MULEs or other DNA TEs with TIRs. According to the first model, the individual MULE TIR is mobile and when a pair of TIRs encompasses a gene sequence, the TIRs and the gene sequence between the TIRs will move as a single unit and form a Pack-MULE [119]. However, we recently demonstrated that the mobility of a single TIR was undetectable [63]. The second model considers that the presence of long TIRs is responsible for sequence acquisition [129]. If the MULE sequence is in a single strand DNA status, the presence of the TIR would lead to the formation of a stem loop and a DNA nick would be created by an endonuclease. At this point, an ectopic sequence would be used as a template to repair the nick, resulting in the introduction of novel sequences into the element [129]. The role of the long TIR in acquisition is consistent with the fact that TIR MULEs more frequently duplicate genes than non-TIR MULEs [38]. A third model assumes that sequence acquisition occurs during the gap repair process after TE excision. The double strand break formed after excision is often repaired using a sister chromatid or a homologous sequence as template via a synthesis-dependent strand annealing (SDSA) mechanism [130,131]. If template switching occurs during this process, novel sequences are introduced to the element [132]. Post-excision sequence rearrangements caused by abortive gap repair have been reported with *Mutator*, *Ac*/*Ds*, and *Spm* elements [133,134,135,136]. In one of the cases, the derivatives of *Ac*, formed at the donor site after the excision of *Ac*, contains short pieces (52 and 96 bp, respectively) of non-TE sequences with unknown origin [135], suggesting that it is possible for TE to acquire non-TE sequences during gap repair. It is worth mentioning that both 2nd and 3rd models predict sequence duplication by a TE without transposition to a new genomic locus. The 3rd model requires the presence of transposase as well as the excision of element whereas transposase does not seem to be essential for the 2nd model.

Although the exact mechanism underlining sequence acquisition by DNA TEs is not well understood, analysis of existing sequence acquisition provides helpful insights. When a *Helitron* element harbors gene fragments from multiple genes, all fragments are located in the same orientation with respect to the transcription of the parental genes [137]. Moreover, the gene fragments are also in the same orientation as the *Helitron* Rep/helicase gene inside the element. This seems to imply that the duplication process is somehow associated with the transcription of genes. Subsequently, it was shown that the biased distribution is likely a combined consequence of preferential acquisition of gene fragments in the same orientation and selection against the presence of the “antisense” gene fragments [55]. A bias for gene orientation was also observed for the *Tpn1* element [124]. In contrast, the orientation of gene fragments inside Pack-MULEs seems to be largely random, although elements carrying gene fragments with different orientation are less likely to be expressed, which is probably the result of selection [58]. The difference in the orientation of acquired fragments demonstrated by Pack-MULEs and *Helitron*s suggests that they employ distinct duplication mechanisms or that selection operates differently on these two types of elements. Moreover, the selection of parental genes for Pack-MULEs is not random, and they preferentially duplicate sequences from genes with high GC content and wide expression [38]. Since high GC content is often associated with open chromatin structure [138], it suggests that Pack-MULE duplicate sequences from the regions that are readily accessible.

In addition to the typical incorporation of genomic sequence into TEs, TEs may duplicate genomic sequences through alternative transposition. This phenomenon was elegantly demonstrated with *Ac*/*Ds* element in maize. When two *Ds* elements are located in nearby regions in the same orientation, an atypical transposition using one terminus from each element causes tandem duplication of flanking sequences [139]. If the alternative transposition occurs at S phase of the cell cycle, it may cause re-replication of TEs and their flanking sequences [140]. The resolution of such duplication/replication is always associated with significant rearrangement of the loci involved. Unfortunately, this type of duplication and rearrangement cannot be confidently inferred without generational materials which harbor currently active elements. Therefore, the abundance of sequence duplication by TEs is significantly underestimated.

4.2.3. Outcome of gene duplication by TEs

Gene duplication can be achieved through a variety of mechanisms including whole genome duplication (polyploidization), tandem duplication (unequal crossover), and other types of segmental duplications. As discussed above, some of the tandem duplications or segmental duplications could be mediated by TEs through atypical transposition but it is difficult, if not impossible, to evaluate the frequency or fraction of such events. Duplicated genes may evolve in three different paths: neofunctionalization (evolution of new functions), subfunctionalization (each copy conducts part of the original function), and pseudogenization (one copy degenerates and eventually becomes a pseudogene) [141]. Compared with other duplication mechanisms, duplication through TEs is distinguished by two significant characteristics. First, in most cases, the duplication only includes gene fragments, not entire genes. TEs frequently combine different gene fragments together and form chimeric ORFs, and this may maximize the opportunity for evolution of new structure or new functions. In maize, as many as ten different gene fragments have been found in a single *Helitron* element [55]. In rice, 42% of the 1235 retrogenes have recruited new exons from flanking sequences, leading to the formation of chimeric ORFs [142]. Likewise, a significant portion of Pack-MULEs contain gene fragments from multiple loci, and they express more frequently than elements with sequence from only one gene [58]. As such, the presence of chimeric ORFs inside TEs may lead to *de novo* gene formation. Second, due to the mobility of TEs, in most cases, the duplicated copy is physically unlinked to the parental copy. The combination of the mobility and shuffling activity enables the duplicated gene to acquire novel expression patterns.

The tomato *SUN* locus provides an excellent example of how duplication of gene through TEs may lead to differential tissue specificity of genes. The wild ancestor of tomato produces round fruit and the elongated shape emerged during domestication of tomato as a result of artificial selection. The *SUN* gene was created by a duplication event mediated by an LTR retrotransposon called *Rider* [143], and it changes fruit shape by increasing cell division in the longitudinal orientation [144]. The parental gene of *SUN* is located on chromosome 10 and it is not expressed in flowers. The *Rider* element duplicated a fragment containing *SUN* and integrated the fragment into a locus on chromosome 7. The *SUN* acquired a new regulatory region which allowed it to express in flowers. Hence the individual plants with *SUN* are associated with elongated fruits.

Considering that in most cases sequences duplicated by TEs represent gene fragments, not entire genes, it is likely that the frequency of pseudogenization of genes duplicated by TEs is higher than that derived from other gene duplication mechanisms. Therefore it is questionable how many of those genes retain coding capacity, especially the ones with only fragments of genes. One way to determine the coding capacity is to calculate the ratio of non-synonymous (Ka) to synonymous (Ks) substitution ratio of the duplicated copy and the parental copy. If the Ka/Ks ratio is significantly smaller than 1, it implies that the duplicated copy has retained the coding capacity of the parental genes or is under purifying selection. In maize, a LTR retrotransposon *Bs1* carries gene fragments from three genes, and the gene fragments are fused with the *Bs1 gag* domain, forming a chimeric ORF, which is transcribed and translated [145]. Interestingly, the Ka/Ks value for two of the gene fragments is smaller than 1, indicating the constraint to maintain the protein sequence of the parental genes. The Ka/Ks value for the third gene fragment is similar to 1, suggesting that particular fragment is under neutral drift. In rice, most of the retrogenes are expressed and some are under purifying selection [142,146]. Among the 3000 Pack-MULEs in rice, over 20% are transcribed and at least 28 are translated [58]. Moreover, about ¼ of the elements appear to be under purifying selection [58]. Taken together, these studies indicate that although the majority of genes duplicated by TEs may not evolve into protein coding genes, a small subset is likely carrying functional coding sequences.

Besides the formation of protein coding regions, the duplicated gene fragments may modify the structure of existing genes. Pack-MULEs frequently insert at the 5′ end of genes, and the transcription is often initiated in the internal region of the Pack-MULE and extends to the adjacent gene. In this case, the internal regions of the Pack-MULEs, which contain the acquired fragment, serve as the 5′ untranslated region and/or coding region in the chimeric transcript. Due to the preference of Pack-MULEs to duplicate GC-rich sequences, the modified transcripts are associated with a negative GC gradient, which means a higher GC content at the 5′ end than that at the 3′ end. The activity of Pack-MULEs partly contributes to the numerous GC-rich genes and genes with negative GC gradient in rice [147]. Interestingly, GC-rich gene fragments are selectively retained over the GC-poor counterparts inside Pack-MULEs [38]. Since both GC-rich and GC-poor sequences are duplicated from codon regions, it suggests the bias is beyond codon capacity.

4.2.4. Distinguishing gene-carrying TEs from non-TE gene sequences

When a TE carries gene fragments, it often shares many features with normal genes, in addition to the presence of coding sequence. Pack-MULEs, for example, are located primarily in gene-rich chromosomal arms instead of the pericentromeric regions where many other TEs are enriched. Most Pack-MULEs have low copy numbers, with an average value of 3 [59]. In addition, a substantial portion of Pack-MULEs are expressed, with detectable selection constraints on coding capability [58]. Accordingly, Pack-MULEs resemble genes in terms of chromosomal distribution, copy number, selection, and expression. Not surprisingly, most (68%) of the 3000 Pack-MULEs were annotated as independent cellular genes (39%) or part of other genes (29%). Obviously, Pack-MULEs are indistinguishable from non-TE genes to gene prediction programs. As a consequence, it remains a great challenge to identify gene-carrying TEs.

Gene-carrying TEs can be distinguished from normal genes based on the following features. First, the size of the coding regions inside TEs is often much smaller than the intact genes. It is relatively rare (but not impossible) for TEs to carry intact genes. As shown in Fig. 3, a Pack-MULE on rice chromosome 6 duplicated a gene fragment from an amine oxidase gene from chromosome 2. The coding region of the parental gene is 2.5 kb, while the acquired fragment inside the Pack-MULE is about 700 bp. As a result, if the gene of interest (GOI) is much smaller than other homologous genes, one of the possibilities is that it is a duplication event mediated by a TE. Second, alignment of gene fragments in a TE with other homologous genes is usually associated with a clear, sharp boundary at the breakpoint of duplication. In contrast, the identity of the alignment between normal gene family members often degrades gradually, without a clear boundary. This is because other gene duplication mechanisms often lead to the duplication of entire genes, and the selection pressure operates differentially at different parts of the gene, forming an identity gradient.

Third, gene fragments in TEs are often flanked by inverted repeats (for most DNA transposons) or direct repeats (for LTR retrotransposon). The presence of TSD flanking the inverted or direct repeat is the unambiguous evidence that the entire structure (TIR or LTR plus the gene fragment in the middle) belongs to a single TE (Fig. 3). Nevertheless, this evidence is only applicable to elements with relatively long TSDs, including retrotransposons, *hAT*s, and MULEs. For elements with short TSDs (CACTA, *PIF*/*Harbinger*, and *Tc1*/*Mariner*), it is rather common for two independent elements to have the same TSD sequences. As a consequence, in this case, the TSD flanking a TIR does not mean the TIR belongs to a single TE. For elements without TIR or LTR, such as *Helitron*s, the gene fragment is still flanked by the element termini. The termini of the element (for all types of elements) should be much more repetitive than the gene fragment (Fig. 3). Moreover, if the gene fragment is indeed inside a TE, the sequences flanking the gene

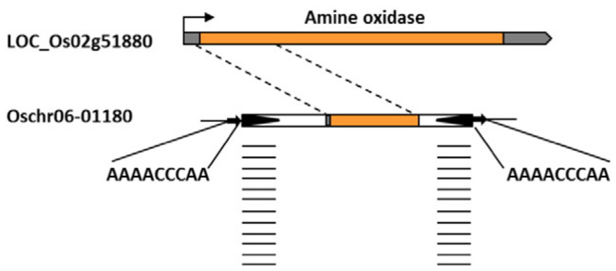

Fig. 3. A Pack-MULE and its parental gene in rice. For the parental gene, colored boxes represent coding regions and grey boxes indicate untranslated regions (UTR). Transcription start site of the parental gene is shown as an arrow over the 5′ UTR. For the Pack-MULE, the captured region from its parental gene is connected with dashed lines. Other sequences inside the Pack-MULE are depicted as white boxes. Terminal inverted repeat (TIR) is depicted as black triangles. Target site duplication (TSDs) are depicted as small black arrowheads flanking the TIR, with its sequence shown below. The lines under the TIR indicate the presence of other homologous TIR sequence in the genome.

fragment, which are the putative termini of the element, should be equally abundant in the genome, and often appear in a pair in the genome. Finally, for any candidate gene-carrying TE, if the relevant structure (a pair of termini flanking a gene fragment(s)) has an additional copy with high identity (>90% at the nucleotide level) and distinct flanking sequences, it indicates that the relevant structure is likely to be formed through transposition. Nonetheless, our study in rice Pack-MULEs indicates that a substantial portion of the Pack-MULEs only have a single copy [58]. Those elements are identified through the presence of long TSDs flanking the element and would have been missed if MULEs had shorter or no TSD. Since we rely on element termini and TSD to identify gene-carrying TEs, it is conceivable that only a subset of the very recent TEs can be recovered with confidence. The implication of this notion is that the number of gene-carrying TEs is significantly underestimated and many of the "normal genes" or parts of the "normal genes" may have been derived from ancient gene-carrying TEs.

4.3. Domestication of TEs

In addition to duplicating, rearranging, recombining, and amplifying gene sequences, TEs contribute to additional coding genes by donating their own sequences, mostly the coding regions of transposase. The process of utilization of TE-derived sequences for cellular functions instead of transposition is called TE domestication. Since transposases are capable of recognizing and binding to TE termini, they are virtually DNA binding proteins, which is the essential feature for transcription factors. In fact, it was proposed that the DNA binding domain of most plant-specific transcription factors were derived from TEs [148].

4.3.1. TE domestication events verified through experimental studies

The first and best-characterized examples of TE domestication in plants are the *FAR1* and *FHY3* genes in Arabidopsis. Mutations in these two genes cause detectable phenotypes in response to far-red light. The *FAR1* and *FHY3* genes exhibit a high degree of similarity to the MULE transposase [149], and a subsequent study demonstrated that each of these genes could act as a *bona fide* transcription factor [150]. Consistent with this notion, a genome-wide study identified over 1000 direct target genes for *FHY3* [151]. The two genes are also involved in regulation of plant resistance to pathogen through chlorophyll synthesis [152]. The *FAR1* and *FHY3* genes belong to the *FAR1*/*FRS* gene family with over ten members in Arabidopsis [153]. Members of the *FAR1* subclade are present in multiple plant species, including both monocots and dicots, suggesting that the domestication event occurred prior to the monocot–dicot split. Similarly, the *MUSTANG* (*MUG*) family of genes was shown to share extensive homology with MULEs without being associated with mobility, which is supported by the fact that they are present at syntenic sites within different plant species [154]. Analysis of mutants in genes of *MUG* family indicates that they play a role in reproductive fitness of flowering plants [155]. The abundance of MULE-related genes is likely due to the presence of the WRKY–GCM1 domain in the transposase. The WRKY–GCM1 domain belongs to a widespread eukaryote-specific group of transcription factors, and MULEs may have been domesticated independently many times over the course of evolution for the production of transcription factor genes [156].

The *DAYSLEEPER* gene in Arabidopsis belongs to the *hAT* superfamily and it is essential for plant growth [157]. The proteins encoded by *DAYSLEEPER* are capable of binding to a motif present in the upstream region of the DNA repair gene *Ku70* and many other genes in Arabidopsis, suggesting that *DAYSLEEPER* is a transcription factor. Like a transposase, the *DAYSLEEPER* protein is mainly located in the nucleus but can also be seen in discrete foci in the cytoplasm [158]. It seems that *DAYSLEEPER* retained the overall structure of *hAT* transposase and such conservation appears to be essential for its cellular function. Yet *DAYSLEEPER* seems to extend its function beyond the nucleus [158]. Moreover, *SLEEPER* genes (*DAYSLEEPER*-like genes) possess the three SLEEPER motifs which are distinguished from *hAT* transposase [159]. As a result, TE-derived genes may evolve new structure or function during domestication. *SLEEPER* genes are unique in angiosperms and mutation of a rice *SLEEPER* gene also causes abnormality of growth [159], suggesting their importance in higher plants.

4.3.2. Identification of domesticated TEs through bioinformatics approaches

The above domestication incidents were identified due to the phenotype of the mutants. However, there are hundreds or thousands of transposase-encoding genes in the genome, and it is unrealistic to knockout all transposase domains in the genome to test whether they have cellular functions. This is just as it is impossible to rely on experimental evidence to annotate every single gene in the genome. So the question is how to distinguish a domesticated TE gene from a regular TE in silico. Several types of evidence are helpful in this process. First of all, most domesticated TEs lost mobility, so they should not be flanked by terminal sequence of TEs such as TIRs or LTRs. Moreover, they should not be repetitive because they lost the ability to amplify. Second, domesticated TE genes should be expressed (transcribed or translated). Third, if the domestication event occurred long time ago, the putative genes should be present in related species and a synteny should be observed between species. Finally, domesticated TE genes tend to form clades by themselves [160]. So if an unknown transposase domain groups together with a known domesticated gene, it is also likely a domesticated TE gene. However, none of the above evidence is essential or sufficient to define a domesticated TE gene. For example, the *ROSINA* gene in snap dragon plays a role in flower development and binds to the promoter of the MADS-box gene *DEFICIENS* [161]. Nonetheless, it is located inside a CACTA element and has multiple copies in the genome [162]. Since *ROSINA* is similar to proteins of CACTA elements from other plant species, it largely rules out the possibility that it was a cellular gene acquired by the TE. As a result, TEs could serve as normal genes without loss of mobility. In addition, expression of domesticated TE genes could be too low to detect or only in some specialized tissues; synteny could be interrupted by rearrangement; a recently domesticated TE genes may group with other TEs instead of other domesticated TE genes. As a result, it is difficult to draw conclusion based on a single piece of evidence, but the presence of multiple lines of evidence certainly increases the confidence of annotation.

To address this issue, a recent study employed an integrative approach which took all evidence into consideration [163]. Using this approach, the authors discovered 36 novel domesticated TE genes with high confidence. Together with the 31 known domesticated TE genes, there is a total of 67 TE-derived genes in the Arabidopsis genome. More than three fourths of those genes are from DNA transposons. Among DNA elements, MULE contributes to most of the domesticated genes, followed by *hAT*, *PIF*/*Harbinger*, and CACTA elements. However, if considering the ratio of domesticated genes to the total number of TEs in the superfamily, *hAT* has highest domestication rate [163]. As the systematic approach applies to more plant genomes, it is expected that many more domesticated TE genes will be discovered.

5. Horizontal transfer and the life cycle of TEs

Given the presence of mechanisms of silencing and sequence elimination, it is surprising to observe that few plant genomes are void of TEs. The only known example is the micro alga *Micromonas pusillii* [164]. Even the minute plant genome of *Utricularia gibba* contains 3% TEs with some intact LTR elements [89]. So there must be some mechanisms to routinely generate new active TEs or new functional transposase. One mechanism to introduce new TEs is through horizontal transfer, which allows TEs to move from one genome to another. Evidence for horizontal transposon transfer (HTT) includes phylogenetic incongruence between the host and TE phylogenies, patchy TE distributions, or a high similarity between TEs from different species [165].

HTT is common in animals, and it is often mediated by parasite–host interactions (reviewed by Schaack et al. [166]). In contrast, there are relatively few HTT events reported from plants. The first unambiguous evidence for HTT event in plants was provided by Diao et al. who showed that a MULE was transferred between *Setaria* and rice [167]. Two successive studies reported multiple HTT events for LTR retrotransposons within the genus *Oryza* [168] and within the Poaceae [169]. In addition, it was proposed that the tomato *Rider* LTR element, which was involved in the formation of the *SUN* locus (see Section 4.2.3), was a result of HTT from Arabidopsis to tomato [170]. Recently, Baidouri et al. annotated LTR retrotransposons in 40 plant genomes and detected 32 HTT events concerning 26 genomes (65%) [171]. HTT events were detected between distantly related species such as palm (*Phoenix dactylifera*) and grape, tomato and bean (*Phaseolus vulgaris*), or poplar (*Populus trichocarpa*) and peach (*Prunus persica*) [171]. This was the first systematic survey about HTT in plants which was enabled by the availability of genomic sequences from multiple plants. It suggests that HTT for LTR retrotransposons is widespread in plants. Compared with retrotransposons, it is unclear how common HTT is for DNA transposons in plants. Given the fact that the content of DNA transposons is extremely low in several plant species (see Section 3.4) and an entire superfamily is missing from some genomes (Section 2), the HTT concerning DNA transposons may not be as frequent as that for retrotransposons.

In summary, the birth and amplification of TEs is a dynamic process where TEs co-evolve with the host genome and other TEs. Different elements employ distinct strategies for success; some non-autonomous TEs amplify by partnering with multiple autonomous elements and others maintain the partnership by continuous sequence exchange with their autonomous counterparts. Transposition is not a true "copying" process. Accompanied with transposition and amplification is the formation of new subfamilies, including large deletion derivatives, very small MITEs or TRIMs, and non-autonomous elements containing host genes or gene fragments (Fig. 4). During amplification, new insertions interact with their adjacent genes or unlinked genes, influence the gene expression, or become a part of the gene. When a TE reaches a certain copy number, it triggers the activity of host defense system, and the element gradually loses transposition activity. Most family members will be degenerated and eliminated from the genome; however, a subset retains the coding capacity. Some of these elements then become the source for new active TEs. Others, both autonomous and non-autonomous TEs, assume cellular functions and likely lose transposition activity (Fig. 4). After loss of their signature sequences such as TIRs or LTRs, the cellular genes derived from autonomous TEs remain recognizable due to the similarity with transposase; genes derived from non-autonomous TEs would be indistinguishable from normal genes. Since non-autonomous elements are much more abundant than their autonomous counterparts, it is likely that many of the "normal" genes were once non-autonomous TEs. With their ability to amplify, TEs provide endless possibilities for modification of existing genes and generation of new genes.

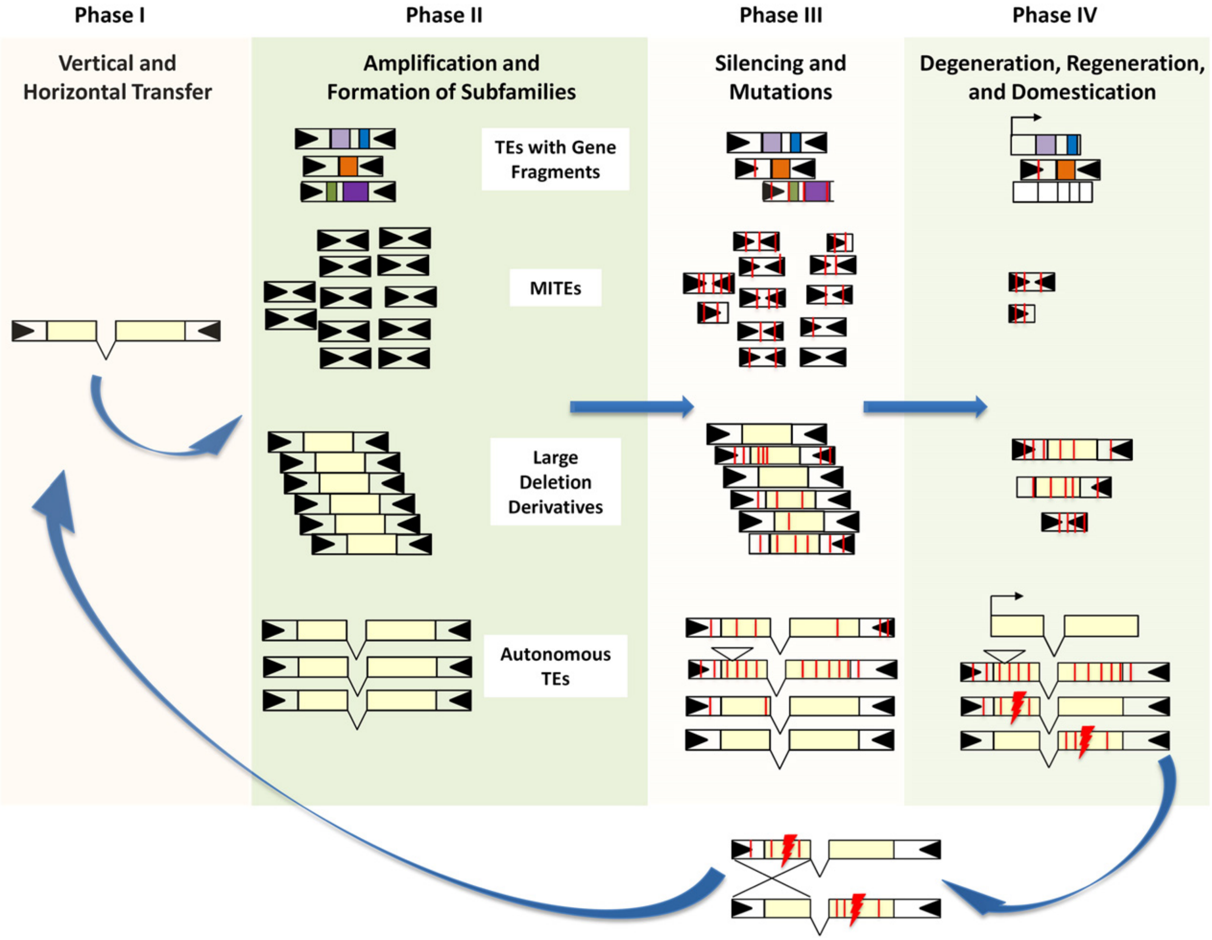

Fig. 4. Proposed life cycle of a generic DNA transposon. A new active DNA transposon forms through horizontal transfer or swapping of existing elements. The element transposes (Phase I) and increases its copy number, during which different subfamilies may be formed (e.g., some retain the original sequences; some may experience deletions of partial or entire internal sequences; some may capture fragments of protein coding genes) (Phase II). The rapid amplification triggers the host surveillance system, which renders them inactive by epigenetically silencing them, especially the autonomous elements (Phase III). After loss of activity, mutations, degradations, and nested TE insertions occur gradually, resulting in copies that are indistinguishable from normal genes, various degrees of truncation and degeneration (Phase III, IV). Sequence swapping between defective copies may result in the generation of a new functional TE, which allows the continuous cycle of birth and death of TEs. Elements are diagrammed as in Fig. 1. Nested insertions are shown as open triangles. Red lines indicate small mutations. Lightening symbols indicate interruption of coding regions. TEs: transposable elements; MITEs: miniature inverted repeat transposable elements.

Acknowledgments

We thank Dr. Wayne Loescher (Michigan State University) for critical reading of the manuscript. Research in Jiang's lab was supported by NSF grants MCB-1121650 and IOS-1126998, USDA National Institute of Food and Agriculture, and AgBioResearch at Michigan State University.

References

[1] B. McClintock, Mutable loci in maize, Carnegie Institution of Washington Yearbook, 47 1948, pp. 155–169.

[2] N. Fedoroff, S. Wessler, M. Shure, Isolation of the transposable maize controlling elements Ac and Ds, Cell 35 (1983) 235–242.

[3] U. Courage-Tebbe, H.P. Doring, N. Fedoroff, P. Starlinger, The controlling element Ds at the shrunken locus in *Zea mays*: structure of the unstable sh-m5933 allele and several revertants, Cell 34 (1983) 383–393.

[4] U. Courage, H.P. Doring, W.B. Frommer, R. Kunze, A. Laird, A. Merckelbach, M. Muller-Neumann, J. Riegel, P. Starlinger, E. Tillmann, et al., Transposable elements Ac and Ds at the shrunken, waxy, and alcohol dehydrogenase 1 loci in *Zea mays* l, Cold Spring Harb. Symp. Quant. Biol. 49 (1984) 329–338.

[5] R.F. Pohlman, N.V. Fedoroff, J. Messing, The nucleotide sequence of the maize controlling element activator, Cell 37 (1984) 635–643.

[6] S.A. Goff, D. Ricke, T.H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchison, C. Martin, F. Katagiri, B.M. Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W.L. Sun, L. Chen, B. Cooper, S. Park, T.C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A. Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R.M. Miller, S. Bhatnagar, N. Adey, T. Rubano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalma, A. Oliphant, S. Briggs, A draft sequence of the rice genome (*Oryza sativa* L. *ssp. japonica*), Science 296 (2002) 92–100.

[7] J. Yu, S. Hu, J. Wang, G.K. Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W. Tong, L. Cong, J. Geng, Y. Han, L. Li, W. Li, G. Hu, J. Li, Z. Liu, Q. Qi, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Xu, K. Zhang, X. Zheng, J. Dong, W. Zeng, L. Tao, J. Ye, J. Tan, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, W. Zhao, P. Li, W. Chen, Y. Zhang, J. Hu, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, M. Tao, L. Zhu, L. Yuan, H. Yang, A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*), Science 296 (2002) 79–92.

[8] International-Rice-Sequencing-Project, The map-based sequence of the rice genome, Nature 436 (2005) 793–800.

[9] N. Jiang, Z. Bao, X. Zhang, H. Hirochika, S.R. Eddy, S.R. McCouch, S.R. Wessler, An active DNA transposon family in rice, Nature 421 (2003) 163–167.

[10] K. Kikuchi, K. Terauchi, M. Wada, H.Y. Hirano, The plant MITE mPing is mobilized in anther culture, Nature 421 (2003) 167–170.

[11] T. Nakazaki, Y. Okumoto, A. Horibata, S. Yamahira, M. Teraishi, H. Nishida, H. Inoue, T. Tanisaka, Mobilization of a transposon in the rice genome, Nature 421 (2003) 170–172.

[12] A. Kumar, J.L. Bennetzen, Plant retrotransposons, Annu. Rev. Genet. 33 (1999) 479–532.

[13] R. Kunze, H. Saedler, W.E. Lonnig, Plant transposable elements, Adv. Bot. Res. 27 (1997) 331–440.

[14] T. Wicker, F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman, A unified classification system for eukaryotic transposable elements, Nat. Rev. Genet. 8 (2007) 973–982.

[15] J.D. Boeke, V.G. Corces, Transcription and reverse transcription of retrotransposons, Annu. Rev. Microbiol. 43 (1989) 403–434.

[16] T.H. Eickbush, V.K. Jamburuthugoda, The diversity of retrotransposons and the properties of their reverse transcriptases, Virus Res. 134 (2008) 221–234.

[17] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.

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

[19] R. Velasco, A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S.K. Bhatnagar, M. Troggio, D. Pruss, S. Salvi, M. Pindo, P. Baldi, S. Castelletti, M. Cavaiuolo, G. Coppola, F. Costa, V. Cova, A. Dal Ri, V. Goremykin, M. Komjanc, S. Longhi, P. Magnago, G. Malacarne, M. Malnoy, D. Micheletti, M. Moretto, M. Perazzolli, A. Si-Ammour, S. Vezzulli, E. Zini, G. Eldredge, L.M. Fitzgerald, N. Gutin, J. Lanchbury, T. Macalma, J.T. Mitchell, J. Reid, B. Wardell, C. Kodira, Z. Chen, B. Desany, F. Niazi, M. Palmer, T. Koepke, D. Jiwan, S. Schaeffer, V. Krishnan, C. Wu, V.T. Chu, S.T. King, J. Vick, Q. Tao, A. Mraz, A. Stormo, K. Stormo, R. Bogden, D. Ederle, A. Stella, A. Vecchietti, M.M. Kater, S. Masiero, P. Lasserre, Y. Lespinasse, A.C. Allan, V. Bus, D. Chagne, R.N. Crowhurst, A.P. Gleave, E. Lavezzo, J.A. Fawcett, S. Proost, P. Rouze, L. Sterck, S. Toppo, B. Lazzari, R.P. Hellens, C.E. Durel, A. Gutin, R.E. Bumgarner, S.E. Gardiner, M. Skolnick, M. Egholm, Y. Van de Peer, F. Salamini, R. Viola, The genome of the domesticated apple (Malus x domestica Borkh.), Nat. Genet. 42 (2010) 833–839.

[20] A. D'Hont, F. Denoeud, J.M. Aury, F.C. Baurens, F. Carreel, O. Garsmeur, B. Noel, S. Bocs, G. Droc, M. Rouard, C. Da Silva, K. Jabbari, C. Cardi, J. Poulain, M. Souquet, K. Labadie, C. Jourda, J. Lengelle, M. Rodier-Goud, A. Alberti, M. Bernard, M. Correa, S. Ayyampalayam, M.R. McKain, J. Leebens-Mack, D. Burgess, M. Freeling, A.M.D. Mbeguie, M. Chabannes, T. Wicker, O. Panaud, J. Barbosa, E. Hribova, P. Heslop-Harrison, R. Habas, R. Rivallan, P. Francois, C. Poiron, A. Kilian, D. Burthia, C. Jenny, F. Bakry, S. Brown, V. Guignon, G. Kema, M. Dita, C. Waalwijk, S. Joseph, A. Dievart, O. Jaillon, J. Leclercq, X. Argout, E. Lyons, A. Almeida, M. Jeridi, J. Dolezel, N. Roux, A.M. Risterucci, J. Weissenbach, M. Ruiz, J.C. Glaszmann, F. Quetier, N. Yahiaoui, P. Wincker, The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants, Nature 488 (2012) 213–217.

[21] J.C. Dohm, A.E. Minoche, D. Holtgrawe, S. Capella-Gutierrez, F. Zakrzewski, H. Tafer, O. Rupp, T.R. Sorensen, R. Stracke, R. Reinhardt, A. Goesmann, T. Kraft, B. Schulz, P.F. Stadler, T. Schmidt, T. Gabaldon, H. Lehrach, B. Weisshaar, H. Himmelbauer, The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*), Nature 505 (2014) 546–549.

[22] R. Ming, R. Vanburen, Y. Liu, M. Yang, Y. Han, L.T. Li, Q. Zhang, M.J. Kim, M.C. Schatz, M. Campbell, J. Li, J.E. Bowers, H. Tang, E. Lyons, A.A. Ferguson, G. Narzisi, D.R. Nelson, C.E. Blaby-Haas, A.R. Gschwend, Y. Jiao, J.P. Der, F. Zeng, J. Han, X.J. Min, K.A. Hudson, R. Singh, A.K. Grennan, S.J. Karpowicz, J.R. Watling, K. Ito, S.A. Robinson, M.E. Hudson, Q. Yu, T.C. Mockler, A. Carroll, Y. Zheng, R. Sunkar, R. Jia, N. Chen, J. Arro, C.M. Wai, E. Wafula, A. Spence, L. Xu, J. Zhang, R. Peery, M.J. Haus, W. Xiong, J.A. Walsh, J. Wu, M.L. Wang, Y.J. Zhu, R.E. Paull, A.B. Britt, C. Du, S.R. Downie, M.A. Schuler, T.P. Michael, S.P. Long, D.R. Ort, J.W. Schopf, D.R. Gang, N. Jiang, M. Yandell, C.W. Depamphilis, S.S. Merchant, A.H. Paterson, B.B. Buchanan, S. Li, J. Shen-Miller, Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.), Genome Biol. 14 (2013) R41.

[23] B. McClintock, The origin and behavior of mutable loci in maize, Proc. Natl. Acad. Sci. U. S. A. 36 (1950) 344–355.

[24] E.S. Coen, R. Carpenter, C. Martin, Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*, Cell 47 (1986) 285–296.

[25] W. McGinnis, A.W. Shermoen, S.K. Beckendorf, A transposable element inserted just 5' to a Drosophila glue protein gene alters gene expression and chromatin structure, Cell 34 (1983) 75–84.

[26] R. Kunze, C.F. Weil, The hAT and CACTA superfamilies of plant transposons, in: N. Craig, R. Craigie, M. Gellert, A. Lambowitz (Eds.), Mobile DNA II, America Society of Microbiology Press, Washington, D.C. 2002, pp. 565–610.

[27] V. Sundaresan, P. Springer, T. Volpe, S. Haward, J.D. Jones, C. Dean, H. Ma, R. Martienssen, Patterns of Gene action in plant development revealed by enhancer trap and gene trap transposable elements, Genes Dev. 9 (1995) 1797–1810.

[28] E. Vollbrecht, J. Duvick, J.P. Schares, K.R. Ahern, P. Deewatthanawong, L. Xu, L.J. Conrad, K. Kikuchi, T.A. Kubinec, B.D. Hall, R. Weeks, E. Unger-Wallace, M. Muszynski, V.P. Brendel, T.P. Brutnell, Genome-wide distribution of transposed dissociation elements in maize, Plant Cell 22 (2010) 1667–1685.

[29] A. Pereira, H. Cuypers, A. Gierl, Z. Schwarz-Sommer, H. Saedler, Molecular analysis of the en/spm transposable element system of *Zea mays*, EMBO J. 5 (1986) 835–841.

[30] U. Bonas, H. Sommer, H. Saedler, The 17-kb Tam1 element of *Antirrhinum majus* induces a 3-Bp duplication upon integration into the chalcone synthase Gene, EMBO J. 3 (1984) 1015–1019.

[31] G. Zabala, L. Vodkin, A putative autonomous 20.5 kb-CACTA transposon insertion in an F3'H allele identifies a new CACTA transposon subfamily in *Glycine max*, BMC Plant Biol. 8 (2008) 124.

[32] Y. Han, S. Qin, S.R. Wessler, Comparison of class 2 transposable elements at superfamily resolution reveals conserved and distinct features in cereal grass genomes, BMC Genomics 14 (2013) 71.

[33] D.S. Robertson, Charaterization of a mutator system in maize, Mutat. Res. 51 (1978) 21–28.

[34] Z. Yu, S.I. Wright, T.E. Bureau, Mutator-like elements in *Arabidopsis thaliana*. Structure, diversity and evolution, Genetics 156 (2000) 2019–2031.

[35] Q. Wang, H.K. Dooner, Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 17644–17649.

[36] A.A. Ferguson, N. Jiang, Mutator-like elements with multiple long terminal inverted repeats in plants, Comp. Funct. Genomics 2012 (2012) 695827.
[37] H. van Leeuwen, A. Monfort, P. Puigdomenech, Mutator-like elements identified in melon, Arabidopsis and rice contain ULP1 protease domains, Mol. Gen. Genomics. 277 (2007) 357–364.
[38] A.A. Ferguson, D. Zhao, N. Jiang, Selective acquisition and retention of genomic sequences by pack-mutator-like elements based on guanine-cytosine content and the breadth of expression, Plant Physiol. 163 (2013) 1419–1432.
[39] T.E. Bureau, S.R. Wessler, Tourist: a large family of small inverted repeat elements frequently associated with maize genes, Plant Cell 4 (1992) 1283–1294.
[40] X. Zhang, C. Feschotte, Q. Zhang, N. Jiang, W.B. Eggleston, S.R. Wessler, P instability factor: an active maize transposon system associated with the amplification of tourist-like MITEs and a new superfamily of transposases, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 12572–12577.
[41] X. Zhang, N. Jiang, C. Feschotte, S.R. Wessler, PIF- and pong-like transposable elements: distribution, evolution and relationship with tourist-like miniature inverted-repeat transposable elements, Genetics 166 (2004) 971–986.
[42] V.V. Kapitonov, J. Jurka, Molecular paleontology of transposable elements from *Arabidopsis thaliana*, Genetica 107 (1999) 27–37.
[43] T.E. Bureau, S.R. Wessler, Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants, Plant Cell 6 (1994) 907–916.
[44] C. Feschotte, L. Swamy, S.R. Wessler, Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs), Genetics 163 (2003) 747–758.
[45] C. Feschotte, S.R. Wessler, Mariner-like transposases are widespread and diverse in flowering plants, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 280–285.
[46] R.H. Plasterk, Z. Izsvak, Z. Ivics, Resident aliens: the Tc1/mariner superfamily of transposable elements, Trends Genet. 15 (1999) 326–332.
[47] O. Jaillon, J.M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyere, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M.E. Pe, G. Valle, M. Morgante, M. Caboche, A.F. Adam-Blondon, J. Weissenbach, F. Quetier, P. Wincker, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, Nature 449 (2007) 463–467.
[48] J.A. Banks, T. Nishiyama, M. Hasebe, J.L. Bowman, M. Gribskov, C. dePamphilis, V.A. Albert, N. Aono, T. Aoyama, B.A. Ambrose, N.W. Ashton, M.J. Axtell, E. Barker, M.S. Barker, J.L. Bennetzen, N.D. Bonawitz, C. Chapple, C. Cheng, L.G. Correa, M. Dacre, J. DeBarry, I. Dreyer, M. Elias, E.M. Engstrom, M. Estelle, L. Feng, C. Finet, S.K. Floyd, W.B. Frommer, T. Fujita, L. Gramzow, M. Gutensohn, J. Harholt, M. Hattori, A. Heyl, T. Hirai, Y. Hiwatashi, M. Ishikawa, M. Iwata, K.G. Karol, B. Koehler, U. Kolukisaoglu, M. Kubo, T. Kurata, S. Lalonde, K. Li, Y. Li, A. Litt, E. Lyons, G. Manning, T. Maruyama, T.P. Michael, K. Mikami, S. Miyazaki, S. Morinaga, T. Murata, B. Mueller-Roeber, D.R. Nelson, M. Obara, Y. Oguri, R.G. Olmstead, N. Onodera, B.L. Petersen, B. Pils, M. Prigge, S.A. Rensing, D.M. Riano-Pachon, A.W. Roberts, Y. Sato, H.V. Scheller, B. Schulz, C. Schulz, E.V. Shakirov, N. Shibagaki, N. Shinohara, D.E. Shippen, I. Sorensen, R. Sotooka, N. Sugimoto, M. Sugita, N. Sumikawa, M. Tanurdzic, G. Theissen, P. Ulvskov, S. Wakazuki, J.K. Weng, W.W. Willats, D. Wipf, P.G. Wolf, L. Yang, A.D. Zimmer, Q. Zhu, T. Mitros, U. Hellsten, D. Loque, R. Otillar, A. Salamov, J. Schmutz, H. Shapiro, E. Lindquist, et al., The Selaginella genome identifies genetic changes associated with the evolution of vascular plants, Science 332 (2011) 960–963.
[49] AmborellaGenomeProject, The Amborella genome and the evolution of flowering plants, Science 342 (2013) 1241089.
[50] K. Turcotte, T. Bureau, Phylogenetic analysis reveals stowaway-like elements may represent a fourth family of the IS630-Tc1-mariner superfamily, Genome 45 (2002) 82–90.
[51] Y.W. Yuan, S.R. Wessler, The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 7884–7889.
[52] V.V. Kapitonov, J. Jurka, Rolling-circle transposons in eukaryotes, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 8714–8719.
[53] M. Morgante, S. Brunner, G. Pea, K. Fengler, A. Zuccolo, A. Rafalski, Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize, Nat. Genet. 37 (2005) 997–1002.
[54] C. Du, N. Fefelova, J. Caronna, L. He, H.K. Dooner, The polychromatic Helitron landscape of the maize genome, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 19916–19921.
[55] L. Yang, J.L. Bennetzen, Distribution, diversity, evolution, and survival of helitrons in the maize genome, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 19922–19927.
[56] W. Xiong, L. He, J. Lai, H.K. Dooner, C. Du, Helitron scanner uncovers a large overlooked cache of helitron transposons in many plant genomes, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 10263–10268.
[57] R.S. Baucom, J.C. Estill, C. Chaparro, N. Upshaw, A. Jogi, J.M. Deragon, R.P. Westerman, P.J. Sanmiguel, J.L. Bennetzen, Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome, PLoS Genet. 5 (2009) e1000732.
[58] K. Hanada, V. Vallejo, K. Nobuta, R.K. Slotkin, D. Lisch, B.C. Meyers, S.H. Shiu, N. Jiang, The functional role of pack-MULEs in rice inferred from purifying selection and expression profile, Plant Cell 21 (2009) 25–38.
[59] N. Jiang, Z. Bao, X. Zhang, S.R. Eddy, S.R. Wessler, Pack-MULE transposable elements mediate gene evolution in plants, Nature 431 (2004) 569–573.
[60] P. SanMiguel, B.S. Gaut, A. Tikhonov, Y. Nakajima, J.L. Bennetzen, The paleontology of intergene retrotransposons of maize, Nat. Genet. 20 (1998) 43–45.
[61] R. Kunze, S. Kuhn, J.D.G. Jones, S.R. Scofield, Somatic and germinal activities of maize Activator (Ac) transposase mutants in transgenic tobacco, Plant J. 8 (1995) 45–54.
[62] M. Heinlein, T. Brattig, R. Kunze, In vivo aggregation of maize Activator (ac) transposase in nuclei of maize endosperm and petunia protoplasts, Plant J. 5 (1994) 705–714.
[63] D. Zhao, A. Ferguson, N. Jiang, Transposition of a rice mutator-like element in the yeast *Saccharomyces cerevisiae*, Plant Cell 27 (2015) 132–148.
[64] C.N. Hancock, F. Zhang, S.R. Wessler, Transposition of the tourist-MITE mPing in yeast: an assay that retains key features of catalysis by the class 2 PIF/harbinger superfamily, Mob. DNA 1 (2010) 5.
[65] J.C. Way, N. Kleckner, Transposition of plasmid-borne Tn10 elements does not exhibit simple length-dependence, Genetics 111 (1985) 705–713.
[66] L.R. Tosi, S.M. Beverley, cis and trans factors affecting Mos1 mariner evolution and transposition in vitro, and its potential for functional genomics, Nucleic Acids Res. 28 (2000) 784–790.
[67] S. Hennig, W. Ziebuhr, A transposase-independent mechanism gives rise to precise excision of IS256 from insertion sites in *Staphylococcus epidermidis*, J. Bacteriol. 190 (2008) 1488–1490.
[68] J. Sakai, R.M. Chalmers, N. Kleckner, Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition, EMBO J. 14 (1995) 4374–4383.
[69] H. Savilahti, P.A. Rice, K. Mizuuchi, The phage mu transpososome core: DNA requirements for assembly and function, EMBO J. 14 (1995) 4893–4903.
[70] G. Yang, D.H. Nagel, C. Feschotte, C.N. Hancock, S.R. Wessler, Tuned for transposition: molecular determinants underlying the hyperactivity of a stowaway MITE, Science 325 (2009) 1391–1394.
[71] C.P. Witte, Q.H. Le, T. Bureau, A. Kumar, Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 13778–13783.
[72] R. Kalendar, J. Tanskanen, W. Chang, K. Antonius, H. Sela, O. Peleg, A.H. Schulman, Cassandra retrotransposons carry independently transcribed 5S RNA, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 5833–5838.
[73] H. Yin, J. Du, L. Li, C. Jin, L. Fan, M. Li, J. Wu, S. Zhang, Comparative genomic analysis reveals multiple long terminal repeats, lineage-specific amplification, and frequent interelement recombination for Cassandra retrotransposon in pear (Pyrus bretschneideri Rehd.), Genome Biol. Evol. 6 (2014) 1423–1436.
[74] D. Gao, Y. Li, B. Abernathy, S. Jackson, Landscape and evolutionary dynamics of terminal-repeat retrotransposons in miniature (TRIMs) in 48 whole plant genomes, bioRxiv, 2014 1–34.
[75] D. Gao, J. Chen, M. Chen, B.C. Meyers, S. Jackson, A highly conserved, small LTR retrotransposon that preferentially targets genes in grass genomes, PLoS One 7 (2012) e32010.
[76] W. Wei, N. Gilbert, S.L. Ooi, J.F. Lawler, E.M. Ostertag, H.H. Kazazian, J.D. Boeke, J.V. Moran, Human L1 retrotransposition: cis preference versus trans complementation, Mol. Cell. Biol. 21 (2001) 1429–1439.
[77] J. Du, Z. Tian, N.J. Bowen, J. Schmutz, R.C. Shoemaker, J. Ma, Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR Swapping in soybean, Plant Cell 22 (2010) 48–61.
[78] N.L. Craig, Target site selection in transposition, Annu. Rev. Biochem. 66 (1997) 437–474.
[79] C.R. Dietrich, F. Cui, M.L. Packila, J. Li, D.A. Ashlock, B.J. Nikolau, P.S. Schnable, Maize Mu transposons are targeted to the 5' untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome, Genetics 160 (2002) 697–716.
[80] S. Liu, C.T. Yeh, T. Ji, K. Ying, H. Wu, H.M. Tang, Y. Fu, D. Nettleton, P.S. Schnable, Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome, PLoS Genet. 5 (2009) e1000733.
[81] K. Naito, F. Zhang, T. Tsukiyama, H. Saito, C.N. Hancock, A.O. Richardson, Y. Okumoto, T. Tanisaka, S.R. Wessler, Unexpected consequences of a sudden and massive transposon amplification on rice gene expression, Nature 461 (2009) 1130–1134.
[82] A. Miyao, K. Tanaka, K. Murata, H. Sawaki, S. Takeda, K. Abe, Y. Shinozuka, K. Onosato, H. Hirochika, Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome, Plant Cell 15 (2003) 1771–1780.
[83] H. Hirochika, K. Sugimoto, Y. Otsuki, H. Tsugawa, M. Kanda, Retrotransposons of rice involved in mutations induced by tissue culture, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 7783–7788.
[84] D. Zhao, N. Jiang, Nested insertions and accumulation of indels are negatively correlated with abundance of Mutator-like transposable elements in maize and rice, PLoS One 9 (2014), e87069.
[85] R. Ming, S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, J.H. Saw, P. Senin, W. Wang, B.V. Ly, K.L. Lewis, S.L. Salzberg, L. Feng, M.R. Jones, R.L. Skelton, J.E. Murray, C. Chen, W. Qian, J. Shen, P. Du, M. Eustice, E. Tong, H. Tang, E. Lyons, R.E. Paull, T.P. Michael, K. Wall, D.W. Rice, H. Albert, M.L. Wang, Y.J. Zhu, M. Schatz, N. Nagarajan, R.A. Acob, P. Guan, A. Blas, C.M. Wai, C.M. Ackerman, Y. Ren, C. Liu, J. Wang, J.K. Na, E.V. Shakirov, B. Haas, J. Thimmapuram, D. Nelson, X. Wang, J.E. Bowers, A.R. Gschwend, A.L. Delcher, R. Singh, J.Y. Suzuki, S. Tripathi, K. Neupane, H. Wei, B. Irikura, M. Paidi, N. Jiang, W. Zhang, G. Presting, A. Windsor, R. Navajas-Perez, M.J. Torres, F.A. Feltus, B. Porter, Y. Li, A.M. Burroughs, M.C. Luo, L. Liu, D.A. Christopher, S.M.

Mount, P.H. Moore, T. Sugimura, J. Jiang, M.A. Schuler, V. Friedman, T. Mitchell-Olds, D.E. Shippen, C.W. dePamphilis, J.D. Palmer, M. Freeling, A.H. Paterson, D. Gonsalves, L. Wang, M. Alam, The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus), Nature 452 (2008) 991–996.

[86] TheTomatoGenomeConsortium, The tomato genome sequence provides insights into fleshy fruit evolution, Nature 485 (2012) 635–641.

[87] S. Huang, R. Li, Z. Zhang, L. Li, X. Gu, W. Fan, W.J. Lucas, X. Wang, B. Xie, P. Ni, Y. Ren, H. Zhu, J. Li, K. Lin, W. Jin, Z. Fei, G. Li, J. Staub, A. Kilian, E.A. van der Vossen, Y. Wu, J. Guo, J. He, Z. Jia, G. Tian, Y. Lu, J. Ruan, W. Qian, M. Wang, Q. Huang, B. Li, Z. Xuan, J. Cao, Asan, Z. Wu, J. Zhang, Q. Cai, Y. Bai, B. Zhao, Y. Han, Y. Li, X. Li, S. Wang, Q. Shi, S. Liu, W.K. Cho, J.Y. Kim, Y. Xu, K. Heller-Uszynska, H. Miao, Z. Cheng, S. Zhang, J. Wu, Y. Yang, H. Kang, M. Li, H. Liang, X. Ren, Z. Shi, M. Wen, M. Jian, H. Yang, G. Zhang, Z. Yang, R. Chen, L. Ma, H. Liu, Y. Zhou, J. Zhao, X. Fang, L. Fang, D. Liu, H. Zheng, Y. Zhang, N. Qin, Z. Li, G. Yang, S. Yang, L. Bolund, K. Kristiansen, S. Li, X. Zhang, J. Wang, R. Sun, B. Zhang, S. Jiang, Y. Du, The genome of the cucumber, *Cucumis sativus* L, Nat. Genet. 41 (2009) 1275–1281.

[88] M. Dassanayake, D.H. Oh, J.S. Haas, A. Hernandez, H. Hong, S. Ali, D.J. Yun, R.A. Bressan, J.K. Zhu, H.J. Bohnert, J.M. Cheeseman, The genome of the extremophile crucifer thellungiella parvula, Nat. Genet. 43 (2011) 913–918.

[89] E. Ibarra-Laclette, E. Lyons, G. Hernandez-Guzman, C.A. Perez-Torres, L. Carretero-Paulet, T.H. Chang, T. Lan, A.J. Welch, M.J. Juarez, J. Simpson, A. Fernandez-Cortes, M. Arteaga-Vazquez, E. Gongora-Castillo, G. Acevedo-Hernandez, S.C. Schuster, H. Himmelbauer, A.E. Minoche, S. Xu, M. Lynch, A. Oropeza-Aburto, S.A. Cervantes-Perez, M. de Jesus Ortega-Estrada, J.I. Cervantes-Luevano, T.P. Michael, T. Mockler, D. Bryant, A. Herrera-Estrella, V.A. Albert, L. Herrera-Estrella, Architecture and evolution of a minute plant genome, Nature 498 (2013) 94–98.

[90] R.K. Slotkin, R. Martienssen, Transposable elements and the epigenetic regulation of the genome, Nat. Rev. Genet. 8 (2007) 272–285.

[91] D. Lisch, R.K. Slotkin, Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts, Int. Rev. Cell Mol. Biol. 292 (2011) 119–152.

[92] M. Rigal, O. Mathieu, A “mille-feuille” of silencing: epigenetic control of transposable elements, Biochim. Biophys. Acta 1809 (2011) 452–458.

[93] J.D. Hollister, L.M. Smith, Y.L. Guo, F. Ott, D. Weigel, B.S. Gaut, Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 2322–2327.

[94] M.A. Beilstein, N.S. Nagalingum, M.D. Clements, S.R. Manchester, S. Mathews, Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 18724–18728.

[95] K.M. Devos, J.K. Brown, J.L. Bennetzen, Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis, Genome Res. 12 (2002) 1075–1079.

[96] J. Ma, K.M. Devos, J.L. Bennetzen, Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice, Genome Res. 14 (2004) 860–869.

[97] Z. Tian, C. Rizzon, J. Du, L. Zhu, J.L. Bennetzen, S.A. Jackson, B.S. Gaut, J. Ma, Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? Genome Res. 19 (2009) 2221–2230.

[98] J.S. Hawkins, C.E. Grover, J.F. Wendel, Repeated big bangs and the expanding universe: Directionality in plant genome size evolution, Plant Sci. 174 (2008) 557–562.

[99] J.S. Hawkins, G. Hu, R.A. Rapp, J.L. Grafenberg, J.F. Wendel, Phylogenetic determination of the pace of transposable element proliferation in plants: copia and LINE-like elements in Gossypium, Genome 51 (2008) 11–18.

[100] J.S. Hawkins, H. Kim, J.D. Nason, R.A. Wing, J.F. Wendel, Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium, Genome Res. 16 (2006) 1252–1261.

[101] J.S. Hawkins, S.R. Proulx, R.A. Rapp, J.F. Wendel, Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 17811–17816.

[102] J.L. Bennetzen, E.A. Kellogg, Do plants have a one-way ticket to genomic obesity? Plant Cell 9 (1997) 1509–1514.

[103] E. Kejnovsky, J.S. Hawkins, C. Feschotte, Plant transposable elements: biology and evolution, in: J.F. Wendel, J. Greilhuber, J. Dolezel, I.J. Leitch (Eds.), Plant Genomes, their Residents, and their Evolutionary dynamics, Springer-Verlag Wien 2012, pp. 17–34.

[104] B. Piegu, R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D.S. Brar, S. Jackson, R.A. Wing, O. Panaud, Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in oryza australiensis, a wild relative of rice, Genome Res. 16 (2006) 1262–1269.

[105] F.D. Sauer, J.K.G. Kramer, The problem associated with the feeding of high erucic acid rapeseed oils and some fish oils to experimental animals, in: J.K.G. Kramer, F.D. Sauer, W.J. Pigden (Eds.), High and Low Erucic Acid Rapeseed Oils, Production, Usage, Chemistry and Toxicological Evaluation, Academic Press, New York 1983, pp. 253–292.

[106] F. Javidfar, B. Cheng, 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 (2013) 142.

[107] F. Zeng, B. Cheng, Transposable element insertion and epigenetic modification cause the multiallelic variation in the expression of FAE1 in *Sinapis alba*, Plant Cell 26 (2014) 2648–2659.

[108] W. Chen, N. VanOpdorp, D. Fitzl, J. Tewari, P. Friedemann, T. Greene, S. Thompson, S. Kumpatla, P. Zheng, Transposon insertion in a cinnamyl alcohol dehydrogenase Gene is responsible for a Brown midrib1 mutation in maize, Plant Mol. Biol. 80 (2012) 289–297.

[109] T. Tsukamoto, N.R. Hauck, R. Tao, N. Jiang, A.F. Iezzoni, Molecular and genetic analyses of four nonfunctional s haplotype variants derived from a common ancestral s haplotype identified in sour cherry (*Prunus cerasus* L.), Genetics 184 (2010) 411–427.

[110] A.D. McCue, S. Nuthikattu, R.K. Slotkin, Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs, RNA Biol. 10 (2013) 1379–1395.

[111] M. Arteaga-Vazquez, J. Caballero-Perez, J.P. Vielle-Calzada, A family of microRNAs present in plants and animals, Plant Cell 18 (2006) 3355–3369.

[112] K. Naito, E. Cho, G. Yang, M.A. Campbell, K. Yano, Y. Okumoto, T. Tanisaka, S.R. Wessler, Dramatic amplification of a rice transposable element during recent domestication, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 17620–17625.

[113] E. Ramirez-Parra, C. Frundt, C. Gutierrez, A genome-wide identification of E2F-regulated genes in Arabidopsis, Plant J. 33 (2003) 801–811.

[114] E. Henaff, C. Vives, B. Desvoyes, A. Chaurasia, J. Payet, C. Gutierrez, J.M. Casacuberta, Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species, Plant J. 77 (2014) 852–862.

[115] M.R. Woodhouse, F. Cheng, J.C. Pires, D. Lisch, M. Freeling, X. Wang, Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 5283–5288.

[116] J.D. Hollister, B.S. Gaut, Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression, Genome Res. 19 (2009) 1419–1428.

[117] D.S. Skibbe, J.F. Fernandes, K.F. Medzihradszky, A.L. Burlingame, V. Walbot, Mutator transposon activity reprograms the transcriptomes and proteomes of developing maize anthers, Plant J. 59 (2009) 622–633.

[118] I. Makarevitch, A.J. Waters, P.T. West, M. Stitzer, C.N. Hirsch, J. Ross-Ibarra, N.M. Springer, Transposable elements contribute to activation of maize genes in response to abiotic stress, PLoS Genet. 11 (2015) e1004915.

[119] L.E. Talbert, V.L. Chandler, Characterization of a highly conserved sequence related to mutator transposable elements in maize, Mol. Biol. Evol. 5 (1988) 519–529.

[120] S. Takahashi, Y. Inagaki, H. Satoh, A. Hoshino, S. Iida, Capture of a genomic HMG domain sequence by the En/Spm-related transposable element Tpn1 in the Japanese morning glory, Mol Gen Genet 261 (1999) 447–451.

[121] K. Turcotte, S. Srinivasan, T. Bureau, Survey of transposable elements from rice genomic sequences, Plant J. 25 (2001) 169–179.

[122] T.E. Bureau, S.E. White, S.R. Wessler, Transduction of a cellular Gene by a plant retroelement, Cell 77 (1994) 479–480.

[123] Y.K. Jin, J.L. Bennetzen, Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bs1 retroelement of maize, Plant Cell 6 (1994) 1177–1186.

[124] S. Kawasaki, E. Nitasaka, Characterization of Tpn1 family in the Japanese morning glory: En/Spm-related transposable elements capturing host genes, Plant Cell Physiol. 45 (2004) 933–944.

[125] D. Holligan, X. Zhang, N. Jiang, E.J. Pritham, S.R. Wessler, The transposable element landscape of the model legume *Lotus japonicus*, Genetics 174 (2006) 2215–2228.

[126] D. Lisch, Pack-MULEs: Theft on a massive scale, BioEssays 27 (2005) 353–355.

[127] J.V. Moran, R.J. DeBerardinis, H.H. Kazazian Jr., Exon shuffling by L1 retrotransposition, Science 283 (1999) 1530–1534.

[128] M. Negroni, H. Buc, Retroviral recombination: What drives the switch? Nat. Rev. Mol. Cell Biol. 2 (2001) 151–155.

[129] J.L. Bennetzen, P.S. Springer, The generation of *mutator* transposable element subfamilies in maize, Theor. Appl. Genet. 87 (1994) 657–667.

[130] W.R. Engels, D.M. Johnson-Schlitz, W.B. Eggleston, J. Sved, High-frequency P element loss in Drosophila is homolog dependent, Cell 62 (1990) 515–525.

[131] N. Nassif, J. Penney, S. Pal, W.R. Engels, G.B. Gloor, Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair, Mol. Cell. Biol. 14 (1994) 1613–1625.

[132] S. Yamashita, T. Takano-Shimizu, K. Kitamura, T. Mikami, Y. Kishima, Resistance to gap repair of the transposon Tam3 in *Antirrhinum majus*: a role of the end regions, Genetics 153 (1999) 1899–1908.

[133] A.P. Hsia, P.S. Schnable, DNA sequence analyses support the role of interrupted gap repair in the origin of internal deletions of the maize transposon, MuDR, Genetics 142 (1996) 603–618.

[134] P. Masson, R. Surosky, J.A. Kingsbury, N.V. Fedoroff, Genetic and molecular analysis of the Spm-dependent a-m2 alleles of the maize a locus, Genetics 117 (1987) 117–137.

[135] E. Rubin, A.A. Levy, Abortive gap repair: underlying mechanism for Ds element formation, Mol. Cell. Biol. 17 (1997) 6294–6302.

[136] X. Yan, I.M. Martinez-Ferez, S. Kavchok, H.K. Dooner, Origination of Ds elements from Ac elements in maize: evidence for rare repair synthesis at the site of Ac excision, Genetics 152 (1999) 1733–1740.

[137] S. Brunner, G. Pea, A. Rafalski, Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize, Plant J. 43 (2005) 799–810.

[138] A.E. Vinogradov, DNA helix: the importance of being GC-rich, Nucleic Acids Res. 31 (2003) 1838–1844.

[139] J. Zhang, T. Zuo, T. Peterson, Generation of tandem direct duplications by reversed-ends transposition of maize ac elements, PLoS Genet. 9 (2013) e1003691.

[140] J. Zhang, T. Zuo, D. Wang, T. Peterson, Transposition-mediated DNA re-replication in maize, Elife 3 (2014) e03724.

[141] J. Zhang, Evolution by gene duplication: an update, Trends Ecol. Evol. 18 (2003) 292–298.

[142] W. Wang, H. Zheng, C. Fan, J. Li, J. Shi, Z. Cai, G. Zhang, D. Liu, J. Zhang, S. Vang, Z. Lu, G.K. Wong, M. Long, J. Wang, High rate of chimeric gene origination by retroposition in plant genomes, Plant Cell 18 (2006) 1791–1802.

[143] H. Xiao, N. Jiang, E. Schaffner, E.J. Stockinger, E. van der Knaap, A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit, Science 319 (2008) 1527–1530.

[144] S. Wu, H. Xiao, A. Cabrera, T. Meulia, E. van der Knaap, SUN regulates vegetative and reproductive organ shape by changing cell division patterns, Plant Physiol. 157 (2011) 1175–1186.

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

[146] H. Sakai, H. Mizuno, Y. Kawahara, H. Wakimoto, H. Ikawa, H. Kawahigashi, H. Kanamori, T. Matsumoto, T. Itoh, B.S. Gaut, Retrogenes in rice (*Oryza sativa* l. *ssp. japonica*) exhibit correlated expression with their source genes, Genome Biol. Evol. 3 (2011) 1357–1368.

[147] N. Jiang, A.A. Ferguson, R.K. Slotkin, D. Lisch, Pack-mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 1537–1542.

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

[149] M.E. Hudson, D.R. Lisch, P.H. Quail, The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway, Plant J. 34 (2003) 453–471.

[150] R. Lin, L. Ding, C. Casola, D.R. Ripoll, C. Feschotte, H. Wang, Transposase-derived transcription factors regulate light signaling in Arabidopsis, Science 318 (2007) 1302–1305.

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

[152] W. Wang, W. Tang, T. Ma, Niu, J.B. Jin, H. Wang, R. Lin, A pair of light signaling factors FHY3 and FAR1 regulates plant immunity by modulating chlorophyll biosynthesis, J. Integr. Plant Biol. (2015).

[153] R. Lin, H. Wang, Arabidopsis FHY3/FAR1 gene family and distinct roles of its members in light control of Arabidopsis development, Plant Physiol. 136 (2004) 4010–4022.

[154] R.K. Cowan, D.R. Hoen, D.J. Schoen, T.E. Bureau, MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms, Mol. Biol. Evol. 22 (2005) 2084–2089.

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

[156] M.M. Babu, L.M. Iyer, S. Balaji, L. Aravind, The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons, Nucleic Acids Res. 34 (2006) 6505–6520.

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

[158] M. Knip, S. Hiemstra, A. Sietsma, M. Castelein, S. de Pater, P. Hooykaas, DAYSLEEPER: A nuclear and vesicular-localized protein that is expressed in proliferating tissues, BMC Plant Biol. 13 (2013) 211.

[159] M. Knip, S. de Pater, P.J. Hooykaas, The SLEEPER genes: A transposase-derived angiosperm-specific gene family, BMC Plant Biol. 12 (2012) 192.

[160] Y. Jiao, X.W. Deng, A genome-wide transcriptional activity survey of rice transposable element-related genes, Genome Biol. 8 (2007) R28.

[161] M. Roccaro, Y. Li, S. Masiero, H. Saedler, H. Sommer, ROSINA (RSI), a novel protein with DNA-binding capacity, acts during floral organ development in *Antirrhinum majus*, Plant J. 43 (2005) 238–250.

[162] M. Roccaro, Y. Li, H. Sommer, H. Saedler, ROSINA (RSI) is part of a CACTA transposable element, TamRSI, and links flower development to transposon activity, Mol. Gen. Genomics. 278 (2007) 243–254.

[163] D.R. Hoen, T.E. Bureau, Discovery of novel genes derived from transposable elements using integrative genomic analysis, Mol. Biol. Evol. 32 (2015) 1487–1506.

[164] A.Z. Worden, J.H. Lee, T. Mock, P. Rouze, M.P. Simmons, A.L. Aerts, A.E. Allen, M.L. Cuvelier, E. Derelle, M.V. Everett, E. Foulon, J. Grimwood, H. Gundlach, B. Henrissat, C. Napoli, S.M. McDonald, M.S. Parker, S. Rombauts, A. Salamov, P. Von Dassow, J.H. Badger, P.M. Coutinho, E. Demir, I. Dubchak, C. Gentemann, W. Eikrem, J.E. Gready, U. John, W. Lanier, E.A. Lindquist, S. Lucas, K.F. Mayer, H. Moreau, F. Not, R. Otillar, O. Panaud, J. Pangilinan, I. Paulsen, B. Piegu, A. Poliakov, S. Robbens, J. Schmutz, E. Toulza, T. Wyss, A. Zelensky, K. Zhou, E.V. Armbrust, D. Bhattacharya, U.W. Goodenough, Y. Van de Peer, I.V. Grigoriev, Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas, Science 324 (2009) 268–272.

[165] G.L. Wallau, M.F. Ortiz, E.L. Loreto, Horizontal transposon transfer in eukarya: detection, bias, and perspectives, Genome Biol. Evol. 4 (2012) 689–699.

[166] S. Schaack, C. Gilbert, C. Feschotte, Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution, Trends Ecol. Evol. 25 (2010) 537–546.

[167] X. Diao, M. Freeling, D. Lisch, Horizontal transfer of a plant transposon, PLoS Biol. 4 (2006) e5.

[168] A. Roulin, B. Piegu, R.A. Wing, O. Panaud, Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon RIRE1 within the genus oryza, Plant J. 53 (2008) 950–959.

[169] A. Roulin, B. Piegu, P.M. Fortune, F. Sabot, A. D'Hont, D. Manicacci, O. Panaud, Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in poaceae, BMC Evol. Biol. 9 (2009) 58.

[170] X. Cheng, D. Zhang, Z. Cheng, B. Keller, H.Q. Ling, A new family of Ty1-copia-like retrotransposons originated in the tomato genome by a recent horizontal transfer event, Genetics 181 (2009) 1183–1193.

[171] M. El Baidouri, M.C. Carpentier, R. Cooke, D. Gao, E. Lasserre, C. Llauro, M. Mirouze, N. Picault, S.A. Jackson, O. Panaud, Widespread and frequent horizontal transfers of transposable elements in plants, Genome Res. 24 (2014) 831–838.