PLANT TRANSPOSABLE ELEMENTS: WHERE GENETICS MEETS GENOMICS

Cédric Feschotte, Ning Jiang and Susan R. Wessler

Transposable elements are the single largest component of the genetic material of most eukaryotes. The recent availability of large quantities of genomic sequence has led to a shift from the genetic characterization of single elements to genome-wide analysis of enormous transposable-element populations. Nowhere is this shift more evident than in plants, in which transposable elements were first discovered and where they are still actively reshaping genomes.

ENDOSPERM
A triploid nutritive tissue in flowering plants.

EPIGENETIC
Any heritable change in gene expression that is not caused by a change in DNA sequence.

Departments of Plant Biology and Genetics, The University of Georgia, Athens, Georgia 30602, USA. Correspondence to S.R.W. e-mail: sue@dogwood.botany.uga.edu

DOI: 10.1038/nrg793

Transposable elements (TEs) are fragments of DNA that can insert into new chromosomal locations and often make duplicate copies of themselves in the process. With the advent of large-scale DNA sequencing, it has become apparent that, far from being a rare component of some genomes, TEs are the single largest component of the genetic material of most eukaryotes. They account for at least 45% of the human genome[1] and 50–80% of some grass genomes[2–4].

TEs were discovered in maize by Barbara McClintock more than a half century ago as the genetic agents that are responsible for the sectors of altered pigmentation on mutant kernels[5]. This discovery and the ensuing characterization of the genetic properties of TEs led to her being awarded a Nobel Prize in 1983, after TEs had been documented in the genomes of *Drosophila melanogaster*, yeast (*Saccharomyces cerevisiae*), *Escherichia coli*, *Caenorhabditis elegans* and humans (reviewed in REF. 6).

The maize kernel was ideal for analysing the interplay of TEs with host genes. An example of a TE-induced mutant phenotype is shown in FIG. 1. Most of what was known about TEs before their molecular characterization came from the analysis of similar unstable phenotypes that are expressed in the aleurone layer of the kernel, a single-cell-thick layer that is derived from the underlying starchy ENDOSPERM and that is the site of pigment biosynthesis. TE-induced alleles of several genes in the anthocyanin (pigment) biosynthetic pathway give rise to distinct patterns with respect to the size, frequency and intensity of sectors (reviewed in REF. 7). Mutant alleles such as these provided the raw material for molecular biologists to isolate active TEs from the maize genome.

As discussed in this review, active elements comprise only a tiny fraction of the TE complement of the genomes of maize and of most other multicellular organisms. However, the genomes of higher eukaryotes are filled with thousands, even millions, of seemingly inactive TEs. In the absence of a mutation caused by a TE insertion, the task of determining whether a particular TE is active, inactive or EPIGENETICALLY silenced presents a new challenge to those studying eukaryotic TEs. In response to this challenge, various protocols have been developed to analyse TEs on a whole-genome basis rather than one at a time. Some of these protocols are described in this review in the context of analysing two types of TEs that are present in very high copy number in plants: miniature inverted-repeat transposable elements (MITEs) and long terminal repeat (LTR) retrotransposons. Circumstantial evidence is rapidly accumulating that plant genomes are remarkably dynamic, largely due to the activity of TEs. Paradoxically, however, the high-copy-number elements that so dramatically contribute to genome evolution at both the interspecific and intraspecific levels seem to be transpositionally inactive. We conclude the review by describing recent studies that offer the promise of reactivating these sleeping giants so that they can be characterized to the same extent as the TEs first described by McClintock.

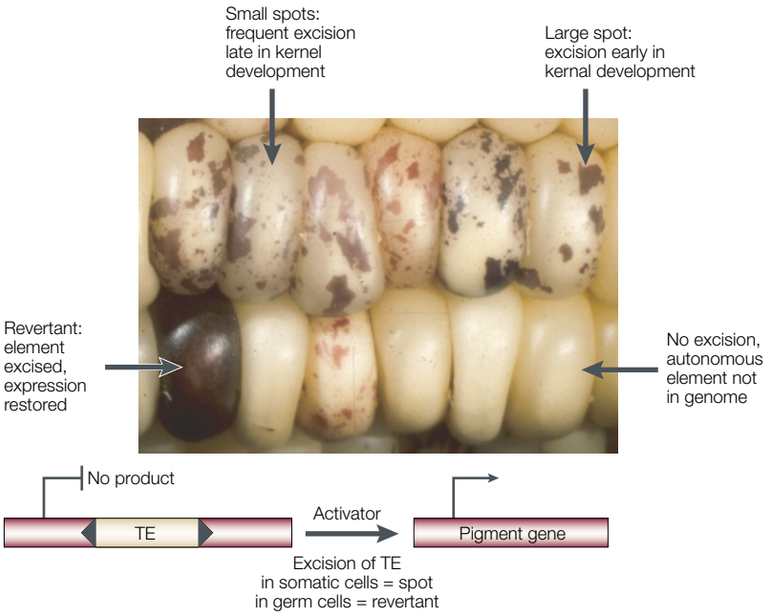


Figure 1 | **Using kernel phenotypes to study transposon behaviour.** Kernels on a maize ear show unstable phenotypes due to the interplay between a transposable element (TE) and a gene that encodes an enzyme in the anthocyanin (pigment) biosynthetic pathway. Sectors of revertant (pigmented) aleurone tissue result from the excision of the TE in a single cell. The size of the sector reflects the time in kernel development at which excision occurred. An understanding of the genetic basis of this and similar mutant phenotypes led to the discovery of TEs and to an amazingly detailed description of the behaviour of what we now call class 2 (DNA) elements (see main text for details).

Characterizing one element at a time

The first element that was recognized to be transposable was a site of chromosome breakage in maize and, as such, was named *Dissociation* (*Ds*). *Ds* could transpose or break chromosomes only in the presence of another genetic locus, called *Activator* (*Ac*), which could also promote its own transposition. Together, *Ac* and *Ds* constitute a TE family that includes autonomous (*Ac*) and non-autonomous (*Ds*) elements (for a review, see REF. 8).

Similar TE families were later found to underlie unstable mutant phenotypes in other plants (for example, snapdragon, petunia, soybean and sorghum) and in animals (such as *Drosophila* and *C. elegans*). All were DNA or class 2 elements (BOX 1a). This group is distinguished by terminal inverted repeats (TIRs) and transposition through a DNA intermediate, in which the element usually excises from one site and reinserts elsewhere in the genome (reviewed in REFS 6,9). The analysis of complex and diverse mutant phenotypes that are induced by insertion and excision of class 2 elements into plant genes revealed a myriad of ways that these elements can modify gene regulation. These include that genes or regulatory regions sustain insertions of transposon footprints (a few extra base pairs) that are usually left behind when elements excise; that some non-autonomous elements function as introns; and that insertions in promoters or other regulatory sequences can alter tissue-specific patterns of expression (for reviews, see REFS 7,10,11).

Whereas unstable mutant alleles yielded class 2 TEs, the cloning of mutations that are responsible for stable mutant alleles in maize led to the isolation of some of the first plant examples of another TE class — LTR retrotransposons (BOX 1b). In these studies, five distinct LTR retrotransposons (*Bs1*, *Stonor*, *Magellan*, *Hopscotch* and *B5/G*; see TABLE 1) were recovered from eight mutant alleles[12–14].

A common feature of all mutagenic TEs (both class 2 elements and LTR retrotransposons) was their relatively low copy number (<100 copies overall) in their host genome. This was unexpected in the light of earlier investigations, which showed, using renaturation kinetics, that highly repetitive DNA predominated in plant and animal genomes[15,16]. What was the nature of the repetitive DNA, if not these mutagenic TEs? One clue came from the analysis of repetitive DNA in humans, in which two TE families, *L1* and *Alu*, were found to be present at very high copy number[17]. *L1* and *Alu* are members of large groups called long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), respectively, and are classified as non-LTR retrotransposons (BOX 1c). Together, these two families make up more than one-quarter of the human genome[1]. LINEs and SINEs are also present in plant genomes, where they have attained very high copy numbers in some species[18,19] (TABLE 1). However, analyses of the growing database of plant genomic sequence soon revealed that plant genomes harbour a huge diversity of TEs, and that two types — MITEs and LTR retrotransposons — have made important and recent contributions to plant genome organization and evolution. For this reason, these two elements are the focus of the rest of this review.

MITEs: origin and amplification

A decade ago, the results of computer-assisted searches for repeated DNA sequences showed that small, non-autonomous elements called MITEs predominate in the non-coding regions of grass genes[20–22]. These elements were structurally reminiscent of active class 2 non-autonomous elements, with their small size (less than 600 bp) and terminal inverted repeats (BOX 1a). However, their high copy number, target-site preference (TA or TAA) and the uniformity of related elements distinguished them from the previously described class 2 elements[23]. After their discovery in plants, MITE families were found in several animal genomes, including *C. elegans*[24], mosquitoes[25–27], fishes[28], *Xenopus*[29] and humans[30,31].

The abundance of MITEs (and of DNA transposons in general) in the gene-rich regions of plant genomes was confirmed after the analysis of the complete genome sequence of *Arabidopsis* (which contains ~1,200 MITEs and ~1,000 other DNA transposons, together constituting ~6% of the genome[32]) and the partial genome sequence of rice (~100,000 MITEs and ~10,000 other DNA transposons, forming ~12% of the genome[33,34]; N.J., C.F. and S.R.W., unpublished observations).

Box 1 | **Structural features and classification of plant transposable elements**

Eukaryotic transposable elements (TEs) are divided into two classes according to whether their transposition intermediate is RNA (class 1) or DNA (class 2). For all class 1 elements, it is the element-encoded transcript (mRNA), and not the element itself (as with class 2 elements), that forms the transposition intermediate. Each group of TEs contains autonomous and non-autonomous elements. Autonomous elements have open reading frames (ORFs; red boxes) that encode the products required for transposition. Non-autonomous elements that are able to transpose have no significant coding capacity but retain the *cis*-sequences necessary for transposition. Integration of almost all TEs results in the duplication of a short genomic sequence at the site of insertion. These target-site duplications (arrows flanking the element) are variable in size and/or sequence among TE superfamilies and families. Several examples of plant TEs are given in TABLE 1.

Class 2 elements

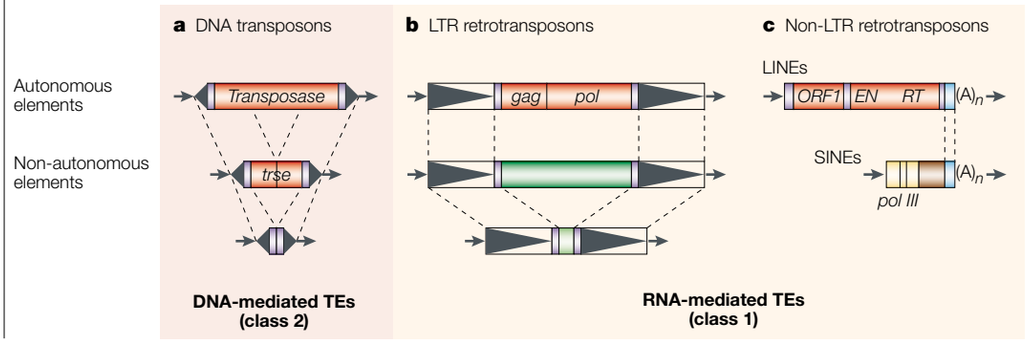
DNA transposons (**a**) have terminal inverted-repeat (black triangles) and target-site (arrows) duplications of conserved length (and sometimes sequence) in superfamilies (for example, 8 bp for *hAT*; TA for Tc1/*mariner*). Non-autonomous family members are usually derived from an autonomous family member by internal deletion.

Class 1 elements

Class 1 elements can be divided into two groups on the basis of transposition mechanism and structure. LTR retrotransposons (**b**) have long terminal repeats (LTRs) in direct orientation (black triangles). Autonomous elements contain at least two genes, called *gag* and *pol*. The *gag* gene encodes a capsid-like protein and the *pol* gene encodes a polyprotein that is responsible for protease, reverse transcriptase, RNase H and integrase activities. Non-autonomous elements lack most or all coding sequence[87–89]. Their internal region (green boxes) can be variable in size and unrelated to the autonomous element (for an example, see REF. 88).

Non-LTR retrotransposons (**c**) are divided into long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Coding regions include: *ORF1*, a *gag*-like protein; *EN*, endonuclease; and *RT*, reverse transcriptase. Both LINEs and SINEs terminate by a simple sequence repeat, usually poly(A). All SINEs described so far are characterized by an internal RNA *pol III* promoter (black stripes) near the 5′ end. The 3′ half of SINEs is of unknown origin, but the extreme 3′ tail of some SINEs share homology with the 3′ tail of LINEs present in the same genome. This indicates that SINEs could parasitize the transposition machinery of partner LINEs[90].

Recent studies have determined that previously described high-copy-number repeats in the genomes of *Arabidopsis* (~2% of the genome), rice and *Caenorhabditis elegans* belong to a new type of TE. Elements called *Helitrons* are unique in structure and are mostly represented by non-autonomous copies. *Helitrons* have been classified as class 2 elements because putative autonomous elements show structural and sequence similarities with bacterial rolling-circle transposons[91]. *Helitrons* are not discussed further in this review because additional data, beyond their initial discovery, has not been reported.



Although descriptions of MITE families have proliferated in the literature, their origin and mechanism of transposition have remained mysterious. Two main obstacles were initially encountered in attempts to classify MITEs with respect to existing transposons. First, none of the available MITE sequences revealed clear-cut relationships with TEs that encode known transposases (autonomous elements). Second, and perhaps more significantly, no MITE family has been shown to be actively transposing. In the absence of coding sequences and activity, it has been difficult to determine how MITEs originate and how they attain their high copy numbers.

This situation has changed markedly in the past few years, as MITE classification and relationships with transposases have become established. First, most of the tens of thousands of MITEs in plant genomes can now be divided into two groups — *Tourist*-like MITEs and *Stowaway*-like MITEs — on the basis of the similarity of their TIRs and target-site duplications (TSDs)[23,33–35]. Second, two approaches have been used to establish relationships between each group and newly discovered plant-transposase families (reviewed in REF. 36). We call these the 'bottom-up' and 'top-down' approaches, as the former uses the MITE sequence to identify the partner transposase, whereas the latter connects an active source of transposase to a MITE family.

Table 1 | **Examples of transposable elements in plants**

Class/subclass/ superfamily	Species	Autonomous member(s)	Non-autonomous member(s)	Copy number of the entire family*	References
Class 1					
Non-LTR retrotransposons					
LINEs; *L1* clade	*Z. mays*	*Cin4*	–	50–100	98
	L. speciosum	*Del2*‡	–	250,000	19
	A. thaliana	*Tal1*‡	–	1–6	99
SINEs	*N. tabacum*	–	*TS*	50,000	18
	B. napus	–	*S1*	500	100
LTR retrotransposons					
copia-like	*N. tabacum*	*Tnt1A*	–	>100	50
	N. tabacum	*Tto1*	–	30 (300)	60
	Hordeum sp.	BARE-1	–	5,000–22,000	3
	O. sativa	*Tos17*	–	2–5 (30)	61
	Z. mays	*Hopscotch*	–	5–8	101
	Z. mays	*Opie-2*‡	–	100,000	4
	Z. mays	–	*BS1*	1–5	87
gypsy-like	*Z. mays*	*Magellan*	–	4–8	102
	Z. mays	*Huck-2*‡	–	200,000	4
	O. sativa	*RIRE2* ?	*Dasheng*	1,200	88
	A. thaliana	*Athila 4*‡	–	22	85
	A. thaliana	*Athila 6*‡	–	11	85
	A. thaliana	*Ta3*	–	1	103
	A. thaliana	*Tar17*	–	2	60
Class 2					
DNA transposons					
hAT	*Z. mays*	*Ac*	*Ds*	50–100	7
CACTA	*Z. mays*	*Spm*	*dSpm*	50–100	104
	A. thaliana	*CAC1*	*CAC2*	4 (20)	83
Mutator	*Z. mays*	*MuDR*	*Mu1*	10–100	105
	A. thaliana	*AtMu1*	–	1 (4)	82
PIF/*Harbinger*	*Z. mays*	*PIFa*	*mPIF*	6,000	40
	Angiosperms	*PIF*-like	*Tourist*-like	Variable	40,41
Tc1/*mariner*	Angiosperms	*MLE*s	*Stowaway*-like	Variable	33,38

*Copy numbers are approximate; those in parentheses result from transpositional activation in cell culture (*Tos17* and *Tto1*) or in mutant backgrounds (*CAC* and *AtMu1*). ‡Copies identified so far seem to be full-length, but do not contain an intact open reading frame. *Ac*, *Activator*; *Ds*, *Dissociation*; LINE, long interspersed nuclear element; LTR, long terminal repeat; *MLE*, *mariner*-like element; *mPIF*, miniature *P* instability factor; *PIF*, *P* instability factor; SINE, short interspersed nuclear element; *Spm*, *Suppressor–Mutator*. Species names: *A. thaliana*, *Arabidopsis thaliana*; *B. napus*, *Brassica napus*; *L. speciosum*, *Lilium speciosum*; *N. tabacum*, *Nicotiana tabacum*; *O. sativa*, *Oryza sativa*; *Z. mays*, *Zea mays*.

Bottom-up approach. This method relies completely on the vast amount of DNA sequence generated by genome projects and the development of database search tools (TABLE 2) to identify potentially autonomous family members that are related to MITE sequences. In the simplest case, a MITE shares sequence similarity over its entire length with a larger, transposase-encoding element. This situation indicates that MITEs, like previously described non-autonomous DNA elements, originate from autonomous elements. In other cases, sequence similarity between the MITE and a DNA transposon is restricted to their TIRs and TSDs (see REF. 36 for further discussion).

The bottom-up approach has revealed complex relationships between dozens of MITE families and the well-characterized Tc1/*mariner* superfamily of transposases[24,31,36,37]. Although Tc1/*mariner* transposons were thought to be rare in plants, recent studies indicate that they are widespread in plant genomes, in which they have given rise to a large fraction of plant MITEs, including those of the *Stowaway* superfamily[33,37,38].

Top-down approach. A different method was used to reveal the origin and transposase source of *Tourist*-like elements, the other principal MITE superfamily in plants. The top-down approach used in the discovery of the transposase source began with the fortuitous discovery that a genetically active DNA transposon family in maize (called *PIF*; *P* instability factor[39]) is related in sequence to a vast group of *Tourist*-like MITEs (called *mPIF*; miniature *PIF*[40]). *PIF* and *mPIF* have identical 14-bp TIRs and similar subterminal sequences (~70% identity over ~100 bp at each end), and have the same target-site preference for the 9-bp palindrome CWCTTAGWG (in which W stands for either A or T). Whereas the extent of sequence similarity alone indicates that *mPIF* was probably derived from the active *PIF* element or from a closely related element, their identical, extended target sites provided the strongest evidence that both elements were mobilized by the same or a related transposase[40]. This transposase was identified through the isolation of an element (called *PIFa*) that co-segregated with *PIF* activity[40].

Table 2 | **Examples of programs used for mining transposable elements in DNA sequences**

Type of search	Application	Advantages and drawbacks	Programs	URL or reference
Homology based				
Query* versus database‡	Primarily for masking§ and annotating repetitive sequences	• All repeats processed with one search • A known repeat library required • Only elements with sequence similarity to known elements will be identified	CENSOR RepeatMasker	http://www.girinst.org/Censor_Server.html http://repeatmasker.genome.washington.edu
Query* versus query	For recovering all types of repetitive sequences	• No previous knowledge about TEs in the query are needed • Computationally intensive • Further classification required	Miropeats REPuter RECON	106 107 http://www.genetics.wust.edu/eddy/recon
Structure based				
	Designed to find MITEs by searching IRs flanked by TSDs of defined length and/or sequence	• Fast processing but only elements with defined features will be found	Find-MITE	http://www.biochem.vt.edu/aedes

*Query refers to sequence of interest. ‡Database refers to an existing repeat database such as Repbase[108] (http://www.girinst.org). §Sequences of repeats identified in the query are replaced by N or X; in this way, they will be ignored in subsequent sequence analyses. IR, inverted repeat; MITE, miniature inverted-repeat transposable element; TE, transposable element; TSD, target-site duplication.

The discovery of *PIFa* in maize led to the recognition of a new superfamily of transposases, called *PIF/Harbinger*[40]. These transposases are probably of ancient origin as they are distantly related to bacterial IS*5* transposases, and have been identified in a wide range of eukaryotes, including various flowering plants, a fungus and nematodes[40,41]. Once *PIF*-like DNA elements were uncovered in the genomic sequences of rice, *Arabidopsis* and nematodes, it was not difficult to identify their associated *Tourist*-like MITEs by searching their respective genomic sequences for non-autonomous members[40,42].

Origin and amplification. On the basis of these recent data, a model for the origin and amplification of MITEs has been proposed[36] (FIG. 2). In this model, MITEs are seen as non-autonomous elements that originated from DNA transposons in a two-step process. First, transposition of an autonomous element gives rise to various, internally deleted, non-autonomous derivatives (FIG. 2). This step is very likely to be dependent on transposase and has been observed for other class 2 families, such as *Ac/Ds* or *P*-elements[43,44]. In a second step, it is proposed that some derivatives can (for unknown reasons) amplify to high copy numbers.

The very high copy numbers attributed to many MITE families might actually result from independent amplifications of different subfamilies in the same genome (diversification step in FIG. 2). This is best seen in rice, in which *Stowaway* MITEs account for more than 2% of genomic DNA, but can in fact be divided into more than 35 different but related subfamilies. Only a small number of these subfamilies have attained copy numbers that are significantly greater than 2,000 (N.J., C.F. and S.R.W., unpublished data). A complex subfamily structure is also observed for *Tourist* MITEs (which make up ~3% of the rice genome). Accordingly, each *Stowaway* and *Tourist* MITE subfamily probably arose from the activity of related, but distinct, *mariner*-like and *PIF*-like autonomous elements, respectively. This hypothesis is supported by the recent discovery that rice and other plant genomes contain a tremendous diversity of *mariner*-like and *PIF*-like transposases[38,40].

The high sequence identity observed for many MITE families indicates that these families might have spread recently throughout their respective host genomes. Consistent with this view is the fact that MITE insertion sites are frequently polymorphic with respect to their presence or absence at a particular locus between individuals of the same species[21,22,35,45,46]. Because MITE excisions seem to be extremely rare, MITE insertion polymorphisms have been successfully exploited as genetic markers in maize and rice[46–48].

MITEs are known to be preferentially located in the vicinity of genes. Many examples have been documented of MITE insertion polymorphisms in promoters, introns or 3′ flanking sequences among ORTHOLOGOUS and PARALOGOUS genes (for example, REFS 20–22,28,29,35,45). Hence, MITEs are an important factor in creating structural allelic diversity. One challenge for the future will be to determine whether, and how frequently, the sequence diversity created by MITE insertions has altered gene expression or gene products.

LTR retrotransposons in plant-genome evolution

Whereas MITEs are the predominant TE in or near plant genes, one of the surprising findings of the past decade has been that LTR retrotransposons, which are located largely in intergenic regions, are the single largest component of most plant genomes[49]. LTR retrotransposons are members of the retroelement or class 1 family, which also includes retroviruses, LINEs and SINEs (BOX 1b). For all class 1 elements, it is the element-encoded transcript (mRNA), and not the element itself (as with class 2 elements), that forms the transposition intermediate (for a review, see REF. 9).

LTR retrotransposons were first discovered in plants as sources of both spontaneous and induced mutations in maize and tobacco[12,13,50]. As with the active class 2 elements that are responsible for unstable mutations, the mutagenic LTR retrotransposons are members of low to moderately repetitive element families[4,50–52]. For example, the *Bs1* element, which was first detected as an insertion in the *alcohol dehydrogenase1* gene (*adh1*), is present in only 1–5 copies in the maize genome[12].

ORTHOLOGOUS GENES
Homologous genes that originated through speciation (for example, human and mouse β-globin).

PARALOGOUS GENES
Homologous genes that originated by gene duplication (for example, human α-globin and β-globin).

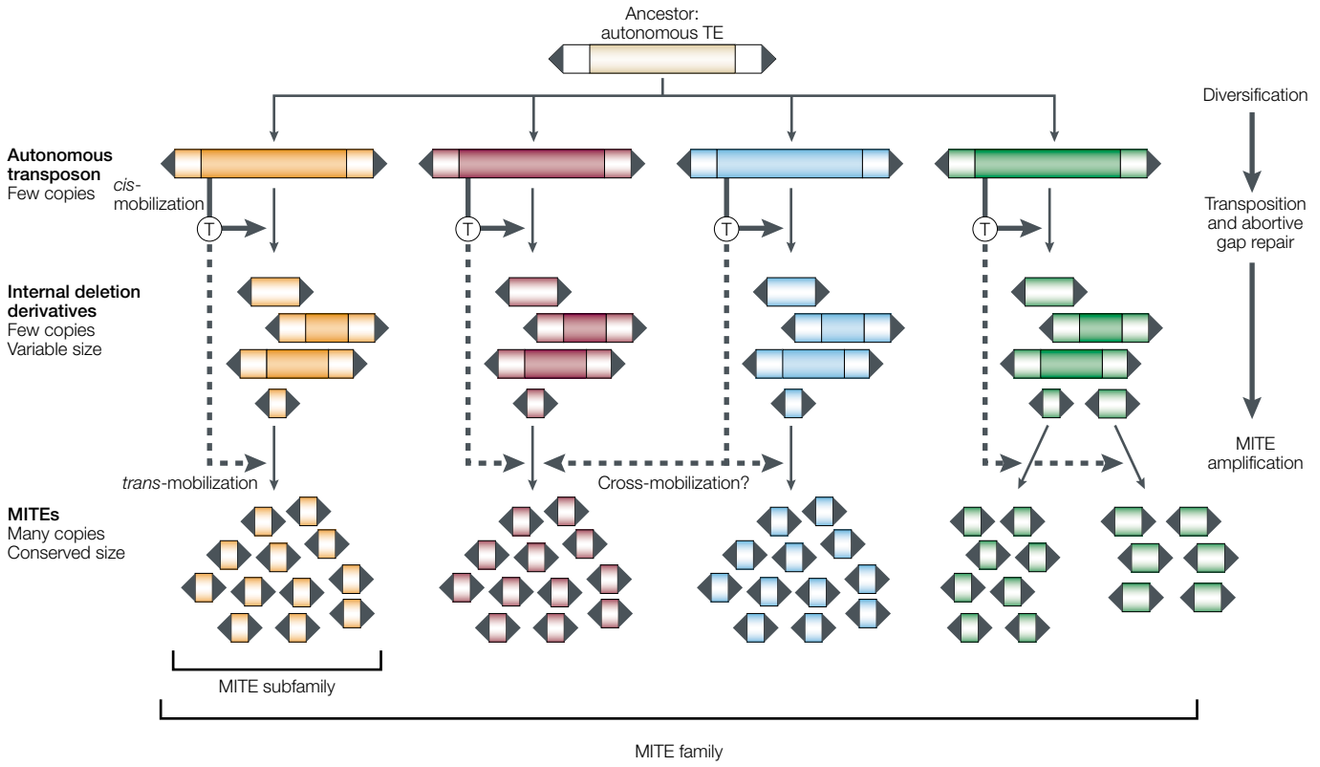


Figure 2 | **Model for the origin and amplification of MITEs.** In this model, the accumulation of miniature inverted-repeat transposable elements (MITEs) within a genome is explained by the activity of numerous related, but distinct, autonomous elements (shown in different colours). Related autonomous elements arise from a single ancestral element, but have diversified to the point at which they only share sequence similarity in their terminal inverted repeats (TIRs; black triangles) and transposase gene (boxes, in darker colour). The activity of each element, mediated by its transposase (circled T), is proposed to form non-autonomous derivatives through mechanisms such as ABORTIVE GAP REPAIR[43,44]. The subsequent amplification of one or a few deletion derivatives gives rise to a group of homogeneous non-autonomous elements (that is, a MITE subfamily). This step is likely to be mediated by the same transposase or one that is produced by a close relative ('*trans*-' or 'cross-mobilization', respectively). See text and REF. 36 for further discussion of the model. TE, transposable element.

It has been known since the late 1980s, however, that both LTR and non-LTR retrotransposons can attain phenomenally high copy numbers in plant species that have large genomes (for a review, see REF. 49). The incredible potential of amplification through retrotransposition in plants was first illustrated by studies on members of the genus *Lilium*. The 14 species that were surveyed have enormous genomes of 30,000–45,000 Mb, the size of which seems to result from massive amplification of retrotransposons[19,53].

An important series of recent studies has shown that differential amplification of LTR retrotransposons largely accounts for the 'C-value paradox' among the agronomically important members of the grass clade. The C-value paradox is the observed lack of correlation between increases in DNA content and the complexity of an organism[54]. This paradox has been documented for both plant and animal species, but so far seems to be 'solved' only for members of the grass family. In this family, the fraction of the genome contributed by LTR retrotransposons increases with genome size from rice, the smallest characterized grass genome (~15% of its 430-Mb genome consists of LTR retrotransposons[34]), to maize (~2,800 Mb, 50–80% retrotransposons[2,4]) and barley (~4,800 Mb, >70% retrotransposons[3]).

Two studies in particular have raised the bar on our concept of the dynamic genome and have positioned the grass clade as a focal point for future studies. In a classic study, Jeff Bennetzen and co-workers[55] analysed a 280-kb region around the maize *adh1* gene and found that nested LTR retrotransposons accounted for most of this sequence. This clustering of LTR retrotransposons in intergenic regions was shown to be representative of the rest of the genome. Their initial observation was dramatically followed up with the demonstration that bursts of LTR retrotransposon activity have doubled the maize genome within the past 6 million years (Myr)[56]. The temporal component to their analysis was made possible by exploiting the fact that the LTRs of a single element are identical on insertion. By comparing the two LTR sequences of a single element, they were able to estimate the insertion time (FIG. 3). This result showed for the first time that TEs could rapidly restructure a genome.

ABORTIVE GAP REPAIR
The double-stranded DNA break left at the site of excision of a class 2 transposon is repaired by the host machinery (gap repair). This gap is sometimes repaired by making an identical copy of the excised transposon, using the element still present on the sister chromatid or the homologous chromosome as a template. The process can be incomplete due to slippage and mispairing events. Such aberrant repairs are commonly responsible for the origin of an internally deleted copy of the excised transposon.

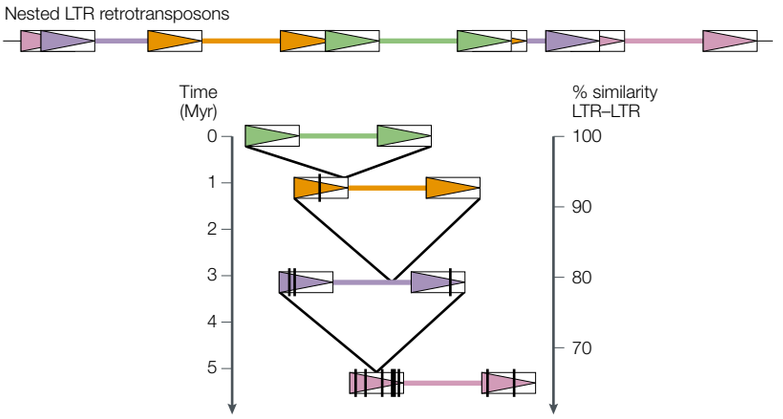


Figure 3 | **Estimating the time of retrotransposon insertion.** At the time of insertion, the long terminal repeats (LTRs) of an element are identical because both are copied from the same template during cDNA synthesis. As time passes, nucleotide changes accumulate in each LTR (represented by vertical bars in the LTRs). If the average rate of nucleotide substitution per year is known for the host organism, then sequence divergence between the LTRs provides an estimate of when insertion occurred. This method has been applied to date the insertions of LTR retrotransposons nested in the intergenic regions that surround the maize *alcohol dehydrogenase1* (*adh1*) gene[56]. See text for details. Myr, million years.

In a second study, Kalendar *et al.* presented a striking example of TE-mediated genome restructuring in populations of the wild barley *Hordeum spontaneum*[57]. In this case, genome restructuring takes the form of pronounced intraspecific genome size variation due to amplification of the BARE-1 LTR retrotransposon. The copy number of BARE-1 among nearby populations that are subjected to different levels of water stress varied between 8,300 and 22,100, corresponding to 1.8–4.7% of the nuclear DNA[57]. The correlation between BARE-1 copy number, genome size and local environmental conditions indicated that a mechanistic connection might exist between the amplification of a particular TE and the adaptive evolution of its host.

Plant TEs: so many elements, so little activity

The phylogenetic analysis of human TEs has revealed that marked amplification ceased almost 35–50 Myr ago[1]. (For a description of some of the methods available for reconstructing TE phylogenies, see BOX 2.) It is, therefore, not surprising that, despite comprising almost 45% of the genome (more than 1,200 Mb), only a handful of human TEs remain active. For example, only 30–60 *L1* elements out of 500,000 in this LINE family are thought to be active at present[1,58]. By contrast, the comparative analysis of closely related plant genomes (particularly those of the grasses) has revealed that TE amplification is responsible for a large fraction of the variation in genome size at both interspecific and intraspecific levels[2,57]. Such circumstantial evidence for massive TE activity in the recent past might lead one to expect that active TEs (especially active LTR retrotransposons) — the 'smoking guns' of transposition — could be routinely isolated from plants. In this regard, it is paradoxical that so few active retroelements have been identified. Here, we address this issue by turning our attention to the regulation of TE activity in plant genomes.

PROTOPLAST
A cell after the removal of its cell wall.

Transcriptional activation of LTR retrotransposons. Some LTR retrotransposons that are largely inactive during development can be transcriptionally activated and transpose under conditions of biotic and abiotic stress. The first active plant retrotransposon, *Tnt1*, was isolated from tobacco by selecting for insertions into the previously cloned nitrate reductase (*NR*) gene; these insertions were identified by the chlorite-resistant phenotype that is characteristic of *NR*-deficient cells or plants. Screening millions of plants for such a rare event is physically impossible. Instead, Grandbastien *et al.*[50] used microbiological techniques to screen millions of PROTOPLASTS (which can be routinely regenerated into healthy tobacco plants) for the chlorite-resistant phenotype. It turned out that *Tnt1* transcription (which is necessary to produce the transposition intermediate, mRNA) is largely quiescent during normal development, but was fortuitously activated when exposed to the fungal extract used to degrade the cell wall to make protoplasts[59]. Subsequent to this pioneering study, two additional tobacco elements, *Tto1* and *Tto2*, and a rice element, *Tos17*, were isolated using a simpler reverse-transcription PCR protocol that was devised to amplify a conserved reverse transcriptase domain in mRNAs that were isolated during cell culture[60,61]. So far, tobacco retrotransposons *Tnt1* and *Tto1* have been shown to be activated by various abiotic and biotic stresses, including wounding, oxidative stress, pathogen infection and microbial elicitors (reviewed in REFS 49,62).

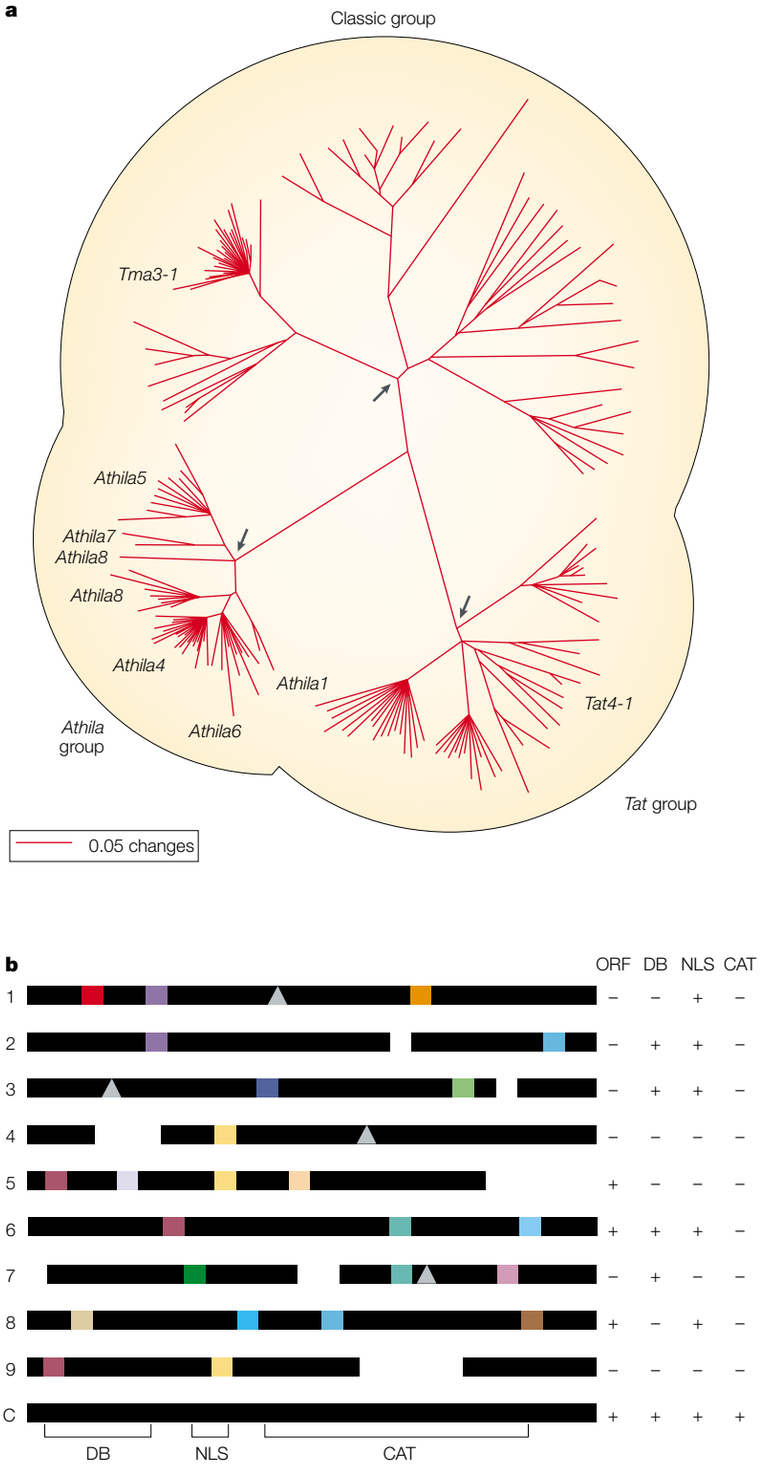
The restricted transcription of LTR retrotransposons in plants is also illustrated by their under-representation in expressed sequence tag (EST) collections. A recent analysis of 407,000 maize ESTs revealed that only 56 are derived from LTR retrotransposons[4]. Furthermore, most of these sequences are derived from the low-to-middle repetitive LTR retrotransposons, and not from the very high copy number elements (including *Huck-2*, ~200,000 copies, and *Opie-2*, ~100,000 copies) that have been responsible for doubling the size of the maize genome in the past 5–6 million years[4,56].

Transcription of LTR retrotransposons does not necessarily correlate with new insertions in the genome (transposition). The replication cycle for LTR retrotransposons includes four steps: transcription, translation, reverse transcription and integration of element cDNA. Regulation at any of these steps can limit the transposition rate. For example, transcripts of the yeast Ty1 retrotransposon are abundant, but new insertions are extremely rare, largely because only one Ty1 cDNA, on average, is made for every 14,000 Ty1 transcripts[63].

The limited data on the activity of plant LTR retrotransposons indicate that transposition is regulated primarily at the level of transcription initiation[64]. After stress-induced transcription, the genomic copy number of the rice LTR retrotransposon *Tos17* increased from 2 to more than 30 copies in some strains, as measured by Southern blots[61]. The relatively high copy number of *Tnt1* in the tobacco genome (>100 copies per haploid genome, before induction)[50] precluded the use of Southern blots to assess whether transcriptional activation led to an increase in genomic copy number.

Instead, a technique known as 'transposon display' (also called sequence-specific amplification polymorphism, or S-SAP[65]) was used to resolve PCR products that were derived from *Tnt1* insertions from tobacco genomic DNA before and after induction (FIG. 4). In this way, Melayah *et al.*[64] showed that 38 out of 41 new DNA bands corresponded to *de novo* insertions of *Tnt1*. It is clear from this study that transposon display is a powerful technique to detect activity among moderate and highly repetitive TE families. Furthermore,

Box 2 | **Phylogenetic analyses of transposable elements**



Access to the genomic sequence of an organism offers unprecedented opportunities to study the evolutionary history of the full transposable element (TE) complement. Many sequence-analysis and phylogenetic programs are freely available on the internet (for a recent compilation, see link to Sue Wessler's lab). Two examples are presented that illustrate some of these procedures and their application to the study of TEs. A third is shown in FIG. 3.

Panel **a** shows a whole-genome phylogenetic analysis of the *Arabidopsis* Ty3/*gypsy* superfamily of LTR retrotransposons (reproduced with permission from REF. 85 © (2002) Cold Spring Harbor Laboratory Press). This phylogenetic tree was constructed from a multiple alignment of all 191 Ty3/*gypsy* reverse transcriptase (RT) amino-acid sequences in the *Arabidopsis* genome (available at http://www.public.iastate.edu/~voytas). The topology of the tree emphasizes: first, the heterogeneity of RT sequences, as reflected by the length of the main branches; second, the clustering of related elements into monophyletic groups that define distinct families (for example, *Tat* or *Athila*) and subfamilies (for example, *Tat4* or *Athila5*); third, the replicative and explosive nature of the transposition process, as revealed by the presence of several subfamilies of highly homogeneous sequence (small groups with very short branch lengths, such as *Athila4*); and fourth, the tempo and succession of family and subfamily amplification. For example, groups with short branch lengths are likely to have been amplified more recently than those with longer branch lengths (for example, the *Athila4* subfamily was amplified more recently than *Athila5*).

Panel **b** illustrates how an active transposase can be reconstructed on the basis of phylogenetic analyses of a subset of inactive TE copies. This approach has been used successfully to reconstruct autonomous elements of the Tc1/*mariner* superfamily[92,93]. A schematic alignment of transposase sequences cloned from a given organism (1–9) and the deduced consensus sequence (C) are shown. After insertion, transposase genes usually evolve as pseudogenes, quickly accumulating substitutions and insertions or deletions. Substitutions might introduce amino-acid changes (coloured squares) or stop codons (grey triangles). This can result in a disrupted open reading frame (ORF) and/or an inactive transposase. The consensus sequence is deduced from a simple majority rule from a multiple alignment of several copies and is likely to reflect the ancestral, active sequence of an element. A consensus transposase sequence can be experimentally reconstructed, expressed and tested *in vitro* and *in vivo* for each of the steps of a 'cut-and-paste' transposition event[93], including: transport into the nucleus mediated by the nuclear localization signal (NLS); sequence-specific binding to the substrate DNA as mediated by the DNA binding domain (DB); and excision and integration of transposon DNA as mediated by the catalytic domain of the transposase (CAT).

Other recent examples of computer-assisted and phylogenetic approaches to study the structural evolution and population dynamics of TEs can be found in REFS 1,32,94–97.

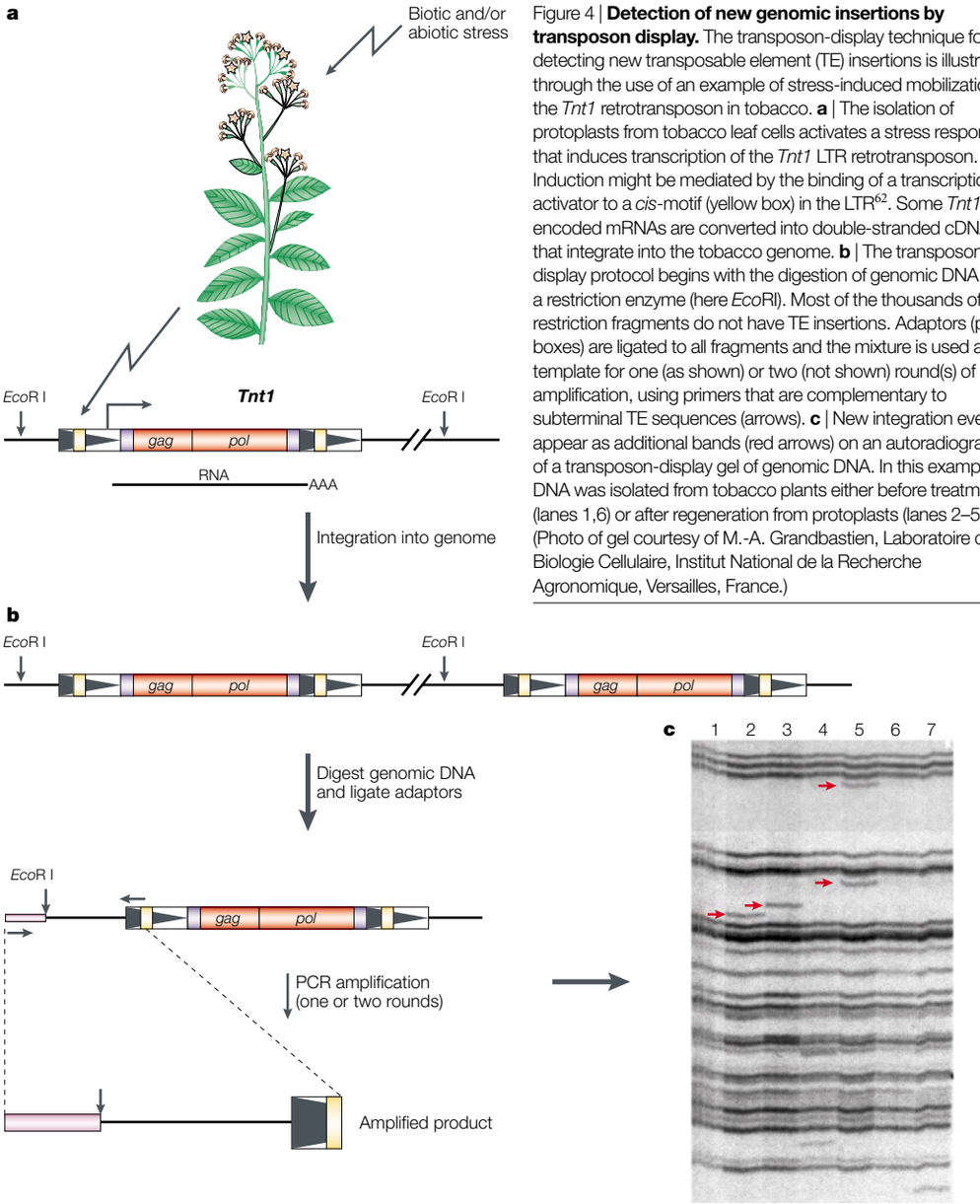


Figure 4 | **Detection of new genomic insertions by transposon display.** The transposon-display technique for detecting new transposable element (TE) insertions is illustrated through the use of an example of stress-induced mobilization of the *Tnt1* retrotransposon in tobacco. **a** | The isolation of protoplasts from tobacco leaf cells activates a stress response that induces transcription of the *Tnt1* LTR retrotransposon. Induction might be mediated by the binding of a transcriptional activator to a *cis*-motif (yellow box) in the LTR[62]. Some *Tnt1*-encoded mRNAs are converted into double-stranded cDNAs that integrate into the tobacco genome. **b** | The transposon-display protocol begins with the digestion of genomic DNA with a restriction enzyme (here *Eco*RI). Most of the thousands of restriction fragments do not have TE insertions. Adaptors (pink boxes) are ligated to all fragments and the mixture is used as template for one (as shown) or two (not shown) round(s) of PCR amplification, using primers that are complementary to subterminal TE sequences (arrows). **c** | New integration events appear as additional bands (red arrows) on an autoradiograph of a transposon-display gel of genomic DNA. In this example, DNA was isolated from tobacco plants either before treatment (lanes 1,6) or after regeneration from protoplasts (lanes 2–5,7). (Photo of gel courtesy of M.-A. Grandbastien, Laboratoire de Biologie Cellulaire, Institut National de la Recherche Agronomique, Versailles, France.)

the ability to identify newly transposed elements (by cutting out and re-amplifying bands) will allow the identification of the few active elements in a large TE family and the recovery of flanking sequences for the analysis of target-site preferences.

To understand the impact of retrotransposition on plant genome evolution, it will be necessary to identify active members of families that are present in very high copy numbers. Among the characterized elements, only two related families offer the promise of activity at present: BARE-1 and OARE-1. As mentioned above, the copy number of BARE-1 shows remarkable variation among species in the genus *Hordeum* and among populations of the wild barley species *Hordeum spontaneum*[3,57]. In addition, the detection of BARE-1 transcripts and virus-like particles is consistent with the activity of some family members[66,67]. A very close relative of BARE-1 (based on sequence identity) is OARE-1, a Ty1/*copia* LTR element in oat, with a copy number of at least 10,000 per haploid genome[68]. The recent demonstration that OARE-1 transcription is activated by biotic and abiotic stress[68] indicates that amplification of this element is responsive to environmental conditions. Unfortunately, new insertions have not yet been detected for either the BARE-1 or OARE-1 family. However, this probably reflects the difficulty that is inherent in distinguishing *de novo* insertion events from the background of existing BARE-1 or OARE-1 insertions. Techniques such as transposon display might prove valuable in detecting such 'needles in a haystack'.

Epigenetic silencing of transposable elements. The paradox of plant TE activity is dramatically illustrated in

maize, where the genome has recently doubled in size through transposition, but not a single active LTR retrotransposon has been identified. Furthermore, the paucity of maize retrotransposon-derived ESTs (REF. 4) indicates that some mechanisms, most likely epigenetic, might have been remarkably effective at repressing the transcription of a large fraction of the genome.

The existence of epigenetic mechanisms to repress TEs goes back to the genetic analysis of unstable mutations in maize. McClintock documented examples of what she called 'changes in phase', which were heritable and reversible (epigenetic) changes in the activity of *Ac* and *Spm* (*Suppressor–Mutator*) elements[69,70]. Characterization of the active and inactive phases of another maize TE family, *Mutator*, provided the first molecular evidence that the activity of members of this family correlated with methylation of cytosine residues[71]. Similar changes in methylation states were also shown to be associated with *Ac* and *Spm* elements that are undergoing phase changes[72,73]. In all cases, genetically inactive elements were hypermethylated (especially at their termini, where the transposase promoter resides in the autonomous family members) relative to their active counterparts, whereas hypomethylation was found to be a hallmark of actively transcribed and transposing class 2 elements. Hypermethylation is also associated with the intergenic clusters of LTR retrotransposons in maize. These regions, which make up at least 50% of the maize genome, are highly condensed and thought to comprise a transcriptionally repressive chromatin environment[4,55].

Although subsequent studies have shown that DNA methylation is required for the normal development of mammals and plants[74,75], its precise role and origin are the subject of speculation. One hypothesis, originally formulated by Bestor for mammals[76] and Flavell for plants[77], is that epigenetic mechanisms of gene regulation have evolved as a host defence to regulate the spread of invasive DNA, including TEs and viruses.

Studies of plant TE methylation were put on the back burner as the attention of researchers turned to the host genes that are involved in the initiation and maintenance of epigenesis. The study of transgene silencing and viral resistance in plants and of TE regulation in animals led to the identification of two distinct epigenetic mechanisms, known as post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS). In PTGS (a process also referred to as co-suppression and related to RNA interference in animals), TE silencing is caused by

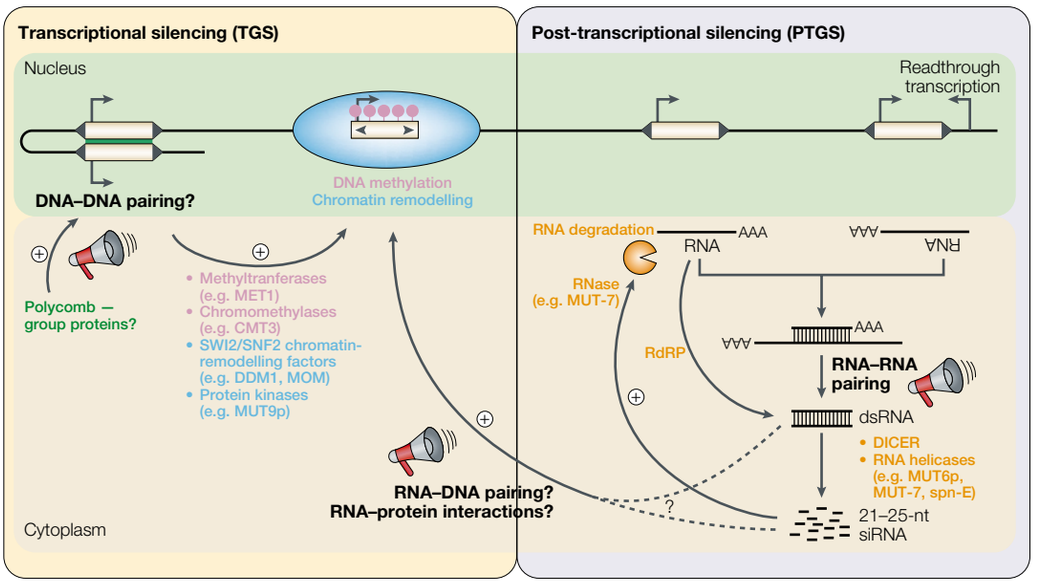


Figure 5 | **Epigenetic silencing of transposable elements.** The figure emphasizes the role of host-encoded proteins in transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) as revealed by mutant analysis (see TABLE 3), and the possible triggers of TE silencing (depicted by sirens). A key determinant for the activation of both TGS and PTGS is the multiple copies of identical or nearly identical TE sequences in the genome. TGS of TEs might be triggered by direct TE–TE interactions, as shown. So far, only TGS has been shown to regulate plant TEs, for which it is associated with the methylation of TE promoters. In PTGS, silencing is caused by sequence-specific degradation of RNAs in the cytoplasm. PTGS is triggered by double-stranded (ds)RNA produced by the pairing of sense and antisense RNAs or from RNAs that contain an inverted repeat (not shown). Antisense RNAs can be produced by readthrough transcription of a TE from an adjacent host gene. Formation of dsRNA might also involve RNA-dependent RNA polymerases (RdRP), as has been shown for the silencing of transgenes and plant viruses. dsRNA is degraded into 21–25-nucleotide (nt) small interfering RNAs (siRNAs) by an enzymatic complex that includes DICER-like proteins and RNA helicases. These small RNAs might act as guides for RNA degradation by RNase(s). siRNAs or dsRNAs might also enter the nucleus and mediate *de novo* methylation of transgenes, possibly through RNA–DNA interactions[80]. RNA-directed methylation occurs along the DNA regions that are complementary with the guide RNAs. It has been proposed that dsRNAs derived from TE promoters could promote methylation and TGS of TEs[80]. Examples of TEs known to be the targets of either TGS and/or PTGS mechanisms are listed in TABLE 3. CMT3, chromomethylase 3; DDM1, decreased DNA methylation 1; MET1, methyltransferase 1; MOM, Morpheus' molecule; spn-E, spindle E.

Table 3 | **Transposable-element targets of epigenetic gene silencing**

Species	Protein involved	Silenced transposable elements	References
Transcriptional gene silencing			
Arabidopsis thaliana	DDM1	LTR retroelements *Tar17**‡, *Athila4,6*‡ DNA transposons *AtMu1*, *CAC1,2**‡§	84,85 82,83
	CMT3	LTR retroelements *Ta3*, *Athila**‡ Non-LTR retrotransposon *Tal1*-like*	109,110 110
	MET1	LTR retroelements *Athila*-like*‡ Non-LTR retrotransposon *Tal1*-like*	110,111 110
	MOM1	LTR retroelements *Athila*-like‡	111,112
Chlamydomonas reinhardtii	MUT9p	LTR retrotransposon *TOC1*‡§ DNA transposon *Gulliver*§	81 81
Post-transcriptional gene silencing			
Caenorhabditis elegans	MUT-7	DNA transposons *Tc1,3,4,5*§	113
Chlamydomonas reinhardtii	MUT6p	LTR retrotransposon *TOC1*‡§ DNA transposon *Gulliver* §	114 114
Drosophila melanogaster	Spn-E	LTR retrotransposons *1731*, *mdg1*‡ Non-LTR retrotransposon F-element‡	115 115
	ND	Non-LTR retrotransposon I factor‖	116
Trypanosoma brucei	ND	Non-LTR retrotransposons *INGI*, *SLACS*¶	117

*Hypomethylation of the promoter region, ‡transcriptional activation and §transpositional activation have been detected in mutant backgrounds; ‖homologous RNA-mediated silencing has been shown; ¶abundant 24–26-nucleotide RNAs that are homologous to these transposable elements have been cloned and detected by RNA blot analysis. CMT3, chromomethylase 3; DDM1, decreased DNA methylation 1; LTR, long terminal repeat; MET1, methyltransferase 1; MOM1, Morpheus' molecule; ND, not determined; Spn-E, spindle E.

degradation of their RNAs. In TGS, TEs are transcriptionally repressed (FIG. 5; see TABLE 3 for references).

The current model for how TE RNA is specifically recognized and degraded in PTGS is briefly described in FIG. 5. So far, PTGS has been found to regulate the activity of TEs in animals that lack (extensive) DNA methylation (for example, *C. elegans* and *D. melanogaster*) and in the green alga *Chlamydomonas reinhardtii*. Although PTGS has been well documented in plants, most notably as a defence against viral replication[78], TGS seems to be the principal pathway to silence plant TEs[79] (FIG. 5; see TABLE 3 for references). Of course, this does not rule out the possibility that TGS and PTGS might overlap and combine to regulate some plant TEs (for further discussion, see REFS 79–81).

As mentioned above, the methylation of TE sequences (especially the promoter of transposase genes) correlates with TE inactivation in maize. Similarly, the methylation of transgene promoter sequences correlates with TGS in plants[79]. Therefore, it is not surprising that recent experiments have shown that endogenous TEs can be activated in mutants that are impaired for the establishment and maintenance of TGS (FIG. 5). For example, in the *Arabidopsis ddm1* (*decreased DNA methylation 1*) mutant, endogenous transposons of the *Mutator* and CACTA (*En/Spm*) superfamilies of class 2 elements are transcriptionally and transpositionally reactivated[82,83]. The endogenous retrotransposons *Tar17*, *Athila4* and *Athila5* were also transcriptionally reactivated in *ddm1* backgrounds, but new insertions were not detected[84,85]. Plants that are homozygous for the *ddm1* mutation have notably decreased CpG methylation. Consequently, the transcriptional derepression of TEs in the mutant strains was accompanied by demethylation of the elements. The *DDM1* gene encodes a protein with strong similarity to SWI2/SNF2 chromatin-remodelling factors[86]. These results indicate a possible functional link between chromatin remodelling, DNA methylation and genome integrity.

Reactivation of DNA transposons in the *ddm1* mutant background might have a direct impact on host fitness, as the reactivated TEs have been found to insert into coding and regulatory sequences of genes[82,83]. It is likely that many of the developmental abnormalities that were observed in *ddm1* plants were induced by the movement of reactivated transposons.

Future directions: waking the sleeping giant

TEs were first discovered and studied as the causative agents of genetically unstable mutations in maize. Many of these mutations had, in fact, been isolated from mutagenized plants or in strains undergoing the BREAKAGE–FUSION–BRIDGE CYCLE. It has been proposed that these, and similar treatments, activated previously quiescent elements that were normal residents of the genome[8].

The recent demonstration that *Arabidopsis* TEs can be reactivated in genetic backgrounds that are deficient in aspects of epigenetic regulation brings this story full-circle and promises to revolutionize the study of TEs, especially in plants that have large genomes. As discussed above, circumstantial evidence strongly implicates the involvement of high-copy-number TEs, such as MITEs and LTR retrotransposons, in the recent restructuring of many plant genomes. However, current activity of these high-copy-number elements, as measured by new insertions in the genome, has not been documented. It is possible that these elements were activated along with the class 2 elements described above. However, given their

BREAKAGE–FUSION–BRIDGE CYCLE
In somatic cells, a cycle of chromosome breakage (at *Ds*, for example), DNA replication, chromatid fusion (forming a dicentric chromosome) and formation of a chromosome bridge occurs during mitotic anaphase. A new round follows when the chromosome breaks yet again as the two centromeres are pulled to opposite poles.

extremely high copy number, activation might have been lethal. Fortunately, modern assays to detect new TE insertions, such as transposon display using DNA from cultured cells, might offer an alternative way to identify elements that are still able to transpose.

It is easy to imagine that epigenetic mutations will soon be available for a wide variety of plant species. Computer-assisted analysis of even a modest amount of genomic DNA sequence from these species should be sufficient to identify high-copy-number TEs and to design family-specific primers for transposon display or genome-wide analysis of TE expression. In this way, it should be possible to isolate active TEs from a variety of plant species. Knowledge of these elements should greatly facilitate studies of the continuing co-evolution of TEs with their hosts.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. SanMiguel, P. & Bennetzen, J. L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **81**, 37–44 (1998).
3. Vicient, C. M. *et al.* Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**, 1769–1784 (1999).
4. Meyers, B. C., Tingey, S. V. & Morgante, M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* **11**, 1660–1676 (2001).
A detailed survey of maize retrotransposons that combines random genomic sample sequencing, hybridization to bacterial artificial chromosomes and data obtained from a large collection of ESTs.
5. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* **16**, 13–47 (1951).
6. Berg, D. E. & Howe, M. H. *Mobile DNA* (American Society for Microbiology Press, Washington, DC, 1989).
7. Wessler, S. R. Phenotypic diversity mediated by the maize transposable elements *Ac* and *Spm*. *Science* **242**, 399–405 (1988).
8. Fedoroff, N. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 375–411 (American Society for Microbiology Press, Washington, DC, 1989).
9. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. *Mobile DNA II* (American Society for Microbiology Press, Washington, DC, 2002).
The newest edition of the transposable element bible.
10. McDonald, J. F. Transposable elements: possible catalysts of organismic evolution. *Trends Ecol. Evol.* **10**, 123–126 (1995).
11. Kidwell, M. G. & Lisch, D. Transposable elements as sources of variation in animals and plants. *Proc. Natl Acad. Sci. USA* **94**, 7704–7711 (1997).
12. Johns, M. A., Mottinger, J. & Freeling, M. A low copy number, *copia*-like transposon in maize. *EMBO J.* **4**, 1093–1102 (1985).
13. Varagona, M. J., Purugganan, M. & Wessler, S. R. Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. *Plant Cell* **4**, 811–820 (1992).
14. Vignols, F., Rigau, J., Torres, M. A., Capellades, M. & Puigdomenech, P. The *brown midrib3* (*bm3*) mutation in maize occurs in the gene encoding caffeic acid *O* methyltransferase. *Plant Cell* **7**, 407–416 (1995).
15. Britten, R. J., Graham, D. E. & Neufeld, B. R. Analysis of repeating DNA sequences by reassociation kinetics. *Methods Enzymol.* **29**, 363–405 (1974).
16. Goldberg, R. B. DNA sequence organization in the soybean plant. *Biochem. Genet.* **16**, 45–68 (1978).
17. Singer, M. F. SINEs and LINEs: highly repeated short and long interspersed sequences in mammals. *Cell* **28**, 433–434 (1982).
18. Yoshioka, Y. *et al.* Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc. Natl Acad. Sci. USA* **90**, 6562–6566 (1993).
19. Leeton, P. R. J. & Smyth, D. R. An abundant LINE-like element amplified in the genome of *Lilium speciosum*. *Mol. Gen. Genet.* **237**, 97–104 (1993).
20. Bureau, T. E. & Wessler, S. R. *Tourist*: a large family of inverted-repeat elements frequently associated with maize genes. *Plant Cell* **4**, 1283–1294 (1992).
21. Bureau, T. E. & Wessler, S. R. *Stowaway*: a new family of inverted-repeat elements associated with genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* **6**, 907–916 (1994).
22. Bureau, T. E., Ronald, P. C. & Wessler, S. R. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl Acad. Sci. USA* **93**, 8524–8529 (1996).
23. Wessler, S. R., Bureau, T. E. & White, S. E. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814–821 (1995).
24. Oosumi, T., Garlick, B. & Belknap, W. R. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J. Mol. Evol.* **43**, 11–18 (1996).
25. Tu, Z. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl Acad. Sci. USA* **94**, 7475–7480 (1997).
26. Tu, Z. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA* **98**, 1699–1704 (2001).
This paper reports the use of new software (available at the author's web site — see links box) that is designed to rapidly identify MITEs on the basis of their structural characteristics.
27. Feschotte, C. & Mouchès, C. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the *Mimo* family. *Gene* **250**, 109–116 (2000).
28. Izsvak, Z. *et al.* Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *J. Mol. Evol.* **48**, 13–21 (1999).
29. Unsal, K. & Morgan, G. T. A novel group of families of short interspersed repetitive elements (SINEs) in *Xenopus*: evidence of a specific target site for DNA-mediated transposition of inverted-repeat SINEs. *J. Mol. Biol.* **248**, 812–823 (1995).
30. Morgan, G. T. Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J. Mol. Biol.* **254**, 1–5 (1995).
31. Smit, A. F. A. & Riggs, A. D. *Tiggers* and DNA transposon fossils in the human genome. *Proc. Natl Acad. Sci. USA* **93**, 1443–1448 (1996).
32. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
33. Turcotte, K., Srinivasan, S. & Bureau, T. Survey of transposable elements from rice genomic sequences. *Plant J.* **25**, 169–179 (2001).
34. Jiang, N. & Wessler, S. R. Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell* **13**, 2553–2564 (2001).
35. Zhang, Q., Arbuckle, J. & Wessler, S. R. Recent, extensive and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* (*Hbr*) into genic regions of maize. *Proc. Natl Acad. Sci. USA* **97**, 1160–1165 (2000).
36. Feschotte, C., Zhang, X. & Wessler, S. R. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) (American Society of Microbiology Press, Washington, DC, in the press).
37. Feschotte, C. & Mouchès, C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Mol. Biol. Evol.* **17**, 730–737 (2000).
38. Feschotte, C. & Wessler, S. R. *Mariner*-like transposases are widespread and diverse in flowering plants. *Proc. Natl Acad. Sci. USA* **99**, 280–285 (2002).
Database searches and PCR experiments with plant-specific degenerate primers were combined to show that *mariner*-like TEs have successfully colonized many angiosperm genomes.
39. Walker, E. L., Eggleston, W. B., Demopulos, D., Kermicle, J. & Dellaporta, S. L. Insertions of a novel class of transposable elements with a strong target site preference at the *r* locus of maize. *Genetics* **146**, 681–693 (1997).
40. Zhang, X. *et al. 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. USA* **98**, 12572–12577 (2001).
Reports the first connection between an active maize DNA transposon system and a *Tourist*-like MITE family. Here, a 'top-down' approach was used in the isolation of an element that is required for transpositional activity and the identification of a new superfamily of eukaryotic transposases.
41. Kapitonov, V. V. & Jurka, J. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* **107**, 27–37 (1999).
42. Le, Q. H., Turcotte, K. & Bureau, T. *Tc8*, a *Tourist*-like transposon in *Caenorhabditis elegans*. *Genetics* **158**, 1081–1088 (2001).
43. Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. & Sved, J. High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* **62**, 515–525 (1990).
44. Rubin, E. & Levy, A. A. Abortive gap repair: underlying mechanism for *Ds* element formation. *Mol. Cell. Biol.* **17**, 6294–6302 (1997).
45. Casacuberta, E., Casacuberta, J. M., Puigdomenech, P. & Monfort, A. Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. *Plant J.* **16**, 79–85 (1998).
46. Wessler, S. R., Nagel, A. & Casa, A. M. in *Proceedings of the Fourth International Rice Genetics Symposium* (eds Khush, G. S., Brar, D. S. & Hardy, B.) 107–116 (International Rice Research Institute, Los Baños, the Philippines, 2001).
47. Casa, A. M. *et al.* The MITE family *heartbreaker* (*Hbr*): molecular markers in maize. *Proc. Natl Acad. Sci. USA* **97**, 10083–10089 (2000).
48. Casa, A. M. *et al.* Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. *Theor. Appl. Genet.* **104**, 104–110 (2002).
49. Kumar, A. & Bennetzen, J. L. Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 (1999).
50. Grandbastien, M.-A., Spielmann, A. & Caboche, M. *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* **337**, 376–380 (1989).
51. Hirochika, H. *et al.* Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. *Plant Cell* **8**, 725–734 (1996).
52. Marillonnet, S. & Wessler, S. R. Extreme structural heterogeneity among the members of a maize retrotransposon family. *Genetics* **150**, 1245–1256 (1998).
53. Joseph, J. L., Sentry, J. W. & Smyth, D. R. Interspecies distribution of abundant DNA sequences in *Lilium*. *J. Mol. Evol.* **30**, 146–154 (1990).
54. Thomas, C. A. The genetic organization of chromosomes. *Annu. Rev. Genet.* **5**, 237–256 (1971).
55. SanMiguel, P. *et al.* Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765–768 (1996).
56. SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. The paleontology of intergene retrotransposons of maize. *Nature Genet.* **20**, 43–45 (1998).
References 55 and 56 show that LTR retrotransposons make up more than half of the maize genome and that amplification of several families is responsible for doubling the genome in the past 5–6 million years.

57. Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. & Schulman, A. H. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl Acad. Sci. USA* **97**, 6603–6607 (2000).
Along with reference 3, this study documents large variations in the copy number of *BARE-1* retrotransposons among and within *Hordeum* species. By correlating this variation with intraspecific variation in genome size and with local changes in environmental conditions, the authors suggest that TE-mediated alterations might be adaptive.
58. Sassaman, D. M. *et al.* Many human *L1* elements are capable of retrotransposition. *Nature Genet.* **16**, 37–43 (1997).
59. Pouteau, S., Grandbastien, M. A. & Boccara, M. Microbial elicitors of plant defense responses activate transcription of a retrotransposon. *Plant J.* **5**, 535–542 (1994).
60. Hirochika, H. Activation of tobacco retrotransposons during tissue culture. *EMBO J.* **12**, 2521–2528 (1993).
61. Hirochika, H., Sugimoto, K., Otsuki, Y. & Kanda, M. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl Acad. Sci. USA* **93**, 7783–7788 (1996).
62. Grandbastien, M.-A. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3**, 181–189 (1998).
63. Curcio, M. J. & Garfinkel, D. J. New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIH. *Trends Genet.* **15**, 43–45 (1999).
64. Melayah, D., Bonnivard, E., Chalhoub, B., Audeon, C. & Grandbastien, M. A. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *Plant J.* **28**, 159–168 (2001).
Illustrates the use of transposon display to detect *de novo* insertions of a moderately high-copy-number retrotransposon following stress induction.
65. Waugh, R. *et al.* Genetic distribution of *BARE1* retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet.* **253**, 687–694 (1997).
66. Suoniemi, A., Narvanto, A. & Schulman, A. H. The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. *Plant Mol. Biol.* **31**, 295–306 (1996).
67. Jaaskelainen, M. *et al.* Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. *Plant J.* **20**, 413–422 (1999).
68. Kimura, Y. *et al.* OARE-1, a Ty1-*copia* retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol.* **42**, 1345–1354 (2001).
69. McClintock, B. The *Suppressor–Mutator* system of control of gene action in maize. *Carnegie Inst. Wash. Yearb.* **57**, 415–429 (1958).
70. McClintock, B. Components of action of the regulators *Spm* and *Ac*. *Carnegie Inst. Wash. Yearb.* **64**, 527–534 (1965).
71. Chandler, V. L. & Walbot, V. DNA modification of a maize transposable element correlates with loss of activity. *Proc. Natl Acad. Sci. USA* **83**, 1767–1771 (1986).
72. Chomet, P., Wessler, S. R. & Dellaporta, S. L. Inactivation of the maize transposable element *Activator* is associated with its DNA modification. *EMBO J.* **6**, 295–302 (1987).
73. Banks, J. A., Masson, P. & Fedoroff, N. Molecular mechanisms in the developmental regulation of the maize *Suppressor–Mutator* transposable element. *Genes Dev.* **2**, 1364–1380 (1988).
74. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
75. Kakutani, T., Jeddeloh, J. A., Flowers, S. K., Munakata, K. & Richards, E. J. Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl Acad. Sci. USA* **93**, 12406–12411 (1996).
76. Bestor, T. H. DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **326**, 179–187 (1990).
77. Flavell, R. B. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl Acad. Sci. USA* **91**, 3490–3496 (1994).
78. Voinnet, O. RNA silencing as a plant immune system against viruses. *Trends Genet.* **17**, 449–459 (2001).
79. Vaucheret, H. & Fagard, M. Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends Genet.* **17**, 29–35 (2001).
80. Matzke, M. A., Matzke, A. J., Pruss, G. J. & Vance, V. B. RNA-based silencing strategies in plants. *Curr. Opin. Genet. Dev.* **11**, 221–227 (2001).
81. Jeong, B. R., Wu-Scharf, D., Zhang, C. & Cerutti, H. Suppressors of transcriptional transgenic silencing in *Chlamydomonas* are sensitive to DNA-damaging agents and reactivate transposable elements. *Proc. Natl Acad. Sci. USA* **99**, 1076–1081 (2002).
82. Singer, T., Yordan, C. & Martienssen, R. A. Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation* (*DDM1*). *Genes Dev.* **15**, 591–602 (2001).
83. Miura, A. *et al.* Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* **411**, 212–214 (2001).
84. Hirochika, H., Okamoto, H. & Kakutani, T. Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* **12**, 357–369 (2000).
85. Wright, D. A. & Voytas, D. F. *Athila4* of *Arabidopsis* and *calypso* of soybean define a lineage of endogenous plant retroviruses. *Genome Res.* **12**, 122–131 (2002).
References 82–85 show that a subset of *Arabidopsis* retrotransposons and DNA transposons are transcriptionally regulated by the SWI2/SNF2 chromatin-remodelling factor DDM1.
86. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genet.* **22**, 94–97 (1999).
87. Jin, Y. K. & Bennetzen, J. L. Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. *Proc. Natl Acad. Sci. USA* **86**, 6235–6239 (1989).
88. Jiang, N. *et al. Dasheng*: a recently amplified non-autonomous LTR element that is a major component of pericentromeric regions in rice. *Genetics* (in the press).
89. Witte, C. P., Le, Q. H., Bureau, T. & Kumar, A. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc. Natl Acad. Sci. USA* **98**, 13778–13783 (2001).
90. Ogiwara, I., Miya, M., Ohshima, K. & Okada, N. Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs. *Mol. Biol. Evol.* **16**, 1238–1250 (1999).
91. Kapitonov, V. V. & Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **98**, 8714–8719 (2001).
92. Lampe, D. J., Churchill, M. E. & Robertson, H. M. A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J.* **15**, 5470–5479 (1996).
93. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvak, Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510 (1997).
References 92 and 93 emphasize the power of phylogenetic analyses to reconstruct an active transposase from a subset of inactive TE copies. Following these breakthrough studies, the reconstructed elements have been further developed into efficient transformation and mutagenic tools for a wide range of organisms.
94. Jordan, I. K. & McDonald, J. F. Comparative genomics and evolutionary dynamics of *Saccharomyces cerevisiae* Ty elements. *Genetica* **107**, 3–13 (1999).
95. Lerat, E., Brunet, F., Bazin, C. & Capy, P. Is the evolution of transposable elements modular? *Genetica* **107**, 15–25 (1999).
96. Malik, H. S. & Eickbush, T. H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res.* **11**, 1187–1197 (2001).
97. Bowen, N. J. & McDonald, J. F. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* **11**, 1527–1540 (2001).
98. Schwarz-Sommer, Z., Le Clercq, L., Gobel, E. & Saedler, H. *Cin4*, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J.* **6**, 3873–3880 (1987).
99. Wright, D. *et al.* Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics* **142**, 569–578 (1996).
100. Deragon, J. M. *et al.* An analysis of retroposition in plants based on a family of SINEs from *Brassica napus*. *J. Mol. Evol.* **39**, 378–386 (1994).
101. White, S. E., Habera, L. F. & Wessler, S. R. Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. *Proc. Natl Acad. Sci. USA* **91**, 11792–11796 (1994).
102. Purugganan, M. D. & Wessler, S. R. Molecular evolution of magellan, a maize Ty3/*gypsy*-like retrotransposon. *Proc. Natl Acad. Sci. USA* **91**, 11674–11678 (1994).
103. Konieczny, A., Voytas, D. F., Cummings, M. P. & Ausubel, F. M. A superfamily of *Arabidopsis thaliana* retrotransposons. *Genetics* **127**, 801–809 (1991).
104. Gierl, A. The *En/Spm* transposable element of maize. *Curr. Top. Microbiol. Immunol.* **204**, 145–159 (1996).
105. Chandler, V., Rivin, C. & Walbot, V. Stable non-mutator stocks of maize have sequences homologous to the *Mu1* transposable element. *Genetics* **114**, 1007–1021 (1986).
106. Parsons, J. D. Miropeats: graphical DNA sequence comparisons. *Comput. Appl. Biosci.* **11**, 615–619 (1995).
107. Kurtz, S. & Schleiermacher, C. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* **15**, 426–427 (1999).
108. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000).
109. Lindroth, A. M. *et al.* Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**, 2077–2080 (2001).
110. Tompa, R. *et al.* Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.* **12**, 65–68 (2002).
A new microarray technique for genome-wide mapping of cytosine methylation reveals that various retrotransposons are the preferential targets of CpNpG methylation in *Arabidopsis*.
111. Steimer, A. *et al.* Endogenous targets of transcriptional gene silencing in *Arabidopsis*. *Plant Cell* **12**, 1165–1178 (2000).
112. Amedeo, P., Habu, Y., Afsar, K., Scheid, O. M. & Paszkowski, J. Disruption of the plant gene *MOM* releases transcriptional silencing of methylated genes. *Nature* **405**, 203–206 (2000).
113. Ketting, R. F., Haverkamp, T. H., Van Luenen, H. G. & Plasterk, R. H. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
114. Wu-Scharf, D., Jeong, B., Zhang, C. & Cerutti, H. Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science* **290**, 1159–1162 (2000).
This paper and reference 81 show that the silencing of a DNA transposon and a LTR retrotransposon in *Chlamydomonas* involve enzymatic components of both PTGS and TGS, but do not seem to require DNA methylation.
115. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
116. Jensen, S., Gassama, M. P. & Heidmann, T. Taming of transposable elements by homology-dependent gene silencing. *Nature Genet.* **21**, 209–212 (1999).
117. Djikeng, A., Shi, H., Tschudi, C. & Ullu, E. RNA interference in *Trypanosoma brucei:* cloning of small interfering RNAs provides evidence for retroposon-derived 24–26-nucleotide RNAs. *RNA* **7**, 1522–1530 (2001).

Acknowledgements

We thank members of the Wessler lab, Z. Bao and E. Pritham for suggestions on the manuscript and stimulating discussions, and H. Cerrutti, S. Eddy, M.-A. Grandbastien, R. Martienssen, J. Tu and D. Voytas for providing helpful information and unpublished data. This work was supported by grants from the National Science Foundation (Plant Genome Initiative), the National Institute of Health, the Department of Energy and the University of Georgia Research Foundation to S.R.W.

Online links

DATABASES

The following terms in this article are linked online to:

MaizeDB: http://www.agron.missouri.edu
Ac | *adh1* | *Bs1* | *Ds* | *Hopscotch* | *Huck-2* | *Magellan* | *Mutator* | *Opie-2* | *Spm* | *Stonor* | *Stowaway* | *Tourist*

The *Arabidopsis* Information Resource:
http://www.arabidopsis.org
ddm1

FURTHER INFORMATION

Barbara McClintock: http://www.nobel.se/medicine/laureates/1983/mcclintock-autobio.html
Encyclopedia of Life Sciences: http://www.els.net
Repetitive DNA: evolution | Transposons; eukaryotic
Sue Wessler's lab: http://www.botany.uga.edu/wesslerlab
Zhijiam Tu's web page: http://www.biochem.vt.edu/aeo
Access to this interactive links box is free online.