How important are transposons for plant evolution?

Damon Lisch

Abstract | For decades, transposable elements have been known to produce a wide variety of changes in plant gene expression and function. This has led to the idea that transposable element activity has played a key part in adaptive plant evolution. This Review describes the kinds of changes that transposable elements can cause, discusses evidence that those changes have contributed to plant evolution and suggests future strategies for determining the extent to which these changes have in fact contributed to plant adaptation and evolution. Recent advances in genomics and phenomics for a range of plant species, particularly crops, have begun to allow the systematic assessment of these questions.

Transposable elements
(TEs). Stretches of DNA that are competent to integrate into new positions in the genome, that are competent to increase their copy number over time and that rely on one or more enzymatic function provided by an autonomous element.

Epigenetic variation
Heritable differences in the expression of a gene in the absence of changes in the DNA sequence of that gene. It is often associated with changes in cytosine methylation and histone modification. Cryptic epigenetic variation refers to epigenetic variation that is only manifest under specific conditions.

Phenomics
The objective and systematic acquisition of high-quality phenotypic data, allowing for phenotypic features to be analysed on a continuum together with molecular data, such as gene expression profiles or causative genomic mutations.

Department of Plant and Microbial Biology, UC Berkeley, Berkeley, California 94720, USA.
e-mail: dlisch@berkeley.edu
doi:10.1038/nrg3374

Evolution cannot proceed in the absence of variation. In plants, as in most eukaryotes, transposable elements (TEs) are by far the most variable parts of the genome (BOX 1). Waves of expansion and contraction in numbers of TEs can result in dramatic differences in the overall architecture of the genomes of even closely related plant species, and TEs make up the majority, probably the vast majority, of all plant DNA[1]. In addition to these gross effects on the overall architecture of genomes, TE activity can cause a broad range of changes in gene expression and function, from subtle quantitative effects to the rewiring of regulatory networks and the evolution of entirely new genes[2,3] (FIG. 1).

Selection against the negative consequences of TE activity, as well as specific strategies used by both TEs and their hosts to ameliorate these consequences, ensures that most of the surviving changes caused by TEs are selectively neutral or mildly deleterious[4]. However, the scope of potentially useful variation generated by TEs suggests that these endogenous mutagens have substantially contributed to the evolution of their hosts. Certainly, it is impossible to fully understand genetic and epigenetic variation in plants (and indeed in any eukaryote) unless the role of TEs is understood. That being said, it is important to acknowledge that many of the data available to date are essentially anecdotal. There are many examples of what TEs can do, but we do not know at this point how often changes caused by TEs have actually contributed to the adaptive evolution of plants. Fortunately, recent advances in genomics and phenomics have made it possible precisely to correlate genotype with phenotype on an unprecedented scale and at an unprecedented level of detail (BOX 2); this will enhance our ability to determine how often molecular changes cause by TEs have been subjected to selection during plant evolution. This Review gives examples of the kinds of genetic changes that TEs can cause, it provides evidence that TEs have actually contributed to plant evolution, and it discusses potentially productive strategies for determining how often this has occurred and how important it has been.

Gene inactivation

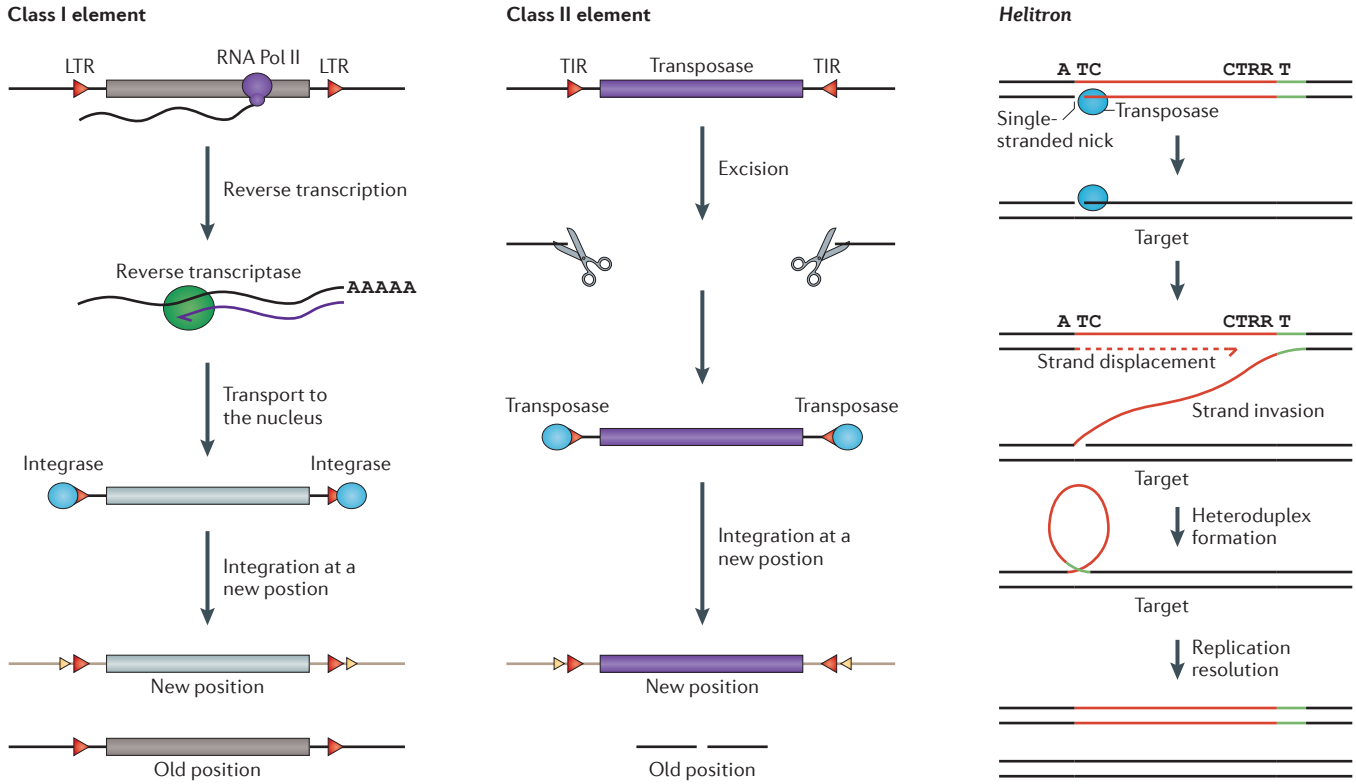
Nothing is so easy to do as to break something. Thus, the simplest and almost certainly the most common type of TE-induced phenotypic change that has been observed results from a simple loss of gene function. The propensity for some classes of TEs to insert into or near genes has been exploited for decades to generate new null mutations[5–7]. TE-induced loss-of-function mutations are qualitatively no different from other mutations that knock out gene function, but the rate at which they occur can dramatically vary depending on levels of TE activity and copy number. Among most experimental populations, TE activity appears to be fairly stable: significant levels of activity are generally associated with mutations in genes required for efficient epigenetic silencing[8,9] or with 'genomic shocks', such as interspecific hybridization[10], chromosomal breakage[11] or tissue culture[12]. Presumably, transient TE-induced spikes in mutation rates in natural populations occur regularly, as is suggested by evidence for recent or current high levels of TE activity in populations of wild sunflowers (*Helianthus anomalus*, *Helianthus deserticola* and *Helianthus paradoxus*) and barley (*Hordeum spontaneum*), as well as some cultivars of rice (*Oryza* spp.)[13–15].

Class I elements
A transposable element that uses a 'copy-and-paste' transposition mechanism involving an RNA intermediate.

Certainly, we have no reason to believe that 'pampered' inbred experimental populations accurately reflect reality in natural populations, which regularly experience various biotic and abiotic stresses that are known to activate TEs[16,17].

TE-induced null mutations have been selected for several times during plant domestication. Perhaps the most striking example of this comes from foxtail millet (*Setaria italica*). Remarkably, all of the 431 landraces that have been examined that are either waxy or that have a low amylose content ('sticky') in this species carry weak or null alleles of the granule-bound starch synthase gene *GBSS1* owing to the insertion of TEs[18]. The TEs are of different types, including both class I elements and class II elements, and at least four alleles arose independently. These data suggest that, at least for this

Box 1 | **Types of transposable elements**



Although there are many kinds of transposable elements (TEs), they fall into a small number of general classes. The three classes of TEs that are present in plant genomes are discussed below and are illustrated in the figure.

Class I elements

These elements are retrotransposons. In plants, they are the most common class of element and can make up the bulk of many genomes. Retroelements transpose via a 'copy-and-paste' mechanism in which mRNA transcribed from the element by RNA polymerase II (RNA Pol II) is converted into a cDNA by reverse transcription and then integrated by an integrase enzyme at a new position in the genome. Class I elements are further divided into long terminal repeat (LTR) retroelements and non-LTR retroelements, which differ in the mechanism of integration (the mechanism for LTR retroelements, which predominate in plants, is shown in the figure). Autonomous retroelements encode all of the necessary proteins for transposition. Nonautonomous retroelements require the presence of enzymes encoded by intact autonomous elements. In plants, the most common nonautonomous retroelements are short interspersed nuclear elements (SINEs).

Class II elements

These elements transpose via a 'cut-and-paste' mechanism in which the element is physically excised from the chromosome and reintegrated at a new location, a process that involves the transposase enzyme encoded by the TE. Effective replication is achieved owing to host repair of the double-strand break caused by excision of the element. In plants, the most common class II elements include members of the hAT (*hobo*, *Activator* and *Tam3*), *CACTA* and *Mutator*-like element (MULE) superfamilies. Nonautonomous versions of these elements carry the minimum sequences necessary for transposition in the presence of autonomous elements. These elements, which often constitute the vast majority of class II elements in a given genome, can be either deletion derivatives of autonomous elements or sequences that share similarity with autonomous elements only at their termini. The most common nonautonomous elements in plant genomes are miniature inverted-repeat transposable elements (MITEs).

Helitrons

Also present in many plant genomes, sometimes in very large numbers, *Helitrons* are a class of elements that are thought to transpose via a 'rolling circle' mechanism. This process involves nicking at the *Helitron* terminus, followed by strand invasion, DNA synthesis, strand displacement and resolution of a heteroduplex by DNA replication. If the initial DNA synthesis and strand displacement proceeds farther than the end of the *Helitron*, flanking sequences (shown in green in the figure) can be co-replicated. Many *Helitrons* are composed of short sequences at their termini and 'filler' sequences composed of fragments of captured genes and gene fragments.

In the figure, R refers to purines (A or G) and TIR stands for terminal inverted repeat.

Class II elements
A transposable element that transposes via a 'cut-and-paste' mechanism in which the DNA of the element is physically transposed from one position in the genome to another.

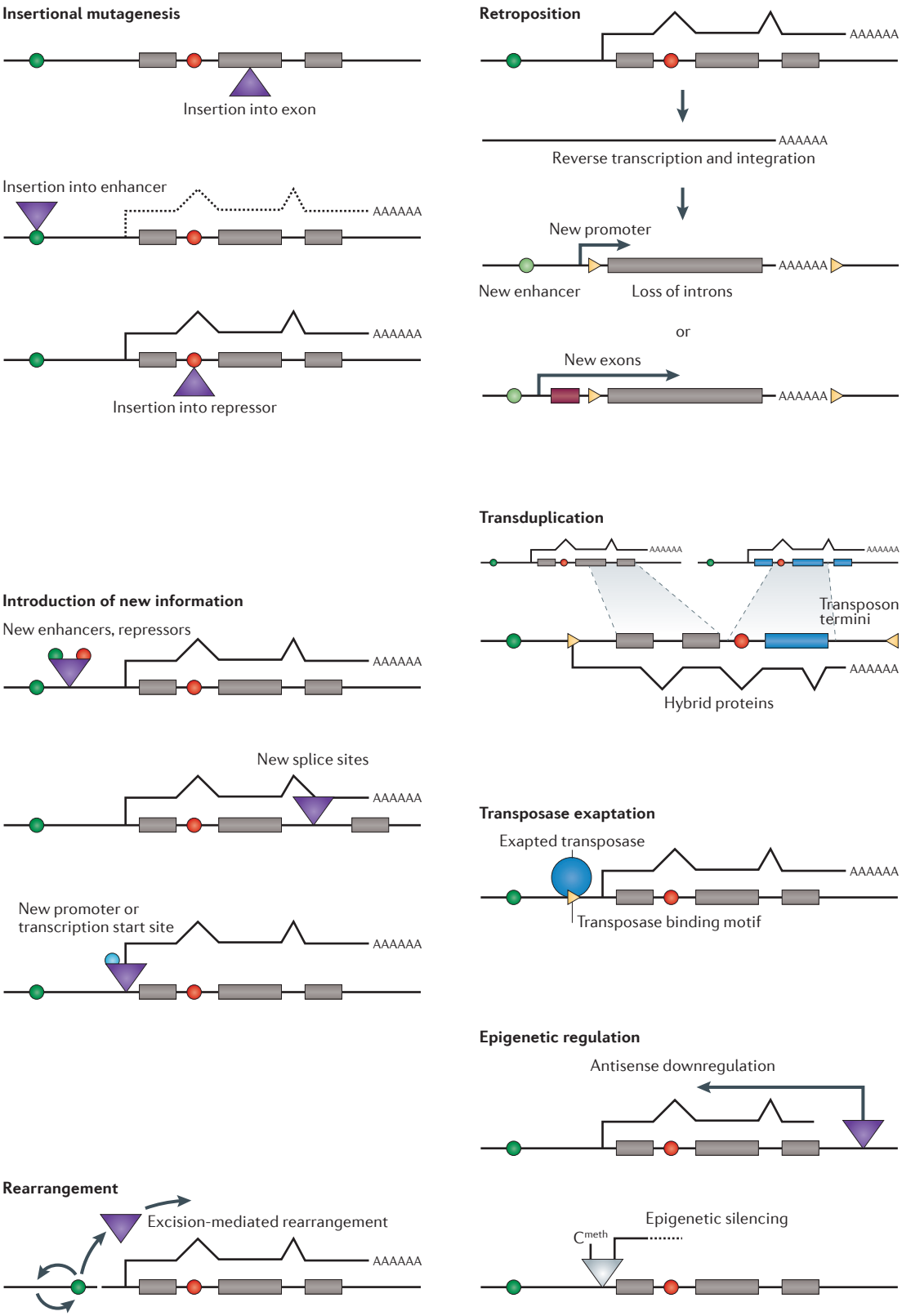


Figure 1 | **Structural and functional changes that can be caused by transposable elements.** This figure summarizes the various kinds of effects that transposable element (TE) insertions can have on gene structure and function. TEs can be involved in changes that include knockout of function, introduction of new functions, changes in the structure of genes, mobilization and rearrangement of gene fragments and epigenetic silencing of genes. Exons are depicted as grey or coloured boxes, enhancers as green circles, repressors as red circles, TEs as purple triangles, transposon termini as yellow triangles and exapted TE proteins as blue circles. C^{meth}, 5-methylcytosine.

Comparative genomics
The discipline devoted to comparing related genomes, focusing on large-scale changes in the overall structure and composition of genomes, chromosomal segments and genes.

Recombinant inbred lines
(RILs). Strains that are derived from crosses between two or more parental strains, followed by recombination of chromosomes and inbreeding to homozygosity. Typically, RILs are carefully genotyped at many loci. A panel of RILs can be a stable resource for quantitative trait locus mapping.

Box 2 | **Resources for assessing the impact of transposable elements on plant evolution**

Genome sequences of related species

Many plant genomes sequences are now available, and many more will become available soon. In addition, transposable element (TE) identification and classification have become increasingly sophisticated over the past several years.

Current versions of all publicly available plant genomes are accessible at PlantGDB, and a nice depiction of the relationship between plant phylogeny and the current state of sequencing is available at CoGePedia. In addition, the 1001 Genomes Project seeks to determine the degree of genetic variation in 1,001 accessions of *Arabidopsis thaliana*.

Transposable element annotations

Annotations of TEs in plants and many other species are available from the Genetic Information Research Institute and the Institute for Systems Biology. A more focused, plant-specific database is available at the Plant Repeat Databases at Michigan State University, and a well-curated and BLAST-searchable set of maize TEs is also available at the Maize TE Database. For currently active TEs in maize and rice, sequence-indexed insertions are available from the UniformMu transposon resource and the International Rice Functional Genomics Consortium. A free Web-based comparative genomics tool is available from CoGe. This platform contains all publicly available genomic sequences and permits users to compare multiple sequences systematically at a range of scales. Finally, a wide range of tools for analysis of large-scale data sets is available from iPlant, which is currently engaged in a major effort to integrate genotype with phenotype using the iPG2P tool.

Mapping resources

Maize HapMap2 represents an attempt to capture and to map the remarkable diversity that is present in maize and related species by characterizing genetic variation present in 103 inbred maize lines, including 60 recombinant inbred lines (RILs) from diverse genetic backgrounds. Combined with analysis of a wide variety of phenotypes, this resource will make it possible to map a wide variety of quantitative traits. Similarly, the Rice Diversity project represents an attempt to determine the genetic basis of phenotypic variation in more than 400 accessions of rice and its wild relatives. In *Arabidopsis thaliana*, more than 60 RILs and near isogenic lines (NILs) are now available at the VAST lab website, with many more on the way. Similar efforts to capture and to characterize the effects of genotype on phenotype are underway in a wide variety of other species, including barley, oat, sorghum, pearl millet and foxtail millet (see Gramene website). An example of high-throughput phenomics that is put into practice is the Australian Plant Phenomics Facility.

Challenges

With respect to the role that TEs may have had in adaptive changes in plants, a major challenge comes from the repetitive nature of many TEs and a widespread assumption that they are unlikely to be functionally relevant. For these reasons, TEs are often filtered from analyses, making it less likely that their role can be assessed. As such, it is important that all TEs are carefully annotated and included in the analysis as a possible source of functional variation. This is particularly important for classes of TEs that are known to carry regulatory information. With this in mind, it is important that experts who are familiar with the biology of TEs be included in the analysis of fine-scale mapping of traits in plants.

gene in this species, TE activity is the primary source of new null mutations. Null mutations caused by TE insertions are also responsible for Mendel's wrinkled peas[19], white wine grape varietals[20–22] (FIG. 2) and several strains of seedless apples[23].

Among natural populations, TE insertions can play an important part in the generation of flower colour polymorphism: this has long been a subject of interest for evolutionary biologists[24]. Some of the most extensive work has been done on the common morning glory (*Ipomoea purpurea*). Nearly all of the colour variation in this and the related Japanese morning glory is due to transposon insertions that cause loss of function or somatically unstable alleles[25]. These insertions are in several genes that are required for colour production, including those that encode chalcone synthase D (CHSD), flavanone 3-hydroxylase (F3H) and the bHLH transcriptional regulator bHLH2, and involve a wide variety of TE types, including the class II miniature inverted repeat transposable elements (MITEs), *CACTA* elements, hAT (*hobo*, *Activator* and *Tam3*) elements and *Mutator*-like transposable elements (MULEs), as well as class I short interspersed elements (SINEs)[24,26]. Perpetuation of colour variation in morning glory populations appears to have involved a complex combination of artificial and natural selection[27].

Clearly, then, TEs have the capacity to generate null mutations, sometimes at high rates, and TE-induced null mutations have certainly played a part in plant domestication. However, it is important to keep in mind that there are many other ways in which null mutations can arise. In both rice and barley, for instance, the *waxy* mutation appears to have arisen only once in each species and did not involve TE insertions[28], and TEs are only one of several documented causes of clonal variation in grapes[29]. In fact, of the few examples of mutants that are known to have been involved in plant domestication, only a minority are due to TE insertions[28]. For instance, none of the variation in the 11 cloned rice domestication genes is known to be due to TE insertions[30].

It is difficult to ascertain the relative importance of TE-induced null mutations over longer periods of time in plants. DNA that does not contribute to fitness is more rapidly lost in plants than in mammals, primarily by deletion[31]. Therefore, if we are to identify large numbers of additional examples of naturally occurring TE-induced null mutations, it will be necessary to carry out careful analysis of closely related plant species for which there is extensive trait characterization and complete genome sequences[32,33]. Of particular use will be maize (*Zea mays*) and rice, both of which have large numbers of well-characterized and polymorphic

TEs, closely related wild species, a wealth of genetic and epigenetic diversity and major resources devoted to characterizing that diversity[34–38] (BOX 2).

Complicating the analysis of the effects of TEs is the fact that they can cause null mutations that are intrinsically unstable owing to the structure of the insertion or owing to ongoing activity. For instance, a screen for reversions of a class I *Mutator*-element-induced null allele of the *alcohol dehydrogenase1* (*Adh1*) gene in maize yielded a new allele that lacked the insertion but that contained a complex rearrangement resulting in a novel pattern of tissue-specific expression[39]. Note that without having the progenitor allele in hand, there would have been no way of knowing that a TE had caused the rearrangement. Among domesticated crops, there are several revertant red grape varietals that were derived from white varietals owing to a recombination event that removed a portion of the long terminal repeat (LTR) retrotransposon that had caused the null mutation in the progenitor allele[20] (FIG. 2). These results suggest that many TE-induced changes will be difficult to detect, although this problem can be ameliorated if well-characterized close relatives carrying progenitor alleles are available.

Reprogramming of gene expression

Insertion into repressors and enhancers. In addition to simply knocking out gene expression altogether, TE insertions can eliminate positive or negative regulatory functions. A classic example of this is the insertion of *Mutator* elements into a conserved non-coding sequence (CNS) in the first intron of the *knotted1* gene in maize[40]. These insertions lead to ectopic expression of this gene in leaves. Given that changes in expression of *knotted1* homologues in various species are associated with differences in the degree of leaf lobing, it will be interesting to see how often these changes are associated with TE insertions into this CNS[41].

Although regulatory sequences are often assumed to be close to coding sequences, there is evidence that they can be quite distant. *Vgt1* is a locus in maize that regulates flowering time and that has been a target of selection[42]. Rather than being a gene, *Vgt1* is a CNS that lies roughly 70 kb upstream of a gene encoding an AP2 transcription factor that is a negative regulator of flowering time[43]. An insertion of a MITE into the conserved portion of *Vgt1* is tightly associated with flowering time variation in maize. It should be emphasized that the identification of this insertion, which is located tens of kilobases upstream of the affected gene, required extremely precise trait mapping; a simple comparison of proximal promoters would not have identified the polymorphism responsible for the expression differences.

Examples such as *Vg1* in maize demonstrate that before we can properly evaluate the importance of TE insertions on regulatory evolution, it will be necessary to utilize the kind of accurate trait-mapping methodologies that are only now becoming available[44] (BOX 2). Given that thousands of CNSs have been identified plants[45], a systematic evaluation of all instances in which a TE has inserted into a CNS may be informative. An insertion into an otherwise well-conserved CNS in a given species that maps tightly to a given trait and that is associated with a novel pattern of expression of a nearby gene would be an excellent candidate for further investigation. Confirmatory data would be provided by experimentally induced TE insertions into the homologous intact CNS in a related species that phenocopies the effect of the naturally occurring insertion[6,7].

Introduction of new regulatory information. In addition to simply interfering with existing gene or gene regulatory functions, TEs can introduce new information. TE expression can be responsive to a wide variety of cues, and the fact that several TEs (such as *Mutator*, Ac and *mPing*) tend to transpose into the 5′ ends of plant genes[46–49] means that promoter and enhancer elements within these TEs can potentially alter gene expression.

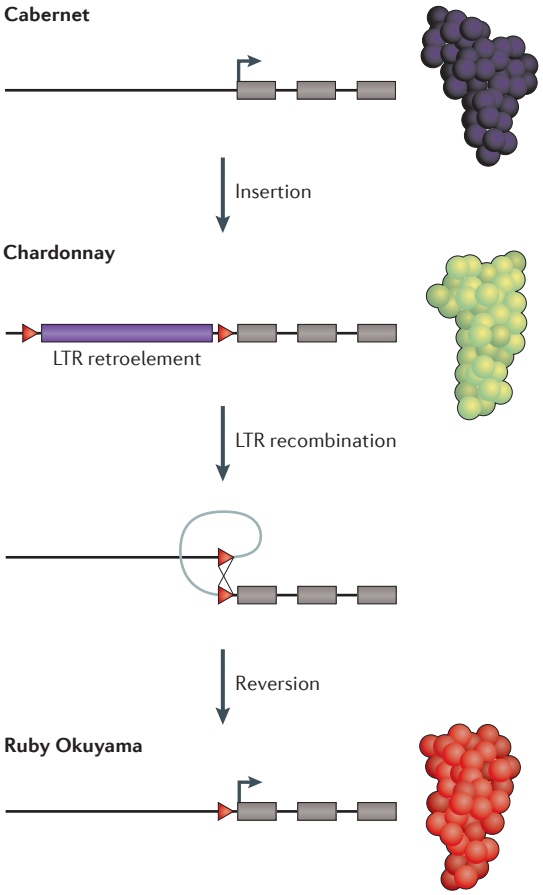


Figure 2 | **Transposable element insertions associated with changes in grape colour.** An initial insertion of a *Gret1* long terminal repeat (LTR) retrotransposon (not present in the Cabernet variety) resulted in a loss-of-function allele of the *Vvmby1A* gene, leading to a loss of colour in the fruit of the Chardonnay variety. A subsequent rearrangement in *Gret1* results in revertant, coloured grapes in varieties such as Ruby Okuyama. Exons are depicted as grey boxes. The LTRs flanking *Gret1* just upstream of the *Vvmyb1A* gene are depicted as red triangles.

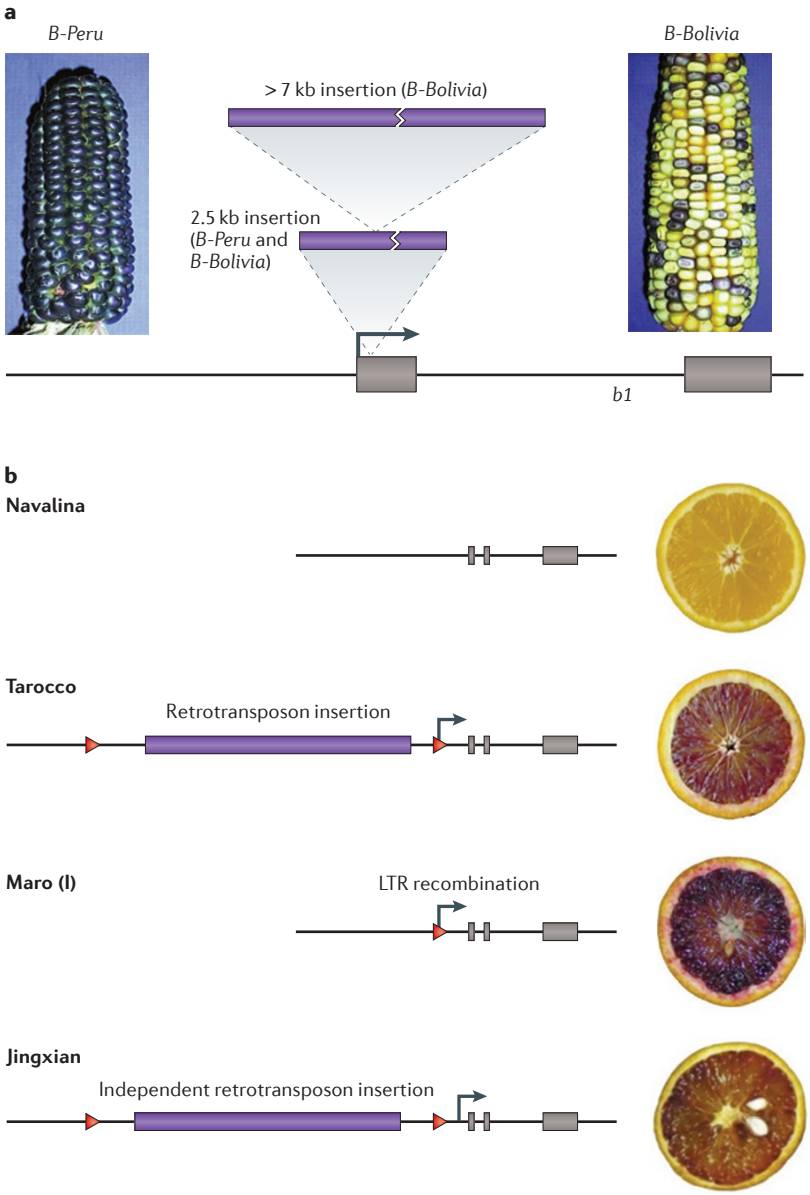


Figure 3 | **Effects of transposable element insertions on expression tissue specificity and cold inducibility.** **a** | Effects of transposable element (TE) insertions on expression of the *b1* gene in maize. TE insertions into the first exon the *b1* gene in the *B-Peru* allele resulted in a shift in expression of this gene from vegetative tissues to the seed. A subsequent insertion in the *B-Bolivia* allele resulted in reduced and variegated expression of this gene. **b** | The role of TE insertions in the development of blood oranges. The insertion of *Rider*, a long terminal repeat (LTR) retrotransposon upstream of the *Ruby* gene resulted in cold-dependent expression of that gene in the fruit. The Navalina variant carries the functional native *Ruby* gene, which shows limited expression in the fruit flesh. Tarocco carries a *Ruby* allele that has an insertion of a retrotransposon immediately upstream of the *Ruby* coding sequence. The LTR of the retroelement provides a novel promoter that drives expression in the flesh of the fruit, resulting in the distinctive red colouration illustrated here. The LTR also confers responsiveness to cold temperatures. Recombination between the LTRs resulted in enhanced expression found in Maro (I). Jingxian is an independently derived blood orange variant that contains a distinct but related LTR retrotransposon that also confers both tissue specificity and cold responsiveness. The LTRs flanking the retrotransposons just upstream of the *Ruby* gene are depicted as red triangles. Exons are depicted as grey boxes. Images of corn in panel **a** are taken, with permission, from REF. 51 © (1999) National Academy of Sciences USA. Images of blood oranges in panel **b** are taken, with permission, from REF. 54 © (2012) American Society of Plant Biologists.

One type of regulatory effect of TE insertions involves changes in tissue specificity, something that is readily visualized by looking at changes in expression of plant colour genes. Anthocyanin expression in maize requires expression of one of two basic helix–loop–helix *myc* genes (either *r1* or *b1*) and one of two *myb* genes (either *c1* or *pl1*)[50]. These pairs of genes have become subfunctionalized so that *r1* and *c1* are required for expression in the seed, and *b1* and *pl1* are required for expression in the plant. There are, however, some alleles of *b1* in that have arisen more recently that confer expression in the seed. In these cases, this new pattern of expression is dependent on a high copy number retroelement that recently inserted into the first exon of the *b1* gene[51,52] (FIG. 3a).

TE activity can also be directly involved in both gene duplication and subfunctionalization. Rearrangements and duplications at the *r1* locus in maize, which are most likely to have been caused by aberrant transposition of a *CACTA* element, resulted in changes in tissue specificity of the duplicate gene copies, resulting in changes in expression patterns of the genes at this locus[53]. What is particularly interesting in this case is that a single TE may have been involved in first producing functional redundancy and then reprogramming gene expression of one of the redundant copies.

There is also evidence that gene reprogramming caused by TE insertion can result in selectable traits. Blood oranges have red fruit flesh owing to tissue-specific and cold-responsive expression of a MYB transcription factor called Ruby, which has been conferred by the insertion of an LTR retrotransposon into the promoter of the *Ruby* gene[54] (FIG. 3b). Additional allelic variation at this locus has been generated owing to recombination between the long terminal repeats (LTRs) flanking the insertion, again demonstrating the ongoing instability that can be introduced by the insertion of TEs.

TEs can also dramatically alter plant phenotype simply by enhancing the pre-existing expression of genes, and this can have dramatic effects on plant phenotypes. It appears that this propensity has been exploited during domestication of maize. There are five loci that are responsible for most of the gross morphological differences between maize and its wild relative teosinte[28]. One of these loci, *teosinte branched1* (*tb1*), encodes a transcription factor that represses branching. Overexpression of *tb1* in maize results in a dramatic reduction in the number of branches relative to the progenitor species. This overexpression is due to the enhancer activity of *Hopscotch*, a retrotransposon that lies roughly 60 kb upstream of the *tb1* coding sequence[55].

Stress induction. Studies of a number of TEs suggest that they can be highly sensitive to various stresses, both biotic and abiotic, including salt[49], wounding[56], cold[49,57] and heat[58], as well as infection by bacteria[59] and viruses[60]. Not surprisingly, in some cases, insertion of TEs upstream of host genes has conferred stress responsiveness on those genes as well.

In rice, we have an excellent opportunity directly to observe the effects of a contemporaneous burst of TE activity on stress induction of gene expression.

Although the reference rice genome contains only a few dozen *mPing* MITE elements, the EG4 cultivar contains more than a thousand elements that are still active[49]. Analysis of several hundred of these new *mPing* element insertions revealed that a number of them have conferred salt or cold inducibility on the genes into which they are inserted. This suggests that in cultivated populations, there may be thousands of new stress-inducible alleles. It will be interesting to see how many of these alleles have been subject to selection. This example demonstrates the important point that the rate at which variation is produced can vary by orders of magnitude in a few generations owing to the activity of TEs, and this rate can be responsive to exactly the kinds of environmental conditions that can impose strong selection pressures.

The evidence from rice suggests that TE activity can increase the overall number of genes that are stress-inducible. This in turn can lead to an assumption that the propensity to reprogram gene expression is in itself selectivity advantageous and that TEs represent a mechanism by which the genome may be poised to respond to stress by reorganizing itself, a view first posited by Barbara McClintock[11]. However, it is important to keep in mind that the vast majority of new insertions that have any effect on gene expression are almost certainly deleterious. Therefore, rice populations with large numbers of active TEs may in fact experience a selective disadvantage under stressful conditions that overwhelms the positive benefits of increased variation.

Although co-segregation of a given insertion with a given phenotype will support the idea that a TE has in fact introduced new regulatory information, conclusive evidence will require transgenic constructs, mutant analysis or analysis of reversions generated by excision events or other rearrangements. When evaluating the effects of TEs over much longer periods of time, comparisons should be restricted to timescales that are short enough to permit identification of a given sequence as being derived from a TE but long enough to permit detection of selection on particular sequences within the TE that is consistent with exaptation of those sequences. The timescale will vary depending on the species owing to variation in mutation rates, but the roughly 50 million years separating, for instance, maize and rice is too long because conservation of non-coding regions is generally restricted to short sequences that have retained function[45]. By contrast, maize and sorghum are closely related enough to retain evidence of a TE origin, but they are distant enough to reveal selection on particular sequences that may have a function[61]. Maize also has the additional advantage of having experienced a whole-genome duplication subsequent to its divergence with sorghum, providing a suite of duplicate genes that can be compared with each other and with their sorghum orthologues[62].

Deletions, rearrangements and gene transposition

Chromosomal rearrangements. TEs can mediate large-scale changes in chromosomal architecture, a phenomenon that has been well documented in maize. Aberrant transposition of linked Ac elements in maize, for instance, can lead to deletions, inversions, translocations and other rearrangements[63]. These observations suggest that class II elements, such as Ac, may have had an important role in the kind of chromosomal rearrangements that can contribute to linkage disequilibrium, changes in gene expression and reproductive isolation in plants and animals[64–67]. Analysis of rice and its close relatives should be particularly illuminating. Rice is host to a large number of TEs of various kinds, and there is a great deal of variation within the *Oryza* genus with respect to TE copy number as well as the distribution of inversions and deletions[68–70]. A careful examination of the junctions of all of the rearrangements observed in rice and its close relatives may reveal the extent to which TE activity is responsible for them. Similar analysis of related inbred lines of maize and close maize relatives[71] or sequenced accessions of *Arabidopsis thaliana* and its relatives[72,73] may also be informative.

Gene movement. Given that gene expression patterns depend on enhancers and repressors, whether nearby or more distant, TE-mediated movement of genes into new chromosomal contexts has the potential to alter their regulation. Gene movement is a common feature of plant genomes that probably involves a number of distinct mechanisms but that can certainly be facilitated by TE activity[74–76]. For instance, a recent analysis of grass microRNA genes reveals that a substantial number of them were mobilized by TEs[77]. It will be interesting to see whether and how transposition has affected the pattern of expression of these and other transposed genes. In the case of a gene that is responsible for the oval shape of Roma tomatoes, this is certainly true. Variation at the *sun* locus is responsible for a substantial proportion of the differences in fruit shape between some varieties of *Solanum lycopersicum* and its round-fruited wild relative *Solanum pimpinellifolium*[78]. The key gene at the *SUN* locus is *IQ domain 12* (*IQD12*), which has been retrotransposed from its original position into a second gene, the promoter of which now drives expression of *IQD12* in the fruit[79] (FIG. 4). The frequency of such events could be assessed by making a systematic comparative analyses of multiple related plant species (the grasses, for instance) in which one clade has a gene that has moved from its original location, is associated with TE sequences and in which the pattern of expression is distinct from that in other clades.

Some genes, particularly those that are involved in disease resistance, such as resistance (*R*) genes, are prone to duplication and transposition[80]. When such genes undergo these processes, they often end up as a part of islands of transposed sequences, which can include both genes and TEs. In some cases, this can lead to novel forms of gene regulation. For instance, proper functioning of an *A. thaliana R* gene that is involved in resistance to *Hyaloperonospora parasitica* requires the presence of a retrotransposon immediately downstream of the coding sequence of that gene[81]. This *R* gene is

Exaptation
The process by which a trait takes on a new function. For a transposable element, this would imply a shift from facilitating replication of the transposable element to providing an adaptive function for the host.

Retroposition
The process by which mRNA from a gene is reverse-transcribed by a retrotransposon-encoded reverse transcriptase and then integrated at a new position by a retrotransposon-encoded integrase. Typically, this process involves the loss of intron sequences and the presence of a new insertion flanked by short target site duplications.

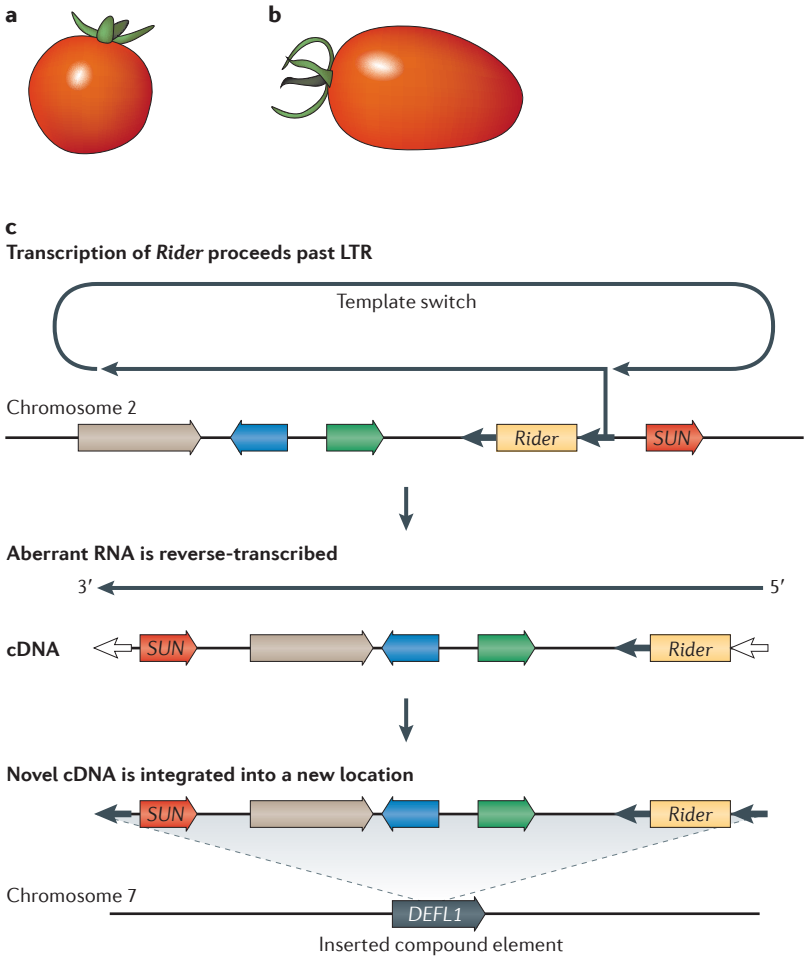


Figure 4 | **Retroposition of the *IQD12* gene at the *SUN* locus results in ectopic expression in the fruit. a** | Most strains of cultivated tomatoes (*Solanum lycopersicum*) retain the round fruit shape characteristic of their wild ancestor *Solanum pimpinellifolium*. **b** | The oval fruits of some strains of cultivated tomatoes, such as Roma, result from retroposition of the *IQ domain 12* (*IQD12*) gene. **c** | This results in tissue-specific expression of *IQD12* owing to the presence of an ectopic promoter via the following three stages. First, transcription of long terminal repeat (LTR) retrotransposons normally initiates within one LTR and terminates at the end of the second LTR. However, in this case it appears that transcription continues through several other genes (green, blue and brown arrows). Template switching then results in continued transcription of the *IQD12*, which is just upstream of the retrotransposon, after which transcription is terminated. The result is a long transcript that includes the retrotransposon as well as several other genes, including *IQD12*. Second, this transcript is reverse-transcribed, resulting in a cDNA. Third, this cDNA is then integrated into a gene, *DEFL1*, which is expressed in developing fruit. This leads to the expression of *IQD12* in the fruit because it is now under the control of the *DEFL1* promoter.

a part of a cluster of tandemly repeated *R* genes for which expression is coordinately and positively regulated by one of these *R* genes and negatively regulated by small RNAs[82]. It is intriguing that there are classes of genes that are mobile, that often increase their copy number over time, that are often found associated with TEs and that can, like TEs, be regulated by small RNAs. Given these similarities, it would be interesting to know whether some genes contain sequences that make them, like TEs, particularly prone to replication.

Transduplication
The process by which a transposable element incorporates a gene or gene fragment and duplicates and transposes it to a new position.

Gene capture. There is also clear evidence that TE activity has had a substantial impact on the evolution of new genes in plants. A notable feature of many TEs is their capacity to capture and to mobilize genes or (more frequently) gene fragments, a process called transduplication. Three thousand Pack-MULEs in rice, for instance, have mobilized fragments of more than 1,000 genes[83]. Many of these gene fragments are likely to be non-functional pseudogenes[84]. However, many of them are expressed, and many of them show signatures of selection, which is consistent with some degree of function[85].

Pack-MULE acquisition of gene fragments is only one of several TE-mediated transduplication processes[84]. Maize, for instance, contains 2,791 gene fragments that have been captured by *Helitron* elements[86], and careful annotation of rice genes has revealed the presence of at least 1,230 primary retrogenes[87]. Expression data, K_A/K_S analysis and the absence of frameshifts and indels in many of these retrogenes suggest continued function. Remarkably, 42% of these retrogenes have recruited new exons, resulting in chimeric genes, roughly one-quarter of which are also found in maize or sorghum. Given the rapid rate at which non-functional genes are lost in plants and the long divergence time between rice and sorghum (~50 million years), this observation strongly supports function. Ultimately, however, mutant analysis of these or any other gene capture events will be required to demonstrate function unambiguously.

Coding-sequence exaptation

In addition to regulatory information, autonomous TEs code for proteins that are necessary for transposition. Evidence from both plants and animals suggests that these proteins can be exapted for additional functions[88]. One particularly striking example involves the exapted MULE proteins. A screen for far-red-insensitive mutants uncovered *FAR-RED-IMPARED RESPONSE1* (*FAR1*), a gene with clear homology to the transposase encoded by *Jittery*, a currently active maize element[89]. This gene, along with its homologue *FAR-RED-ELONGATED HYPOCOTYLS 3* (*FHY3*), is a DNA-binding transcription factor that directly stimulates the transcription of *FHY1* and *FHY1-LIKE* (*FHL*), two genes that are required for phyA-dependent light responses[90,91]. Binding-site analysis as well as mutant analysis suggests that these transposase-derived transcription factors may also be involved in a wide variety of other processes as well[92]. Members of the related MULE-derived MUSTANG family of proteins are essential for flower development and fitness[93]. Given that there are at least two families (FAR1 and MUSTANG) composing six distinct clades of MULE-related proteins that appear to predate the monocot–dicot divergence, MULE transposases are likely to have been exapted early and at least twice during the evolution of higher plants[93,94]. Other class II plant transposases that are known to have been exapted are the hAT elements *Gary1* in the grasses and *DAYSLEEPER* in *Arabidopsis thaliana*[88,95].

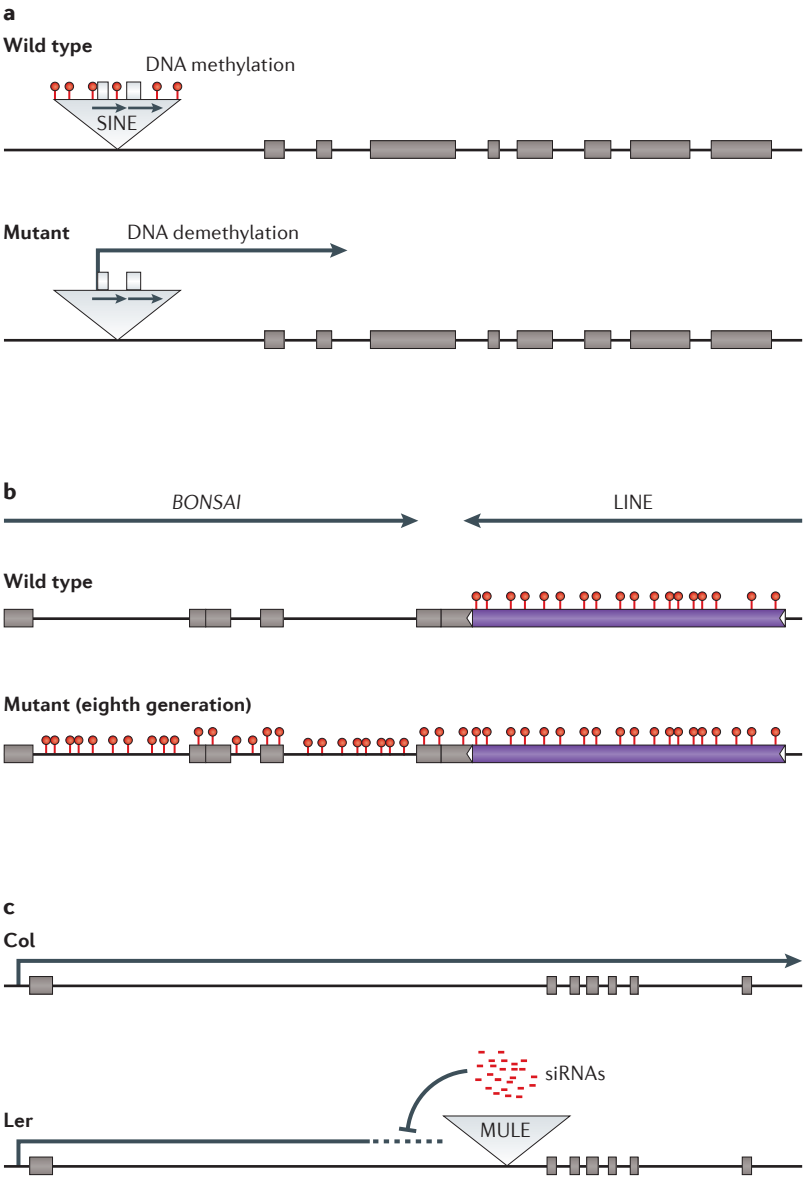


Figure 5 | **Epigenetic regulation of plant genes.** **a** | The *FLOWERING WAGENIGEN* (*FWA*) locus in *Arabidopsis thaliana*. Transcription from this gene initiates in a short interspersed element (SINE) that contains a short tandem duplication (black arrows). Expression of this transcript in vegetative tissues is prevented owing to negative regulation by small RNAs and DNA methylation (red lollipops). Mutants that alter small RNA processing or DNA methylation can result in ectopic expression of *FWA*, resulting in a late flowering phenotype that can be heritably propagated even in the absence of the mutations. The white boxes above the SINE insertion represent exons in the element. The grey boxes represent the original genic exons. **b** | The *BONSAI* (*BNS*) locus in *A. thaliana*. Normally, this gene is unmethylated and expressed, and the LTR retrotransposon immediately adjacent to it is repressed, a process that requires the activity of *DEFICIENT IN DNA METHYLATION1* (*DDM1*), a chromatin-remodelling protein required for TE repression and DNA methylation. Several generations in a *ddm1*-mutant background resulted in the spread of DNA methylation from the long interspersed element (LINE) retrotransposon into *BNS*, resulting in transcriptional silencing of that gene and severe dwarfing. **c** | The *FLOWERING LOCUS C* (*FLC*) locus in *A. thaliana*. The Col allele of this gene lacks an insertion in the first intron, and *FLC* expresses normally. The Ler allele has an insertion of a *Mutator*-like element (MULE) transposable elements into the first intron of the *FLC* gene. The resulting transcript, which contains MULE sequences, is targeted by silencing machinery and is processed by short interfering RNAs (siRNAs), resulting in loss of functional transcript.

As additional plant genomes are sequenced, it is likely that additional exaptation events will be detected. In this respect, the propensity for plant genomes to purge DNA that does not contribute to fitness (including TEs) can be an advantage to researchers; any recognizable TE coding sequence retained for longer than a few million years at the same chromosomal position is likely to provide some function to the host. A systematic analysis of such syntenic sequences in fairly distantly related plants (rice and sorghum, for instance) should provide a wealth of clues.

There is also evidence that portions of TEs can be combined with plant host genes to form new chimeric genes[96]. It is important, however, to be cautious when referring to these events as examples of exaptation. Particularly, when transposons are involved, plant genomes can be messy places, and the simple presence of an open reading frame (ORF), or even a transcript, does not necessarily mean that a given sequence has been subjected to selection. This is particularly true when confronting linage-specific events, many of which are likely to be selectively neutral. Ultimately, strong signals of selection or experimental evidence of function are essential.

Epigenetic effects

Despite the occasional selectively advantageous TE insertion, the vast majority of TE activity is almost certainly at best selectively neutral and, at worst, fatal[97,98]. Thus, it is not surprising that plants devote considerable resources to transposon control. This is achieved in large measure by epigenetic silencing of autonomous elements, a process that involves small RNAs, DNA methylation and various histone modifications[99,100]. Given that most genes appear to function as well in large genomes as in much smaller genomes (bearing in mind the fairly subtle global effects discussed below), it appears that plants are quite adept at balancing the requirements of keeping TEs inactive and genes active. However, there are exceptions that show that epigenetic regulation of a transposable element can have dramatic effects on gene expression.

Perhaps the best-studied example is the *FLOWERING WAGENINGEN* (*FWA*) locus in *A. thaliana*. In this case, a SINE element immediately upstream of the *FWA* coding sequence ensures that this gene is epigenetically silenced in vegetative tissues, a process that involves small RNAs and DNA methylation[101,102] (FIG. 5a). Ectopic expression of *FWA* in mutant backgrounds results in a late flowering phenotype. *fwa* was just one of several epialleles that arose in lines that are mutant for *DEFICIENT IN DNA METHYLATION1* (*DDM1*), which encodes a chromatin-remodelling protein that is required for efficient DNA methylation[103]. A second epiallele, *bonsai* (*bns*), arose in the *ddm1* mutant background as a consequence of derepression of a non-LTR retroelement downstream of the *BNS* coding sequence[104] (FIG. 5b). Expression of that element resulted in the production of a transcript that is antisense relative to the *BNS* gene, resulting in methylation and transcriptional silencing of that gene. Although epimutations recovered in a mutant background may not reflect natural variation, it is worth remembering

Helitron
A class of transposable element that transposes by a 'rolling circle' mechanism, a process that is frequently associated with the capture of short portions of host coding sequence.

Retrogenes
Genes that have arisen through the reverse transcription of an mRNA.

K_A/K_S
K_A and K_S are the rates of substitution at nonsynonymous sites and synonymous sites, respectively. The ratio K_A/K_S is often used to infer selection: $K_A/K_S < 1$ indicates a functional constraint; $K_A/K_S = 1$ indicates a lack of functional constraint; and $K_A/K_S > 1$ indicates positive Darwinian selection.

Transposase
An enzyme that is encoded by a gene carried by an autonomous class II transposable element and that catalyses the transposition reaction.

Syntenic
A term that describes the presence of collinear homologous DNA sequences in related chromosomal regions, implying a common ancestor. The presence of syntenic, conserved sequences over long periods of time suggests continued function, particularly in plants, which exhibit a high rate of DNA elimination.

Autonomous element
A transposable element that encodes the genes necessary for its own transposition.

Epialleles
Alternative chromatin states at a given locus, defined with respect to individuals in the population for a given time point and tissue type. Epialleles vary greatly in their stability and they affect gene expression levels or patterns rather than gene products.

Genome-wide association study
A study in which associations between genetic variation and a phenotype or trait of interest are identified by genotyping cases and controls for a set of genetic variants that capture variation across the entire genome.

that various biotic and abiotic factors can induce expression of TEs. Thus, to the extent that TEs expression can induce epigenetic changes in nearby genes, there exists the possibility for a great deal of cryptic epigenetic variation in plants.

Among natural accessions of *A. thaliana*, it also appears that the epigenetic modification of TEs may act to attenuate, rather than simply to eliminate gene expression. Epigenetic silencing of a TE inserted into the first intron of the *FLOWERING LOCUS C* (*FLC*) gene is associated with attenuation of expression of this gene in natural populations (FIG. 5c), a property that may have been selected for in late flowering accessions[105,106].

It might be expected that plant genomes with more TEs would exhibit more of this kind of variation. Indeed, the rice genome, which is considerably larger than the *A. thaliana* genome, shows evidence for a number of TEs that appear to regulate nearby genes by antisense transcription. In plants, LTR retroelements are generally methylated and transcriptionally inactive[107–109]. However, unmethylated LTRs can become transcriptionally active, and this can result in expression of transcripts originating in the retrotransposon and extending into flanking sequences[104,110]. *Dasheng* is a recently amplified retroelement that is polymorphic in a number of species and subspecies of rice[111,112]. Expression of some *Dasheng* elements can result in downregulation of adjacent rice genes[113]. Remarkably, this hypomethylation can also be tissue-specific, such that these genes have become regulated in a tissue-specific manner owing to differences in TE methylation. Given that stress can alter expression of LTR retroelements, it would be fascinating to see whether stress affects epigenetic regulation of genes that are adjacent to *Dasheng* as well. Re-examination of recently obtained global methylation and expression data for *Oryza sativa* subsp. *japonica* and *Oryza sativa* subsp. *indica* with these results in mind should be informative[114].

Finally, it is worth considering the global effects of silenced TEs on gene expression. Unrestrained TE activity is clearly deleterious, but efficient silencing of TEs can also have negative consequences, particularly when the silenced TEs are close to genes. A recent analysis of levels of gene expression in *A. thaliana* revealed that genes with TEs that are epigenetically silenced have a lower level of expression on average[115], an effect that varies in intensity depending on the species of *Arabidopsis* examined[116]. Similarly, a recent analysis of MITEs in rice (a major target of short interfering RNAs (siRNAs) in this species) suggests that genes with MITE insertions are more likely, on average, to be expressed at a lower level[117]. Together, these data suggest that the load imposed by TEs on their hosts may be more severe than previously appreciated. Although examples such as *FWA* suggest that epigenetic regulation can be co-opted by the host, efficient epigenetic regulation of TE activity may also impose a substantial cost. A burst of TE activity may introduce a great deal of variation, but it may also end up having a generally repressive effect on gene expression and thus may reduce the overall fitness of a population or a species.

Conclusion: limits and opportunities

On the basis of what we know now, it is clear that TEs are capable of causing many kinds of genetic variation, and it is also clear that they have had important effects on the course of plant evolution. Doubtlessly, many more examples of TE involvement in the evolution of new adaptive traits will emerge over the next few years. For those of us who study TEs, each such instance provides additional evidence that TEs represent an endogenous system that provides a degree of evolvability that would not be available otherwise. The implicit assumption is that the trajectory of plant evolution would be very different in their absence and that a substantial portion of the genetic variation that has been necessary for plant adaptation has been due to TE activity. But wishing it does not make it so, and we are faced at this point with some always-difficult evolutionary questions. First, how much has it actually happened and how often has it mattered? And second, has the ability for TEs to cause very rapid and programmatic changes — for example, simultaneous shifts in both tissue-specificity and cold-sensitivity — played an essential part in plant evolution?

We are constrained in answering these questions by the rapid rate of plant genome evolution; even if a TE were exapted for stress responsiveness, for instance, evidence of its TE origins would be rapidly erased, as only sequences that are important for function would be retained after only a few million years. This is probably the reason that much of the best evidence to date for a role for TEs in plant evolution comes from analysis of domestication, which occurred only recently and involved selection for a wide variety of very well-characterized traits. For these reasons, domesticated plants may be an excellent place to begin to address these questions systematically. Recent advances in phenomics, genomics and trait mapping using the genome-wide association study approach provide us with the means to determine how many of the mutations that have been important for domestication and subsequent trait modification have been caused by TE activity[33,44,118,119]. Further, many of the tools used for mapping traits in domesticated plant species can and will be used for mapping traits in natural plant populations, such as wild relatives of domesticated crop species, which will extend our reach from artificial to natural selection.

On deeper timescales, we must depend on evidence of selection by comparing homologous sequences at various timescales. For instance, the maintenance of exapted transposases at the same chromosomal position in multiple grasses lineages provides excellent, almost irrefutable evidence for function. In a similar vein, any recognizable TE sequence that is retained at the same position over long periods of time is likely to have been exapted and is worth further investigation. A systematic comparison of large numbers of related plant genomes with careful attention paid to TE sequences will be informative, as will be a comparison of genes that are duplicated as a result of the whole-genome duplication events that have periodically occurred during plant evolution[120]. If these types of data were combined with large-scale trait mapping, we would have a far better handle on the role that TEs have had.

Certainly, much of this work will be done regardless of whether or not TEs are involved; reducing phenotype to genotype and understanding the origin during evolution of phenotypic change are at the core of all of biology. However, there are important computational challenges when dealing with repetitive elements. For a given TE in maize, for instance, a given short sequencing read can match hundreds of loci, making both genome sequencing and quantitative assessment of expression levels difficult but not impossible[121]. Along with an assumption that TEs are selectively neutral, these problems may have led to a systematic bias against identification of TEs that have had important effects on phenotype (indeed, it is no coincidence that the most popular tool for identification of TEs is called 'RepeatMasker'[122]).

TEs represent a rich and constantly changing pool of genetic and epigenetic variation on which selection can operate. The examples presented here are suggestive, but conclusions based on anecdotal evidence are inevitably subject to confirmation bias. An honest assessment of the relative importance of TEs for adaptive evolution will require an unbiased survey of all sources of genetic variation that collectively contribute to that process. TEs may in fact have had an essential role, but we really will not know until we look. And now, for the first time, we can.

1. Bennetzen, J. L., Ma, J. & Devos, K. M. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* **95**, 127–132 (2005).
2. Dooner, H. K. & Weil, C. F. Give-and-take: interactions between DNA transposons and their host plant genomes. *Curr. Opin. Genet. Dev.* **17**, 486–492 (2007).
3. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9**, 397–405 (2008).
4. Tollis, M. & Boissinot, S. The evolutionary dynamics of transposable elements in eukaryote genomes. *Genome Dyn.* **7**, 68–91 (2012).
5. Candela, H. & Hake, S. The art and design of genetic screens: maize. *Nature Rev. Genet.* **9**, 192–203 (2008).
6. Hsing, Y. *et al.* A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.* **63**, 351–364 (2007).
7. Settles, A. M. *et al.* Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics* **8**, 116 (2007).
8. 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).
9. Tsukahara, S. *et al.* Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* **461**, 423–426 (2009).
10. Wang, H. Y. *et al.* Transpositional reactivation of two LTR retrotransposons in rice-Zizania recombinant inbred lines (RILs). *Hereditas* **147**, 264–277 (2010).
11. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
12. Lin, C. *et al.* Dramatic genotypic difference in, and effect of genetic crossing on, tissue culture-induced mobility of retrotransposon Tos17 in rice. *Plant Cell Rep.* **31**, 2057–2063 (2012).
13. Naito, K. *et al.* Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl Acad. Sci. USA* **103**, 17620–17625 (2006).
14. Ungerer, M. C., Strakosh, S. C. & Zhen, Y. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* **16**, R872–873 (2006).
15. 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).
16. Wessler, S. R. Turned on by stress. Plant retrotransposons. *Curr. Biol.* **6**, 959–961 (1996).
17. Maumus, F. *et al.* Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* **10**, 624 (2009).
18. Kawase, M., Fukunaga, K. & Kato, K. Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol. Genet. Genom.* **274**, 131–140 (2005).
This comprehensive analysis of hundreds of accessions of foxtail millet demonstrates that all known 'sticky' accessions in this species are the result of TE-induced mutations of the *waxy* locus.
19. Bhattacharyya, M. K., Smith, A. M., Ellis, T. H., Hedley, C. & Martin, C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**, 115–122 (1990).
20. Kobayashi, S., Goto-Yamamoto, N. & Hirochika, H. Retrotransposon-induced mutations in grape skin color. *Science* **304**, 982 (2004).
21. Cadle-Davidson, M. M. & Owens, C. L. Genomic amplification of the Gret1 retroelement in white-fruited accessions of wild vitis and interspecific hybrids. *Theor. Appl. Genet.* **116**, 1079–1094 (2008).
22. Shimazaki, M., Fujita, K., Kobayashi, H. & Suzuki, S. Pink-colored grape berry is the result of short insertion in intron of color regulatory gene. *PLoS ONE* **6**, e21308 (2011).
23. Yao, J., Dong, Y. & Morris, B. A. Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proc. Natl Acad. Sci. USA* **98**, 1306–1311 (2001).
24. Clegg, M. T. & Durbin, M. L. Tracing floral adaptations from ecology to molecules. *Nature Rev. Genet.* **4**, 206–215 (2003).
25. Clegg, M. T. & Durbin, M. L. Flower color variation: a model for the experimental study of evolution. *Proc. Natl Acad. Sci. USA* **97**, 7016–7023 (2000).
26. Park, K. I. *et al.* A bHLH regulatory gene in the common morning glory, *Ipomoea purpurea*, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed trichome formation. *Plant J.* **49**, 641–654 (2007).
27. Glover, D., Durbin, M., Huttley, G. & Clegg, M. T. Genetic diversity in the common morning glory. *Plant Species Biol.* **11**, 41–50 (1996).
28. Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
29. Pelsy, F. Molecular and cellular mechanisms of diversity within grapevine varieties. *Heredity* **104**, 331–340 (2010).
30. Izawa, T., Konishi, S., Shomura, A. & Yano, M. DNA changes tell us about rice domestication. *Curr. Opin. Plant Biol.* **12**, 185–192 (2009).
31. Freeling, M. *et al.* Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* **15**, 131–139 (2012).
32. Weigel, D. Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* **158**, 2–22 (2012).
33. Hamilton, J. P. & Buell, C. R. Advances in plant genome sequencing. *Plant J.* **70**, 177–190 (2012).
34. Feuillet, C. & Eversole, K. Plant science. Solving the maze. *Science* **326**, 1071–1072 (2009).
35. Doi, K., Yasui, H. & Yoshimura, A. Genetic variation in rice. *Curr. Opin. Plant Biol.* **11**, 144–148 (2008).
36. Huang, X. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
This is an example of the scale of analysis that is now possible using modern sequencing technology. Sequencing data from hundreds of accessions of wild and domesticated rice has made high-resolution genome-wide association studies (GWASs) for many agronomic traits in cultivated and wild rice possible.
37. Xu, X. *et al.* Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotech.* **30**, 105–111 (2012).
38. He, Z. *et al.* Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet.* **7**, e1002100 (2011).
39. Kloeckener-Gruissem, B. & Freeling, M. Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles. *Proc. Natl Acad. Sci. USA* **92**, 1836–1840 (1995).
40. Greene, B., Walko, R. & Hake, S. Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations. *Genetics* **138**, 1275–1285 (1994).
41. Bharathan, G. *et al.* Homologies in leaf form inferred from *KNOXI* gene expression during development. *Science* **296**, 1858–1860 (2002).
42. Salvi, S. *et al.* Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol.* **48**, 601–613 (2002).
43. Salvi, S. *et al.* Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl Acad. Sci. USA* **104**, 11376–11381 (2007).
This is the first demonstration that insertion of a TE into a conserved non-coding sequence can affect an agronomically important trait. This analysis also demonstrates that even TE insertions that are distant from plant genes can have important effects on gene expression.
44. Chia, J. M. *et al.* Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genet.* **44**, 803–807 (2012).
This paper provides a comprehensive view of extant genetic variation in maize and relates it to variation in agronomically important traits in a species that has an order of magnitude more genetic diversity than humans. Because of the density of the data now available, this analysis makes it possible for high-resolution GWASs to be carried out in maize for the first time.
45. Freeling, M. & Subramaniam, S. Conserved noncoding sequences (CNSs) in higher plants. *Curr. Opin. Plant Biol.* **12**, 126–132 (2009).
46. Liu, S. *et al.* Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet.* **5**, e1000733 (2009).
47. Vollbrecht, E. *et al.* Genome-wide distribution of transposed Dissociation elements in maize. *Plant Cell* **22**, 1667–1685 (2010).
48. Pan, X., Li, Y. & Stein, L. Site preferences of insertional mutagenesis agents in *Arabidopsis*. *Plant Physiol.* **137**, 168–175 (2005).
49. Naito, K. *et al.* Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461**, 1130–1134 (2009).
This study demonstrates that rapid and recent amplification of a family of TEs can have substantial effects on expression of nearby genes by introducing stress-responsiveness. It also provides evidence that MITEs may have adapted to minimize the negative effects of transposition by avoiding insertion into exons.
50. Dooner, H. K., Robbins, T. P. & Jorgensen, R. A. Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* **25**, 173–199 (1991).
51. Selinger, D. A. & Chandler, V. L. Major recent and independent changes in levels and patterns of expression have occurred at the *b* gene, a regulatory locus in maize. *Proc. Natl Acad. Sci. USA* **96**, 15007–15012 (1999).

52. Selinger, D. A. & Chandler, V. L. *B-Bolivia*, an allele of the maize *b1* gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. *Plant Physiol.* **125**, 1363–1379 (2001).
53. Walker, E. L., Robbins, T. P., Bureau, T. E., Kermicle, J. & Dellaporta, S. L. Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. *EMBO J.* **14**, 2350–2363 (1995).
54. Butelli, E. *et al.* Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* **24**, 1242–1255 (2012).
This paper provides an example of the programmatic changes that can be introduced by TE insertion. In this case, insertion of a retroelement provided both cold inducibility and tissue specificity.
55. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genet.* **43**, 1160–1163 (2011).
This paper demonstrates that a key mutation permitting domestication of maize involved a retroelement insertion many kilobases upstream of a gene that enhanced that gene's expression.
56. Mhiri, C. *et al.* The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* **33**, 257–266 (1997).
57. Ivashuta, S. *et al.* Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (*Medicago sativa*). *Plant J.* **31**, 615–627 (2002).
58. Ito, H. *et al.* An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**, 115–119 (2011).
59. Grandbastien, M. *et al.* Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* **110**, 229–241 (2005).
60. Buchmann, R. C., Asad, S., Wolf, J. N., Mohannath, G. & Bisaro, D. M. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol.* **83**, 5005–5013 (2009).
61. Wang, Q. & Dooner, H. K. Dynamic evolution of bz orthologous regions in the Andropogoneae and other grasses. *Plant J.* **72**, 212–221 (2012).
62. Swigonova, Z. *et al.* Close split of sorghum and maize genome progenitors. *Genome Res.* **14**, 1916–1923 (2004).
63. Yu, C., Zhang, J. & Peterson, T. Genome rearrangements in maize induced by alternative transposition of reversed ac/ds termini. *Genetics* **188**, 59–67 (2011).
64. Navarro, A. & Barton, N. H. Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science* **300**, 321–324 (2003).
65. Guillen, Y. & Ruiz, A. Gene alterations at *Drosophila* inversion breakpoints provide prima facie evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics* **13**, 53 (2012).
66. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
67. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500 (2010).
68. Ammiraju, J. S. *et al.* Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell* **20**, 3191–3209 (2008).
69. Hurwitz, B. L. *et al.* Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus *Oryza*. *Plant J.* **63**, 990–1003 (2010).
70. Piegu, B. *et al.* Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**, 1262–1269 (2006).
71. Vielle-Calzada, J. P. *et al.* The *Palomero* genome suggests metal effects on domestication. *Science* **326**, 1078 (2009).
72. Ledford, H. Halfway point for 1,001 genomes quest. *Nature* **477**, 14 (2011).
73. Hu, T. T. *et al.* The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genet.* **43**, 476–481 (2011).
74. Woodhouse, M. R., Pedersen, B. & Freeling, M. Transposed genes in *Arabidopsis* are often associated with flanking repeats. *PLoS Genet.* **6**, e1000949 (2010).
75. Yang, S. *et al.* Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLoS Genet.* **4**, e3 (2008).
76. Bhutkar, A., Russo, S. M., Smith, T. F. & Gelbart, W. M. Genome-scale analysis of positionally relocated genes. *Genome Res.* **17**, 1880–1887 (2007).
77. Abrouk, M. *et al.* Grass microRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication. *Plant Cell* **24**, 1776–1792 (2012).
78. van der Knaap, E., Sanyal, A., Jackson, S. A. & Tanksley, S. D. High-resolution fine mapping and fluorescence *in situ* hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. *Genetics* **168**, 2127–2140 (2004).
79. Xiao, H., Jiang, N., Schaffner, E., Stockinger, E. J. & van der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**, 1527–1530 (2008).
This study provides the first evidence in plants that retrotransposon-mediated retrotransposition can result in functional reprogramming patterns of gene expression.
80. Freeling, M. *et al.* Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. *Genome Res.* **18**, 1924–1937 (2008).
81. Wang, Y. H. & Warren, J. T. Jr. Mutations in retrotransposon AtCOPIA4 compromises resistance to *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. *Genet. Mol. Biol.* **33**, 135–140 (2010).
82. Yi, H. & Richards, E. J. A cluster of disease resistance genes in *Arabidopsis* is coordinately regulated by transcriptional activation & RNA silencing. *Plant Cell* **19**, 2929–2939 (2007).
83. Jiang, N., Bao, Z., Zhang, X., Eddy, S. R. & essler, S. R. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431**, 569–573 (2004).
In this paper, it is shown that TEs can capture, duplicate and combine thousands of gene fragments in rice, raising the possibility that TEs may be an important source of new genes in plants.
84. Juretic, N., Hoen, D. R., Huynh, M. L., Harrison, P. M. & Bureau, T. E. The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res.* **15**, 1292–1297 (2005).
85. Hanada, K. *et al.* The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. *Plant Cell* **21**, 25–38 (2009).
86. Du, C., Fefelova, N., Caronna, J., He, L. & Dooner, H. K. The polychromatic *Helitron* landscape of the maize genome. *Proc. Natl Acad. Sci. USA* **106**, 19916–19921 (2009).
87. Wang, W. *et al.* High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **18**, 1791–1802 (2006).
88. Muehlbauer, G. J. *et al.* A hAT superfamily transposase recruited by the cereal grass genome. *Mol. Genet. Genom.* **275**, 553–563 (2006).
89. Hudson, M. E., Lisch, D. R. & Quail, P. H. The *FHY3* and *FAR1* genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. *Plant J.* **34**, 453–471 (2003).
90. Li, G. *et al.* Coordinated transcriptional regulation underlying the circadian clock in *Arabidopsis*. *Nature Cell Biol.* **13**, 616–622 (2011).
91. Lin, R. *et al.* Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. *Science* **318**, 1302–1305 (2007).
This is the first demonstration that a plant transposase can take on a new functional role as a transcription factor. This study also showed that a family of exapted transposases has a key role in light signalling in plants.
92. Ouyang, X. *et al.* Genome-wide binding site analysis of *FAR-RED ELONGATED HYPOCOTYL3* reveals its novel function in *Arabidopsis* development. *Plant Cell* **23**, 2514–2535 (2011).
93. Joly-Lopez, Z., Forczek, E., Hoen, D. R., Juretic, N. & Bureau, T. E. A. Gene family derived from transposable elements during early angiosperm evolution has reproductive fitness benefits in *Arabidopsis thaliana*. *PLoS Genet.* **8**, e1002931 (2012).
94. Cowan, R. K., Hoen, D. R., Schoen, D. J. & Bureau, T. E. MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. *Mol. Biol. Evol.* **22**, 2084–2089 (2005).
95. Bundock, P. & Hooykaas, P. An *Arabidopsis* hAT-like transposase is essential for plant development. *Nature* **436**, 282–284 (2005).
96. Donoghue, M. T., Keshavaiah, C., Swamidatta, S. H. & Spillane, C. Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. *BMC Evol. Biol.* **11**, 47 (2011).
97. Charlesworth, B., Borthwick, H., Bartolome, C. & Pignatelli, P. Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. *Genetics* **167**, 815–826 (2004).
98. Brookfield, J. F. The ecology of the genome — mobile DNA elements and their hosts. *Nature Rev. Genet.* **6**, 128–136 (2005).
99. Lisch, D. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* **60**, 43–66 (2009).
100. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nature Rev. Genet.* **8**, 272–285 (2007).
101. Kinoshita, Y. *et al.* Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* **49**, 38–45 (2007).
102. Fujimoto, R. *et al.* Evolution and control of imprinted FWA genes in the genus *Arabidopsis*. *PLoS Genet.* **4**, e1000048 (2008).
103. 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).
104. Saze, H. & Kakutani, T. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J.* **26**, 3641–3652 (2007).
105. Liu, J. He, Y., Amasino, R. & Chen, X. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.* **18**, 2873–2878 (2004).
106. Strange, A. *et al.* Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions. *PLoS ONE* **6**, e19949 (2011).
107. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
108. Gehring, M., Bubb, K. L. & Henikoff, S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**, 1447–1451 (2009).
109. Zemach, A. *et al.* Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl Acad. Sci. USA* **107**, 18729–18734 (2010).
110. Kashkush, K., Feldman, M. & Levy, A. A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genet.* **33**, 102–106 (2003).
This paper shows that read-out transcription from retroelements in plants can have substantial effects on gene expression. It also demonstrates that naturally occurring variation in methylation of retroelements can result in differences in expression and tissue specificity of genes.
111. Jiang, N., Jordan, I. K. & Wessler, S. R. Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. *Plant Physiol.* **130**, 1697–1705 (2002).
112. Ammiraju, J. S. *et al.* Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. *Plant J.* **52**, 342–351 (2007).
113. Kashkush, K. & Khasdan, V. Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics* **177**, 1975–1985 (2007).
114. He, G. *et al.* Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**, 17–33 (2010).
115. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).
116. Hollister, J. D. *et al.* Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl Acad. Sci. USA* **108**, 2322–2327 (2011).
Evidence is provided in this paper that silencing of TEs near genes can result in an overall reduction in gene expression. This suggests that although there is a benefit to silencing TEs, there is also a substantial potential cost, and it raises the possibility that TEs can have global effects on host gene function.

117. Lu, C. *et al.* Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol. Biol. Evol.* **29**, 1005–1017 (2012).
118. Furbank, R. T. & Tester, M. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **16**, 635–644 (2011).
119. Brachi, B., Morris, G. P. & Borevitz, J. O. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**, 232 (2011).
120. Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L. & Vandepoele, K. The flowering world: a tale of duplications. *Trends Plant Sci.* **14**, 680–688 (2009).
121. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Rev. Genet.* **13**, 36–46 (2012).
122. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–4.10.14 (2009).

Acknowledgements
The author apologizes to the many authors whose work he was unable to include owing to space constraints. The author is supported by a grant from the US National Science Foundation (DBI 031726).

Competing interests statement
The author declares competing financial interests: see Web version for details.

FURTHER INFORMATION

Damon Lisch's homepage: http://nature.berkeley.edu/dlisch/Mutator/Mutator_Home.html
1001 Genomes Project: http://www.1001genomes.org/index.html
The Australian Plant Phenomics Facility: http://www.plantphenomics.org.au
CoGe: ANKoCG: http://genomevolution.org/CoGe/index.pl
Genetic Information Research Institute: http://www.girinst.org
Genotype to Phenotype (iPG2P): http://www.iplantcollaborative.org/challenge/iplant-genotype-phenotype
Gramene QTL Database: http://www.gramene.org/qtl
International Rice Functional Genomics Consortium: http://irfgc.irri.org/index.php?option=com_content&task=view&id=68&Itemid=9
iPlant Collaborative: http://www.iplantcollaborative.org
Maize transposable element database: http://maizetedb.org/~maize
Panzea: http://www.panzea.org
Plant Repeat Databases at Michigan State: http://plantrepeats.plantbiology.msu.edu/index.html
PlantGDB: http://plantgdb.org
RepeatMasker homepage: http://www.repeatmasker.org
Rice Diversity: http://www.ricediversity.org
Sequenced plant genomes — CoGePedia: http://genomevolution.org/wiki/index.php/Sequenced_plant_genomes
The UniformMu Transposon Resource: http://www.maizegdb.org/documentation/uniformmu
The VAST lab: Variation and Abiotic Stress Tolerance — RIL populations resource: http://www.inra.fr/vast/RILs.htm

ALL LINKS ARE ACTIVE IN THE ONLINE PDF