Transposable elements employ distinct integration strategies with respect to transcriptional landscapes in eukaryotic genomes

Xinyan Zhang\textsuperscript{1,2}, Meixia Zhao\textsuperscript{3}, Donald R. McCarty\textsuperscript{4} and Damon Lisch\textsuperscript{2,*}

\textsuperscript{1}Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China, \textsuperscript{2}Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA, \textsuperscript{3}Department of Biology, Miami University, Oxford, OH 45056, USA and \textsuperscript{4}Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA

Received November 25, 2019; Revised April 24, 2020; Editorial Decision April 27, 2020; Accepted April 29, 2020

ABSTRACT

Transposable elements (TEs) are ubiquitous DNA segments capable of moving from one site to another within host genomes. The extent distributions of TEs in eukaryotic genomes have been shaped by both \textit{bona fide} TE integration preferences in eukaryotic genomes and by selection following integration. Here, we compare TE target site distribution in host genomes using multiple \textit{de novo} transposon insertion datasets in both plants and animals and compare them in the context of genome-wide transcriptional landscapes. We showcase two distinct types of transcription-associated TE targeting strategies that suggest a process of convergent evolution among eukaryotic TE families. The integration of two precision-targeting elements are specifically associated with initiation of RNA Polymerase II transcription of highly expressed genes, suggesting the existence of novel mechanisms of precision TE targeting in addition to passive targeting of open chromatin. We also highlight two features that can facilitate TE survival and rapid proliferation: tissue-specific transposition and minimization of negative impacts on nearby gene function due to precision targeting.

INTRODUCTION

Transposable elements (TEs) are ubiquitous mobile DNA sequences that can make additional copies of themselves and integrate into new positions in host genomes. Because of these properties, TEs can make up the majority, sometimes the vast majority, of eukaryotic genomes \cite{1}. Indeed, the overall architecture of many genomes is determined in large measure by the quantity and distribution of TEs, which in turn is determined by both integration preferences and selection following integration.

According to their structural and biochemical features, TEs can be classified as retrotransposons (Class I TEs) or DNA transposons (Class II TEs). Both Class I and Class II TEs can be either autonomous or non-autonomous. Autonomous elements encode proteins that are capable of mobilizing both autonomous and non-autonomous elements. Non-autonomous elements can only transpose in the presence of their cognate autonomous elements. Retrotransposons duplicate via a ‘copy-and-paste’ mechanism involving reverse transcription as a step in the process of replication. DNA transposons transpose via a ‘cut-and-paste’ mechanism, in which an element is physically excised from one position and reintegrated at a second position.

TEs have a significant impact on genome evolution. Increases in TE copy numbers result in increased genome size, complexity and instability \cite{2}. TE transposition is associated with chromosomal structural variation \cite{3} and can also affect expression of individual genes \cite{4}. TEs have served as a rich source of novel mutations on which selection can operate and have contributed to gene evolution and phenotypic diversification \cite{5}. Despite this, changes induced by TEs are primarily neutral or deleterious to their hosts \cite{6}, and TEs are kept under strict control by host immune systems. Overlapping transcriptional and post-transcriptional silencing mechanisms have evolved in plants and animals as layered defenses that have evolved to repress TE expression and amplification \cite{7,8}. Although this system is highly efficient and results in epigenetic silencing of most TEs in most genomes, it is clear that TEs can also undergo rapid increases in copy number, and currently or recently active TEs have been identified in a wide variety of organisms \cite{9}.

The consequences of TE activity depend largely on where TEs integrate. While TE integration sites in eukaryotic genomes are broadly distributed, different TEs adopt distinct integration strategies, resulting in dramatically differ-
ent insertion profiles (10). There is ample evidence that both DNA transposons and retrotransposons insert non-randomly in host genomes. For instance, in maize, Activator elements preferentially target linked gene regions (11), and maize Mutator (Mu) elements target unlinked open chromatin regions near recombination hot spots, which tend to be in the 5′ ends of genes (12). P elements in Drosophila integration has been associated with replication origins, which are also in the 5′ end of genes (13). Integration of L1 retrotransposons in the human genome also appears to be influenced by DNA replication, and is not targeted to either actively transcribed regions or to open chromatin (14, 15). Some other retrotransposons, such as Ty1 in yeast, target nucleosome-bound DNA near the H2A/H2B interface upstream of Pol III-transcribed genes due to physical interaction between the transposase and RNA polymerase III subunits (16–19). In contrast, gene space is a preferred target of many TEs (10), presumably because these are regions of relatively open chromatin, which may facilitate both integration and subsequent expression of autonomous elements (20, 21). However, this is not always the case. Tall, for instance, targets centromeres (22) and Ty5 targets silenced heterochromatin (23). Given that transposases in many cases are recruited to particular genomic niches via physical interaction with pre-seated proteins (16, 17, 23–27), a tethering model has been proposed for TE targeting (10, 26, 27). In addition, the timing of transposition is critical for copy number increase of cut-and-paste TE element and is therefore important for TE amplification. For instance, transposition of Activator primarily takes place after replication of the donor site but before replication of the target site, resulting in a net increase in copy number following cell replication (28).

Although there has been a great deal of analysis of de novo insertions of TEs in a variety of species, there have been few broad comparisons of TE insertion profiles in both animals and plants. Gene expression is often associated with open, accessible chromatin, which in turn is associated with insertion of a number of elements in both plants and animals (12, 13). Further, RNAseq gene expression data are available from a broad range of tissues in multiple species, making direct comparisons relatively easy (29–31). With this in mind, we performed a comparison of multiple TEs in multiple species and examined TE distribution in the context of genome-wide transcriptional landscapes using multiple relatively unselected de novo transposon insertion datasets collected by many groups, including our own. We identify two distinct types of Pol II-associated TE targeting strategies, as well as those that are independent of Pol II transcription, and we provide evidence for convergent evolution among plant and animal TE families. We also provide data that suggests that TEs have evolved strategies to minimize their effects on host gene expression, even when those TEs specifically target gene space.

**MATERIALS AND METHODS**

**Collecting de novo transposon coordinates and RNA-seq datasets**

The TE families that were analyzed in this study have all been employed as efficient mutagens and each has been used for sequence-indexed mutant library construction. The UniformMu, Dissociation (Ds)-GFP, Transposon of Oryza sativa 17 (Tos17), Ds, Suppressor-mutator (Spm), P, piggyBac (Pb) and Minos (Mb) de novo transposon insertion collections in maize, rice and Drosophila are publicly available. Coordinates of TE insertions were retrieved from relevant websites and databases (32–38). Somatic Mu elements (SomaticMu elements) have been generated by performing Mutu-seq with leaves collected from high-copy Mu-active maize seedlings and coordinates were called using the same pipeline as was used for the germinally inserted UniformMu collection (39). In the >320,000 SomaticMu insertions analyzed, a few hundred germininal insertions (ancient and germinally transmitted background insertions) were not removed as their impact on the overall distribution of SomaticMu insertions is negligible. We also collected 1358 annotated Pack-MULEs in the maize genome (ftp://ftp.grame.org/pub/gramene/release61/gff3/zea_mays/repeat_annotation/B73v4.TE.filtered.gff3.gz) and 2959 Pack-MULEs in the rice genome from the literature (40). The TE coordinates in each organism were made consistent with current genome assembly versions (AGPv4 B73 for maize, Oryza sativa.IRGSP-1.0.42 for rice and FB2014.03, R5.57 for fly). All coordinates for all insertions in each species are provided in Supplemental Tables S1–S5. Coordinates for insertions near tRNA and rRNA genes are provided in Supplemental Table S6.

Raw FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values of publicly available RNA-sequencing experiments for AGPv4 genes from maize were retrieved from the Maize Genetic Resource database (http://maize.plantbiology.msu.edu) (29). PCA analysis was performed based on the average FPKM values generated from RNAseq datasets of different maize tissues using the R package FactoMineR (41). The rice and Drosophila RNAseq data were retrieved from the Rice Expression Database and FlyAtlas 2, respectively (30, 31). Expression levels for all genes used this analysis are available in Supplemental Table S7.

**Meta-analysis of transposon distributions near transcription start sites (TSSs) and transcription termination sites (TTSs) of genes at various transcriptional landscapes**

For a given gene set (all genes in a genome or a gene subset), each de novo and ancient TE insertion was classified as being either genic or intergenic and distances from transcriptional start sites (TSSs) and transcriptional termination sites (TTSs) were calculated for each insertion event. Genic insertions were plotted along the positive X-axis and intergenic insertions were plotted along the negative X-axis relative to the TSS and TTS in comparison to randomly selected genomic loci. A total of 421 280 random insertions were in-silico generated on the maize chromosomes at a density of one insertion per 5 kb, which is comparable to the largest insertion dataset (SomaticMu). The densities of random insertions in both rice and Drosophila were set to be one insertion per 400 bp on average given that these genomes are relatively small and gene-rich. For each TE insertion or randomly selected locus within intergenic regions, its distance to the TSS and the TTS of both the nearest upstream gene
and the nearest downstream gene were counted in the meta-
profiling plots. In a number of gene rich regions, a small
proportion of TE insertions or randomly distributed loci
are <4 kb from both upstream and downstream genes, so
they were counted twice. This would be expected to cause
a mildly uneven distribution of a subset of the random se-
lected loci.

The metaprofiles for intergenic and genic transposon in-
sertions were plotted separately, using normalized insertion
numbers in sliding 30-basepair (bp) windows centered on
each position. In order to compare the enrichment of TEs
near all annotated genes (or subsets of genes), normaliza-
tion of the insertion numbers was performed by calculating
the number of insertions per 30 bp window per 100 000 in-
sertions per 10 000 genes at each position surrounding TSSs
or TTSs. For both the TSSs and TTSs, genic transposon in-
sertions were plotted along the positive X-axis while the in-
tergenic insertions were plotted along negative X-axis.

For CHH islands near the 5′ or 3′ ends of genes, the rel-
ative position of these islands (which are 100 bp in length)
and gene TSSs or TTSs was unified in such a way that the
CHH islands were located on an interval [−50, 49] on X-axis
of each plot, and adjacent genes were placed downstream
of both 5′-end CHH and 3′-end CHH on the positive X-axis.
Insertion numbers at each position are normalized to 100
000 insertions and 10 000 CHH islands. Given the repetitive
nature of tRNA, and particularly rRNA genes, special care
was taken to ensure that only independent insertions were
counted by using polymorphisms between sequences flank-
ing the insertions in these genes. The ~30% of insertions
into tRNA and rRNA genes that lacked sufficient polymor-
phism were not included in our analysis.

Independent RNAseq experiments were treated as repli-
cates in each organism. For each experiment, genes were
placed into 20 bins based on their relative level of expres-
sion, with bin 1 representing the lowest level of expression
and bin 20, the highest. The percentage of TSS-associated
insertions (<2 kb upstream of TSSs) in each bin were calcu-
lated for each RNAseq dataset, and the averaged percent-
ages in all datasets were plotted along the X-axis. For each
experiment, a bin represents a categorical level of gene ex-
pression. That is to say, a bin is not always the collection
of identical genes in independent RNAseq experiments, but
rather it contains a set of genes whose ranks based on their
expression level fall in the same category in each particular
experiment.

Sequencing-based transposon profiling and sequencing-based
allele frequency analysis
Miseq-based Mu element profiling was performed as de-
scribed previously using F1 hybrid progeny seedlings (42).
The B73 parent was carried Mutator activity that had been
introduced into the B73 genetic background. The Mo17 parent lacked active Mu elements. Thus, all new in-
sertions were into the B73 genome. Genomic DNA was
extracted from 6-day-old seedlings of B73/Mo17 hybrid
plants. Amplicon-based enrichment of Mu flanking DNA
was then performed. The purified PCR products were sub-
ject to Miseq-based Wideseq pipeline at Purdue Genomics
Core Facility (https://www.purdue.edu/hla/sites/genomics/
wideseq-2/). Wideseq reads were mapped to the B73 refer-
cence genome as described previously (42). By identifying the
Mu target site duplications (TSDs), a set of genes targeted by
Mu insertions that segregated in hybrid progeny was
obtained and those containing B73/Mo17 SNPs in their
mRNA sequences were used for allele-specific expression
analysis. Because new insertions were into the B73 genome,
the effect of these insertions would be expected to be specific
in all cases to the B73 allele. To quantify the allele frequency,
we performed RT-PCR followed by Wideseq from the ident-
ical shoot tissues of the hybrid seedlings mentioned above.
Total RNA was extracted using the RNA Extraction Kit
(Zymo) and cDNAs were synthesized using Promega M-
MLV Reverse Transcriptase. For a subset of 16 genes that
carried SNPs, RNA fragments containing B73/Mo17 SNPs
were amplified by RT-PCR. Primers used for this analysis
were then sequenced by the ‘WideSeq’ pipeline. We also
performed RNAseq on endosperms of four individual hy-
brid seeds. Preliminary processing of RNAseq reads and
transcriptome mapping were carried out as described previ-
ously (43). As above, segregating Mu insertions were iden-
tified using Wideseq in these four hybrid individuals. For
16 genes with segregating Mu insertions, SNPs are avail-
able so that the allele frequency in the B73/Mo17 hybrid
transcripts could be called. Fold changes of gene expression
caused by Mu insertions for each gene were calculated by
comparing the B73 allele frequency in individuals contain-
ing Mu insertions with those without Mu insertions which
were further normalized using the Mo17 allele transcript
frequency in plants that lacked an insertion in either B73 or
Mo17.

RESULTS
Distribution of de novo transposons near TSSs and TTSs of
host genes
To understand bona fide target preferences of transposons,
we examined the target site distribution of nine de novo in-
sertion datasets. These included the UniformMu (44), So-
maticMu (this report) and Ds-GFP collections in maize
(32,33), the Tos17, Ds and Spm collections in rice (34–36),
as well as the P-element, Pb and Mb insertion collections
in Drosophila (37,38). All TEAs examined here are DNA
transposons with the exception of Tos17, which is a low
copy number LTR retrotransposon. UniformMu is primar-
ily composed of germlinally transmitted Mu insertions and
SomaticMu elements are inferred to be derived primarily
from somatic insertions due to the relative low number of
reads obtained relative to the insertions that segregated in
the families examined.

A comparative analysis of insertion profiles relative to
randomly selected loci revealed dramatic similarities and
differences between different elements in different species
with respect to their association with TSSs or TTSs. The
difference between the distribution pattern of TE insertions
and that of randomly selected loci reveals a dramatic enrich-
ment of TE insertions near TSSs or TTSs for some elements.
Both the Mu element (UniformMu and SomaticMu) and
P element insertions were vastly enriched near TSSs (peak
shift < 50 bp), but were largely missing near TTSs, indicating a tight TSS-specific association with both transposases (Figure 1A, B, D, E, Supplementary Figure S1A, B, D, E). In addition, the distribution curves of Mu element insertions decrease rapidly upstream of the TSSs but are reduced more gradually downstream of the TSSs, particularly within 1 kb (Supplementary Figure S1A). Given the extreme bias in integration of Mu and P elements, we refer to these elements as precision-targeting elements. In contrast, enrichment of Ds and Pb insertions was observed near both the TSSs and TTSs, with much wider and lower peaks, suggesting less specificity than Mu and P elements (Figure 1A, B, D, E, Supplementary Figure S1A, B, D, E). There is no enrichment of Mb, Spm and Tos17 insertions near either TSSs or TTSs. Indeed, Tos17 insertions are actually somewhat enriched in the gene body relative to these sites (Figure 1C, F, Supplementary Figure S1C, F).

To determine whether Mu and P element targeting is specific to Pol II-dependent transcription or is actually associated with any RNA polymerase, we examined the distribution of de novo Mu element and P element insertions in maize and Drosophila, respectively, near rRNA and tRNA genes, which are transcribed by RNA Pol I or III, respectively. To minimize Pol II TSS-associated TE enrichment, we filtered the rRNA and tRNA gene set in maize based on their distance to Pol II TSSs and obtained a list of 1610 genes over 5 kb away from the TSSs of any annotated genes transcribed by Pol II. Few transposon insertions are into the gene body of the 1610 rRNA and tRNA genes, as can be observed along the positive X-axis, likely due to the small size of these genes. In regions both downstream of TSSs and upstream of TSSs of these genes, UniformMu and SomaticMu resemble random selected genomic loci in their distribution except that the UniformMu shows a less smooth distribution curve than the SomaticMu (Figure 2A, B), which can be attributed to the fact that there are ~3.5 times as many SomaticMu insertions as there are UniformMu insertions. Similar to Mu element insertions, P elements insert into or near the TSSs and TTSs of rRNA and tRNA genes at a frequency comparable to randomly selected loci (Supplementary Figure S2).

Plants have two plant-specific RNA polymerases, Pol IV and Pol V, that are required for cytosine methylation in asymmetrical (CHH, where H is A, T or C) sequence contexts (45). A large number of CHH islands are located immediately upstream of the 5’ ends or downstream of the 3’ ends of genes in maize (46). We observed no enrichment of Mu element insertions in CHH islands; the vast majority of insertions are adjacent to the 5’ CHH islands, where Pol II transcripts are initiated at TSSs. These results indicate that Pol IV and Pol V transcription start or stop sites are not notable targets for Mu element insertions (Figure 2C, D).

**Distribution of TE target sites near host genes with different expression levels**

Given that the target sites of several TE families examined are TSS- or TTS-associated, we hypothesized that the transposases of some families are recruited to TSSs or TTSs in a manner that is dependent on the level of transcription. To test this hypothesis, we examined the correlation between transposon targeting frequency and relative expression levels of host genes. We extracted a subset of TSS-associated TE insertions that are located near (<2 kb) the TSSs and retrieved publicly available RNAseq datasets from 79 tissues in maize, 38 tissues in Drosophila and 59 tissues in rice (29–31). For each dataset, genes were binned into 20 equal sized groups based on ranked FPKM values in each RNAseq experiment, where bin 1 contains the lowest expressed 5% genes and bin 20 contains the highest expressed 5% genes. While randomly selected genomic loci (the control datasets) were evenly distributed near genes expressed at various levels, Mu and P elements preferentially target highly expressed genes, as indicated by the upward sloping curves (Figure 3A, B). Targeting frequency of Pb also positively correlates with gene expression, but to a lesser extent than that of P and Mu elements (Figure 3B). Interestingly, the distribution of Ds (in both maize and rice), Spm and Tos17 transposon insertion hotspots are all overrepresented in the medium expression bins (Figure 3A, C), suggesting that genes expressed at these levels are preferred targets for these elements. In contrast, the Mb elements, which show a mild enrichment at TSSs but not TSSs, actually target the lowest expression bins (Figure 3B).

Using the two available Mu insertion datasets, we tested the hypothesis that genes frequently targeted by precision-targeting TE are associated with specialized biological functions or processes by performing gene ontology (GO) analysis of 4225 Mu element hotspot genes (>3 UniformMu and >10 SomaticMu insertions <2 kb from TSSs, Supplementary Table S1) using the online webserver agriGO (47). The seven GO terms enriched for Mu hotspot genes are related to a wide range of general, rather than specialized, biological processes and molecular functions (Supplementary Figure S3A). Genes categorized by these seven GO terms, on average, were expressed at markedly higher levels than the total gene set (Kolmogorov-Smirnov test, P values as indicated next to the box plots) (Supplementary Figure S3B), consistent with our observation that Mu elements preferentially target highly expressed genes. Moreover, in each of these seven gene sets, Mu hotspot genes are expressed at a significantly higher level than the non-hotspot genes in each set (Kolmogorov-Smirnov test) (Supplementary Figure S3C). These observations suggest that these GO terms are enriched not because they are associated with particular processes, but because they tend to express at higher levels than the average gene.

**Distribution of TE insertions in meristematic and differentiated tissues**

Given that the targeting frequency of several TE families associates with host gene expression levels in a collection of tissues, we hypothesized that tissues showing the strongest correlation between the two would be the tissues in which transposition occurs most frequently, assuming transposition has tissue-specificity. We explored the tissue specificity of TE transposition by using Mu and Ds elements in maize as examples.

To explore the major factors that contribute to transcriptome variation in different maize tissues, we first performed principle component analysis (PCA) on the above-
Figure 1. Characterization of distribution of de novo transposon insertions that are near genes. The figure shows normalized numbers of insertions around TSSs and TTSs of annotated genes. The insertion numbers at each distance was smoothed by computing the mean in a 30-bp rolling window. For both TSS and TTS plots, normalized numbers of genomic insertions were plotted on the positive X-axis coordinates and normalized numbers of intergenic insertions were plotted on the negative coordinates. Thus, positive values for both plots represent insertions into the gene body. Randomly selected loci were used as a background control to take into account the fact that although insertions were plotted up to 4 kilobase pairs (kb) away from the TSSs and TTSs, many gene bodies are shorter than 4 kb, and many genes are more or less than 4 kb downstream or upstream from a neighboring gene. Panels in the upper and bottom rows show metaprofiles of transposon insertions in maize (A, D), Drosophila (B, E) and rice (C, F) for regions surrounding the gene TSS and TTS, respectively.

mentioned 79 RNAseq datasets and found that the first principle component (17.8% of the variance) separated tissues of meristematic and differentiated identities well (Supplementary Figure S4A). The distribution curves of Mu and Ds element insertions in low-to-high expression bins showed distinct patterns in the six meristematic and six differentiated tissues (Figure 4A, Supplementary Figure S4B). The two curves differed most in the medium (bins 8–12, ranked between 35% and 60%) and highly (bins 16–20, ranked between 75% and 100%) expressed genes (Figure 4A). To address whether tissue-specific gene expression in meristematic and differentiated tissues associated with this shift of the distribution curve, we identified genes that were expressed at high levels in the meristem-enriched tissues and at medium levels in the differentiated tissues (meristematic-dominant genes), and those that were expressed at high levels in differentiated tissues and medium levels in meristem enriched tissues (differentiated-dominant genes). We obtained a set of 746 meristematic-dominant genes that were present in bins 16–20 in the majority (no less than five) of six meristematic tissues and in bins 8–12 in the majority (no less than five) of six differentiated tissues. We also obtained a set of 723 differentiated-dominant genes that were present in bins 16–20 in the majority (no less than five) of six differentiated tissues and in bins 8–12 in the majority (no less than five) of six meristematic tissues (Supplementary Figure S5). We found a much higher enrichment of Mu element insertions (both UniformMu and SomaticMu) that were near TSSs of the meristematic-dominant genes than were near TSSs of differentiated-dominant genes (Figure 4B). Such enrichment is specific to TSSs, but not TTSs, consistent with our previous observations (Figure 4C). We also found that Ds enrichment in the medium expression bins 8–12 is higher in meristematic tissues (Figure 4A), likely due to a preference for medium expressed genes by Ds elements. In line with this observation, we observed a lower level of Ds enrichment near TSSs, and to a lesser extent, TTSs in the meristematic-dominant gene set (Figure 4B, C), suggesting that Ds elements insert at a higher frequency in genes that express at a medium level in meristematic tissues (the differentiated-dominant gene set). These results suggest that both Mu and Ds elements insert most frequently in genes that express at targeted levels in meristematic or rapidly dividing cells. In the case of Mu, these are
Figure 2. Distribution of de novo Mu insertions near target sites of RNA Polymerases I, III, IV and V. (A and B) Metaprofiles of UniformMu and SomaticMu insertions surrounding TSS (A) and TTS (B) of tRNA and rRNA genes. The figure shows normalized numbers of insertions around TSSs and TTSs of annotated tRNA and rRNA genes. The insertion numbers at each distance were smoothed by computing the means in 30-bp rolling windows. Normalized numbers of genic insertions were plotted on the positive X-axis coordinates and those of intergenic insertions were plotted on the negative coordinates. (C and D) Distribution of UniformMu and SomaticMu insertions near CHH islands in maize located at 5′ (C) and 3′ ends (D) of maize genes. The 100-bp CHH islands are centered on position zero and range from −50 to 49 bp on the X-axis. Directions of CHH islands were unified so that nearby genes are located on the right of both 5′ and 3′ CHH islands.

Re-evaluation of the mutagenic capability of precision-targeting transposons

TE insertions in regions in or near genes may or may not disrupt host genes depending on where the TEs integrate. To evaluate the mutagenic capability of the precision-targeting transposons, we examined the enrichment of Mu and P elements in 5′ and 3′ proximal regions as well as sub-genic features of annotated maize and fly genes, respectively. Consistent with the meta-analysis above (Figure 1), insertions of both Mu and P element show strong enrichment in the 5′ ends of genes (particularly 200 bp upstream of the TSSs and 5′ UTRs) (Table 1). Mu element insertions are enriched in 5′ UTRs (64.4-fold for SomaticMu; 90.1-fold for UniformMu) relative to insertions into coding sequences (8.8-fold for SomaticMu; 7.0-fold for UniformMu) (all P values less
Figure 3. De novo transposon insertions near genes expressed at different levels. Percentage of transposon insertions distributed in each of the 20 expression bins were shown for maize (A), Drosophila (B) and rice (C) transposons. The figures show the correlation between TE targeting frequency with categorized gene expression levels. For each RNA-seq dataset, the whole annotated gene set was grouped to 20 expression bins from low to high independently, each containing an equal number of genes. Allocation of genes in each bin was independently performed for each RNA-seq experiment. Each RNA-seq dataset was regarded as one replicate and the averaged percentages in each categorized expression levels were calculated. Error bars represent standard deviations of the percentages in each bin based on independent RNAseq datasets.

Over a longer time-scale, purifying selection would be expected to purge insertion mutations that have only weak deleterious effects. To evaluate the selection pressures on older Mu element insertions, we examined a class of Mutator-like elements called Pack-MULEs, many of which are ancient insertions in the genomes that have diverged terminal inverted repeats (TIRs) (48). Profiling the distribution of 1358 Pack-MULEs in maize and 2959 Pack-MULEs in rice surrounding TSSs show that in both species, PackMULE occupancy peaks just upstream of TSSs and is reduced to nearly background levels >1 kb upstream of TSSs (Figure 5A, B). This is quite similar to our observation of de novo insertions in maize. However, there is a sharp decline of Pack-MULE insertions into gene bodies (Figure 5A, B), indicating selection pressure against older genic insertions into this region. Consistent with this observation, and consistent for selection against insertion into genes, underrepresentation of genic transposon insertions downstream of TSSs were found in other DNA transposon families in the maize genome, including hAT, Mariner, CACTA, Harbinger and Helitrons and LTR retrotransposons (Figure 5C-H). In contrast, we found that P element annotated in the genomes of wild D. melanogaster accessions exhibit an identical distribution of the de novo P element insertions (Figures 1B and 5I). Presumably, this is due to the fact that P elements have only been in the D. melanogaster genome for a relatively short period of time and are unlikely to be fixed or homozygous in wild populations (49).

DISCUSSION

It has been suggested that genomes resemble ecological systems, and that different TE families occupy distinct niches, presumably because there are multiple ways to be a successful genomic parasite (50). In each case, TE targeting represents a balance between successful amplification of the TE
Figure 4. Comparative analysis of the distribution of \textit{Mu} and \textit{Ds} elements in meristematic or differentiated tissues. (A) Percentage of random genomic loci, Uniform\textit{Mu}, Somatic\textit{Mu} and \textit{Ds} element insertions distributed in each of 20 expression bins in meristematic and differentiated tissues. An identical set of randomly selected loci that was used in Figure 1A (and 1D) was used as a background control. Error bars represent standard deviations of the percentages in each bin based on independent RNAseq datasets. (B and C) Metaprofiles of random genomic loci, Uniform\textit{Mu}, Somatic\textit{Mu} and \textit{Ds} elements in maize surrounding TSSs (B) and TTSs (C) of meristematic-dominant and differentiated-dominant gene sets, respectively. For both TSS and TTS plots, normalized numbers of genic insertions were plotted on the positive X-axis coordinates and normalized numbers of intergenic insertions were plotted on the negative coordinates. The insertion numbers at each distance were smoothed by computing the means in 30-bp rolling windows. An identical set of randomly selected loci used in Figure 1A (and 1D) was used as a background control.

In contrast to type-A strategy, sites targeted by TE families employing the type-B strategy (\textit{Ds} and \textit{Pb}) are enriched at both TSSs and TTSs of medium expressed genes, and the enrichment levels of TE insertions near TSSs are lower for type-B than for type-A elements (Figure 6A, B). The type-A and type-B strategies have been named according to the single-peak and double-peak shapes of representative TE distribution curves near annotated genes reminiscent of the single- and double-hump nature of Arabian camels and Bactrian camels, respectively. Given that transposases encoded by the different TE families we have analyzed are distantly related phylogenetically (51), the transcription-associated type-A and type-B strategies suggest a process of convergent evolution among different TE families in plants and animals. Although it is formally possible that targeting of these different classes of TEs predated their divergence, we suggest that it is more likely that selection independently favored similar targeting strategies. To our knowledge, this...
Figure 5. Metaplots of ancient transposons near the TSS of host genes. Normalized numbers of insertions around TSSs and TTSs of annotated genes. Genic insertions were plotted on the positive X-axis coordinates and intergenic insertions were plotted on the negative coordinates. The insertion numbers at each distance were smoothed by computing the means in 30-bp rolling windows. Identical sets of randomly selected loci used in Figure 1 were used as a background control.

is the first comparative study showing such relationships between TEs in both plants and animals.

Those TEs lacking an association with either Pol II transcription initiation or termination fall in the type-C group. Pol II-independent integration can be transcription-independent or transcription-associated. Reminiscent of the Sleeping Beauty transposons reported previously (52), Tos17 insertions do not show an association with either TSSs or TTSs, but are enriched in gene bodies of genes that express at a moderate level (Figure 3C, Supplementary Figure S1C and S1F). In contrast, M6 insertions show a mild association with TTSs, but not TSSs (Figure 1B and E) and actually exhibit a negative correlation between targeting frequencies and gene expression levels (Figure 3B). Spm also
Figure 6. Classification of TE targeting strategies in context of host transcription. A proposed model depicting occupation of TEs at genomic locations with different expression characteristics. Type-A and type B transposons primarily target open chromatin regions where active transcription occurs. Integration of the type-A transposons associates with RNA Pol II transcription initiation. Factors that directly mediate transposase-recruitment to TSSs are yet to be identified. Targeting of the type-B transposons associates with both RNA Pol II transcription initiation and termination.

show targeting enrichment near genes that are expressed at medium-to-high levels, but this TE preferentially targets intergenic regions in host genomes. Overall, the type-C strategy may involve genome targeting mechanisms that are only indirectly related to host transcriptional activities.

Previous reports provided evidence for an ‘open-chromatin’ targeting model for some DNA transposons (12,53). Our results challenge the universality of this model because it does not fully account for the type-A integration strategy, although the insertion sites of precision-targeting type-A TEs certainly do co-localize to some extent with a set of chromatic modifications associated with open chromatin (12). The majority of open chromatin regions lie just upstream of TSSs and downstream of TTSs in both plants and animals (54,55). Occupancy of Pol II at or near both TSSs and TTSs has also been reported in both maize and Drosophila (56,57). The distribution of Mu and P element insertions is reminiscent of Pol II occupancy near TSSs, but not near TTSs, indicating that integration of the typical type-A TEs is strongly associated with transcription initiation, and not simply with the occupancy of Pol II. Further, we demonstrated that the type-A transposition strategy is specific to Pol II, but not other RNA polymerases like Pol I, Pol III or plant-specific Pol IV and Pol V (Figure 2). P integration near TSSs has been attributed to enrichment of replication origins in those regions, suggesting that an association with Pol II transcription initiation is indirect (13). The proposed model involves targeting of unfired replication origins by transposons in combination with homologous repair of excision sites following replication. While this model provides an attractive mechanism for increasing element copy number, an association with replication origins does not immediately account for the observed correlations of both P and Mu targeting with levels of gene expression (Figure 3A, B). Instead, the extreme specificity of P and Mu elements suggests that they are targeted via some form of tethering. This is consistent with the observation that type-A Tj1 retrotransposons in fission yeast are known to target the 5' end of Pol II-transcribed genes via interaction between the integrase and the DNA binding protein Sap1, which causes replication fork arrest (26,27). Similarly, a comparative analysis of insertion site profiles has revealed that the Mouse Leukemia Virus (MLV) and the piggyBac transposon in human cell lines are targeted specifically to acetylated histones near TSSs via a tethering mechanism dictated by chromatin-bound bromodomain and extraterminal (BET) domain proteins that bind to acetylated H3 and H4 near TSSs (24). Future characterization of proteins or chromatin features associated with Mu and P element transposases will provide mechanistic insights into precision targeting of these elements as well.

Our analysis highlights two features that can facilitate TE survival and rapid proliferation: tissue-specific transposition and minimization of negative impacts on nearby gene function due to precision targeting. Certainly, tissue-specific transposition is true for P elements, which only express functional transposase in the germline (58). In plants, rapid and heritable amplification of TEs would also be facilitated by meristematic-tissue-specific transposition because actively dividing plant cells (particularly in floral tissues and meristems) are more likely to be transmitted to the next generation than those that are not. In this regard, we have observed clear targeting preferences of Mu and Ds elements for genes that express at targeted levels (highly expressed for Mu and expressed at a medium level for Ds) in meristem-enriched tissues (Figure 4B). Given that the very large number of somatic Mu insertions we have identified showed in-
### Table 1. A summary of the numbers and percentages of TE insertions, values below 1 mean a depletion of TE insertions in or near genes. For random insertions, enrichment fold is defined as 1; values above 1 mean an enrichment of insertion preferences similar to germinally transmitted Mu insertions. Mu elements may be primarily avoiding insertions into genes that express at high levels primarily in terminally differentiated cells rather than targeting ‘germinal’ lineages.

The type-A (TSS-targeting) strategy employed by Mu and P elements endows these TEs with the capacity to exploit a permissive environment with respect to transcription of autonomous elements. This is particularly important for the survival of TEs in heterochromatin-rich genomes such as the maize genome. This strategy has the potential to cause deleterious effects on host gene expression and function. Indeed, we found that Mu insertions into CDS regions, which are most likely to be disruptive, were more frequent by 7–9-fold than random insertions. Despite of this, there was a much higher enrichment (64–90-fold) for Mu insertions near TSSs, and the majority of Mu insertions in promoters and 5’ UTRs have minimal to no effect on gene expression. This suggests that Mu elements are actually much less mutagenic than one might expect given their propensity to target genic regions because the reduction of host fitness is minimized due to a tight association between Mu element insertions and TSSs. Given this, and given that the vast majority of genic Mu element insertions are into the 5’UTR, we suggest that researchers who use Mu as a genetic resource treat these insertions with some caution, as they are unlikely to be knockouts. P elements also rarely insert into CDSs, likely because they are also precisely targeted to TSSs, although the effects caused by promoter and 5’ UTR targeting by P elements require future evaluation. Similarly, MITE TE insertions, although they tend to be into or near genes, also have a minimal effect on gene expression, although this may be in part due to their small size (59). Collectively, these data suggest that for some TEs selection has favored insertions that are into genes but that are minimally disruptive. This historical view of Mu and P elements as highly effective mutagens has likely been shaped by the fact that many of the insertion mutations caused by these elements were identified in screens for mutant phenotypes (60,61). More broadly, nearly all of the known active TEs in higher eukaryotes were first identified due to their mutagenic effects. It may well be that there are many additional active TEs in natural populations that have yet to be identified because they only rarely cause visible mutations.

Our analysis of older MULE insertions in both maize and rice suggests that 5’ UTR insertions of these elements are eventually purged from the genome, indicating that TEs that target 5’ UTRs are subject to purifying selection in the long run. Interestingly, the purging appears to be much less efficient for proximal promoter MULE insertions, which are still present in high numbers in both species.

TE families occupy distinct genomic niches by employing distinct strategies for integration. This, in turn has influenced the degree to which TEs have affected host gene function and, ultimately, host genome evolution. TEs have also proved to be invaluable tools, both as mutagens and as transformation vectors (62). A deeper understanding of the ways in which TEs target particular regions of the genome for integration promise to make those tools both more effective and more precise.
Table 2. Re-evaluation of mutagenic capability of Mu insertions by RNAseq and sequencing RT-PCR products via Miseq

| Affected_gene | Insertion Site | B73 Insertion | Mo17_pr insertion | B73 pr insertion | Mo17_pr insertion | Knockdown_index | Consequence |
|---------------|---------------|---------------|-------------------|------------------|-------------------|----------------|-------------|
| Zm00001d020901 | five_pr UTR   | 0             | 247               | 461              | 519               | %              | knockout    |
| Zm00001d027590 | intron        | 2             | 517               | 221              | 285               | 200.5          | knockout    |
| Zm00001d039733 | five_pr UTR   | 39            | 960               | 553              | 480               | 28.4           | knockdown   |
| Zm00001d044466 | five_pr UTR   | 31            | 703               | 148              | 144               | 23.3           | knockdown   |
| Zm00001d028712 | five_pr UTR   | 290           | 510               | 192              | 38                | 12.1           | knockdown   |
| Zm00001d006460 | five_pr UTR   | 86            | 102               | 50               | 5                 | 11.9           | knockdown   |
| Zm00001d017424 | five_pr UTR   | 6             | 34                | 819              | 1065              | 4.4            | weak knockdown |
| Zm00001d005587 | five_pr UTR   | 10            | 60                | 81               | 122               | 4              | weak knockdown |
| Zm00001d029892 | five_pr UTR   | 78            | 488               | 181              | 463               | 2.4            | weak knockdown |
| Zm00001d008642 | five_pr UTR   | 119           | 1235              | 361              | 1782              | 2.1            | weak knockdown |
| Zm00001d006126 | intron        | 83            | 68                | 144              | 50                | 2.4            | weak knockdown |
| Zm00001d006768 | promoter-65   | 63            | 93                | 79               | 62                | 1.9            | unchanged   |
| Zm00001d033167 | promoter-117  | 154           | 69                | 140              | 112               | 0.6            | unchanged   |
| Zm00001d039683 | five_pr UTR   | 206           | 95                | 166              | 86                | 0.9            | unchanged   |
| Zm00001d012182 | promoter-78   | 732           | 405               | 239              | 138               | 0.9            | uncharged   |
| Zm00001d039156 | five_pr UTR   | 455           | 838               | 542              | 676               | 1.5            | uncharged   |
| Zm00001d047761 | five_pr UTR   | 161           | 393               | 50               | 83                | 1.5            | uncharged   |
| Zm00001d018461 | five_pr UTR   | 402           | 818               | 223              | 383               | 1.2            | uncharged   |
| Zm00001d0022153 | five_pr UTR   | 320           | 208               | 122              | 79                | 1              | uncharged   |
| Zm00001d034867 | five_pr UTR   | 924           | 854               | 288              | 302               | 0.9            | uncharged   |
| Zm00001d034191 | five_pr UTR   | 138           | 96                | 36               | 30                | 0.8            | uncharged   |
| Zm00001d028784 | five_pr UTR   | 27            | 20                | 355              | 349               | 0.8            | uncharged   |
| Zm00001d050081 | five_pr UTR   | 53            | 53                | 31               | 42                | 0.7            | uncharged   |
| Zm00001d039253 | five_pr UTR   | 30            | 184               | 16               | 133               | 0.7            | uncharged   |
| Zm00001d018788 | five_pr UTR   | 233           | 178               | 497              | 563               | 0.7            | uncharged   |
| Zm00001d002956 | intron        | 189           | 314               | 104              | 96                | 1.8            | uncharged   |
| Zm00001d050163 | intron        | 8             | 11                | 127              | 110               | 1.6            | uncharged   |
| Zm00001d053452 | intron        | 963           | 1324              | 236              | 310               | 1              | uncharged   |
| Zm00001d049619 | intron        | 70            | 260               | 251              | 1275              | 0.7            | uncharged   |
| Zm00001d006630 | five_pr UTR   | 336           | 292               | 258              | 232               | 1.1            | uncharged   |
| Zm00001d002909 | five_pr UTR   | 244           | 131               | 46               | 59                | 0.4            | unchanged/weak activation |
| Zm00001d022122 | intron        | 43            | 29                | 2                | 9                 | 0.1            | unchanged/weak activation |

Table 3. Classification of integration strategies used by TE families analyzed in this study based on their distribution near genes and the association of TE targeting frequency with gene expression levels

| Strategy | Typical TE family | Host | Associated region | Expression levels of target genes |
|----------|-------------------|------|-------------------|----------------------------------|
| A        | Mu                | Maize| TSS               | High                             |
| A        | P                 | Drosophila| TSS            | High                             |
| B        | Ac/Ds             | Maize, rice | TSS and TTS | Medium-to-high                  |
| B        | Piggybac          | Drosophila| TSS and TTS | Higher than Ac/Ds, lower than Mu and P |
| C        | Minos             | Drosophila| Not observed | Low-to-medium                    |
| C        | Tnt1              | Medicago   | Gene body       | Low-to-medium                    |
| C        | Tos17             | Rice     | Gene body       | Medium-to-high                   |
| C        | Spm               | Rice     | Not observed    | Medium-to-high                   |

DATA AVAILABILITY
The RNA-seq data of B73/Mo17 hybrids and Museq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) data bank, accession codes PRJNA56108 and GSE146647.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank Phillip SanMiguel for helpful discussions on high-throughput-sequencing-based insertion profiling, Nathan Springer for sharing the CHH island data and Gur-mukh Johal for providing the Mu-active line in the B73 background.

Authors contributions: D.L. and X.Z. conceived the study and designed the experiments. X.Z. performed most of the experiments. D.L. and X.Z. wrote the manuscript. X.Z. and M.Z. performed insertion profiling and analyzed the sequencing data. D.M. and M.Z. created the Mu insertion datasets and contributed to the Mu distribution analysis. The majority of the work was done when X.Z. was working at Purdue University. Data organization, manuscript preparation and submission were partly done after X.Z. moved to the AGIS. All authors read and approved the final manuscript.

FUNDING
National Science Foundation [IOS-1237931]; Purdue Center for Plant Science SEED grant (to D.L.). Funding for open access charge: Lisch Startup Funds.

Conflict of interest statement. None declared.

REFERENCES
1. Wendel,J.F., Jackson,S.A., Meyers,B.C. and Wing,R.A. (2016) Evolution of plant genome architecture. Genome Biol., 17, 37.
2. Naville,M., Henriet,S., Warren,I., Sumic,S., Reeve,M., Volff,J.N. and Chourrout,D. (2019) Massive changes of genome size driven by expansions of non-autonomous transposable elements. Curr. Biol., 29, 1161–1168.
3. Zhang,J., Yu,C., Krishnaswamy,L. and Peterson,T. (2011) Transposable elements as catalysts for chromosome rearrangements. Methods Mol. Biol., 701, 315–326.
transposable elements preferentially transpose to replication origins. Mol. Cell, 74, 555–570.

16. Cheung, S., Ma, L., Chan, P.H., Hu, H.L., Mayor, T., Chen, H.T. and Measday, V. (2016) Tyl integrase interacts with RNA polymerase III-specific subcomplexes to promote insertion of Tyl elements upstream of polymerase (Pol) III-transcribed genes. J. Biol. Chem., 291, 6396–6411.

17. Brider-Nahmais, A., Tchelikian-Cossou, A., Baller, J.A., Menouni, R., Fayol, H., Flores, A., Sauib, A., Werner, M., Votayt, D.F. and Lesage, P. (2015) Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposition integration. Science, 348, 585–588.

18. Baller, J.A., Gao, J., Stamenova, R., Curcio, M.J. and Votayt, D.F. (2012) A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Tyl1 retrotransposon. Genome Res., 22, 704–713.

19. Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S.J. and Bocke, J.D. (2012) Retrotransposition Tyl integration targets specifically positioned asymmetric nucleosomal DNA segments in rRNA hotspots. Genome Res., 22, 693–703.

20. Sigman, M. and Lutkin, R. (2016) The first rule of plant transposable element silencing: location, location, location. Plant Cell, 28, 304–313.

21. Zamudio, N. and Bour’his, D. (2010) Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity (Edinb), 105, 92–104.

22. Tsukahara, S., Kawabe, A., Kobayashi, A., Ito, T., Aizu, T., Shin, T., Toyoda, A., Fujiyama, A., Tarutani, Y. and Kakutani, T. (2012) Centomere-targeted de novo integrations of an LTR retrotransposon in Arabidopsis lyrata. Genes Dev., 26, 705–713.

23. Xie, W., Gai, J., Yu, Z., Zappulla, D.C., Sterngray, R. and Votayt, D.F. (2001) Targeting of the yeast Tyl retrotransposon to silent chromatin is mediated by interactions between inteGrase and Sir4p. Mol. Cell. Biol., 21, 6606–6614.

24. Gogol-Dorling, A., Ammar, I., Gupta, S., Bunsch, M., Miskey, C., Chen, W., Uckert, W., Schulz, T.F., Ivsik, Z. and Ivsic, Z. (2016) Genome-wide profiling reveals remarkable parallels between insertion site selection properties of the MLV retrovirus and the piggyBack transposable element silencing in primary human CD4(+)-T cells. Mol. Ther., 24, 592–606.

25. Yieh, L., Hassavesis, G., Geiduscheck, E.P. and Sandmeyer, S.B. (2000) The Bef and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIIB mediate position-specific integration of the gypsy-like element, Ty3. J. Biol. Chem., 275, 29880–29807.

26. Jabon, I., Rosado-Lugo, J.D., Cram-Mileva, S., Caccagnone, K.M., Tournier, V. and Varatigou, M. (2015) Arrested replication forks guide retrotransposition integration. Science, 349, 1549–1553.

27. Hickey, A., Ensaull, C., Majumdar, A., Chatterjee, A.G., Iben, J.R., McQueen, P.G., Yang, A.X., Mizuguchi, T., Grewal, S.I. and Levin, H.L. (2015) Single-nucleotide-specific targeting of the Tfi retrotransposon promoted by the DNA-binding protein Sapl of Schizosaccharomyces pombe. Genes, 201, 905–924.

28. Chen, J., Greenblatt, J.M. and Dellaporta, S.L. (1987) Transposition of Ac from the P locus of maize into unreplicated chromosomal sites. Genetics, 117, 109–116.

29. Hoopes, G.M., Hamilton, J.P., Wood, J.C., Esteban, E., Pasha, A., Vaillancourt, B., Provert, N.J. and Buel, C.R. (2019) An updated gene atlas for maize reveals organ-specific and stress-induced genes. Plant J., 97, 1154–1167.

30. Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A. and Dow, J.A.T. (2018) FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res., 46, D809–D815.

31. Xia, Z., Zou, D., Sang, J., Xu, Y., Yin, H., Li, M., Wu, S., Huo, H., Lao, L. and Zhang, Z. (2017) Rice Expression Database (RED): an integrated RNA-Seq-derived gene expression database for rice. J. Genet. Genomics, 44, 235–241.

32. McCarthy, D.R., Settles, A.M., Suzuki, M., Tan, B.C., Latshaw, S., Porch, T., Robin, K., Baier, J., Avigne, W., Lai, J. et al. (2005) Steady-state transposon mutagenesis in inbred maize. Plant J., 44, 52–61.

33. Li, Y., Segal, G., Wang, Q. and Dooner, H.K. (2013) Gene tagging with engineered Ds elements in maize. Methods Mol. Biol., 1057, 83–99.

34. Miya, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and Hirochika, H. (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell, 15, 1771–1780.

35. Upadhyaya, N.M., Zhou, Q.H., Zhou, X.R., Eamens, A.L., Hoque, M.S., Ram, K., Shivakumar, R., Smith, K.F., Pan, S.T., Li, S. et al. (2006) Dissociation (Ds) constructs, mapped Ds launch pads and a transiently-expressed transposase system suitable for localized insertional mutagenesis in rice. Theor. Appl. Genet., 112, 1326–1341.

36. Kumar, C.S., Wing, R.A. and Sundaresan, V. (2005) Efficient insertional mutagenesis in rice using the maize En/Spm elements. Plant J., 44, 879–892.

37. Bellen, H.J., Levi, R.W., He, Y.C., Carlson, J.W., Evans-Holm, M., Bae, E., Kim, J., Metaxakis, A., Savakis, C., Schulze, K.L. et al. (2011) The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics, 188, 731–U341.

38. Thibault, S.T., Singer, M.A., Miyaizaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Densky, M., Fawcett, R., Francis-Lang, H.L. et al. (2004) A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nat. Genet., 36, 283–287.

39. McCarty, D.R., Latshaw, S., Wu, S., Suzuki, M., Hunter, C.T., Avigne, W.T. and Koch, K.E. (2013) Mu-seq: sequence-based mapping and identification of transposition induced mutations. PLoS One, 8, e71712.

40. Zhao, D., Hamilton, J.P., Vaillancourt, B., Zhang, W., Eizenga, G.C., Cui, Y., Jiang, J., Buel, C.R. and Jiang, N. (2018) The unique epigenetic features of Pack-MULEs and their impact on chromosomal base composition and expression spectrum. Nucleic Acids Res., 46, 2380–2397.

41. Le,S., Josse,J. and Husson,F. (2008) FactoMineR: an R package for multivariate analysis. Nucleic Acids Research, 2020, Vol.48, No.12, 6697.
by bidirectional cytoplasmic RNA decay in Arabidopsis. Science, 348, 120–123.
44. McCarty, D.R., Suzuki, M., Hunter, C., Collins, J., Avigne, W.T. and Koch, K.E. (2013) Genetic and molecular analyses of UniformMu transposon insertion lines. Methods Mol. Biol., 1057, 157–166.
45. Tucker, S.L., Reece, J., Ream, T.S. and Pikaard, C.S. (2010) Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. Cold Spring Harb. Symp. Quant. Biol., 75, 285–297.
46. Li, Q., Gent, J.I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C.D., Hirsch, C.N., Dow, R.K., Madzima, T.F., McGinnis, K.M. et al. (2015) RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc. Natl. Acad. Sci. U.S.A., 112, 14728–14733.
47. Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W. and Su, Z. (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res., 45, W122–W129.
48. Jiang, N., Ferguson, A.A., Slotkin, R.K. and Lisch, D. (2011) Pack-Mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition. Proc. Natl. Acad. Sci. U.S.A., 108, 1537–1542.
49. Anxolabehere, D., Kidwell, M.G. and Periquet, G. (1988) Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. Mol. Biol. Evol., 5, 252–269.
50. Brookfield, J.F.Y. (2005) The ecology of the genome - Mobile DNA elements and their hosts. Nat. Rev. Genet., 6, 128–136.
51. Arkhipova, I.R. (2017) Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. Mobile DNA-UK, 8, 19.
52. de Jong, J., Akhtar, W., Budhai, J., Rust, A.G., Rad, R., Hilkens, J., Berns, A., van Lohuizen, M., Wessels, L.F. and de Rидder, J. (2014) Chromatin landscapes of retroviral and transposon integration profiles. PLoS Genet., 10, e1004250.
53. Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S.J. and Craig, N.L. (2010) DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. Proc. Natl. Acad. Sci. U.S.A., 107, 21966–21972.
54. Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zamstein, K., Woodhouse, M., Bubb, K., Dorrity, M.W. et al. (2018) Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell, 30, 15–36.
55. Rogers-Melnick, E., Vera, D.L., Bass, H.W. and Buckler, E.S. (2016) Open chromatin reveals the functional maize genome. Proc. Natl. Acad. Sci. U.S.A., 113, E3177–E3184.
56. Lozano, R., Booth, G.T., Omar, B.Y., Li, B., Buckler, E.S., Lis, J.T., Jannink, J.-L. and del Cardo, D.P. (2018) RNA polymerase mapping in plants identifies enhancers enriched in causal variants. bioRxiv doi: https://doi.org/10.1101/376640, 24 July 2018, preprint: not peer reviewed.
57. Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fardo, D.C., Kwak, H., Adelman, K. and Lis, J.T. (2012) Defining the status of RNA polymerase at promoters. Cell Rep., 2, 1025–1035.
58. Laski, F.A., Rio, D.C. and Rubin, G.M. (1986) Tissue-specificity of Drosophila P-element transposition is regulated at the level of messenger-RNA splicing. Cell, 44, 7–19.
59. Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okamoto, Y., Tanisaka, T. and Wessler, S.R. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature, 461, 1130–1134.
60. Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., Mozden, N., Misra, S. and Rubin, G.M. (1999) The berkeley drosophila genome project gene disruption project: single p-element insertions mutating 25% of vital Drosophila genes. Genetics, 153, 135–177.
61. Candela, H. and Hake, S. (2008) The art and design of genetic screens: maize. Nat. Rev. Genet., 9, 192–203.
62. Munoz-Lopez, M. and Garcia-Perez, J.L. (2010) DNA transposons: nature and applications in genomics. Curr. Genomics, 11, 115–128.