Saturation mutagenesis using maize transposons

Virginia Walbot

Transposon mutagenesis facilitates gene discovery by tagging genes for cloning. New genomics projects are now cataloging transposon insertion sites to define all maize genes. Once identified, transposon insertions are ‘hot spots’ for generating new alleles that are useful in functional studies.

Introduction

The analysis of mutant phenotypes yields an understanding of gene function in a whole-organism context. Consequently, the generation, evaluation, maintenance and distribution of seed containing verified mutations or seed with populations of mutations suitable for high-throughput screening are essential in the modern genetic analysis of maize and other plants.

Saturation mutagenesis of maize genes involves two ambitious goals: to define all of the genes and to find the phenotype of every individual gene. This review will outline how transposon mutagenesis can be used to achieve both goals. Transposons are already the primary tool for tagging and cloning maize genes. Studies so far demonstrate that specific transposons are well suited for either global mutagenesis and gene discovery or multiple rounds of mutagenesis at a defined target gene [1]. New genomics approaches, employing strategies for screening by PCR, and for plasmid rescue, are now providing indexed collections of transposon insertion sites to define all maize genes. Once identified, transposons can be used in most recent mutagenesis experiments [1–3].

Insertional mutagenesis across the genome: primer on key transposon properties

The Ac/Ds and MuDR/Mu maize transposons have been used in most recent maize mutagenesis experiments [1–3]. In plants with the transcriptionally active regulatory elements Ac or MuDR, family members with essential transposase-binding sites in the terminal inverted repeats (TIRs) are mobilized to create new insertion mutations. During a gene-tagging experiment, transposon activity is conveniently monitored with a reporter allele that has a phenotype that becomes visible after the excision of a transposon (Figure 1a). Saturation mutagenesis of all genes requires a robust mutagenized population. An estimate of the distinct units (genes, for example) and the confidence level of the coverage are used to calculate the population size required to provide one insertion per unit, assuming random mutagenesis (Table 1).

<table>
<thead>
<tr>
<th>Gene estimate</th>
<th>Probability (%)</th>
<th>Population size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000</td>
<td>95</td>
<td>~91,000</td>
</tr>
<tr>
<td>50,000</td>
<td>95</td>
<td>~150,000</td>
</tr>
<tr>
<td>50,000</td>
<td>99</td>
<td>~230,000</td>
</tr>
<tr>
<td>70,000</td>
<td>95</td>
<td>~214,000</td>
</tr>
</tbody>
</table>

Using the formula \( (1 – \text{probability}) = n \log (\text{not desired class}) \), the required population size \( (n) \) can be calculated for a specific probability of finding at least one mutation in each unit. Because the number of functional units in maize is currently unknown, the estimate of ~50,000 genes is often used; this figure is about twice the gene estimate for Arabidopsis, to accommodate the tetraploidy of maize (3B).

Table 1

Using Ac or Ds elements spaced at 10–20 cM intervals across the recombination map would facilitate intensive mutagenesis of the entire maize genome [9]. Because mobile Ac/Ds copy number is low, confirmation that a particular transposon insert is the cause of a mutant phenotype is easily obtained [3]. There are several strategies for exploiting Ac/Ds in maize that depend on enriching for transposed copies, which may be linked or unlinked to the original location [3] (Figure 2).
Transposon properties. (a) Somatic excision of a transposon from a gene of the anthocyanin pathway can result in spots of pigmentation in cell images with a functional allele. This visible phenotype demonstrates somatic mutability and transposon activity. A range of phenotypes is observed, because revertant alleles differ in sequence and function. (b) Cut and paste transposition. The transposase liberates a double-stranded transposon from the original site by executing four strand breaks (arrowed at top left). Transposase creates staggered nicks at the target site (8 bp apart) and inserts the double-stranded transposon into a new site. The host chromosome is repaired (with modest changes) at the old site, and an 8-bp host sequence duplication is created by DNA synthesis at the insertion site during repair synthesis. If timed to coincide with chromosome replication, Ac/Ds copy number can increase. It is often observed that plant sectors or kernels on an ear differ in Ac copy number, in twin sectors. If the transposon is excised from an old site after its DNA replication and is inserted into a new site before DNA replication, one daughter cell receives two copies of Ac (old site and new site) and the other receives one copy (new site). (c) Replicative transposition. Transposase makes two single-strand nicks next to the 5' ends of the transposon in the original site. In the model, transposase makes two staggered nicks in the target (9 bp apart for Ac elements). A single strand (dotted line) is, in effect, transferred to a new site. Massive DNA replication is required to make both the old and new site double-stranded. A host sequence duplication is also created at the new insertion site.

These strategies have also been implemented in transgenic events expressing maize transposons [10–12]; with genetically engineered Ac/Ds gene and enhancer trap transposons have also been developed [10–12].

\textit{MoDR/Mu} transposons are active late in tissue development, resulting in tiny somatic sectors from ‘cut and paste’ events. \textit{MoDR/Mu} undergo duplicative transposition in pre-meiotic and gametophytic cells (Figure 1c), consequently, few germinal revertants exist (excision frequency $<10^{-5}$, which is less than a new germinal insertion) [1]. Mutator stocks typically contain several copies of \textit{MoDR} and many \textit{Mu} elements [16], resulting in a high forward mutation frequency ($10^{-4}$ range $10^{-3}$–$10^{-5}$) in any gene [1,16]. When \textit{Mu} element copy number is low, the mutation frequency is similar to that of \textit{Ac/Ds} [17]. Because epigenetic silencing is stochastic and frequent, somatic instability (Figure 1a) and unmethylated \textit{Mu} TIRs should be verified before mutagenesis [2].

\textit{Mu} insertions occur preferentially into low-copy DNA; there is no apparent local bias [16]. Recent analyses of hundreds of somatic sectors (unpublished data) and germinale \textit{Mu} element insertions (K Edwards, personal communication) confirm insertion into or near ex-on-like sequences. Although high copy numbers make \textit{MoDR/Mu} efficient mutagens, the resulting multiple mutant genes complicate the assignment of phenotypes to specific genes [2].

Transposon tagging in maize is much more efficient than would be predicted from the large genome size, $2.3 \times 10^9$ base pairs (bp) [18]. More than half of the maize genome, however, consists of inactive retrotransposons found in complex arrays; these create the spacers containing 20–200 thousand base pairs (kb) between maize genes [19]; consequently, insertions would be expected to affect only one gene. With short introns, maize genes are as compact as those of Arabidopsis [20]. Thus, transcription, recombination [21] and DNA transposon insertion are all highly biased for the genes, and for these important processes absolute genome size is of relatively little importance.

**DNA sequencing strategies for genomic DNA next to transposon insertions:** methods used to analyze transposons in a single plant

Initial applications of transposon tagging in maize relied on correlating the inheritance of a plant phenotype with a band on a DNA hybridization blot. Particularly for the multi-copy \textit{Mu} elements, demonstrating tight linkage required some luck in restriction enzyme choice and in hybridization probes to resolve specific \textit{Mu} elements (\textit{Mu1}, \textit{Mu2}, \textit{Mu3}, \textit{Mu4} and \textit{MoDR}); the \textit{Mu} TIR probe detects too many bands to be useful. Proof that the correct gene was closed after recovery of the transposon-tagged genomic fragment requires the analysis of independent mutants, or a complementation test. Given the difficulty of maize transformation, complementation has so far been useful only for traits that can be assayed in transient expression assays with protoplasts or tissues amenable to particle gun bombardment [22]. Several techniques for amplifying and sequencing genomic DNA next to transposon ends are amenable to genomics approaches. PCR primers anchored on transposons are ‘read out’ into genomic DNA [23,24,25••] with a wide choice of strategies for priming from the flanking genomic DNA of unknown or specified sequence (Figure 2) [26•]. Initial methods relied on the separation of the PCR products by size, with manual recovery and analysis of bands that segregated with a phenotype. Nowadays, simplifications include...
shotgun cloning combined with high-throughput sequencing. Selective PCR primer strategies can examine a subset of transposon family members or a subset of insertion sites (with primers spanning the joint with genomic DNA). Because transposon insertions result in characteristic host sequence duplications [9•], selective PCR primer strategies can examine a subset of insertion sites after artificial joining of the genomic DNA flanking the left and right ends of the transposon. Transposon insertions into a target gene or motif by using a PCR primer that reads out of the left and right ends of the transposon. Several restriction digestion and ligation steps are required to achieve this structure [26••]. One restriction enzyme ligates to an adapter. In ‘panhandle’ PCR, in which transposon TIRs form intramolecular duplexes in single-stranded DNA; the flanking gene is in the pan and the transposon is the handle. Several restriction digestion and ligation steps are required to achieve this structure [26••]. One restriction enzyme ligates to an adapter. In ‘panhandle’ PCR, in which transposon TIRs form intramolecular duplexes in single-stranded DNA; the flanking gene is in the pan and the transposon is the handle. Several restriction digestion and ligation steps are required to achieve this structure [26••]. One restriction enzyme ligates to an adapter.

Figure 2
Identifying transposon insertions. (a) Enriching for transposed Ac/De. In this diagram, kernels with excision sectors are white with spots of gray. Early somatic excisions that restore gene function result in a sector of gray kernels. By selecting seeds in a sector of kernels with a functionally reverted allele at the reporter gene (gray circles), investigators identify a somatic excision event from the original site, these kernels should contain the same transposed Ac at a new location in the genome. Because Ac shares a negative dosage effect on the frequency of somatic excision, kernels with fewer spots probably contain both the original donor site and a new transposed Ac [3]. (b) The amplification of sequences next to transposons typically uses a ‘read out’ primer (arrows) complementary to the TIRs and a strategy to allow priming from the contiguous genomic DNA (not) such as a restriction site overhang ligated to an adapter. In ‘panhandle’ PCR, TIRs form intramolecular duplexes in single-stranded DNA; the flanking gene is in the pan and the transposon is the handle. Several restriction digestion and ligation steps are required to achieve this structure [26••]. One restriction enzyme ligates to an adapter.

Indexed mutant collections and parallel searches for transposon insertions
Traditional genetics starts with the mutant individual and seeks the gene, whereas genomics methods start with the DNA sequence and search for the phenotypes. The first genomics approach, TUSC (Trait Utility System for Corn), was developed by Pioneer Hi-Bred International; it is available to academic researchers who sign material transfer agreements to receive the seeds containing putative ‘hits’ to user-supplied sequence motifs. TUSC contains DNA samples and progeny seed from ~44,000 Mutator plants that probably contain >106 independent Mu insertions. DNA samples, from single plants or post-preparation pools, are screened for Mu insertions into a target gene or motif by using a PCR primer that reads out of Mu TIRs and a target-specific primer. All mobile Mu elements share a very highly conserved sequence in the terminal 25 bp of the TIRs [16], consequently, a single ‘read-out’ primer suffices. TUSC has been used successfully by both Pioneer Hi-Bred International [29] and by academic researchers [30,31]. One drawback of a PCR strategy with a high cycle number is that both germinal and somatic insertions are found; ~10–20% genuine germinal insertions are verified among candidate insertions from TUSC screening [26••]. A convenient field organization into rows and columns for collecting pooled plant (and hence DNA) samples was implemented at about the same time as TUSC [32]; this method overcomes the identification of ‘false positive’ somatic insertions. DNA samples representing pooled leaf punches from a row of plants and separate sets of leaf punches from each column of plants are screened by PCR; two successful reactions define the row and the column and hence a specific plant that has the germinal insertion mutation (Figure 3). The Maine Targeted Mutagenesis (MTM) project, funded in the NSF plant genomics program, is now offering screening of a Mutator collection of ~30,000 plants and providing seed for the phenotypic evaluation of ‘hits’ [33•]. Project personnel are evaluating plant phenotypes at the kernel, seedling and adult plant stages to build a database of maize mutant phenotypes. To use MTM services,
Row and column pooling strategies for high-throughput transposon insertion site screening. Plant samples such as leaf punches are collected from every plant in a row (individuals 1 to n in the horizontal direction). A single DNA sample is prepared from each row for PCR screening or plasmid rescue. Similarly, plant samples are collected from every column, the vertical direction of the grid; in this example column 4 is in bold and boxed. Positive PCR results from a row and a column define a specific plant with the germinal insertion of interest, for example, row 3 and column 8 (circled). Somatic insertions are amplified only in a row or only in a column in Mutator grids because somatic excisions and insertions occur mainly late in development.

Saturation mutagenesis within a gene

One mutation, no matter how instructive, is rarely sufficient to lead to a comprehensive understanding of a gene’s function. Most transposon insertions into exons or introns are ‘knockouts’ of gene function; alternative splicing events that use sequences in the ends of the transposon can result in modest expression in a few cases [1]. Transposon insertions, unlike most mutations induced by physical agents or Agrobacterium inserts, are ‘hot spots’ for secondary mutations. Of greatest current use in maize are Ac/Ds and Spm/En mutants from which germinal revertants are readily recovered; these typically contain one or a few base changes in addition to the host sequence duplication and, more rarely, larger deletions, ‘filler DNA additions’ or rearrangements. Of particular utility is the presence of Ac/Ds for local transposition, an Ac/Ds that transposes nearly in readily transpose back to the target gene, providing many new types of mutant allele [34,35]. For example, the promoter can be saturated with insertions at different sites followed by selection for minor alterations that affect regulation, or the requirements for splicing can be explored from insertion sites in or near the conserved intron motifs.

An important feature of all maize transposons is that they are somatically instable in the presence of the transposase-encoding element. Consequently, somatic tissue is a mosaic of mutant and revertant cells of many different phenotypes. To produce somatic tissue of a single phenotype, revertants can be selected, but not all of these are ‘knockout’ alleles. Alternatively for Ac/Ds, individuals lacking Ac can be recovered to stop somatic excisions. This strategy is virtually impossible for multi-copy MuDR lines; however, epigenetic silencing of transposons can occur spontaneously and is frequent in Mutator lines [16]. In the case of Mu1 insertions in promoters, methylation after silencing can activate a ‘read out’ promoter in the TIRs, in effect restoring gene expression [36]. Stabilization of a fully mutant phenotype is possible by selecting for deletions, which occur with about 10⁻² frequency from the ends of Mu1 elements [37].

Conclusions

Transposon-induced phenotypes have long provided geneticists with beautiful materials and insights into gene expression, development and chromosome mechanics. Cloned transposons have facilitated gene discovery and cloning. In the genomics era, maize transposons have emerged as the premier method for gene discovery and sequencing, as well as the phenotypic analysis of gene expression in a whole-organism context. Transposons allow simultaneous effort in both phases of genomics, gene discovery and functional studies. Transposons are more efficient than a sequential approach to gene sequencing followed by the design of tools to study gene expression.

Acknowledgements

I thank Michelle Facette for her comments on the manuscript. Research in Minutiae and preparation of this article were supported by grants from the NSF (Maize Gene Discovery, DNA Sequencing and Phenotypic Analysis grant 9872657) and the National Institutes of Health (GM61665).

References and recommended reading

Pages of particular interest: published within the annual period of review, have been highlighted as:
* of special interest
** of outstanding interest


The control of maize spiklet meristem fate by the PRL gene of maize: a database of phenotypic information on large populations of mutator plants. Mutagenesis Database. URL http://mtm.cshl.org


A database of phenotypic information on large populations of mutator plants. The website includes instructions on how to request PCR screens for insertions in specific motifs.


Saturation mutagenesis using maize transposons Walbot 107