the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2009) **58**, 883–892 doi: 10.1111/j.1365-313X.2009.03821.x

TECHNICAL ADVANCE

High-throughput linkage analysis of *Mutator* insertion sites in maize

Gibum Yi[1,2], Diane Luth[3], Timothy D. Goodman[2,4], Carolyn J. Lawrence[2,3,4,5] and Philip W. Becraft[1,2,3,*]

[1]*Plant Biology Program, Iowa State University, Ames, IA 50011, USA,*

[2]*Genetics, Development & Cell Biology Department, Iowa State University, Ames, IA 50011, USA,*

[3]*Agronomy Department, Iowa State University, Ames, IA 50011, USA,*

[4]*Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, USA, and*

[5]*CICGRU, USDA-ARS, Ames, IA 50011, USA*

Received 20 October 2008; revised 27 December 2008; accepted 21 January 2009; published online 3 March 2009.

*For correspondence (fax +1 515 294 6755; e-mail becraft@iastate.edu).

SUMMARY

Insertional mutagenesis is a cornerstone of functional genomics. High-copy transposable element systems such as *Mutator* (*Mu*) in maize (*Zea mays*) afford the advantage of high forward mutation rates but pose a challenge for identifying the particular element responsible for a given mutation. Several large mutant collections have been generated in *Mu*-active genetic stocks, but current methods limit the ability to rapidly identify the causal *Mu* insertions. Here we present a method to rapidly assay *Mu* insertions that are genetically linked to a mutation of interest. The method combines elements of MuTAIL (thermal asymmetrically interlaced) and amplification of insertion mutagenized sites (AIMS) protocols and is applicable to the analysis of single mutants or to high-throughput analyses of mutant collections. Briefly, genomic DNA is digested with a restriction enzyme and adapters are ligated. Polymerase chain reaction is performed with TAIL cycling parameters, using a fluorescently labeled *Mu* primer, which results in the preferential amplification and labeling of *Mu*-containing genomic fragments. Products from a segregating line are analyzed on a capillary sequencer. To recover a fragment of interest, PCR products are cloned and sequenced. Sequences with lengths matching the size of a band that co-segregates with the mutant phenotype represent candidate linked insertion sites, which are then confirmed by PCR. We demonstrate the utility of the method by identifying *Mu* insertion sites linked to seed-lethal mutations with a preliminary success rate of nearly 50%.

Keywords: transposon, tagging, mutant, transposon display, gene cloning, method.

INTRODUCTION

Insertional mutagenesis is a powerful method for the identification of genes. In plants, T-DNAs and several transposable element (TE) systems have been extensively used. In maize (*Zea mays*), endogenous TEs are the most effective and widely used insertional mutagens. Mutant screens are conducted in lines containing active TEs of known sequence, and the TE is then used as a tag to isolate genomic DNA flanking the insertion site, which corresponds to the disrupted gene of interest. Several TE systems have been extensively used and each has advantages and disadvantages. The two systems most commonly used in maize are *Activator/Dissociation* (*Ac*/*Ds*) and *Mutator* (*Mu*) (Brutnell, 2002).

Activator (*Ac*) is a low-copy system: forward mutation rates are low making mutant generation laborious, but subsequent identification and isolation of the mutated gene is relatively easy. *Ds* is a related non-autonomous element that requires the transposase encoded by *Ac* for its transposition. This provides convenient control of *Ds* activity. Both *Ac* and *Ds* transpose preferentially to linked sites, so genes in proximity to one of these elements may experience a relatively high mutation rate (Dooner and Belachew, 1989).

Mu is a high-copy system with typically 50 to 200 copies per individual genome (Walbot and Warren, 1988). Forward mutation rates are high, making mutant generation efficient. For example, when *Mu* activity was introduced into the W22 inbred line, plants contained an average of 57 *Mu* elements and generated new seed mutants with a frequency of 7% per generation, with transpositions randomly scattered throughout the genome (McCarty *et al.*, 2005). However, subsequent identification of the particular *Mu* element responsible for the mutation of interest is challenging.

The development of efficient methods for analyzing *Mu* elements would greatly facilitate their utility in gene isolation efforts, and several PCR strategies have been developed to allow the identification and recovery of genomic sequences flanking *Mu* insertion sites. Amplification of insertion mutagenized sites (AIMS) is a ligation-mediated method (Frey *et al.*, 1998; Wang *et al.*, 2008). Briefly, genomic DNA is digested with a restriction enzyme and an adapter is ligated to the ends of the restriction fragments. A PCR is then performed using a *Mu* primer that recognizes conserved sequences in the terminal inverted repeats (TIRs) of *Mu* elements and a primer that recognizes adapter sequences. In the first round of amplification the *Mu* primer is biotinylated, and the products are purified with streptavidin beads to enrich for *Mu*-containing fragments. In the second round of PCR the *Mu* primer is labeled with radioisotope or a fluorescent tag. Fragments are run on a sequencing gel, and if a fragment co-segregates with the mutant phenotype of interest it is cut from the gel and cloned. This approach is generalizable to other transposable element families as well (Yephremov and Saedler, 2000).

Another strategy for amplifying sequences flanking insertion sites is thermal asymmetrically interlaced (TAIL) PCR (Liu *et al.*, 1995). MuTAIL is a recent adaptation of this approach to the analysis of *Mu* insertion sites (Settles *et al.*, 2004, 2007). This approach relies on nested specific primers to *Mu* TIRs in combination with arbitrary degenerate primers to amplify unknown flanking sequences. Preferential amplification of fragments containing *Mu* insertions is achieved by using primers with different melting temperatures. The *Mu* TIR-specific primers have a high melting temperature while the degenerate primers have low melting temperatures. Multiple cycles run at high annealing temperatures allow linear amplification only from the *Mu*-specific primers. Interspersed cycles at low annealing temperatures allow exponential amplification from the degenerate primers.

While the AIMS approach has proven successful, the drawback for implementing this approach in a high-throughput manner is the amount of labor involved. Genomic DNA must be digested and ligated, then after a set of PCR cycles the biotinylated *Mu*-containing fragments are purified with streptavidin beads. After segregation analysis, fragments of interest must be excised and recovered from the sequencing gel, reamplified, and finally cloned and sequenced. Another limitation is that fragment sizes generally are small, ranging from 100 to 400 base pairs (Frey *et al.*, 1998), although with the preference of *Mu* elements for genic sequences (McCarty *et al.*, 2005, and references therein) and the advent of a fully sequenced maize genome the small fragment size is likely to no longer pose a significant problem.

The MuTAIL approach is amenable to high-throughput utilization in that the DNA requires no manipulations prior to PCR. Additionally, the TAIL-PCR regimen strongly enriches for *Mu*-containing fragments, allowing amplification products to be directly cloned and sequenced. MuTAIL also produces a greater range of fragment sizes, up to 2000 base pairs (Settles *et al.*, 2004). The TAIL method has been used effectively for extensive analysis of *Mu* insertion sites in maize lines (Settles *et al.*, 2004; McCarty *et al.*, 2005). In principle, MuTAIL could be adapted to undertake co-segregation analyses of mutants of interest, but the large fragments go beyond the size that can be readily resolved on standard sequencing equipment, and the complexity is too great to be resolved on agarose gels. Furthermore, a battery of up to 12 different degenerate primers is required to achieve an estimated 80% representation of the *Mu* insertions present in a genome.

Here we combine elements of both MuTAIL and AIMS into a procedure we call MuTA, which is an efficient method for co-segregation analysis of *Mu*-induced mutants and is suitable for high-throughput use.

RESULTS

Overview of the method

Like AIMS, MuTA is based on the principles of amplified fragment length polymorphism (AFLP; Frey *et al.*, 1998; Vos *et al.*, 1995; Wang *et al.*, 2008). The genomic DNA is digested with a restriction enzyme to generate fragments and an adapter is ligated onto the ends of the fragments. The PCR is then performed using one primer to the *Mu* TIRs and one to the adapter. However, to enrich for *Mu*-containing fragments, instead of using a biotinylated MuTIR primer and streptavidin purification we incorporated principles of TAIL PCR (Liu *et al.*, 1995). The MuTIR primer we used had a melting temperature (T_m) of 65°C while the adapter primer had a T_m of 50°C. The PCR protocol consists of multiple cycles with an annealing temperature of 60°C interspersed with a single cycle at 47°C (Figure 1). The high annealing temperatures allow linear amplification from *Mu* elements while the low temperature allows annealing of both primers and exponential amplification.

To detect the *Mu*-containing products, the MuTIR primer used in the second round of PCR contains a fluorescent label. Products are analyzed on a capillary sequencer, such as an ABI3100 (Mitchelson, 2001), and the chromatograms

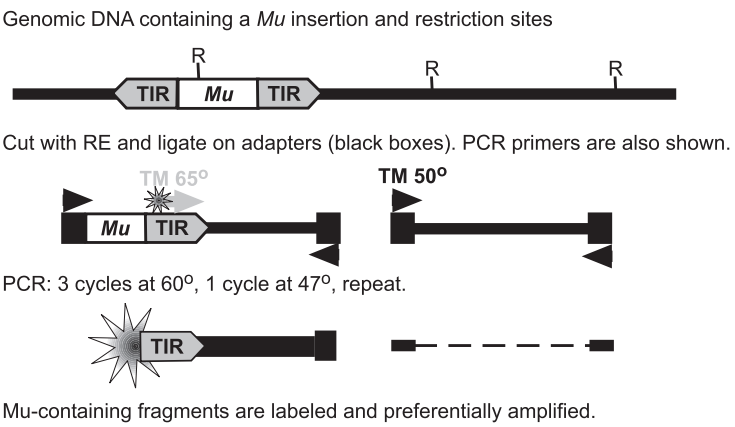

Figure 1. Overview of the MuTA PCR strategy. RE, restriction endonuclease; TIR, terminal inverted repeat.

are analyzed with genotyping software such as the publicly available Genographer (http://hordeum.msu.montana.edu/genographer/), or other commercial packages. For example, fragments can be conveniently visualized and sized on a pseudogel image (see Figures 3 and 5).

Co-segregation with the mutant of interest is used to identify candidate fragments that are genetically linked to the mutation. If a transposon is responsible for a mutation, absolute linkage is expected between the transposon insertion and the mutant phenotype. A PCR product that is present in all the mutants and none of the wild types is a candidate for a causal insertion site (see Figures 3 and 5). We identified candidate fragments by analyzing small segregating families consisting of four mutant and four homozygous wild-type individuals. The genotyping software provides an accurate estimate of the size of the candidate fragment.

To recover a candidate fragment, a library of cloned PCR products from one of the reactions containing the desired target is generated. We used a commercial topoisomerase cloning kit and then picked 96 clones and sequenced the inserts. The target clone is recognized by comparing the number of sequenced bases with the size estimate of the desired fragment provided by the genotyping software. The sequence analysis has been automated by software we developed called MuTAlyzer, described in the Experimental Procedures. The candidate sequence is then used for database searches and to design gene-specific PCR primers to verify the genetic linkage between the insertion and the mutant of interest.

Identifying a *Mu* insertion genetically linked to a mutant phenotype

We initially developed the method using an empty pericarp mutant, *emp*02S-0422* (Figure 2), obtained from the UniformMu collection (McCarty *et al.*, 2005; Settles *et al.*, 2007). Because the mutation is homozygous lethal, it must be maintained as a heterozygote. We crossed a heterozygous mutant to the wild-type W22 inbred line to generate a family segregating 1:1 heterozygous mutants and

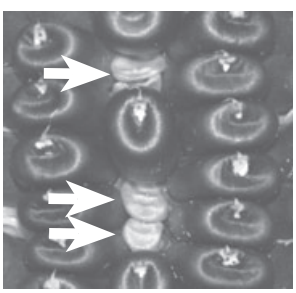

Figure 2. The seed lethal 'empty pericarp' mutant *emp*02S-0422*. Mutant segregants on the self-pollinated ear are indicated with arrows.

homozygous wild types. Tissue was isolated from progeny plants, which were then self-pollinated to determine the genotypes; the heterozygotes segregated *emp*02S-0422* mutant kernels on their ears.

DNA was isolated from four heterozygous mutants and four wild types and subjected to the MuTA analysis. Six PCR templates were prepared by digesting the DNA with *Ase*I, *Bst*BI, *Nde*I, *Bsa*HI, *Bfa*I, or *Hin*P1I (6, 6, 6, 5, 4 and 4-cutters, respectively). All these enzymes produce either a 5′ TA or a 5′ CG overhang. Two sets of adapters were prepared that were identical except for the TA or CG overhang. Adapters were ligated to the ends of genomic restriction fragments and PCR was performed as described. The PCR reactions on the *Nde*I and *Hin*P1I templates were performed with FAM-labeled MuTIR primers, *Bsa*HI and *Bst*BI with HEX, and *Ase*I and *Bfa*I with NED. Following PCR, portions of the *Nde*I, *Bsa*HI, and *Ase*I reactions from a given individual were combined in a well of a 96-well plate. Thus, these samples occupied one column (eight wells) of the plate. The *Hin*P1I, *Bst*BI, and *Ase*I reactions were likewise combined to occupy a second column of the plate. The samples were analyzed on an ABI3100 capillary sequencer (Applied Biosystems ABI3100 Genetic Analyzer, http://www3.appliedbiosystems.com) at the Iowa State University DNA Facility. Pseudogel images were produced from the resultant chromatograms using Genographer software. As shown in Figure 3, the *Bsa*HI reactions produced a band that was observed in all

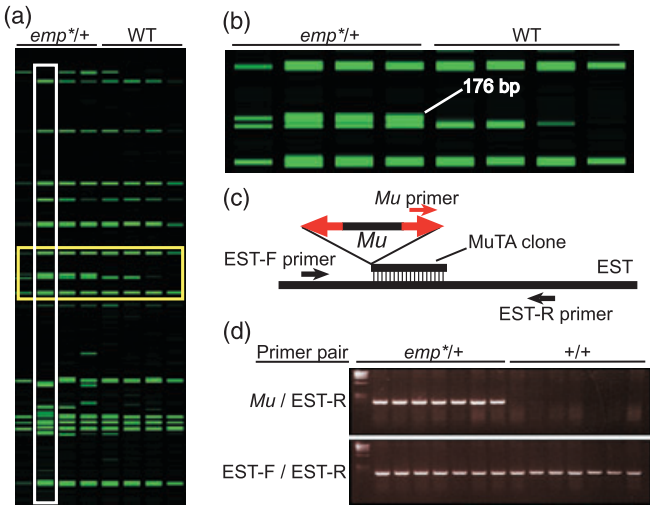

Figure 3. Identification of a *Mu* insertion linked to mutant *emp*02S-0422*.
(a) Pseudogel image produced by Genographer® from the chromatograms obtained by running the MuTA PCR products on an ABI3100 capillary sequencer. The yellow box is enlarged in (b) and the white box identifies the reaction from which a clone library was prepared.
(b) Close-up of the inset shown in (a). A band co-segregates with the mutant. Genographer provided a size estimate of 176 bp.
(c) Diagram depicting the relation of the *Mu* insert, the cloned PCR product, the expressed sequence tag (EST) identified in database searches, and the primers used in PCR reactions to verify the identity of the recovered band and linkage of the *Mu* insertion site to the mutation.
(d) The PCR results showing the presence of the *Mu* insertion in all individuals carrying the mutation. Genomic DNA from segregating individuals was used as the template for PCR reactions using the primer pairs depicted in (c).

four mutants and none of the wild types, as would be expected for a *Mu* insertion responsible for causing the mutant phenotype. The band associated with the mutants was estimated by Genographer to be 176 base pairs.

To recover the desired DNA fragment, PCR products from mutant individual 2 were shotgun cloned using the Strataclone topoisomerase cloning kit (Stratagene, http://www.stratagene.com/). Transformed bacteria were plated on LB medium containing ampicillin and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). Ninety-six white colonies were picked and the plasmid inserts sequenced. One clone contained an insert of 176 bases.

Another PCR experiment was performed to verify that the recovered clone corresponded to the PCR product associated with the mutant individuals. The clone sequence was used in BLAST searches (Altschul *et al.*, 1990) of the GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) High Throughput Genomic Sequence (HTGS), Genome Survey Sequences (GSS), and Expressed Sequence Tag (EST) databases (Benson *et al.*, 2008) and showed a 100% match to a maize EST (GenBank accession number DV491867). A gene-specific primer was designed against the EST and used for PCR on DNA from additional segregants of the mutation. In combination with a MuTIR primer, only heterozygous mutant individuals produced a product, while in combination with a second gene-specific primer, all individuals produced a product (Figure 3c,d). This demonstrated that both the recovered clone and the EST correspond to the *Mu* insertion associated with the mutation. To date, we have assayed over 80 segregants and have observed complete linkage between this *Mu* insertion and the mutant phenotype. This demonstrates the feasibility of this approach for identifying *Mu* elements tightly linked to mutant genes of interest.

High-throughput identification of linked *Mu* insertion sites

To test the adaptability of this method to high-throughput analysis, we assembled 12 seed-lethal mutants for which we had sufficient segregating material. Because efficient amplification of a *Mu* insertion site is dependent on the proximity of the restriction site used for template preparation, the likelihood of identifying a linked *Mu* insertion for a given mutant is maximized by performing multiple analyses on templates prepared with different restriction enzymes. For high throughput this must be balanced with the increased labor and cost of performing additional reactions on a given mutant. We decided on three digests per mutant: *Bfa*I, *Csp*6I and *Msp*I. Each digest was used as template for MuTA PCR using a MuTIR primer labeled with a different fluorophore, NED, FAM, or HEX. Multiplexing was used to reduce lane costs in the genotyping analysis; the products of the three separate reactions were combined in a single lane for analysis on the ABI3100. When the products of the three reactions are combined, each mutant occupies one column of a 96-well plate, allowing the electrophoretic analysis of 12 mutants per plate (Figure 4).

Of the 12 mutants we analyzed, we were able to identify a band associated with the mutant individuals for six (Figure 5). In three cases, a co-segregating product

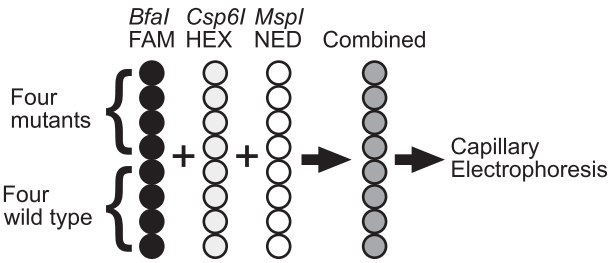

Figure 4. Multiplexing of three independent reactions. A mutant occupies a single column of a 96-well plate.

was identified in two of the different enzyme digests (Figure 5a) and in one case a linked product was detected in all three digests.

To recover the candidate-linked DNA fragment for a given mutant, the PCR products from one mutant reaction were cloned and 96 clones sequenced. For example, for mutant 2 we cloned the products from the reaction displayed in lane 2 of the *Bfa*I reactions (Figure 5). In every case, we were able to recover a correctly sized fragment by sequencing 96 cloned products. For five cases, there was a unique product of the correct size. For mutant 12, the estimated size of the linked product was 240 bp and three different candidate clones had sizes of 242, 244, and 245 bp.

The sequences of all the candidate fragments were used in database searches to identify genomic sequences from which gene-specific primers were designed. When the gene-specific primers were used in combination with a MuTIR primer, a product was produced from mutant individuals of lines 2, 6, 8, 10, and 11, but not wild-type individuals. This confirmed that the recovered clone corresponded to the co-segregating PCR product observed on the pseudogel image. In the case of mutant 12, PCR primers were designed against all three candidate clones and only the 244-base sequence produced a product specific to the mutant DNA samples.

Verification of linkage between isolated *Mu* insertions and mutant phenotypes

Further analysis of linkage in larger segregating populations was undertaken. For each mutation, DNA was isolated from individuals in lines segregating 1:1 homozygous wild type and heterozygous mutants, and PCR was performed using the same gene-specific primer and MuTIR primer pairs described. Each plant was also self-pollinated to determine whether the plant carried the mutant allele. Linkage was verified for all seven insertions. In six of the lines, there was perfect correspondence between mutant phenotype and the candidate *Mu* insertion (Table 1 and Figure S1 in Supporting Information). A total of 88, 37, 39, 42, 40, and 56 segregants were analyzed for each of these six lines, indicating that the identified *Mu* insertions were tightly linked to all these mutations.

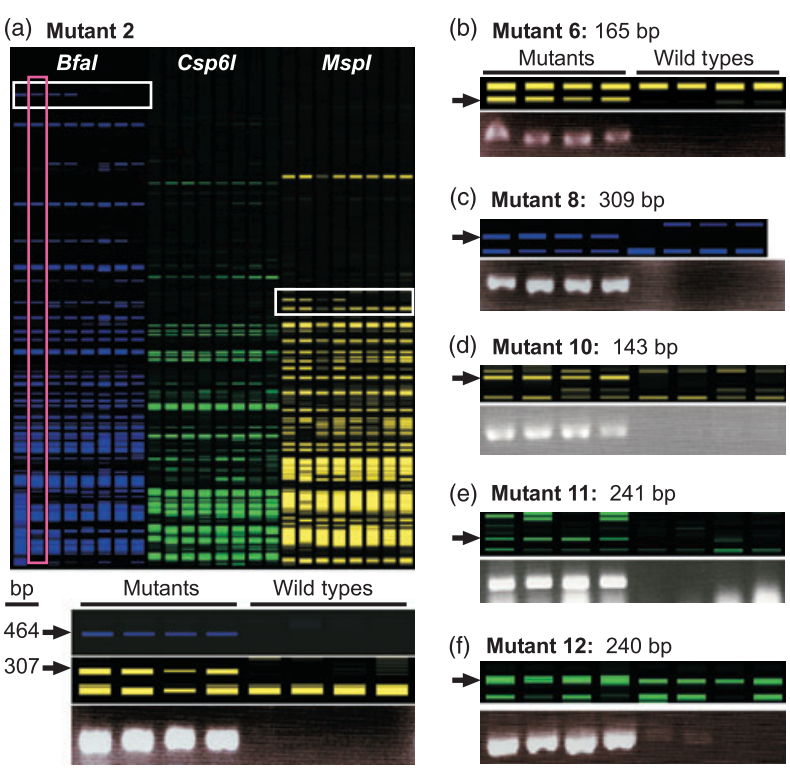

Figure 5. Results of a high-throughput analysis of 12 mutants. Linked insertions were found for six.
(a) Typical output of three different reactions on samples of eight segregating individuals for seed-lethal mutant 2. Co-segregating bands were found in the *Bfa*I and *Msp*I reactions (white boxes). The red vertical box is the reaction from which a library was prepared. Below are shown enlargements of the co-segregating fragments from the ABI3100 output and the ethidium bromide stained gel showing the PCR verification that the desired fragment had been identified, similar to Figure 3 (c,d).
(b)–(f) *Mu* insertions co-segregating with five additional seed-lethal mutations. Fluorescent bands represent the ABI3100 output and the ethidium bromide gels verify recovery of the linked fragment. The respective fragment sizes are indicated for each.

Table 1 Linkage data including the original eight segregants plus additional individuals sampled

Mutant line	WT lacking insertion/WT	Mutants with insertion/mutants
*emp*02S-0422*	42/42	46/46
2	16/16	21/21
6	20/20	19/19
8	20/20	22/22
10	21/21	19/19
11	28/28	22/25
12	28/28	28/28

WT, wild type.

Mutant 11 showed several discrepancies between the presence of the *Mu* insertion and the mutant phenotype. Of 53 segregants analyzed, three produced mutant progeny but did not contain the candidate *Mu* insertion. Two segregating lines were analyzed for this mutant and all the discrepancies occurred in the same line. We suspect this line might actually carry a second empty pericarp mutation that could account for the discrepancies, but this remains to be tested. Existing evidence indicates that this *Mu* insertion is genetically linked to the respective mutation, but does not support it being causal.

Analysis of cloned sequences

For each mutant with a candidate band, a micro library of 96 clones was sequenced. Sequences that meet the 50-bp required minimum length have been deposited in the GenBank GSS collection (accession numbers FI495664–FI496078), and a complete file of raw sequences is available in the Supporting Information (Appendix S1). A quality measure of the protocol is provided by the analysis of all the sequences: 95.5% of the sequences were 'good', where good was defined as a high-quality sequence that included a *Mu* TIR and an adapter at the termini. 'Bad' included colonies that failed to grow, empty vectors, failed reactions, multiple clones per reaction, and clones with two adapters and no *Mu* TIR. Among the 550 'good' sequences represented in the latter six libraries, all but 17 had hits when used in BLAST searches (Altschul *et al.*, 1990) of maize sequences in the GenBank HTGS, GSS, and EST databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Benson *et al.*, 2008). Of those that failed, 14 were too short to perform a search. Of the searchable sequences, all but 10 had matches in the HTGS database. Only three sequences of sufficient length to perform a search had no matches among the maize HTGS, GSS, or EST GenBank accessions.

Table 3 Summary of the sequence analysis for all six libraries combined

Total insertion sites	93
Mean no. of insertion sites per library	21
Insertion sites shared among >1 library	21
Insertion sites unique to a library	72
Insertion sites with both flanks recovered	27
Insertions per library with both flanks recovered	5.2

The analysis of the sequenced insertion sites among the latter six libraries is summarized in Tables 2 and 3. Most clones had multiple isolates. An average of 21 independent insertion sites were identified per library, with an average of five represented by both flanks. A total of 93 unique insertion sites were identified, with 21 shared among multiple libraries and 72 unique to a given library. Some of the shared insertions reflect the common parentage among the lines in that four were derived from the UniformMu population. However, mutants 11 and 12 were derived from different materials and, surprisingly, 11 insertion sites from these two libraries were shared with the other four UniformMu libraries.

Of the 93 insertions identified, 79 had hits in EST database searches. Of the six clones identified that were linked to mutant phenotypes, five had direct hits in ESTs (Table 2). For mutant 11, genomic sequence corresponding to the PCR product was identified and used to search the EST database. A 5′ EST was identified that mapped 197 bp 3′ to the *Mu* insertion site.

DISCUSSION

The protocol described here is simple and reproducible. The method can be applied to either a single mutant of particular interest or adapted for high-throughput use. Thus far, we

Table 2 Summary of the sequence analyses for the libraries associated with six seed lethal mutants

Library	Mutant	No. of insertion sites sequenced	No. of isolates of linked site	Corresponding EST
2	*emp*02S-0288*	23	3	CF631304
6	*emp*02S-0254*	11	3	EB815794
8	*emp*02S-0089*	23	1	DY241608
10	*emp*02S-0390*	19	1	EB640046
11	*dek*0285*	37	1	DY623734
12	*o5-PS3038*	18	2	EE189213

EST, expressed sequence tag.

have analyzed 15 mutants with this method and co-segregating fragments have been identified and recovered for seven of them (Figures 3 and 5). In all seven cases, we were able to recover a correctly sized fragment by sequencing 96 cloned PCR products from one of the reactions containing the band of interest. Database searches identified maize genomic sequences for all seven clones. These were used to design PCR primers that were used in combination with a MuTIR primer to verify that the desired fragment was recovered from each reaction. Examination of a larger segregating population verified the linkage between the *Mu* insertion and the respective mutation in all seven cases, and in six cases the linkage was perfect, suggesting the insertion could be responsible for the mutation.

The general protocol could be easily adapted to various laboratory circumstances. For example, alternative means of labeling the *Mu* primer are feasible and products could be analyzed on acrylamide sequencing gels. Products of interest could be recovered by excising bands from the gel to circumvent the need for sequencing. While we recovered the fragment of interest with an average frequency of 1.7 isolates per 96 clones, the fact that four of the seven libraries produced only a single isolate suggests that sometimes 96 clones would not be sufficient. In contrast, recovery by excision from an acrylamide gel would allow the more direct recovery of a desired product. This might be a preferable approach for an analysis focused on a small number of mutants of particular interest. However, the increased labor involved in running acrylamide gels makes the automated sequencing approach preferable for high-throughput gene cloning.

Factors affecting the identification of linked *Mu* insertion sites

Thus far we have experienced a success rate of recovering insertions linked to mutations derived from *Mu* lines of slightly less than a 50%. Of the factors affecting success, the most critical is whether the mutation is caused by a *Mu* insertion. Mutants derived from *Mu*-active lines have been shown to be caused by other transposons (Patterson *et al.*, 1991) or by deletions (Robertson and Stinard, 1987; Robertson *et al.*, 1994; Das and Martienssen, 1995). Anecdotal evidence suggests that approximately half the mutants derived from *Mu* stocks might not be *Mu*-tagged. Thus, only a subset of mutants derived from *Mu* stocks are amenable to cloning by any *Mu*-tagging approach, including this one.

Whether a particular *Mu*-containing genomic region amplifies with MuTA depends on the proximity of the nearest site for the particular restriction endonuclease used in the template preparation, as well as the PCR conditions. Any enzyme that produces a constant overhang can be used and no one enzyme will work for every *Mu* insertion. In the case of a single mutant of particular interest, the likelihood of identifying a linked *Mu* insertion is maximized by performing the analysis with several different restriction enzymes. We created two adapters that differed only in their 5′ overhangs, TA versus CG (Table 4). These can be used in conjunction with any restriction enzyme that produces a compatible overhang. There are at least 15 restriction enzymes that produce one of these two overhangs (although *Taq*I is not useful because it cleaves most *Mu* TIRs and several other enzymes are too expensive to be practical). Thus, these adapters allow substantial versatility in template preparation.

The average fragment size is dependent on the recognition site; enzymes with 4-base sites cut an average of once every 256 bases, 6-base sites cut an average of every 4096 bases. The upper limit of useful size for PCR products in our system is 1200 bases, which is the upper limit of the available size standards for the capillary sequencer. The lower limit is determined by the ability to unequivocally identify the cognate sequence in a database search. Several variables influence this, including whether the *Mu* insertion site is in a unique sequence. In our experience, 20 bases are sometimes sufficient but 50 have a good probability of identifying a unique target. The range of PCR products we have obtained thus far is generally below 700 bases, regardless of restriction endonuclease, suggesting that PCR conditions are currently limiting efficient amplification of longer fragments.

We recovered an average of 21 insertion sites from each 96 cloned PCR products sequenced. This appears consistent with the number of bands observed in the pseudogel images. Some of the genetic materials we used include isolates from the UniformMu population, which contains an average of 57 *Mu* elements per individual (McCarty *et al.*, 2005). Assuming this value, we appear to amplify approximately 37% of the insertion sites with a given

Table 4 Primer and adapter oligonucleotide sequences

Name	Sequence
Adapter 1	5′ TAGAAAGAATTCGGATCCAATTATAT 3′ 3′ CTTTCTTAAGCCTAGGTTAATATA 5′
Adapter 2	5′ CGGAAAGAATTCGGATCCAATTATAT 3′ 3′ CTTTCTTAAGCCTAGGTTAATATA 5′
Adapter primer	ATATAATTGGATCCGAATTCTTTC
MuTIR1	AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC
MuTIR2	TCGTCGAATCCCBTYCBCTCTTCKTCYATAAT
MuTIR3	ATCCCGTCCGCTCTTCGTCTATAA
Net-for.	AGCTCGCGGTACAGCTTTCCA
Net-rev.	AAGCGGAGAAAGAAGCTCCTGA
B2B-D11	CAATCCTGCACGAGCACAAGC
C6D-A6	TGGCAGTCGATCAAGCTGTTG
C8C-G10	GCCCAGCACATCTAGCATCAA
M10B-E5	GATAGGGAAACCGTGGAAGGA
C11C-D3	CCGCTTGAAACGAAGACCTTTGGA
C12D-E3	TGGGCGTCCTTCTTCTCCAAT

enzyme. To reduce the complexity of fragments displayed, AIMS adds a selective base to the adapter primer (Frey *et al.*, 1998). We have not found this necessary, perhaps due to the outstanding resolution afforded by capillary sequencers.

The size of the segregating family affects the statistical ability to detect linkage. Our experiments have been conducted on segregating families consisting of four heterozygous mutants and four wild types, which is smaller than typical for linkage studies. Since we are interested in insertions showing absolute linkage to a mutant, any element showing linkage in a large sample will also be apparent in a small sample. But what is the likelihood of false positives? With the current sample size, the probability of a false positive (an element present in all four mutants and none of the wild types) for an unlinked element is $(0.5)^8$, or 0.004. However, the UniformMu lines contain an average of 57 *Mu* elements (McCarty *et al.*, 2005). With that number, there is an approximately 50% chance of having an element within 20 cM of any given locus, and a 20% chance of having one within 5 cM (Briggs and Beavis, 1994). In these cases, a sample size of eight would have an approximately 0.08 or 0.13 chance of producing a false positive, respectively. For example, the probability of a false positive (no recombinant) with 20 cM linkage would be $(1 - 0.2)^8$, or 0.168, multiplied by 0.5 (the probability of having a Mu element within 20 cM) equals 0.84. At 5 cM, the calculation is $(0.95)^8 \times 0.2 = 0.133$. If the sample size were increased to 12, these likelihoods of false positives would drop to 0.03 and 0.11, respectively – not a dramatic improvement given the extra labor and expense. Because further analysis is required for verification of candidate insertion sites regardless, this is an acceptable false positive rate for the first pass. Furthermore, the Poisson distribution of *Mu* elements in these lines (McCarty *et al.*, 2005) argues that the majority of co-segregating insertions will be responsible for the linked mutation.

Quality and reproducibility

The banding patterns generated in the pseudogel images clearly show products in common among related individuals of a segregating line, as well as bands unique to various individuals. The conserved banding patterns among related individuals demonstrate the reproducibility of the method. The analysis of sequenced PCR products indicated that this method is an effective means of reliably identifying *Mu* insertion sites. Approximately 95% of the products sequenced produced high-quality sequences containing a *Mu* TIR and an adapter. Among the flanking sequences analyzed, only three showed no match in BLAST searches of existing maize database accessions. This reflects the efficacy of the method for amplifying bona fide *Mu* insertion sites from maize, and also reflects the degree of coverage in maize sequence databases. As database resources continue to improve and the maize genome sequence becomes more complete, sequence availability will pose less of a limitation on methods such as this that rely on sequence searches. In summary we conclude that this is an effective method for rapidly identifying candidate *Mu* insertions linked to mutants of interest. This method could be readily adapted to other transposon systems or other species.

EXPERIMENTAL PROCEDURES

Genetic stocks

Plants of lines segregating seed lethal mutations (Figure 2) were self-pollinated for progeny tests to identify wild-type and heterozygous mutant individuals.

Template preparation

The DNA was isolated from leaf tissue by the Iowa State DNA facility using an AutoGen 740® (http://www.autogen.com/). Half to one microgram of genomic DNA was digested with an appropriate restriction endonuclease in a 10-μl reaction volume according to the manufacturer's specifications. If the subsequent adapter ligation re-created the restriction site, the restriction enzyme was heat killed after the digestion was complete. One microliter of the digested genomic DNA was added to 1 μl of 50 μM adapter DNA in a 10-μl ligation reaction. Adapter sequences are shown in Table 4. Ligations were performed overnight at 4°C.

Polymerase chain reaction

Two rounds of PCR were performed. For the first round, 1 μl of template DNA from the ligation reaction was added to a 20-μl PCR reaction containing 20 pmol each of the MuTIR1 primer and adapter primer (Table 4), GoTaq polymerase and buffer (Promega, http://www.promega.com/), 1.5 mM $MgCl_2$ and 0.5 mM deoxyribonucleotides (dNTPs). The PCR program is shown in Table 5.

For the second round, PCR products of the first round were diluted 1:1000 in water and 1 μl added to a 30-μl reaction containing

Table 5 Cycling parameters for the first round of PCR

Step	Condition	Time
1	95°C	3 min
2	94°C	20 sec
3	60°C	1 min
4	72°C	2 min 30 sec
5	Go to 2	Twice
6	94°C	20 sec
7	47°C	1 min
8	72°C	2 min 30 sec
9	Go to 2	12 times
10	72°C	7 min 30 sec
11	4°C	Forever

30 pmol each of MuTIR2 and the adapter primer (Table 4). Other components were identical to the first round reaction mix. MuTIR2 is a 1:1 mix of labeled and unlabeled primer because the 5′ fluorescent tag interferes with cloning. The PCR cycling is identical to the first round except the 60°C annealing temperature in step 3 is reduced to 57°C to accommodate the lower melting temperature of MuTIR2.

PCR product analysis

A total volume of 1.5 μl of the second reaction products was analyzed. When performing multiple analyses on a given sample (i.e. multiple restriction enzymes), the PCR for each digest was performed with a different fluorescent label on the MuTIR2 primer. We used FAM, HEX, and NED fluorophores. Half a microliter from each of three reactions was combined in a single well. Samples were submitted to the Iowa State University DNA Facility for analysis on an ABI3100. Chromatograms were downloaded and analyzed using GENOGRAPHER software (http://hordeum.msu.montana.edu/genographer/) to produce pseudogel images. Images were scrutinized for co-segregating bands and size estimates of candidate bands of interest were provided by the software.

Fragment recovery

To recover the desired DNA fragment, PCR products from one of the mutant individuals containing a fragment of interest were shotgun cloned using the Strataclone topoisomerase cloning kit (Stratagene). Transformed bacteria were plated on LB medium containing ampicillin and X-gal. Ninety-six white colonies were picked and the plasmid inserts sequenced by the Iowa State University DNA Facility.

Sequences were analyzed to identify the *Mu* TIR and the adapter using Vector NTI (Invitrogen, http://www.invitrogen.com/) or the custom software we developed called MuTAlyzer (described below). Clone inserts corresponding in size to the fragment of interest represented candidates for the linked *Mu*. The sequence in between the *Mu* TIR and adapter represents the genomic sequence flanking the insertion site. Flanking sequences were used in BLAST (Altschul *et al.*, 1990) searches of maize sequence databases.

MuTAlyzer software

Custom software to automate selection of sequenced fragments where the sequence length matched fragment sizes of interest was coded in PHP. The tool is available online at http://www.lawrencelab.org/mutalyzer/. To use the tool, FASTA-formatted sequences are pasted into a window and the button marked 'Submit!' is pressed. Results are returned such that the insert size is reported, adaptors and *Mu* TIRs are color-coded, and the insert itself is in bold within the full context of the sequence.

Fragment verification

Database sequences were used to design gene-specific primers that were used in combination with the MuTIR3 primer to verify that the desired fragment had been recovered and was indeed linked to the mutation. The primers B2B-D11, C6D-A6, C8C-G10, M10B-E5, C11C-D3, and C12D-E3 (Table 4) were used with mutants 2, 6, 8, 10, 11, and 12, respectively.

ACKNOWLEDGEMENTS

This research was supported by funds from NSF Plant Genome Award 0077676, and from the Genetics, Development, and Cell Biology Department at Iowa State University. We thank Trent Seigfried (formerly with the USDA-ARS), for help with the sequence analysis software described herein and the Iowa State University DNA Facility for technical support.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Verification of genetic linkage.

Appendix S1. Sequence reads.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2008) GenBank. *Nucleic Acids Res.* **36**, D25–D30.

Briggs, S.P. and Beavis, W.D. (1994) How RFLP loci can be used to assist transposon-tagging efforts. In *The Maize Handbook* (Freeling, M. and Walbot, V., eds). New York: Springer-Verlag, pp. 653–659.

Brutnell, T.P. (2002) Transposon tagging in maize. *Funct. Integr. Genomics*, **2**, 4.

Das, L. and Martienssen, R. (1995) Site-selected transposon mutagenesis at the hcf106 locus in maize. *Plant Cell*, **7**, 287–294.

Dooner, H.K. and Belachew, A. (1989) Transposition Pattern of the Maize Element Ac from the Bz-M2(ac) Allele. *Genetics*, **122**, 447–457.

Frey, M., Stettner, C. and Gierl, A. (1998) A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). *Plant J.* **13**, 717–721.

Liu, Y.G., Mitsukawa, N., Oosumi, T. and Whittier, R.F. (1995) Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463.

McCarty, D.R., Settles, A.M., Suzuki, M. *et al.* (2005) Steady-state transposon mutagenesis in inbred maize. *Plant J.* **44**, 52–61.

Mitchelson, K.R. (2001) The application of capillary electrophoresis for DNA polymorphism analysis. *Methods Mol. Biol.* **162**, 3–26.

Patterson, G.I., Harris, L.J., Walbot, V. and Chandler, V.L. (1991) Genetic Analysis of B-Peru, a Regulatory Gene in Maize. *Genetics*, **127**, 205–220.

Robertson, D.S. and Stinard, P.S. (1987) Genetic Evidence of Mutator-Induced Deletions in the Short Arm of Chromosome 9 of Maize. *Genetics*, **115**, 353–361.

Robertson, D.S., Stinard, P.S. and Maguire, M.P. (1994) Genetic evidence of Mutator-induced deletions in the short arm of chromosome 9 of maize. II. wd deletions. *Genetics*, **136**, 1143–1149.

Settles, A.M., Latshaw, S. and McCarty, D.R. (2004) Molecular analysis of high-copy insertion sites in maize. *Nucleic Acids Res.* **32**, e54.

Settles, A.M., Holding, D.R., Tan, B.C. *et al.* (2007) Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics*, **8**, 116.

Vos, P., Hogers, R., Bleeker, M. *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414.

Walbot, V. and Warren, C. (1988) Regulation of Mu element copy number in maize lines with an active or inactive Mutator transposable element system. *Mol. Gen. Genet.* **211**, 27–34.

Wang, Y., Yin, G., Yang, Q., Tang, J., Lu, X., Korban, S.S. and Xu, M. (2008) Identification and isolation of Mu-flanking fragments from maize. *J Genet Genomics*, **35**, 207–213.

Yephremov, A. and Saedler, H. (2000) Technical advance: display and isolation of transposon-flanking sequences starting from genomic DNA or RNA. *Plant J.* **21**, 495–505.

GenBank sequence accessions: DV491867, CF631304, EB815794, DY241608, EB640046, DY623734, EE189213, FI495664 – FI496078.

