The Plant Journal (2000) **22**(4), 557–566

TECHNICAL ADVANCE

Identification of transposon-tagged genes by the random sequencing of *Mutator*-tagged DNA fragments from *Zea mays*

Steven Hanley[1], David Edwards[1], David Stevenson[1], Stephen Haines[1], Matthew Hegarty[1], Wolfgang Schuch[2] and Keith J. Edwards[1,*]

[1]IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol BS41 9AF, UK, and

[2]Zeneca Wheat Improvement Centre, John Innes Centre, Norwich NR4 7UH, UK

Received 27 March 2000; revised 25 May 2000; accepted 8 June 2000.
*For correspondence (fax +44 1275 394 281; e-mail keith.edwards@bbsrc.ac.uk).

Summary

We have used a universal adaptor amplification procedure to isolate random *Mutator*-tagged fragments from *Mutator*-active maize plants. Direct sequence characterization of 761 *Mutator*-tagged fragments indicated that a significant number were homologous to sequences within the public databases. The ability of *Mutator*-tagged fragments to detect homology was not related to the length of the sequence within the range 100–400 bp. However, fragments above this size did show an increased chance of detecting homology to either expressed sequence tags or genes. Characterization of the insertion sites of the *Mutator* elements suggested that while it does target transcribed regions, *Mutator* does not appear to have any site preference within the transcription unit. Hybridization of previously unidentified *Mutator*-tagged fragments to arrayed cDNA libraries confirmed that many of these also showed homology to transcribed regions of the genome. Examination of back-crossed progeny confirmed that all the insertions examined were germinal; however, in all but one case, selfing five individual *Mutator*-tagged lines failed to reveal an obvious phenotype. This study suggests that the random sequencing of *Mutator*-tagged fragments is capable of producing both a significant number of interesting transposon tagged genes and mutant plant lines, all of which could be extremely valuable in future gene discovery and functional genomics programmes.

Keywords: *Mutator*, transposon tagging, *Zea mays*, sequencing.

Introduction

The maize genome is relatively complex, presumably due to its ancient allotetraploid origins (Gaut and Doebley, 1997) and the presence of large numbers of both retro-elements and transposons (Edwards *et al.*, 1996; San Miguel *et al.*, 1996). The number and variety of transposable elements within the maize genome have greatly facilitated its use in insertional mutagenesis programmes, whilst its allotetraploid-based gene redundancy has allowed the characterization of mutants which may be lethal in a diploid species. These features, which limit the use of maize as a model for cereal genome structure, make it an eminently suitable model for the study of cereal gene function (O'Sullivan *et al.*, 2000).

Transposons were first identified in maize in the late 1940s by Barbara McClintock, as mobile pieces of genetic material acting as mutable elements (McClintock, 1950). Since then a variety of different elements have been characterized and used as molecular tools. Early studies primarily worked with *Ac* and *Spm* transposable element systems (Wessler, 1988).

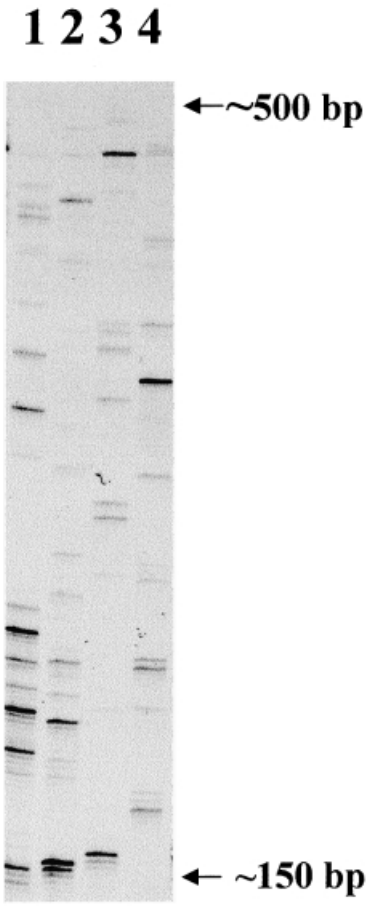

Figure 1. Analysis of amplified *Mu*-tagged border fragments on 5% denaturing polyacrylamide gels.
Mu-tagged DNA fragments were prepared from four independent *Mu*-active maize lines by digestion with *Mlu*I and *Mse*I, followed by ligation to both a *Mlu*I-biotinylated adaptor and a non-biotinylated *Mse*I adaptor. *Mu*-tagged fragments were selected by amplification with the ^{33}P-labelled *Mutator*-specific primer MuB, and an *Mse*I adaptor-specific primer. Fragments at the bottom of the gel have a size of approximately 150 bases and those at the top approximately 500 bases. Some fragments at the lower end of the gel are present as multiple bands (usually two). This is due to the incomplete addition of a T nucleotide to the 3′ end of the amplified fragment by the Taq DNA polymerase.

However, the *Mutator* (*Mu*) family of elements have since become the transposons of choice for maize gene mutagenesis, for several reasons: they are present in high copy number (10–100); they transpose via replication; and they appear to insert preferentially within and around genes, leading to a forward mutation frequency to an unlinked gene by up to 100 times greater than that observed with *Ac*/*Spm* elements (Walbot, 1992). *Mu*-based screens to identify transposon-tagged genes have until recently been phenotype-based, with large numbers of transposon-active plants being examined for an altered morphological or biochemical feature (McLaughlin and Walbot, 1987). Phenotype-based screens require the production of plants that are homozygous for specific *Mu*-insertions. Transposons linked genetically to the phenotype are then identified through successive generations, using molecular methods such as hybridization-based cloning and PCR. Although essentially random in its approach, and therefore requiring large numbers of plants, this procedure has led to the cloning and characterization of several genes of agronomic importance (Bennetzen *et al.*, 1987; Wise *et al.*, 1996) and is still a powerful tool with which to identify genes with novel phenotypes.

More recently, site-selected transposon mutagenesis (the so-called 'gene machine'; Meeley and Briggs, 1995) has allowed the identification of plants containing insertions within specific genes (Bensen *et al.*, 1995; Das and Martienssen, 1995). Because site-selected transposon mutagenesis is based upon genotype rather than phenotype, heterozygous *Mu* insertions can be detected, removing the need to generate homozygous plants. As an extension to the site-selected transposon mutagenesis procedure, Frey *et al.* (1998) reported that *Mu*-tagged fragments could be isolated from individual plants using PCR in conjunction with a modified adaptor-amplification protocol. This procedure has the advantage that a large number of *Mu*-specific fragments can be generated without any prior knowledge of the sequence that flanks the element. Our interest in generating a public functional genomics resource for maize led us to investigate whether the procedure developed by Frey *et al.* (1998) could be used to generate a large number of *Mu*-tagged fragments for direct sequencing. Sequence characterization of a large number of *Mu*-tagged fragments would not only provide a reservoir of *Mu*-tagged genes for use in future functional genomics studies, but would also help confirm or reject the hypothesis that *Mu* targets low-copy, gene-rich regions. In this report we describe both an improved method for generating large numbers of *Mu*-tagged fragments, and the results obtained from sequence characterization of several hundred *Mu*-tagged fragments. Our results suggest that the majority of *Mu* elements do insert into low copy-number, gene-rich regions of the maize genome, and a large number of these *Mu*-flanking regions show significant homology to known expressed sequence tags (ESTs) or genes.

Results and Discussion

Adaptor amplification using a modified protocol of Frey *et al.* (1998) produced from 35 to 50 *Mu*-tagged fragments per *Mu*-active plant. The fragments ranged in size from 100 to >500 bp (Figure 1). Using genomic DNA prepared from 55 plants, sequence information was generated for 761 *Mu*-tagged fragments. To generate these fragments we found it necessary to modify the protocol of Frey *et al.* (1998). Firstly we reduced the number of amplifications before electrophoresis from two to one, and secondly we

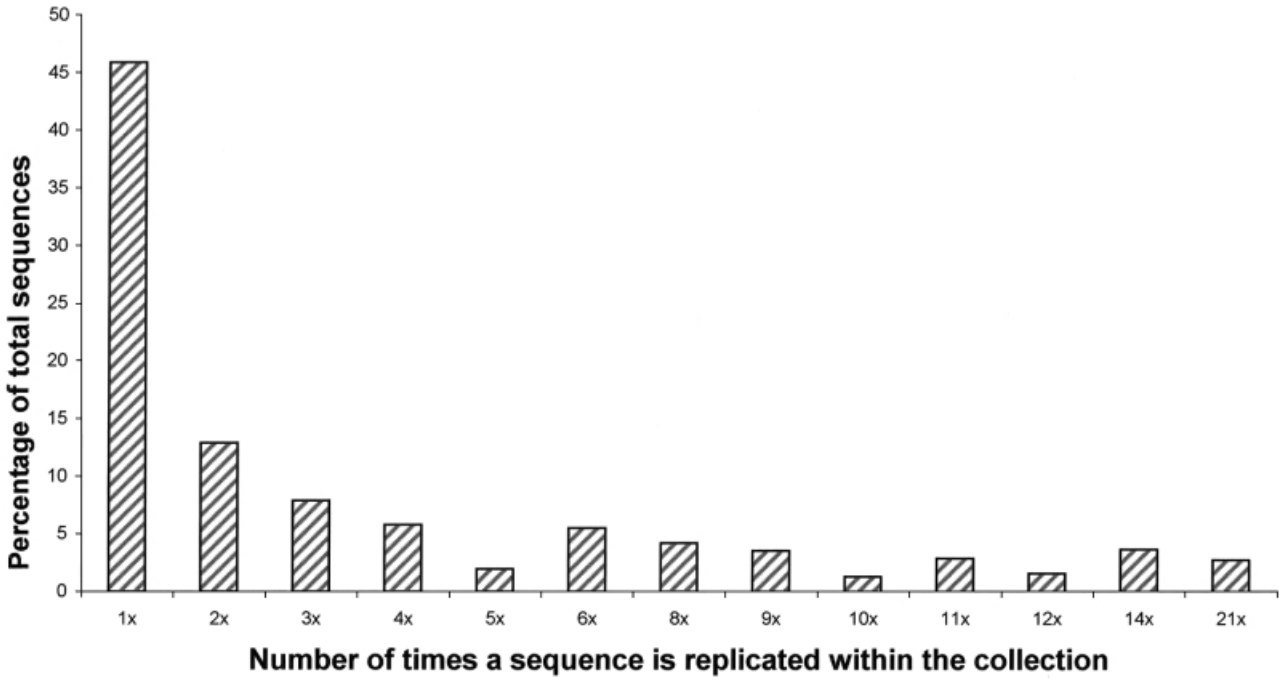

Figure 2. Representation of the number of *Mu*-tagged sequences present multiple times within the sequence collection.
The 761 *Mu*-tagged sequences generated as part of this program were compared with each other using the AUTOASSEMBLER (Perkin Elmer) software package. The number of times a sequence is replicated in the collection is presented as a histogram in which the number of times a sequence was found to be present were grouped with the numbers of sequences, as a percentage, within that category.

utilized the *Mlu*I restriction site present in the inverted repeats of the *MuDR*, *Mu1, Mu2* and *Mu8* elements, to preselect genomic DNA enriched for *Mu*-tagged fragments before amplification. In our hands, these modifications resulted in greater reproducibility of the *Mu*-tagged fragment profile. However, because the modified protocol relies on the presence of a *Mlu*I site in the inverted repeat, only *Mu*-tagged elements belonging to the *MuDR*, *Mu1*, *Mu2*, *Mu8* classes are amplified. Additionally, *Mlu*I is sensitive to CpG methylation and copies of *Mu* which have undergone methylation, and may therefore be inactive and will not produce *Mu*-tagged fragments by this procedure. In the work described here we chose to use *Mse*I as the four-base recognition restriction enzyme; however, using a modified adaptor, any four- or six-base restriction enzyme which does not recognize a sequence within the *Mu*-element terminal repeat could be used.

Characterization of the 761 sequences using the AUTOASSEMBLER program revealed that 389 (54%) were represented multiple times, with some sequences being represented 21 times (Figure 2). There are two reasons why specific sequences are represented multiple times. First, the nature of the *Mu* transposition event (replication rather than excision) leads to parental insertions being over-represented in the progeny when compared to new insertion events (Alleman and Freeling, 1986). Secondly, fragments within the size range amenable for both amplification and separation on denaturing polyacrylamide gels (50–600 bp) will be preferentially recovered. Removal of the duplicate sequences left 450 fragments, consisting of 350 unique sequences and 100 sequences chosen to represent those which occurred multiple times. A comparison between the size distribution of the unique and multiple sequences indicated that there was no significant difference between the two. The 450 sequences were used to screen the DNA and protein databases using both the BLASTN and BLASTX programs (Altschul *et al.*, 1997). Using a cut-off *e* value of 2e-04, a total of 64 (14.2%) and 48 (10.6%) sequences detected putative homologous sequences in the BLASTN and BLASTX searches, respectively. 24 (5.3%) of the sequences detected significant homologies with both BLASTN and BLASTX searches, meaning that 88 sequences (19.5%) of the 450 unique sequences could be tentatively identified. Table 1 contains a list of the *Mu*-tagged fragments which had significant homology to entries within the public sequence databases. Of those sequences which showed homology to entries within the public database via the BLASTN program, 56.2% (36 out of 64) were either maize ESTs (produced by the maize gene discovery programme) or rice ESTs (produced by various sequencing programmes; Sasaki, 1998; http://www.zmdb.iastate.edu/zmdb/EST_project.html). In addition to maize and rice ESTs, the BLASTN program detected homology to six entries related to either the *Mu5* transposon (sequence l47) or

Table 1. Mutator fragments with homology to entries within the public databases[a]

Sequence ID	Plant	BLASTN hits	*e* value	Size (bp)	Probable insertion site
A4	Mu3	*Zea mays* Rad51B	1e-07	241	5′ untranslated
A17	101107	C74482 Rice panicle >3 cm *Oryza sativa* cDNA	9e-23	317	Intron
A49	kje-1	*Z. mays* mRNA gs1-4 for glutamine synthetase	1e-67	241	Intron
C13	SH6	*Z. mays* globulin-1 gene, promoter region	6e-05	317	Promoter
E19	101206	C28153 rice callus cDNA *O. sativa* cDNA	4e-58	294	Intron
E38	101212	*Z. mays* cosmid IV.1E1 22 kDa zein protein 21	2e-08	384	Non-coding
E40	101212	*Sorghum bicolor* mitochondrial gene	1e-150	286	Coding
E48	101214	AI657302 *Z. mays* leaf primordia cDNA	9e-48	439	Coding
E50	101214	AI820302 *Z. mays* endosperm cDNA	4e-43	336	Coding
E51	101214	AW066018 *Z. mays* early embryo cDNA	2e-98	290	Coding
F28	105209	AI714888 *Z. mays* ear tissue cDNA	1e-108	331	Intron
G12	105317	AI947576.1 *Z. mays* stressed root cDNA	1e-146	289	Coding
GF-2	kje-1	AI944262 *Z. mays* root cDNA library	1e-138	279	5′ untranslated
GF-7	kje-1	AW000013 *Z. mays* root cDNA library	1e-56	304	Coding
H38	105125	*Hordeum vulgare L.* (Alexis) serine carboxypeptidase	2e-04	236	Coding
H42	105301	*Z. mays* retrotransposon Cinful-2	6e-11	284	Coding
H49	105302	AI665004 *Z. mays* endosperm cDNA	1e-116	407	Coding
I11	105306	*Arabidopsis thaliana* chromosome 5	4e-29	621	unknown
I40	105309	*Z. mays* copia-like retrotransposon Stonor	3e-36	674	Intron (amidase)
I42	105309	AI737372 *Z. mays* ear tissue cDNA	1e-77	394	Intron
I46	105309	*Z. mays* nitrate-induced NOI protein gene	3e-92	178	Promoter
I47	105310	*Z. mays* Mu5 transposable element	4e-69	510	Transposon
I55	105310	*Z. mays* ACCase//intron containing colonist 1 and 2	1e-19	564	Various retro-elements
I63	105312	Putative 26S proteasome subunit athMOV34	2e-09	582	5′ untranslated
I75	105313	AI691241 *Z. mays* ear tissue cDNA library	0.0	415	Coding
I87	105314	Chromosome 5, P1 clone[*Arabidopsis*]	4e-17	618	Non-coding
J23	105216	AI677607 *Z. mays* endosperm cDNA	3e-58	570	Outside existing EST
J31	105217	AI783084 *Z. mays* root cDNA library	1e-49	288	Intron
KE-6	kje-1	C72881 rice cDNA, partial sequence	3e-11	363	5′ untranslated
L37	105224	AI782966 *Z. mays* ear tissue cDNA	1e-28	93	Coding
L63	105103	*Z. mays* A1 gene for 40.1 kDa A1 protein	1e-13	168	Intron
L72	105104	Rice mRNA for ADP-ribosylation factor	2e-44	198	Coding
M7	105106	AU068551 *O. sativa* cDNA	2e-07	212	Coding
M76	105117	*Z. mays* zeta-carotene desaturase precursor	6e-28	148	Coding
N7	DE41	AI770851 *Z. mays* ear tissue cDNA	1e-16	137	Intron
N22	DE42	AI737406 *Z. mays* ear tissue cDNA	8e-36	101	Coding
P26	102304	I33112 Sequence 3 from patent US 5589610	3e-28	149	Promoter
P65	103115	AI665091 *Z. mays* endosperm cDNA	1e-46	282	Outside existing EST
Q44	103123	*Z. mays* PK1 gene//receptor-like protein kinase	1e-18	256	Intron
Q92	105309	C26026 Rice callus cDNA *O. sativa* cDNA	8e-07	235	Intron
R9	SH3	C24922 Rice green shoot *O. sativa* cDNA	4e-08	184	Coding
R27	101101	*Z. mays* triosephosphate isomerase 1 gene	2e-22	204	5′ untranslated
R42	101102	*Z. mays* clone Zm-Rab2-A GTP binding protein	3e-33	192	Coding
R54	101103	AI615233 *Z. mays* leaf primordia cDNA	3e-14	166	Intron
R59	101104	AI649598 *Z. mays* leaf primordia cDNA	2e-35	233	Coding
R75	101108	*Z. mays* retrotransposon Cinful-1	2e-71	161	Coding
S2	DE50	AU063869 *O. sativa* cDNA	1e-06	382	Outside existing EST
S19	104323	*Z. mays* triose phosphate/phosphate translocator	8e-30	426	5′ untranslated
S39	104319	AI712245 *Z. mays* endosperm cDNA	1e-19	315	Intron
S42	104319	AW129804 *Z. mays* mixed adult tissue cDNA	4e-18	213	Outside existing EST
S80	104315	Unknown protein [*A. thaliana*]	4e-41	311	Coding
R81	101108	RICS12715A rice green shoot *O. sativa* cDNA	9e-07	238	Intron
S92	103124	*Z. mays* eIF-5 gene, exons 1–2	8e-19	221	Coding
T25	103402	AI670571 *Z. mays* endosperm cDNA	2e-39	433	Outside existing EST
T44	103405	C20327 rice panicle at ripening stage cDNA	2e-30	302	Outside existing EST
T46	103405	AI770425 *Z. mays* ear tissue cDNA	9e-20	274	Outside existing EST
T53	103405	AI668519 *Z. mays* endosperm cDNA	7e-37	119	5′ untranslated
T87	103412	AI491543 *Z. mays* leaf primordia cDNA	1e-120	300	3′ untranslated
T90	103412	AW091134 *Z. mays* root cDNA library	1e-110	224	Coding

Sequence ID	Plant	BLASTN hits	*e* value	Size (bp)	Probable insertion site
U1	107313	*Z. mays* DNA for Fd VI, complete cds	5e-38	184	Coding
U14	107313	*Cratero-stigma plantagineum* tkt7 gene for transketolase	1e-06	120	Coding
U68	107307	*Z. mays* Ama gene, RNA polymerase	3e-06	219	Intron
U55	107308	AU057865 *O. sativa* cDNA	6e-12	243	Outside existing EST
2065-34	2065	AI770596 *Z. mays* ear tissue cDNA	8e-14	283	Intron
A4*	Mu3	*Z. mays* Rad51	2e-27	241	5′ untranslated
A49*	kje-1	*Z. mays* glutamine synthetase root isozyme 4	2e-20	241	Intron
C8	SH4	Putative protein [*A. thaliana*]	1e-10	228	Intron
E19*	101206	Putative protein [*A. thaliana*]	6e-10	294	Intron
E40*	101212	Hypothetical 267 kDa protein	4e-18	286	Coding
F28*	105209	Putative protein [*A. thaliana*]	4e-05	331	Intron
GF-2*	kje-1	O-methyltransferase [*Prunus dulcis*]	1e-10	279	5′ untranslated
H38*	105125	Serine carboxypeptidase II-3 precursor	9e-22	236	Coding
H42*	105301	Prpol [*Z. mays*]	6e-26	284	Coding
I8	105305	Unknown protein [*A. thaliana*]	9e-30	235	Coding
I10	105306	Peroxidase [*A. thaliana*]	1e-05	653	Intron
I12	105306	Similarity to ANK repeat region of Fowlpox virus	3e-16	567	Intron
I13	105306	Hypothetical protein [*A. thaliana*]	1e-32	516	Coding
I23	105307	Lysosomal Pro-X carboxypeptidase-like protein	5e-16	495	5′ untranslated
I40	105309	Putative amidase [*A. thaliana*]	9e-22	674	Intron
I63*	105312	Putative 26S proteasome subunit athMOV34	3e-20	582	5′ untranslated
I75*	105313	Strictosidine synthase [*A. thaliana*]	6e-22	415	Coding
KE14	SH3	Human alpha-mannosidase II	1e-11	326	Intron
KE18	kje-1	Reverse transcriptase [*Ginkgo biloba*]	2e-14	202	Coding
L63*	105103	Integral membrane protein [*Beta vulgaris*]	2e-04	168	Intron
L72*	105104	ADP-ribosylation factor homologue GTP-bp	4e-17	198	Coding
M7*	105106	Phosphatidylinositol-4-phosphate-5-kinase	4e-05	212	Coding
M28	105109	JM23 [*Homo sapiens*]	8e-09	367	5′ untranslated
M76*	105117	Zeta-carotene desaturase [*A. thaliana*]	2e-05	148	Coding
N22*	DE41	Dem [*Lycopersicon esculentum*]	3e-09	107	Coding
P35	102306	R30923-1 [*H. sapiens*]	6e-06	199	Coding
Q75	DE47	Receptor-like protein kinase 5 precursor	5e-07	347	Coding
Q92*	DE48	Putative amidase [*A. thaliana*]	1e-19	235	Intron
R9*	SH3	Contains similarity to protein phosphatase 2C	1e-21	184	Coding
R11	SH3	1-aminocyclopropane-1-carboxylic acid oxidase	9e-05	138	Coding
R42*	101102	*Z. mays* RAS-related protein RAB-2-A	7e-09	192	Coding
R56	101104	Putative monosaccharide transporter 1 [*Petunia*]	1e-05	239	Intron
R73	101105	Putative protein [*A. thaliana*]	1e-09	184	Intron
R75*	101108	Prpol [*Z. mays*]	2e-25	161	Coding
S18	104323	Unknown protein [*A. thaliana*]	5e-11	431	Intron
S19*	104323	*Z. mays* chloroplast triose phos-translocator	9e-12	426	5′ untranslated
S31	104321	Cinnamyl Co-A reductase [*A. thaliana*]	2e-05	243	Coding
S66	104317	Putative protein [*A. thaliana*]	1e-05	177	Coding
S79	104315	ATT20K18 *A. thaliana* DNA	6e-18	342	Coding
S80*	104315	*A. thaliana* chromosome II BAC T11A7	2e-04	311	Coding
S91	103124	Unknown protein [*A. thaliana*]	1e-41	315	Coding
T2	104324	ATT29A15 *A. thaliana* DNA	1e-34	462	Coding
T20	103401	T7N9.20 [*A. thaliana*]	1e-12	287	Intron
T90*	103412	Hypothetical protein [*Cicer arietinum*]	1e-29	224	Coding
U14*	107313	Transketolase 1 [*Capsicum annuum*]	7e-18	120	Coding
U43	107311	Copia-type pol polyprotein [*Z. mays*]	1e-44	156	Coding
U57	107308	Putative carbonyl reductase [*A. thaliana*]	4e-08	304	Intron
2065-34*	2065	Hypothetical protein [*A. thaliana*]	2e-08	283	Intron

[a]The latest version of this table can be found at http://www.maize.bbsrc.ac.uk/mutator/random_sequencing.html.
*Sequences which also have significant BLASTN hits.

sequences associated with retro-elements, including Stoner (sequence I40); Cinful 1 (sequence R75); Cinful 2 (sequence H42); copia-type pol polyprotein (U43); and reverse transcriptase (KE18). In all six cases the *Mu* transposon had inserted into a coding region, either directly within or closely associated with the retro-element. For

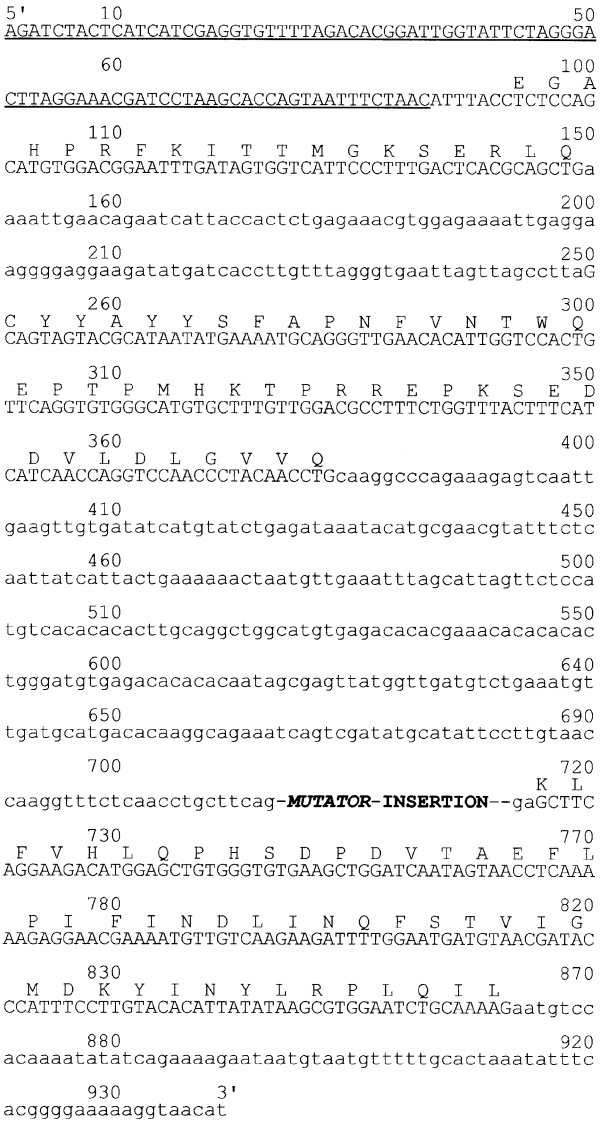

Figure 3. Characterization of the I40 and Q92 *Mu*-tagged sequences.
The position of the *Mu* element within the putative amidase gene and its relationship to the Stoner element are indicated. Sequences having a strong BLASTN homology (*e* value = 5e-36) to the Stoner element are underlined (bp 1–86). Known or suspected introns within the amidase gene are in lower case (bp 150–249, 378–715, 864–939). Known or suspected exons are represented in upper case, together with the amino-acid sequence. I40 sequences (bp 1–715) are 5′ to the indicated *Mutator* insertion, whereas Q92 sequences (bp 716–939) are 3′. In this example the non-coding strand is presented.

instance, in the case of both Cinful retro-elements the *Mu* transposon had inserted into the coding region of the *Prpol* gene (San Miguel *et al.*, 1996). At the amino-acid level, sequence I40 showed homology to a putative *Arabidopsis* amidase (accession number AC003028; unpublished results). Further examination of the I40 sequence showed that the *Mu* element had inserted into the putative amidase gene, 3′ of the insertion site of a Stoner retro-element (accession number AF082134). Further sequencing also identified the second flanking region of this insertion (sequence Q92). This second sequence confirmed that in this example, the *Mutator* element had inserted into an intron, close to the intron–exon boundary site (Figure 3). Given that retro-elements make up most of the high copy-number repetitive DNA of the maize genome (San Miguel *et al.*, 1996), it was surprising that <7% (six out of 88) of the *Mu*-tagged fragments showed homology to repetitive sequences such as retro-elements. There are two possible reasons for this observation: firstly, in our protocol the choice of restriction enzymes would preclude the recovery of methylated sequences such as those known to exist in highly repetitive regions; and secondly, it has been suggested that *Mu* targets low copy sequences in gene-rich regions (Cresse *et al.*, 1995). Either one, or a combination of, these two possibilities would lead to the results described here. However, our observation that in all six cases examined the *Mu* element had inserted into a transcribed region associated with the retro-element does suggest that, irrespective of the copy number characteristics of the target site, *Mu* preferentially inserts into transcriptionally active DNA.

In a previous study of 16 *Mu*-insertion sites, Cresse *et al.* (1995) identified a consensus sequence present at the *Mu*-insertion site. We repeated this study using all 450 unique sequences. Our results were in close agreement with the previous study (previous consensus: 5′-G-T-T-G-G/C-A-G-A/G-G-3′); however, our work suggests that the consensus sequence is more flexible than previously suggested (our consensus: 5′-G-T/C-T/C-T/G-G/C-A > T-G-A > G-*G* > C-3′).

To investigate if the size of the *Mu*-tagged fragment had an effect on its ability to identify homologous sequences, the size of the *Mu*-tagged fragments was plotted against their ability to detect homologous sequences within the public databases (Figure 4). This analysis suggested that for both BLASTN and BLASTX searches, fragments in the size range 50–400 bp had an approximately equal chance of detecting homologous sequences. However, for fragments >400 bp the data suggested that there was an increased probability of detecting homologous sequences using both BLASTN and BLASTX programs. To investigate this phenomenon further, the insertion sites for each of the 88 sequences identified in the BLASTN and BLASTX searches were examined. Given the limited availability of full-length maize gene sequences, it was only possible to determine the exact location of the *Mu* insertion for 80 of the 88 BLASTN /BLASTX hits (column 6 in Table 1). The various *Mu* insertions were then divided into those that were associated with genes but not translated (promoters, 3′, 5′, transcribed regions and introns); and those associated with translated regions (coding regions). This study showed that 44 (55%) of the insertions were within non-translated regions, and 36 (45%) were in translated

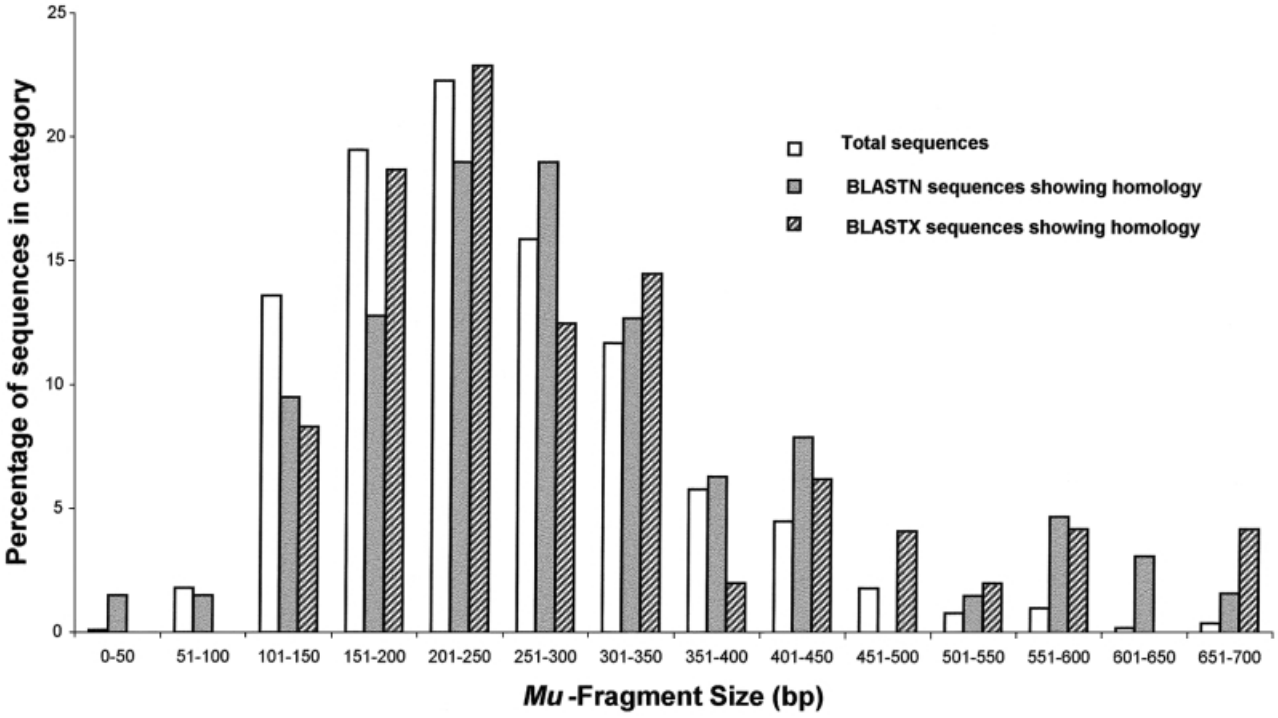

Figure 4. Comparison of the size of *Mu*-tagged fragments with their ability to detect homologous sequences in the public databases.
The 450 unique sequences generated in this program were compared to sequences within the Genbank and EMBL databases (as of December 1999) via both BLASTN and BLASTX programs, using the default values. The results are presented as a histogram in which the sizes of fragments were compared to their ability to detect significant homologies.

Table 2. Maize leaf cDNAs with homology to random *Mutator*-tagged fragments

EST sequence ID	BLASTX hits	*e* value
1	LRR-like protein [*Arabidopsis thaliana*]	1e-05
3	Adenylosuccinate lyase [*Escherichia coli*]	3e-35
4	Cell division protein [*Homo sapiens*]*	4e-44
5	ADP-ribosylation factor [*Zea mays*]*	8e-32
6	Photosystem I reaction centre subunit [*Z. mays*]	4e-16
7	Nicotianamine synthase [*H. vulgaris*]	7e-23
8 + 9	Glycine-rich RNA-binding protein [*Z. mays*]	4e-38
10	Hypothetical 15.9 kDa protein YE72_SCHPO	2e-06

*Control sequences with known homology.

regions. This observation suggests that whilst *Mu* may target genes, within genes it does not appear have any site preference. However, if the observation that *Mutator* preferentially targets genes is correct, then the majority of our *Mu*-tagged fragments should be homologous to genes. Yet this is contrary to our observation that only 19.5% of the *Mu*-tagged sequences show homology to known ESTs or genes. To confirm that a significant number of our unidentified *Mu*-tagged fragments were also derived from genes, 3000 arrayed maize leaf cDNA clones were hybridized with 20 randomly chosen, unidentified *Mu*-tagged fragments, and with two *Mu*-tagged fragments previously shown to share homology with either the maize ADP-ribosylation factor (L72) or the human JM23 protein (M28). Following autoradiography, 10 hybridizing clones were identified. Sequence analysis of these clones revealed nine unique sequences. Sequence alignment of the cDNAs and the corresponding *Mu*-tagged fragments showed that, while there was significant homology (ranging from 85 to 95% at the nucleotide level) between the two, in none of the nine cases examined was there 100% homology. This suggests that, while the cDNA

inserts and analogous *Mu*-tagged fragments probably represent members of the same gene family, they do not represent the same genes. When used to screen the public database via the BLASTX program, all the cDNA clones showed significant homology (*e* value >1e-05) to entries within the public databases (Table 2). One cDNA clone, clone 5, showed homology to a member of the maize ADP-ribosylation factor gene family, and another, clone 4, showed homology to the human cell-division protein, which itself has significant homology to JM23. While we had no previous information on the expression profile of the maize JM23 protein homologue, previous work by Regad *et al*. (1993) showed that plant ADP-ribosylation factor is expressed in leaf tissue. In all cases, except for the ADP-ribosylation factor and the JM83 sequences, the region of homology between the relatively large cDNA inserts (1–1.5 kb) and the sequences uncovered in the corresponding BLASTX screen did not overlap with the original *Mu*-tagged fragment. It was therefore not surprising that the original *Mu*-tagged fragment failed to identify the same sequences as the corresponding cDNA–BLASTX search. It was remarkable that the 20 random *Mu*-tagged fragments identified as many as seven (nine hybridizing clones minus two previously known clones) cDNA clones from the 3000 clones screened. We believe that this result is possibly due to the ability of the *Mu*-tagged fragments to identify (via hybridization) all the members of their respective gene families which have >85% nucleotide sequence homology. However, it is also possible that, considering the relatively small number of fragments used, screening the array with a larger and more diverse population of *Mu*-tagged sequences might produce fewer hits per fragment. Whatever the reason for the relatively large number of hybridizing clones, our results suggest that screening further cDNA clones with *Mu*-tagged fragments could identify even more homologies. Taken together with the results from the direct sequencing, these observations suggest that eventually most of the *Mu*-tagged fragments identified in this study will be confirmed as being derived from, or close to, transcribed sequences.

A fundamental principle of transposon tagging is that insertion of the element into a gene results in an inheritable phenotype. Somatic insertions have been identified as a significant problem with transposon tagging based on the *Mutator* system (Qin *et al*., 1991). To confirm the status of the *Mu* insertion described here, putative heterozygotes for insertions A49 (glutamine synthetase 1–4); KE14 (alpha mannosidase II); M76 (zeta-carotene desaturase precursor); Q75 (receptor-like protein kinase 5 precursor); and U14 (transketolase) were back-crossed with the inbred line B73 for four generations. In all cases, B73 was used as the pollen source. At each generation, the transmission of the specific *Mu* insertion was monitored via PCR using an insertion-specific primer and the MuB primer. In all cases, *Mu*-tagged fragments were transmitted at or near the expected frequency (50 : 50; data not shown). These observations confirmed that the original *Mu* insertions were germinal in origin. We believe that the reason lies in our strategy of isolating individual *Mu* bands from denaturing polyacrylamide gels. We believe that distinct *Mu* bands isolated from the gel were more likely to be germinal insertions, whereas somatic insertions were more likely to be represented by the faint background smear present on the polyacrylamide gel.

Following four generations of back-crossing to B73, amplification-positive plants were selfed and the progeny examined for any obvious phenotype. From the five separate insertions examined, only M76 (zeta-carotene desaturase precursor) had an obvious phenotype. As previously described, this zeta-carotene desaturase *Mu*-tagged line produced weakly viviparous kernels. Further characterization of this line has confirmed that it represents a new allele of the *vp5* phenotype (Wurtzel, 1992).

The method for generating *Mu*-tagged DNA fragments described here is relatively simple to perform. Our results suggest that it is amenable to being scaled up to include several thousand *Mu*-active maize plants. Significantly, our results also suggest that the majority of these sequences will represent germinal insertion events. Sequence analysis of such a large number of germinally derived *Mu*-tagged fragments, derived from independent plants, has the potential to reduce the bottleneck in the elucidation of gene function.

Experimental procedures

Growth and maintenance of maize plants

Mutator active maize seed in a W22 background was a kind gift of Dr Martienssen of the Cold Spring Harbor Laboratory. Plants were grown as described by Neuffer (1993). *Mu*-active plants were maintained as heterozygote families by random mating within the *Mu*-active pool. Southern blot hybridization (Southern, 1975) of DNA prepared from this material indicated that individual plants contained between 35 and 50 *Mu* elements when probed with a ^{32}P-labelled *Mu*-inverted repeat (Maniatis *et al*., 1982).

Isolation and characterization of Mu*-tagged DNA fragments*

Mu-tagged fragments were isolated by a modification of the method of Frey *et al*. (1998). Briefly, 100 ng of total genomic DNA was digested with 2 units of *Mlu*I (a restriction enzyme with the six-base recognition sequence A_CGCGT) and 4 units of *Mse*I (a restriction enzyme with the four-base recognition sequence T_TAA) in a volume of 50 µl for 1 h at 37°C. Following the restriction digestion, 50 ng of a biotinylated *Mlu*I adapter (consisting of a biotinylated 17mer: biotin-5′-CTCGTAGACTGCGTACC-3′ and a complementary 15mer: 5′-CGCGGTTACGCAGTC-3′) and 50 ng of a non-biotinylated *Mse*I adaptor (consisting of a 16mer: 5′-GACGATGAGTCCTGAG-3′ and a complementary

14mer: 5′-TACTCAGGACTCAT-3′), together with 5 μl 10 mM ATP and 1 unit T_4 DNA ligase, were added to the digestion mix to make a total volume of 60 μl. The ligation was allowed to proceed for 2 h at 37°C. Genomic fragments linked to the *Mlu*I adaptors were isolated and purified using 10 μl Dynal (NY, USA) magnetic streptavadin beads (Tong and Smith, 1992.). Following purification (Mathes *et al*., 1998;), a 1 μl aliquot of the DNA–bead complex was subjected to amplification using a [^{33}P]5′-labelled primer specific for a conserved region of the *Mutator* inverted repeat (5′-CAGAATTCCATAATGGCAATTATCTC-3′) and a primer specific for the *Mse*I adaptor (5′-GACGATGAGTCCTGAGTAA-3′) in 25 μl with 1 unit of Taq DNA polymerase, using the cycling conditions previously described by Mathes *et al*. (1998). Following amplification the entire reaction was electrophoresed in a 5% denaturing polyacrylamide gel. Individual bands were visualised via autoradiography and those >50 bp excised and eluted as described by Maniatis *et al*. (1982). Individual bands were then re-amplified using the original *Mlu*I and *Mse*I primers, and subjected to direct sequencing using the Perkin Elmer (Cheshire, UK) Applied Biosystems BIG DYE terminator kit and the original *Mse*I-specific primer.

Screening plasmid cDNA libraries with Mu*-tagged fragments*

50 ng each of 20 randomly chosen *Mu*-tagged fragments which did not have homology to sequences within the databases, one *Mu*-tagged fragment with homology to the maize ADP-ribosylation factor, and one *Mu*-tagged fragment with homology to the human cell-division protein JM23, were pooled and labelled with α[^{32}P]dCTP according to the method of Feinberg and Vogelstein (1983). The labelled fragments were used to probe 3000 arrayed colonies containing plasmid with cDNAs derived from maize leaf mRNA. Following hybridization at 65°C in 6 × SSC, 5 × Denhardt's and 2% SDS for 16 h, the filters were washed three times at 65°C in 0.2 × SSC, 1% SDS. The washed filters were subjected to autoradiography with a single intensifying screen for 3 days at –70°C. Hybridizing clones were purified to single colonies and used to prepare plasmid DNA. Purified plasmids were sequenced using the Perkin Elmer Applied Biosystems BIG DYE terminator kit and the universal forward M13 primer.

Database screening

Primer and vector sequences were removed using AUTOASSEMBLER (Perkin Elmer). Corrected sequences were screened against all of the GenBank and EMBL databases using both BLASTN and BLASTX programs (Altschul *et al*., 1997). Both BLASTN and BLASTX programs were used with the default settings. Only homologies with an *e* value >2e-04 were recorded as being significant.

Acknowledgements

IACR-Long Ashton receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. S. Hanley was funded by a Zeneca Plant Science award to K.J.E.; S. Haines was funded by a fruit and cider studentship in collaboration with Advanta; M. Hegarty was funded by an Industrial CASE studentship with Zeneca Plant Sciences; D. Stevenson was funded by a BBSRC Agri-Foods award; and D. Edwards was funded by the Framework IV programme ZEROPA. Seed for the various *Mu* insertions described here are available for research applications by writing to Keith J. Edwards.

References

Alleman, M. and Freeling, M. (1986) The *Mu* transposable elements of maize: evidence for transposition and copy number regulation during development. *Genetics*, **112**, 107–119.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, E.W. and Lipman, D.L. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.

Bennetzen, J.L., Cresse, A., Brown, W.E. and Lee, L. (1987) Molecular cloning of maize genes by transposon tagging with *Mutator*. In *Plant Gene Systems and their Biology, UCLA Symposia on Molecular and Cellular Biology New Series*, Vol. 62 (Key, J.L. and McIntosh, L., eds). New York: Alan R. Liss, pp. 183–204.

Bensen, R.J., Johal, G.S., Crane, V.C., Tossberg, J.T., Schnable, P.S., Meeley, R.B. and Briggs, S.P. (1995) Cloning and characterization of the maize *An1* gene. *Plant Cell*, **7**, 75–84.

Cresse, A.D., Hulbert, S.H., Brown, W.E., Lucas, J.R. and Bennetzen, J.L. (1995) *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics*, **140**, 315–324.

Das, L. and Martienssen, R.A. (1995) Site-selected mutagenesis at the Hcf106 locus in maize. *Plant Cell*, **7**, 287–294.

Edwards, K.J., Veuskens, J., Rawles, H., Daly, A. and Bennetzen, J.L. (1996) Characterization of four dispersed repetitive DNA sequences from *Zea mays* and their use in constructing contiguous DNA fragments using YAC clones. *Genome*, **39**, 811–817.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.

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.

Gaut, B.S. and Doebley, J.F. (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl Acad. Sci. USA*, **94**, 6809–6814.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory.

Mathes, M., Daly, A. and Edwards, K.J. (1998) Amplified fragment length polymorphisms. In *Molecular Tools for Screening Biodiversity: Plants and Animals* (Karp, A., Isaac, P.G. and Ingram, D.S., eds). London: Chapman & Hall, pp. 181–192.

McClintock, B. (1950) The origin and behaviour of mutable loci in maize. *Proc. Natl Acad. Sci. USA*, **36**, 344–355.

McLaughlin, M. and Walbot, V. (1987) Cloning of the *Bronze2* locus of maize by transposon tagging and differential hybridisation. *Genetics*, **117**, 771–776.

Meeley, R.B. and Briggs, S.P. (1995) Reverse genetics for maize. *Maize Genet. Newsl.* **69**, 67–82.

Neuffer, M.G. (1993) Growing maize for genetic studies. In *The Maize Handbook* (Freeling, M. and Walbot, V., eds). New York: Springer Verlag, pp. 197–208.

O'Sullivan, D., Edwards, D. and Edwards, K.J. (2000) Maize Genomics. *Agbiotechnet*, **2**, 1–6.

Qin, M., Robertson, D.S. and Ellinboe, A.H. (1991) Cloning of the *Mutator* transposable element *MuA2*: a putative regulator of somatic mutability of the *a1-Mum2* allele in maize. *Genetics*, **129**, 845–854.

Regad, F., Bardet, C., Tremousaygue, D., Moisan, A., Lescure, B. and Axelos, M. (1993) cDNA cloning and expression of an Arabidopsis GTP-binding protein of the ARF family. *FEBS Lett.* **316**, 133–136.

San Miguel, P., Tikhonov, A., Jin, Y.K. *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.

Sasaki, T. (1998) The rice genome project in Japan. *Proc. Natl Acad. Sci. USA*, **95**, 2027–2028.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.

Tong, X. and Smith, L.M. (1992) Solid phase method for the purification of DNA sequencing reactions. *Anal. Chem.* **64**, 2672–2677.

Walbot, V. (1992) Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Annu. Rev. Plant Phys. Plant Mol. Biol.* **43**, 49–82.

Wessler, S.R. (1988) Phenotypic diversity mediated by the maize transposable elements Ac and Spm. *Science*, **242**, 399–405.

Wise, R.P., Dill, C.L. and Schnable, P.S. (1996) *Mutator*-induced mutations of the *rf1* nuclear fertility restorer of T-cytoplasm maize alter the accumulation of *T-urf13* mitochondrial transcripts. *Genetics*, **143**, 1383–1394.

Wurtzel, E.T. (1992) Use of a Ds chromosome-breaking element to examine maize *vp5* expression. *J. Hered.* **83**, 109–113.