Maize *Mu* Transposons Are Targeted to the 5′ Untranslated Region of the *gl8* Gene and Sequences Flanking *Mu* Target-Site Duplications Exhibit Nonrandom Nucleotide Composition Throughout the Genome

Charles R. Dietrich,*,† Feng Cui,†,‡ Mark L. Packila,§ Jin Li,†, Daniel A. Ashlock,†† Basil J. Nikolau‡‡,§§ and Patrick S. Schnable*,†,‡,§,**,§§,1**

*Interdepartmental Plant Physiology Program, †Department of Zoology and Genetics, ‡Bioinformatics and Computational Biology Program, **Interdepartmental Genetics Program, ††Department of Mathematics, ‡‡Department of Biochemistry, Biophysics and Molecular Biology, §Department of Agronomy, and §§Center for Plant Genomics, Iowa State University, Ames, Iowa 50011

Manuscript received October 19, 2001
Accepted for publication November 8, 2001

ABSTRACT

The widespread use of the maize *Mutator* (*Mu*) system to generate mutants exploits the preference of *Mu* transposons to insert into genic regions. However, little is known about the specificity of *Mu* insertions within genes. Analysis of 79 independently isolated *Mu*-induced alleles at the *gl8* locus established that at least 75 contain *Mu* insertions. Analysis of the terminal inverted repeats (TIRs) of the inserted transposons defined three new *Mu* transposons: *Mu10*, *Mu 11*, and *Mu12*. A large percentage (>80%) of the insertions are located in the 5′ untranslated region (UTR) of the *gl8* gene. Ten positions within the 5′ UTR experienced multiple independent *Mu* insertions. Analyses of the nucleotide composition of the 9-bp TSD and the sequences directly flanking the TSD reveals that the nucleotide composition of *Mu* insertion sites differs dramatically from that of random DNA. In particular, the frequencies at which C's and G's are observed at positions −2 and +2 (relative to the TSD) are substantially higher than expected. Insertion sites of 315 *RescueMu* insertions displayed the same nonrandom nucleotide composition observed for the *gl8-Mu* alleles. Hence, this study provides strong evidence for the involvement of sequences flanking the TSD in *Mu* insertion-site selection.

ABOUT a dozen families of maize transposons have been identified (Nevers *et al.* 1985; Peterson 1988; Capy *et al.* 1998). Each family consists typically of two categories of transposons, autonomous and nonautonomous. Autonomous transposons encode all nonhost factors required for their own transposition. In contrast, the transposition of nonautonomous transposons is dependent upon factors encoded by autonomous transposons of the same family. The *Mutator* (*Mu*) family consists of the autonomous transposon *MuDR* (Schnable and Peterson 1986; Hershberger *et al.* 1991; Chomet *et al.* 1991; Qin *et al.* 1991; Hsia and Schnable 1996) and at least seven classes of nonautonomous transposons: *Mu1*, *Mu2*, *Mu3*, *Mu4*, *Mu5*, *Mu7/rcy*, and *Mu8* (for reviews see Walbot 1991; Chandler and Hardeman 1992). *Mu* transposons contain ~215-bp terminal inverted repeats (TIRs) that are highly conserved and are thought to be recognized by the *MuDR*-encoded transposase (Benito and Walbot 1997; Raizada and Walbot 2000). In addition to these *Mu* transposons, a novel *Mu* TIR (GenBank accession no. AF231940), recently identified as part of a *Mu* insertion in the *rf2* gene, was defined as the left TIR of *Mu10* (X. Cui, A. Hsia, D. A. Ashlock, R. P. Wise and P. S. Schnable, unpublished data).

Some transposons exhibit nonrandom patterns of insertion. For example, miniature inverted repeat transposable elements such as the *Tourist* and *Stowaway* described by Bureau and Wessler (1992, 1994), have a preference for insertion into genic regions (Zhang *et al.* 2000). In addition, they exhibit a preference for 2- to 3-bp A/T-rich target sites. Transposons and insertion sequences from bacteria show a wide range of insertion specificity ranging from recognition of specific target sequences to preferences for insertion into A/T-rich sequences of promoter regions (Berg and Howe 1989). The most dramatic example of nonrandom insertion within genes is the preference of the *P* element of Drosophila to insert into the 5′ untranslated regions (UTRs) of genes (reviewed by Spradling *et al.* 1995). Spradling compiled data for 56 insertions (from 49 different genes) for which sequence data were available and thereby demonstrated that all the insertions had occurred in the 5′ halves of the affected genes and over one-half of all the insertions had occurred in the 5′ UTRs.

As expected, *Mu* insertions that are responsible for

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF348367, AF348368, AF348369, AF348370, AF302098, AF247740, AF247741, AF247742, AF302101, and AF302099.

[1] *Corresponding author:* Department of Agronomy, Iowa State University, Ames, IA 50011. E-mail: schnable@iastate.edu

mutations are located in genes—usually in exons, but in some cases in noncoding regions (reviewed by Bennetzen 1996). However, even *Mu* transposons that were not preselected for being responsible for a mutation seem to be preferentially located in gene-like, low copy, hypomethylated regions of the genome. No obvious target site or secondary structure has been identified to explain this preference (Bennetzen *et al.* 1993). More recently, PCR-based techniques such as amplification of insertion mutagenized sites (Frey *et al.* 1998) and the use of a transgenic modified *Mu1* transposon, *RescueMu* (Raizada *et al.* 2001), have allowed for the large-scale isolation of DNA sequences that flank random *Mu* insertions. Analyses of such fragments have revealed that these sequences have significant levels of nucleotide identity to expressed sequence tags (ESTs) at frequencies higher than would be expected for random genomic DNA (J. Vogel, personal communication; Hanley *et al.* 2000; Raizada *et al.* 2001). This preference for *Mu* transposons to insert into genes, combined with their high transposition rate (Alleman and Freeling 1986), results in mutation rates that are 50-fold higher than the spontaneous rate (Robertson 1978) and makes *Mu* a powerful tool for tagging and cloning genes (for reviews see Bennetzen *et al.* 1987; Shepherd *et al.* 1988; Walbot 1992).

Despite the widespread use of *Mu* transposons for gene cloning, relatively little is known of *Mu* transposon insertion preference within genes. There is evidence to suggest *Mu* transposons insert nonrandomly within at least some genes (reviewed by Bennetzen *et al.* 1993). For example, the mapping of 24 *bz-Mu* alleles via DNA gel blotting revealed that as many as 19 are clustered into an ~600-bp region around intron 1 (Taylor *et al.* 1986; Brown *et al.* 1989; Hardeman and Chandler 1989, 1993; Schnable *et al.* 1989; Britt and Walbot 1991; Doseff *et al.* 1991). Of these *bz1* alleles, 5 were subsequently sequenced and 4 were found to have *Mu* insertions in a narrow region just 3′ of intron 1 (Schnable *et al.* 1989; Britt and Walbot 1991; Doseff *et al.* 1991; Chandler and Hardeman 1992). Indeed, three of these independent insertions occurred at exactly the same nucleotide position. Similarly, the 4 characterized *adh1-Mu* alleles have insertions in the first intron (Barker *et al.* 1984; Rowland and Strommer 1985; Chen *et al.* 1987). Analysis of a collection of *Mu*-induced dominant alleles at the *kn1* locus revealed that all 9 have *Mu* insertions within a 310-bp region of the third intron (Greene *et al.* 1994). In addition, it has been suggested that *Mu* transposons may have a preference for insertion into promoters on the basis of the fact that a number of insertions have been isolated from promoter regions (Bennetzen *et al.* 1993; Bennetzen 1996).

Although some evidence suggests that *Mu* transposons may insert at preferred sites within genes, to date this hypothesis has been tested only via the analysis of relatively few mutant alleles that were generated from multiple, and often unrelated, *Mu* stocks. It is therefore difficult to draw firm conclusions regarding the specificity of *Mu* insertions from the extant data. In the current study, each member of a large collection of *Mu*-induced *glossy8* (*gl8*) alleles generated from genetically related *Mu* stocks was characterized by PCR amplification to determine whether the locus was disrupted by a *Mu* insertion. Subsequently, sequence analyses of the resulting PCR products established the exact *Mu* insertion sites in 75 of the 79 *gl8-Mu* alleles. These data demonstrate that *Mu* transposons have a strong preference for inserting into the 5′ UTR of the *gl8* gene. Analysis of sequences flanking the 9-bp TSD has revealed a highly significant conservation of nucleotide composition in the positions directly flanking the 9-bp TSD. Analysis of insertion sites from 315 *RescueMu* transposons demonstrated that the nonrandom nucleotide composition flanking the *gl8-Mu* insertion sites is not unique to insertions in the *gl8* gene.

MATERIALS AND METHODS

Genetic stocks: The *Mu* transposon stocks used to generate 75 of the *Mu*-tagged *gl8* alleles have been described previously (Stinard *et al.* 1993). Of the remaining alleles, *gl8-Mu 94-1480-26* and *gl8-Mu 94-1641-25* were isolated from a different *Mu* stock maintained by the Schnable laboratory, and *gl8-Mu 90-2940-A* and *gl8-Mu 90-3230-5* were isolated from F_2 *Mu* families provided by S. Briggs, when he was at Pioneer Hi-Bred International (Johnston, IA). Alleles *gl8-Mu 89g-5348-24* and *gl8-Mu 91-2145* (Schnable *et al.* 1994) were found to have *Mu8* insertions at the exact site and orientation as the *Mu8* insertion allele *gl8-Mu 88-3142*. Because of the way the random-tagged alleles *gl8-Mu 89g-5348-24* and *gl8-Mu 91-2145* were isolated, it was not possible to confirm their independence from *gl8-Mu 88-3142* and they were therefore not included in this study.

The *gl8* locus was originally defined by a spontaneous mutation (Emerson *et al.* 1935) that was designated as the reference allele (*gl8-ref*). The *gl8-ref pr* stock used in the directed-tagging experiment was provided by D. Robertson (Iowa State University; Schnable accession no. 552). The Aet/LineC genetic stock has the genotype *Gl8 pr/Gl8 pr*.

All inbred lines were maintained by selfing and/or sib mating. The inbred lines Q66 and Q67 (Schnable accession nos. 111 and 113, respectively) were originally obtained from A. Hallauer (Iowa State University). The inbred lines B77 and B79 (Schnable accession nos. 403 and 404, respectively) were originally obtained from D. Robertson (Iowa State University). The inbred line W64A (Schnable accession no.142) was provided by D. Pring.

Isolation and sequencing of the *gl8* genomic clones: A B73 genomic library constructed in λDash II (Stratagene, La Jolla, CA) by J. Tossberg (Pioneer Hi-Bred International) was screened by DNA hybridization (Maniatis *et al.* 1982). A partial 0.8-kb *gl8* cDNA was used as a probe in the initial library screen. In subsequent λ purification steps, a 140-bp fragment isolated from the 5′ end of the 1.4-kb apparent full-length *gl8* cDNA clone pgl8 (Xu *et al.* 1997) was used as the probe (probe A, Figure 1). The *gl8* genomic clone, λ1512-38, was isolated and determined to contain a 6.8-kb *Hin*dIII fragment containing the entire *gl8* gene. This 6.8-kb *Hin*dIII fragment was sequenced at the Iowa State University Nucleic Acid Facility on

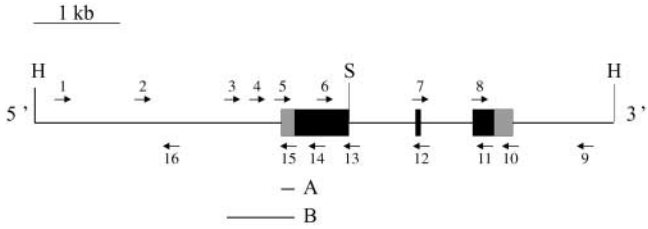

FIGURE 1.—*gl8* gene structure. Coding regions are represented as solid boxes and 5′ and 3′ UTRs are shown as shaded boxes. The approximate positions and orientation of 16 *gl8*-specific primers used to characterize *gl8-Mu* alleles are indicated as arrows on a 6.8-kb *Hin*dIII fragment isolated from the B73 genomic clone λ1512-38. Positions of probes A and B are shown below the genomic clone. Primers are designated as follows: 1, gl8a54; 2, gl8a61; 3, gl8a58; 4, gl8a62; 5, 8a2840; 6, gab457; 7, gab812; 8, g24he.p4; 9, 8a2637; 10, gl8ain; 11, gab869; 12, gab830; 13, xx022; 14, mcd696; 15, gl8a51; and 16, gl8a59. Indicated restriction sites: H, *Hin*dIII; S, *Sac*I.

an ABI 373A automated DNA sequencer (Applied Biosystems, Foster City, CA). Sequence analyses were performed using the Sequencher Version 3.0 software package (Gene Codes, Ann Arbor, MI). The 3.6-kb *Hin*dIII/*Sac*I fragment from the 5′ half of the *gl8* genomic clone λ1512-38 was subcloned into pBSK to make the clone pgl83.6. Probe B (Figure 1) was obtained by PCR amplification of pgl83.6 with primers gl8a58 and gl8a51.

Isolation of genomic DNA: F_2 families segregating for *gl8-Mu* alleles were grown in greenhouse sand benches for 7 days, at which time individual glossy plants (*gl8-Mu/gl8-Mu*) were identified by the "water-beading" phenotype (SCHNABLE *et al.* 1994). DNA was extracted from each of these plants using a version of the ROGERS and BENDICH (1985) hexadecyltrimethyl-ammonium bromide (CTAB) extraction protocol modified to allow a substantially higher throughput (eight 96-well plates can easily be completed in a single work day). In the high-throughput version of the protocol, sections of seedlings with fully expanded leaves one and two were harvested from just above the tip of the coleoptile to the top of the first leaf sheath. The first leaf sheath was removed from this section and the remaining tissue was inserted into 2-ml strip tubes in 96-well format and freeze dried. Approximately 20 1.7- to 2.5-mm glass beads (MO-Sci Corporation, Rolla, MO, catalog no. MS-302/GL-01915) were added to each tube and dried tissue was pulverized for 5 min using a paint shaker (Red Devil Equipment, Brooklyn Park, MN, model no. 5400). A total of 600 μl of CTAB extraction buffer [1% CTAB, 100 mM Tris (pH 7.5), 0.7 M NaCl, 10 mM EDTA, 100 mM 2-mercaptoethanol] was added to the resulting powder and incubated at 60° for 60 min. Samples were allowed to cool to room temperature before adding 300 μl chloroform/octanol (24:1). Tubes were then mixed by inverting for 5 min and centrifuged at 5000 × *g* for 10 min. The aqueous phase was withdrawn and mixed with an equal volume of isopropanol to precipitate the PCR-ready DNA.

For 11 of the *gl8-Mu* alleles derived from a directed-tagging experiment, F_2 seed was not available. For these alleles, 10 individual seedlings from crosses such as cross 2 or cross 3 (see RESULTS) were pooled and DNA was extracted via the method of DELLAPORTA *et al.* (1983). DNA samples from the maize inbred lines Q66, Q67, B77, and B79 and a stock homozygous for the *gl8-ref* allele were extracted according to methods of SAGHAI-MAROOF *et al.* (1984).

***Mu* transposons:** TIR sequences for *Mu1*, *Mu3*, *Mu4*, *Mu5*, *Mu7/rcy*, *Mu8*, and *MuDR* were obtained from GenBank accession nos. X13019, U19613, X14224, X14225, X15872, X53604, and M76978, respectively. The left and right TIRs were defined as the TIRs that were listed first and second, respectively, in the appropriate GenBank entry. The TIRs of *Mu2* were obtained from TAYLOR and WALBOT (1987). The single TIR from *Mu10* was obtained from GenBank accession no. AF231940.

PCR amplification of *Mu*-flanking regions: PCR was performed using a primer in the conserved region of the *Mu* TIR, primer *Mu*-TIR (5′ AGA GAA GCC AAC GCC A(AT)C GCC TC(CT) ATT TCG TC 3′), in combination with individual *gl8*-specific primers to amplify the *gl8* sequences flanking each *Mu* insertion. PCR amplification reactions were performed with a PTC-200 (MJ Research, Waltham, MA) thermal cycler with the following conditions: denature at 94° for 1 min, anneal at 62° for 1.5 min, and extend at 72° for 2 min for 40 cycles, followed by a final extension at 72° for 5 min. PCR products obtained by amplification with *Mu*-TIR and a *gl8*-specific primer upstream or downstream of the *Mu* transposon within the *gl8* gene were called 5′ or 3′ products, respectively. To amplify the 3′ product of *Mu10* and *Mu12* insertions, it was necessary to lower the annealing temperature to 55°. To compensate for the high GC content in the 5′ half of exon 1 of the *gl8* gene, DMSO was added to a final concentration of 10% for PCR reactions involving primers xx022, 8a2840, or mcd696. PCR products were purified using a QIAGEN (Valencia, CA) PCR purification kit (catalog no. 28104) and sequenced. PCR reactions were performed using the following *gl8*-specific primers. The approximate location of each primer is shown in Figure 1.

gl8a54: 5′ GCC ACC CGG ACT AAA ACC TG 3′
gl8a59: 5′ TAA TGG CCT CGC TGT CAC 3′
gl8a61: 5′ AGC AGC AGC GAT CAC CTC AG 3′
gl8a51: 5′ TGT GCC TGC CCC TGT GTC 3′
gl8a58: 5′ AAG AGT GTG GCG CGT GCT ATG 3′
gl8a62: 5′ AAG TGA GAA AGA AAG GTT GTC C 3′
gl8a64: 5′ TTT CGA ATA TTT GTC CTA CTG TTA G 3′
8a2840: 5′ CCA CCC ACC ACC GGA TAT AGG TCA TG 3′
mcd696: 5′ CGC ACC TCG GGG ACC TTG G 3′
xx022: 5′ CGG ATC AGA AGG CAC GAC GGA G 3′
gab457: 5′ GGT GGA CGA GGA GCT GAT G 3′
gab830: 5′ CAT TGC ACA TCA ATA CCC TTG CTC TTG TAC TC 3′
gab812: 5′ TCA AGA TGC CTC TAT GTT GAG TAC AAG AGC AAG 3′
gl8ain: 5′ CTC AGG AGG TAA TGG TAG 3′
gab869: 5′ GCC AGC CCC TTC TTG CGG ATC TTA ATG 3′
g24he.p4: 5′ CCT ATG CTC GTG CTG CCG TTC GTC 3′
8a2637: 5′ GTG GCG ACA AAG CTT GCA TCT ATC AGG AAG TCT 3′

Isolation of the 5′ UTR region from *Gl8* progenitor alleles: A portion of the *gl8* gene containing the 5′ UTR region was sequenced from each of the *Mu* stock progenitors. This region of the *Gl8-B77*, *Gl8-B79*, and *Gl8-Q67* alleles was PCR amplified with primers gl8a58 and mcd696 (GenBank accession nos. AF348367, AF348368, AF348369, respectively). Because the *Gl8-Q66* allele could not be amplified with gl8a58, it was amplified using the primer pair gl8a62 and mcd696 (GenBank accession no. AF348370). The resulting PCR products were purified and cloned into the TOPO TA cloning vector (Invitrogen, Carlsbad, CA, catalog no. K4500-40) and a bulk of 10 individual clones from each *gl8* allele was sequenced.

Genetic algorithm for alignment of sequences: The genetic algorithm (GOLDBERG 1989) used to align the *gl8-Mu* and *RescueMu* sequences specifies a single alignment as a collection of either forward- or reverse-complement orientation for each insertion-site sequence. An initial random population of 2000

such alignments was used in each run of the genetic algorithm. The fitness of any given alignment was taken to be the sum, over columns and over the four bases, of the squared deviation of the nucleotide counts from the background distribution of C's, G's, T's, and A's of the sequence data. This function was maximized by selection. Selection was performed by shuffling the population randomly into 500 groups of four alignments. In each group of four alignments, the two lower-scoring alignments were discarded and the higher-scoring alignments were copied in their place. Suffixes of the copies, selected uniformly at random, were exchanged (this is a one point crossover) and then the copies were mutated. Uniform mutation with positional mutation probability of $1/n$ (the number of insertion sites being aligned) was used. Uniform mutation proceeds down the string of forward-reverse specifications that define an alignment and independently flips each sequence with a fixed probability, $1/n$ in this case. Each run of the genetic algorithm was permitted to continue for 250 selection steps (generations). For each collection of insertion sites, the maximized consensus sequence was found in a majority of 100 runs of the genetic algorithm, suggesting that it is the true optimum alignment. Each of the 100 runs used a different initial random population of alignments.

RESULTS

The *gl8* gene structure: The *gl8* gene was previously cloned using a *Mu*-tagged allele (*gl8-Mu 88-3142*) that has a *Mu8* insertion near the gene's start codon (Xu *et al.* 1997). A B73 λ genomic clone, λ1512-38, was isolated (see MATERIALS AND METHODS) and restriction analysis identified a 6.8-kb *Hin*dIII fragment that contained the entire *gl8*-coding region. This 6.8-kb fragment was completely sequenced (GenBank accession no. AF302098). Comparisons between this B73 genomic sequence and the apparent full-length *gl8* cDNA clone (pgl8) described by Xu *et al.* (1997) revealed that the *gl8* gene contains two introns of 829 and 583 bp separated by a 70-bp exon (Figure 1). The sequenced 6.8-kb *Hin*dIII genomic *gl8* fragment includes 2871 bp upstream of the 5′ end of the *gl8* cDNA clone and 1193 bp 3′ of its polyadenylation site. Analysis of the *gl8* genomic sequence revealed that within the first exon there is a stop codon 50 bp 5′ of the start of the pgl8 sequence. As is true for many plant genes, there are no obvious TATA or CCAAT boxes in the 5′ region of the *gl8* gene.

Isolation of *gl8-Mu* alleles: The *gl8* gene product is the β-keto acyl reductase component (Xu *et al.* 2002) of the long-chain fatty acid elongase complex involved in the production of cuticular waxes. The inability of the *gl8* mutant to produce a normal wax load gives the leaves of mutant seedlings a "glossy" appearance and permits the ready identification of mutants. As reported previously, (SCHNABLE *et al.* 1994; Xu *et al.* 1997) genetic screens were used to isolate 58 *gl8-Mu* alleles. This collection included alleles generated via both random transposon tagging (10 alleles) and direct transposon tagging (46 alleles). Randomly tagged mutants were identified as new glossy mutant plants (*gl*/gl**) in F_2 progenies of plants that carried an active *Mu* system. Subsequent allelism tests established which of these *gl** alleles were allelic to *gl8*. Further analysis of the random-tagged *gl** alleles described by SCHNABLE *et al.* (1994), and additional alleles, has resulted in the identification of 5 more randomly tagged *gl8-Mu* alleles, bringing the total to 15. Further analysis of the progeny from the *gl8* directed-tagging experiment previously described by SCHNABLE *et al.* (1994) and Xu *et al.* (1997) has increased the number of directly tagged alleles isolated by the Schnable laboratory from 46 to 64. Hence, a total of 79 *gl8-Mu* alleles (randomly and directly tagged) were available for the current study. Directly tagged alleles were generated via cross 1 (in all crosses the female parent is listed first).

Cross 1: *Mu Gl8 Pr/Gl8 Pr* × *gl8-ref pr/gl8-ref pr*: Ears from cross 1 were individually shelled and kernels from each ear were planted in greenhouse sand benches such that family structures were maintained. Rare glossy seedlings from this cross carried newly generated *gl8-Mu* alleles. Because each *gl8-Mu* allele was isolated from a different ear, each allele must necessarily represent an independent mutational event. The exceptional glossy seedlings, which had the genotype *gl8-Mu Pr/gl8-ref pr*, were transplanted to pots and crossed to a *Gl8 pr* stock (cross 2) to facilitate the genetic separation of the *gl8-Mu* and *gl8-ref* alleles.

Cross 2: *gl8-Mu Pr/gl8-ref pr* × *Gl8 pr/Gl8 pr*: Kernels from cross 2 segregated one purple (*Pr/pr*):one red (*pr/pr*). Because *pr* is genetically tightly linked to *gl8* (<1 cM; STINARD and SCHNABLE 1993), >99% of the purple kernels will have the genotype *gl8-Mu Pr/Gl8 pr*.

In some instances, glossy progeny from cross 1 were crossed onto inbred lines such as *W64A* (*Gl8 Pr/Gl8 Pr*) instead of the *Gl8 pr* stock (cross 3).

Cross 3: *gl8-Mu Pr/gl8-ref pr* × *Gl8 Pr/Gl8 Pr* (*W64A* or other inbreds): Because all of the colored kernels from cross 3 were purple, it was not possible to use phenotypic selection to identify kernels that carried *gl8-Mu* alleles. Instead, in these families F_2 analysis was used to distinguish between *gl8-Mu Pr/Gl8 Pr* (F_2 will not segregate red kernels) and *gl8-ref pr/Gl8 Pr* (F_2 will segregate red kernels) progeny of cross 3.

Identification of *Mu* insertion sites: To identify the *Mu* insertion site for each of the 79 *gl8-Mu* alleles, PCR was performed on each allele using a *gl8*-specific primer in combination with a primer located in the highly conserved TIRs of *Mu* transposons. The resulting PCR products that hybridized to the *gl8* genomic sequence were purified and sequenced. For each *gl8-Mu* allele, as many as 16 *gl8*-specific primers (Figure 1) spanning a 6.0-kb interval containing the *gl8* gene were used to identify a primer that in combination with the *Mu*-TIR primer would amplify the *gl8/Mu*-flanking DNA. Amplification products were obtained from 75 of the 79 *gl8-Mu* alleles analyzed. Although *Mu*-flanking PCR products were obtained from both sides of the *Mu* transposon for most of these alleles, sequence analysis of only one side was sufficient to determine the transposon insertion site. These analyses revealed that 62 of these 75 *Mu* insertions

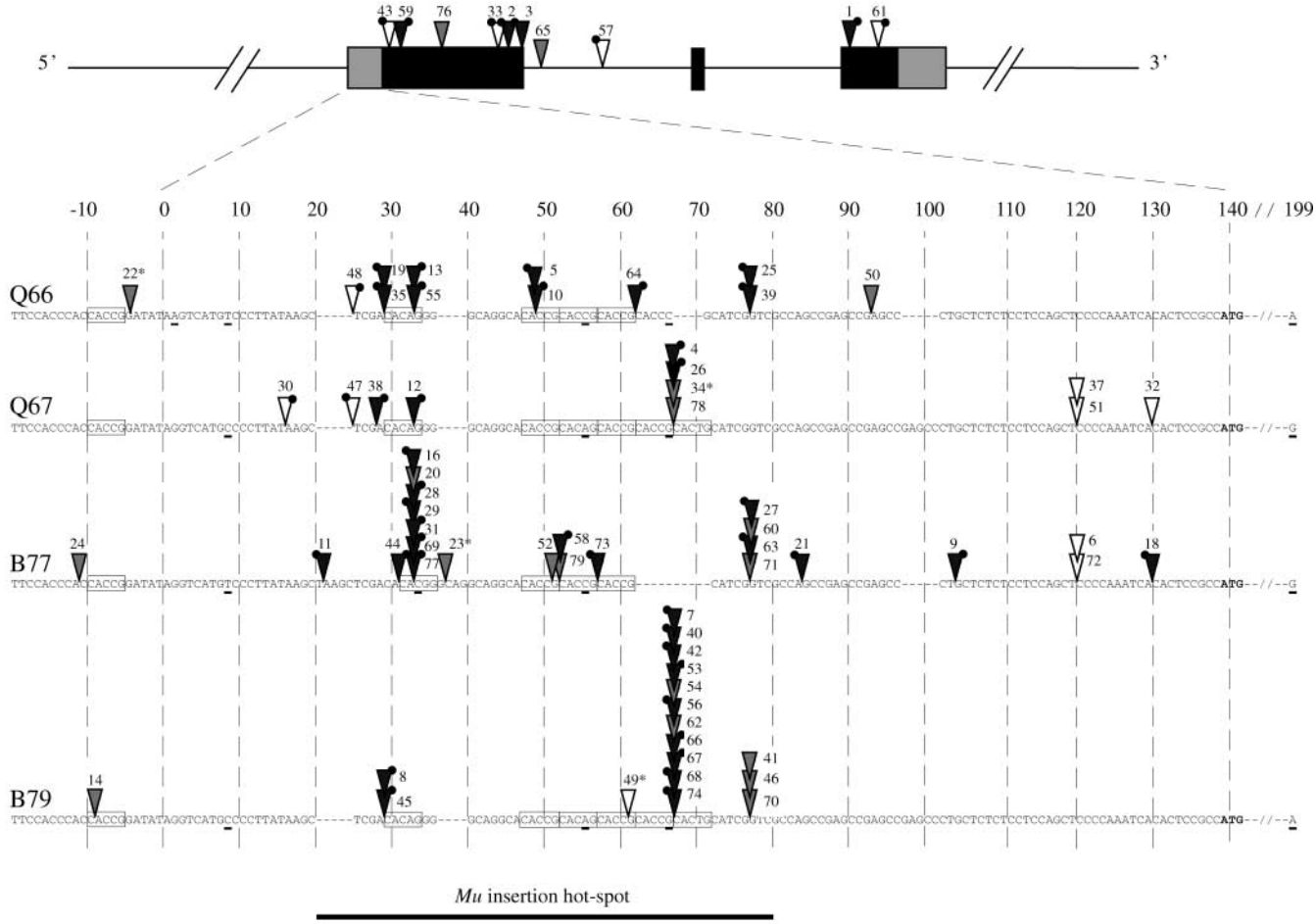

FIGURE 2.—*Mu* transposon insertion sites in 75 independent *gl8-Mu* insertion alleles. The *gl8* gene structure shown is based upon the *Gl8-B73* allele with corresponding 5′ UTR sequences provided for each of the four progenitor alleles that exist in the Schnable laboratory *Mu* stock. Triangles indicate *Mu* transposon insertion sites; the 62 insertion sites within the 5′ UTR and the three insertions between position −11 and −4 are shown at the 5′ side of the 9-bp TSD. *MuDR* insertions are indicated by shaded triangles and *Mu1* or **Mu1* (see Table 1) are indicated by solid triangles. Open triangles designate transposon insertions other than *Mu1* or *MuDR*. Solid circles indicate the positions of the right TIRs for those transposons for which orientation was determined. Numbers directly above or to the right of insertions identify alleles as listed in Table 1. Alleles with noncanonical insertion sites are indicated by asterisks and only their approximate insertion sites are indicated. Nucleotide positions are numbered with respect to the beginning of an apparently full-length *gl8-B73* cDNA clone (XU *et al.* 1997). The translation start codon is shown in boldface type near position 140. Gaps (—) were inserted where necessary to optimize the alignment of the four sequences. These gaps represent IDPs and underlined bases designate single nucleotide polymorphisms. A silent substitution at nucleotide 199 introduces an additional polymorphism that was used to identify the progenitor alleles of the 5′ insertion alleles. CACNG motifs in the 5′ UTRs are boxed. A solid bar below position 20–80 indicates the region of highest *Mu* insertion activity.

(>80%) had occurred in the 5′ UTR to the *gl8* gene and 52 (69%) of those occurred within an ~60-nucleotide interval of the *gl8* 5′ UTR (Figure 2). Ten positions within the 5′ UTR experienced multiple *Mu* insertions. One position in particular was host to 15 independent insertion events. Each of the alleles associated with a multiple insertion site was reamplified and sequenced from DNA extracted from an independent batch of seedlings. In all cases the results from the second analysis were in agreement with those from the first.

Because the *Mu* stocks used to generate most of these mutant alleles were maintained via crosses to the F_1 hybrids B77 × B79 and Q67 × Q66 (STINARD *et al.* 1993), the *gl8-Mu* alleles were derived from four potential *Gl8* progenitor alleles. The 5′ UTR sequence of each of these four *Gl8* progenitor alleles was obtained by PCR amplification with two *gl8*-specific primers (see MATERIALS AND METHODS). Several allele-specific insertion/deletion polymorphisms (IDPs) 5′ of the start codon were identified among the four parental alleles (Figure 2 and data not shown). These IDPs, along with a silent substitution at position 199 of exon 1, permitted the identification of the progenitor alleles of each of the 65 *gl8-Mu* alleles that arose via a *Mu* insertion 5′ of the *gl8* start codon (Figure 2 and Table 1).

To ensure that the *gl8-ref* allele used in the directed-tagging experiments did not interfere with the PCR-based mapping of *Mu* transposons, *gl8-ref* was character-

ized by PCR amplification and sequence analysis. No *gl8*-hybridizing PCR products were obtained with DNA from *gl8-ref/gl8-ref* individuals using the *Mu*-TIR primer in combination with any of 16 *gl8*-specific primers (data not shown). In addition, no DNA sequence polymorphisms were found between the *gl8-ref* allele and the wild-type *Gl8-B73* allele in the region defined by PCR primers gl8a58 to gl8ain, which includes the entire coding region.

Analysis of *Mu* TIR sequences: The *gl8/Mu*-flanking PCR products contained 39 nucleotides of *Mu* TIR sequence terminal to the *Mu*-TIR primer annealing site. Comparing the sequence of these 39 nucleotides from each *gl8-Mu* allele to the left and right TIRs of the previously defined *Mu* transposons (Figure 3) identified most of the *Mu* transposons. Insertions corresponding to six *Mu* transposons (*MuDR*, *Mu1*, *Mu2*, *Mu4*, *Mu8*, and *Mu10*) were identified in this fashion. Several novel TIR sequences were also recovered. The novel TIRs identified from the 5′ and 3′ PCR products of *gl8-Mu 91g211* define the left and right TIRs of the *Mu11* transposon (GenBank accession nos. AF247740 and AF247741, respectively). An additional novel *Mu* transposon was identified from *gl8-Mu 91g241* and was designated *Mu12*. Sequences from the 5′ and 3′ PCR product from *gl8-Mu 91g241* revealed that the terminal 39 bp from the TIRs of *Mu12* are perfect inverted repeats and were designated as the left and right TIRs of *Mu12* (GenBank accession nos. AF247742 and AF302101, respectively). Analysis of the sequence of the 5′ PCR product from *gl8-Mu 91g209* identified the inserted transposon as *Mu10*. Because only the left TIR of *Mu10* had previously been reported, the 3′ PCR product from *gl8-Mu 91g209* was sequenced to obtain the right TIR of *Mu10* (GenBank accession no. AF302099).

Characterization of the four *gl8-Mu* alleles in which *Mu* insertions were not detected: *Mu* insertions were not identified in 4 of the 79 *gl8* alleles analyzed in this study. These 4 alleles were subjected to further analyses. On the basis of PCR amplification and sequence analysis it was determined that the progenitor of the directly tagged allele *gl8-Mu 91g215* was *Gl8-Q67* (data not shown). DNA gel blot analysis was then performed using probe B (Figure 1). Although a restriction fragment length polymorphism (RFLP) exists between *gl8-Mu 91g215* and *Gl8-Q67* (Figure 4, lanes 5 and 1, respectively), *gl8-Mu 91g215* is indistinguishable from *gl8-ref* (Figure 4, lane 3) in this hybridization experiment. Therefore, it is likely that during cross 2 the *gl8-Mu 91g215* allele was replaced by the *gl8-ref* allele as the result of a crossover between *gl8* and *pr*.

DNA isolated from a plant homozygous for *gl8-Mu 91g159* failed to hybridize to probe B (Figure 4, lane 4). Hybridization with an unrelated single-copy probe established that lane 4 in Figure 4 contains DNA of sufficient quality and quantity to yield hybridization signals to single-copy genes (data not shown). A subsequent experiment using the same blot and the 1.4-kb *gl8* cDNA also failed to detect the *gl8* gene in lane 4 (data not shown). This result suggests that *gl8-Mu 91g159* consists of a deletion of the entire coding region of the *gl8* gene.

The *gl8-Mu 94-1641-25* allele was not generated from a *Mu* stock that contains Q66-, Q67-, B77-, and B79-derived alleles (see MATERIALS AND METHODS). Hence, it is not possible to determine with certainty its progenitor allele. However, the 5′ UTR of *gl8-Mu 94-1641-25* was PCR amplified and sequenced from plants homozygous for this allele and was found to be identical to *Gl8-B73*. In addition, DNA gel blot analysis using probe B revealed that *gl8-Mu 94-1641-25* was indistinguishable from *Gl8-B73* (Figure 4, lanes 6 and 2, respectively). These results suggest that *gl8-Mu 94-1641-25* either contains a minor rearrangement not detectable via RFLP analysis or has an insertion (or other mutation) outside of the 6.8-kb *Hin*dIII fragment detected by the RFLP analysis.

Although it was not possible to identify a *Mu* insertion in *gl8-Mu 93B227*, RFLP analysis of this allele shed some light on the nature of its molecular lesion. Lane 9 of Figure 4 contains DNA from a pool of 10 progeny resulting from a cross between the inbred line W64 and a plant with the genotype *gl8-Mu 93B227/gl8-ref* (cross 3). Analysis with probe B revealed that this DNA sample contains two RFLP fragments. A comparison between lanes 3 and 10 revealed that the *gl8-ref* and *Gl8-W64* alleles are indistinguishable in this hybridization experiment and account for the smaller RFLP signal in lane 9. Hence, the larger RFLP signal in lane 9 must be derived from *gl8-Mu 93B227*. This RFLP differs from each of the four possible progenitor alleles of *gl8-Mu 93B227*: *Gl8-Q67*, *Gl8-66*, *Gl8-B77*, and *Gl8-B79* (lane 1 and data not shown). Hence, *gl8-Mu 93B227* may contain a novel *Mu* transposon that could not be amplified because it contains divergent TIR sequences or because sequence divergence near the insertion site of *gl8-Mu 93B227* relative to the *Gl8-B73*-derived primers may have impeded amplification. Alternatively, this allele may have resulted from another type of molecular rearrangement.

Noncanonical target-site duplications: For most alleles, sequence analysis of the *gl8/Mu* PCR product derived from one side of the *Mu* transposon was sufficient to determine the *Mu* insertion site and *Mu* identity. Nonetheless, to better define a novel transposon or to confirm apparent sequence anomalies, the *gl8/Mu*-flanking PCR products were sequenced from both sides of the *Mu* transposons associated with 14 alleles. Analysis of these 14 sequences revealed that 4 alleles (*gl8-Mu 91g168*, *gl8-Mu 91g169*, *gl8-Mu 91g213*, and *gl8-Mu 91g239*) did not have the characteristic 9-bp TSD (Figure 5). Each of these alleles was reamplified and sequenced from an independent batch of seedlings. In all instances results from the second analysis were in

TABLE 1

Summary of 79 *gl8-Mu* insertion alleles

Allele no.	*gl8-Mu* allele	Tagging method[a]	Progenitor allele	*Mu* insertions site[b]	*Mu* insertions[c]	*Mu* orientation[d]	Genic region[e]
1	75-5074-22	R	ND	*AAGATGGCAT*::CCATCAGGA::AGTCTTCCTT	*Mu1*	F	Exon 3
2	77-3134-48	R	ND	ATCCGCTCTA::CTCCGTCTA::CGCCGCTACC	*Mu1*	R	Exon 1
3	78-1841-47	R	ND	GTCTACGCCG::CTACCAAAG::*CGTGAGCTCC*	*Mu1*	R	Exon 1
4	83-8123-7	R	Q67	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*Mu1*	F	5′ UTR
5	87-2250-8	R	Q66	*GGCAGGCACA*::CCGCACCGC::ACCGCACCCG	**Mu1*	R	5′ UTR
6	88-3142-4	R	B77	TCCTCCAGCT::CCCCAAATC::ACACTCCGCC	*Mu8*	R	5′ UTR
7	90-2940-A	R	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	**Mu1*	R	5′ UTR
8	90-3230-5	R	B79	*ATAAGCTCGA*::CACAGGGGC::AGGCACACCG	*Mu1*	F	5′ UTR
9	91-2136-36	R	B77	GCCGAGCCCT::GCTCTCTCC::TCCAGCTCCC	*Mu1*	F	5′ UTR
10	91g-6079-25	R	Q66	*GGCAGGCACA*::CCGCACCGC::ACCGCACCCG	*Mu1*	F	5′ UTR
11	92-1205-80	R	B77	*CTTATAAGCT*::AAGCTCGAC::ACACGGGCAG	**Mu1*	R	5′ UTR
12	92-1253-64	R	Q67	*GCTCGACACA*::GGGGCAGGC::ACACCGCACA	*Mu1*	F	5′ UTR
13	92g-4908-27	R	Q66	GCTCGACACA::GGGGCAGGC::*ACACCGCACC*	**Mu1*	F	5′ UTR
14	94-1480-26	R	B79	TCCACCCACC::ACCGGATAT::*AGGTCATGCC*	*MuDR*	ND	5′ region*
15	94-1641-25	R	ND	No insertion detected			
16	91g 157	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	**Mu1*	R	5′ UTR
17	91g 159	D	ND	No insertion detected			
18	91g 160	D	B77	CCCCAAATCA::CACTCCGCC::ATGGCCGGCA	*Mu1*	R	5′ UTR
19	91g 161	D	Q66	*ATAAGCTCGA*::CACAGGGGC::AGGCACACCG	**Mu1*	R	5′ UTR
20	91g 163	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	*MuDR*	ND	5′ UTR
21	91g 164	D	B77	TCGGTCGCCA::GCCGAGCCG::*AGCCCTGCTC*	**Mu1*	R	5′ UTR
22	91g 168	D	Q66	See Figure 5	*MuDR*	ND	5′ region*
23	91g 169	D	B77	See Figure 5	*MuDR*	ND	5′ UTR
24	91g 170	D	B77	*GTTCCACCCA*::CCACCGGAT::ATAGGTCATG	*MuDR*	ND	5′ region*
25	91g 171	D	Q66	*ACCCGCATCG*::GTCGCCAGC::CGAGCCGAGC	**Mu1*	R	5′ UTR
26	91g 203	D	Q67	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*Mu1*	F	5′ UTR
27	91g 204	D	B77	*CACCGCATCG::GTCGCCAGC::CGAGCCGAGC*	**Mu1*	R	5′ UTR
28	91g 205	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	*Mu1*	F	5′ UTR
29	91g 208	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	**Mu1*	R	5′ UTR
30	91g 209	D	Q67	TGCCCCTTAT::AAGCTCGAC::ACAGGGGCAG	*Mu10*	F	5′ UTR
31	91g 210	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	*Mu1*	F	5′ UTR
32	91g 211	D	Q67	CCCCAAATCA::CACTCCGCC::ATGGCCGGCA	*Mu11*	F	5′ UTR
33	91g 212	D	ND	CTTCTGATCC::GCTCTACTC::*CGTCTACGCC*	*Mu11*	R	Exon 1
34	91g 213	D	Q67	See Figure 5	*MuDR*	ND	5′ UTR
35	91g 214	D	Q66	*ATAAGCTCGA*::CACAGGGGC::AGGCACACG	**Mu1*	R	5′ UTR
36	91g 215	D	Q67	No insertion detected			
37	91g 216	D	Q67	TCCTCCAGCT::CCCCAAATC::*ACACTCCGCC*	*Mu12*	ND	5′ UTR
38	91g 217	D	Q67	*TATAAGCTCG*::ACACAGGGG::CAGGCACACC	*Mu1*	F	5′ UTR
39	91g 219	D	Q66	*ACCCGCATCG*::GTCGCCAGC::CGAGCCGAGC	**Mu1*	R	5′ UTR
40	91g 220	D	B79	*CACCGCACCG::CACTGCATC::GGTCGCCAGC*	**Mu1*	R	5′ UTR

(*continued*)

TABLE 1

(Continued)

Allele no.	*gl8-Mu* allele	Tagging method[a]	Progenitor allele	*Mu* insertions site[b]	*Mu* insertions[c]	*Mu* orientation[d]	Genic region[e]
41	91g 228	D	B79	*CACTGCATCG*::GTCGCCAGC::CGAGCCGAGC	*MuDR*	ND	5′ UTR
42	91g 229	D	B79	CACCGCACCG::CACTGCATC::GGTCGCCAGC	*Mu1*	R	5′ UTR
43	91g 230	D	ND	*GAGTTCCTCC*::GCGCGCAGC::CGGCGTGGGC	*Mu11*	R	Exon 1
44	91g 231	D	B77	*AAGCTCGACA*::CACGGGCAG::GCAGGCACAC	*Mu1*	F	5′ UTR
45	91g 232	D	B79	*ATAAGCTCGA*::CACAGGGGC::AGGCACACCG	*Mu1*	F	5′ UTR
46	91g 233	D	B79	*CACTGCATCG*::GTCGCCAGC::CGAGCCGAGC	*MuDR*	ND	5′ UTR
47	91g 235	D	Q67	*CCTTATAAGC::TCGACACAG::GGGCAGGCAC*	*Mu11*	R	5′ UTR
48	91g 238	D	Q66	CCTTATAAGC::TCGACACAG::GGGCAGGCAC	*Mu11*	F	5′ UTR
49	91g 239	D	B79	See Figure 5	*Mu4*	ND	5′ UTR
50	91g 240	D	Q66	*TCGCCAGCCG*::AGCCGAGCC::CTGCTCTCTC	*MuDR*	ND	5′ UTR
51	91g 241	D	Q67	TCCTCCAGCT::CCCCAAATC::ACACTCCGCC	*Mu12*	F	5′ UTR
52	91g 248	D	B77	*CAGGCACACC*::GCACCGCAC::CGCATCGGTC	*MuDR*	ND	5′ UTR
53	93B 141	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*Mu1*	F	5′ UTR
54	93B 142	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*MuDR*	ND	5′ UTR
55	93B 144	D	Q66	*GCTCGACACA*::GGGGCAGGC::ACACCGCACC	*Mu1*	F	5′ UTR
56	93B 145	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	**Mu1*	R	5′ UTR
57	93B 146	D	Q66	GCAAAAAACT::TTCACTGTA::*AGGCGTACTG*	*Mu8*	R	Intron 1
58	93B 147	D	B77	*AGGCACACCG*::CACCGCACC::GCATCGGTCG	*Mu1*	F	5′ UTR
59	93B 148	D	ND	*GGCCGTGCGC*::GCCGCCGCC::CGCTTCGCGC	*Mu1*	F	Exon 1
60	93B 150	D	B77	*CACCGCATCG*::GTCGCCAGC::CGAGCCGAGC	*MuDR*	ND	5′ UTR
61	93B 151	D	ND	*CCGAGTCCCT*::TATCGACAG::CGTGCGCCTG	*Mu2*	F	Exon 3
62	93B 152	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*MuDR*	ND	5′ UTR
63	93B 153	D	B77	*CACCGCATCG*::GTCGCCAGC::CGAGCCGAGC	**Mu1*	R	5′ UTR
64	93B 154	D	Q66	*CACCGCACCG*::CACCCGCAT::CGGTCGCCAG	*Mu1*	F	5′ UTR
65	93B 155	D	ND	GATTGGCTCG::ATTCTAGGA::*TAGATGGCTT*	*MuDR*	ND	Intron 1
66	93B 156	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*Mu1*	F	5′ UTR
67	93B 158	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	*Mu1*	F	5′ UTR
68	93B 159	D	B79	*CACCGCACCG*::CACTGCATC::GGTCGCCAGC	**Mu1*	R	5′ UTR
69	93B 161	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	*Mu1*	F	5′ UTR
70	93B 162	D	B79	CACTGCATCG::GTCGCCAGC::*CGAGCCGAGC*	*MuDR*	ND	5′ UTR
71	93B 164	D	B77	CACCGCATCG::GTCGCCAGC::*CGAGCCGAGC*	*MuDR*	ND	5′ UTR
72	93B 181	D	B77	*TCCTCCAGCT*::CCCCAAATC::ACACTCCGCC	*Mu8*	R	5′ UTR
73	93B 189	D	B77	CACCGCACCG::CACCGCATC::*GGTCGGCAGC*	*Mu1*	R	5′ UTR
74	93B 226	D	B79	CACCGCACCG::CACTGCATC::*GGTCGCCAGC*	*Mu1*	R	5′ UTR
75	93B 227	D	ND	No insertion detected			
76	93B 228	D	ND	CTGATGCGCA::CCCTCATCC::*GGGTCAACGT*	*MuDR*	ND	Exon 1
77	93B 606	D	B77	*GCTCGACACA*::CGGGCAGGC::AGGCACACCG	*Mu1*	F	5′ UTR
78	93B 607	D	Q67	CACCGCACCG::CACTGCATC::GGTCGCCAGC	*MuDR*	ND	5′ UTR
79	93B 609	D	B77	*AGGCACACCG*::CACCGCACC::GCATCGGTCG	*MuDR*	ND	5′ UTR

(*continued*)

agreement with the first set of analyses. The sizes of the 5′ and 3′ *Mu*-flanking PCR products from *gl8-Mu 91g168* were not in agreement with the predicted product sizes from any of the 4 progenitor alleles, suggesting that this allele contained a deletion. Sequence analysis of both of these PCR products established the progenitor of *gl8-Mu 91g168* as *Gl8-Q66* and revealed an apparent deletion of 232 bp from the *Mu* insertion site (Figure 5A). However, the PCR-based characterization performed here could not distinguish between a deletion and two closely linked (232 bp) *Mu1* insertions. Such a structure was observed in the *hcf106-mum2* allele in which a second *Mu1* transposon inserted 244 bp downstream of the original *Mu1* insertion of the *hcf106-mum1* allele (Das and Martienssen 1995). To distinguish between these two possibilities, PCR was performed using primers *Mu*-TIR and gl8a64, which lies in the predicted deletion. The observation that this PCR reaction failed to amplify a *gl8*-hybridizing product provides evidence that the *gl8-Mu 91g168* allele contains a 232-bp deletion rather than a double *Mu* insertion. Deletions adjacent to a *Mu* insertion have been observed to arise subsequent to initial insertion events at frequencies near 1% and are thought to result from aborted transposition events or illegitimate recombination between the transposon and the gene (Levy and Walbot 1991; reviewed by Das and Martienssen 1995; Raizada *et al.* 2001). Given this frequency, during the propagation of 79 alleles over multiple generations, the recovery of such a deletion would not be unexpected.

Sequence data from one side of the *Mu* insertion of alleles *gl8-Mu 91g169*, *gl8-Mu 91g213*, and *gl8-Mu 91g239* established polymorphisms relative to each of the four progenitor alleles. For these alleles, sequence data were also obtained from the other side of the inserted *Mu* transposon. The progenitor of *gl8-Mu 91g169* could be identified as *Gl8-B77* but contained a deletion of seven nucleotides from the 3′ TSD followed by the insertion of four C's directly flanking the *Mu* insertion (Figure 5B). Analysis of the *Mu*-flanking sequence from *gl8-Mu 91g213* identified the progenitor as *Gl8-Q67*, but the 3′ flanking sequence contains an A-to-C transversion in the second nucleotide position of the TSD (Figure 5C). Sequence analysis of the 3′ PCR product from *gl8-Mu 91g239* identified the progenitor as *Gl8-B79* and the *Mu* transposon as either *Mu3* or *Mu4*. Sequence obtained from the 5′ PCR product established that the inserted *Mu* transposon was *Mu4*. However, the TSDs on the two sides of this transposon were not identical due to an apparent deletion of a G from the 3′ TSD (Figure 5D).

Statistical analysis of target-site sequences: *Mu* insertions occurred most frequently in an ~60-bp region between nucleotide positions 20 and 80 in the 5′ UTR of the *gl8* gene. Depending on the progenitor allele, and after removal of gaps that were inserted to allow alignment of the sequences, the actual length of this region varies from 47 to 51 nucleotides. A motif search of the *B73-gl8* genomic sequence identified a five-base motif, CACNG, which appears frequently in this region of the 5′ UTR that experienced the highest *Mu* insertion frequency. Also depending on the progenitor allele, the CACNG motif appears between four and six times in this ~60-bp region with as many as five motifs arranged in tandem. Excluding the targeted *Mu* insertion region of the 5′ UTR, the CACNG motif appears at the expected frequency (~1/256 bp) in the 3045 nucleotides of the *B73-gl8* sequence between primers 3 and 10 (Figure 1). Therefore the nonrandom distribution of CACNG motifs in the *gl8* gene mirrors the nonrandom insertion pattern of *Mu* transposons in the *gl8* gene.

A genetic algorithm was used to identify the sequence alignment that would provide the best consensus sequence for the *Mu* insertion sites identified in this study. Genetic algorithms create a population of solutions to

TABLE 1

(Continued)

[a] Alleles were generated via either random (R) or direct (D) transposon-tagging experiments.

[b] *Mu* insertion sites are shown as they would appear in the corresponding progenitor allele (see Figure 6A). The sequence of the 9-bp TSD is enclosed by double colons flanked on either side by 10 nucleotides immediately upstream and downstream of the TSD. For alleles in which only one side of the *Mu* transposon was sequenced, the TSD represents a predicted duplication. For these alleles, *gl8* sequences from the nonsequenced side of the *Mu* transposon are based on the progenitor sequence for that allele and are shown in italics. For insertions in which transposons inserted in the coding region (and intronic insertion allele no. 65) progenitor alleles were not determined and flanking sequences are derived from the *G18-B73* allele. Underlined TSDs indicate sequences that were obtained from both sides of a *Mu* transposon and were confirmed to be exact duplications.

[c] The terminal 39 nucleotides of the left TIR of *Mu1* and the right TIR of *Mu2* are identical. Hence it was not possible to distinguish between *Mu1* and *Mu2* insertions in *gl8-Mu* alleles in which only the *Mu1* left TIR sequence was obtained. Such alleles are designated **Mu1* in Table 1. Since *Mu1* transposons are present in considerably higher numbers than *Mu2* transposons in *Mu* stocks (reviewed by Bennetzen *et al.* 1993), the orientation of **Mu1* insertions is given with respect to *Mu1*.

[d] Terminal *Mu* TIR sequences allowed for determination of transposon orientation for most transposons. Orientation was defined as forward (F) if the transposon inserted with TIR arranged left to right with respect to *gl8* and reverse (R) if the transposon inserted in the opposite direction.

[e] Three alleles (allele nos. 14, 22, and 24), marked by asterisks, have *Mu* insertions 4–11 nucleotides 5′ of the transcription start site of the B73 allele. Because transcription start sites were not determined for each of the progenitor alleles, the insertions in these three alleles may be in the 5′ UTR.

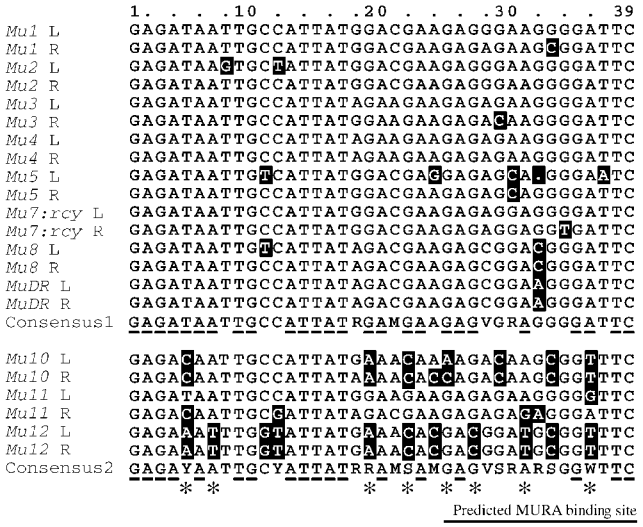

FIGURE 3.—Comparison of the terminal 39 nucleotides of 22 *Mu* TIR sequences. Consensus1 was derived from the TIRs of *Mu1*, *Mu2*, *Mu3*, *Mu4*, *Mu5*, *Mu7*, *Mu8*, and *MuDR* only. Consensus2 was derived from all 22 TIR sequences. Nucleotides in the consensus lines are underlined if that position is invariant among the TIRs included in the consensus line. Nucleotides in the consensus line that are not underlined are well conserved (*i.e.*, no other nucleotide is present more than twice at that position). Positions that do not have an invariant or a well-conserved nucleotide are indicated in the consensus line by their IUB ambiguity codes (M, R, S, V, W, and Y). Nucleotides present only one or two times at a position were not considered when assigning ambiguity codes. Nucleotides that do not conform to consensus1 are shaded. Asterisks indicate eight positions that are invariant in the TIRs of *Mu1*, *Mu2*, *Mu3*, *Mu4*, *Mu5*, *Mu7/rcy*, *Mu8*, and *MuDR* (consensus1) but are no longer invariant when the TIRs of the newly defined *Mu10*, *Mu11*, and *Mu12* transposons are considered (consensus2).

a problem and use an iterative selection and variation process to identify good solutions. In this case, the goal is to determine, for each insertion site, which of the two possible orientations provides the best alignment with all other insertion sites. The genetic algorithm does this by determining for each insertion site the orientation that maximizes the divergence from the background distribution of nucleotides. This maximization is performed by selecting alignments with high divergence scores in pairs and then applying variation operations (crossover and mutation) to the alignment to generate similar and possibly superior alignments (see MATERIALS AND METHODS).

The 35 unique insertion sites as defined by the 10 nucleotides 5′ of the TSD, the 9-bp TSD, and the 10 nucleotides 3′ of the TSD were analyzed in the manner described above. Although each of the 71 *Mu* insertion sites listed in Table 1 is from an independent insertion event, to avoid biasing the data from sites with multiple insertion events, only a single instance of any given 29-bp insertion site was used in this analysis. Figure 6 shows the orientations of the sequence alignments as selected by the genetic algorithm.

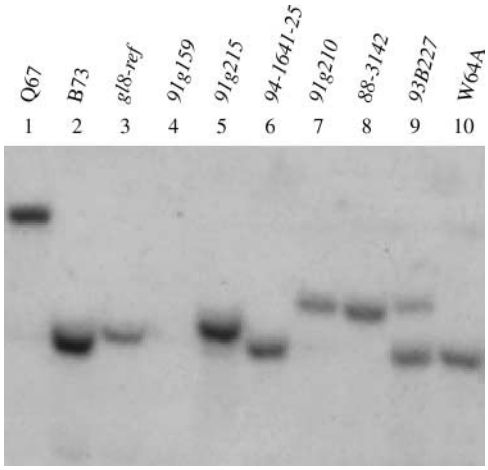

FIGURE 4.—DNA gel blot analysis of six *gl8-Mu* alleles. Approximately 10 μg of *Hin*dIII-digested genomic DNA was loaded per lane and hybridized with probe B (Figure 1). Lanes 1, 2, and 10 contain DNA from inbred lines Q67, B73, and W64A, respectively. Lane 3 contains DNA from a stock homozygous for the *gl8-ref* allele. Lanes 4–8 contain DNA from stocks homozygous for the designated *gl8-Mu* alleles. Lane 9 contains DNA from *gl8-Mu 93B227* isolated from a bulk of 10 progeny from cross 3.

As shown in Table 2, the 35 insertion sites are GC rich and particularly low in T. This may be due, at least in part, to the GC richness of the *gl8* gene. To normalize for this high GC content an expected nucleotide frequency was calculated based on the GC content of the combined insertion sites. Chi-square analyses were performed to identify positions that had significant deviations from the expected nucleotide composition. Within the TSD, T's, G's, and C's appeared at positions 2, 8, and 9, respectively, at significantly higher-than-expected frequencies. Position 4 has a lower-than-expected frequency of A. The conserved nucleotides at positions 2, 8, and 9 and the weak consensus nucleotides at the other TSD positions are consistent with the reverse complement of consensus sequences reported by CRESSE *et al.* (1995) and HANLEY *et al.* (2000) (Table 3).

Chi-square analyses of the 20 positions flanking the *gl8* TSDs reveals a more significant nucleotide conservation than is observed within the 9-bp TSD itself. Five of the six positions directly flanking the 9-bp TSD have a nucleotide composition that differs from the expected nucleotide composition at the 99% confidence interval (Table 2 and Figure 6). These include conserved C's, A's, and G's at positions −2, +1, and +2, respectively, and significantly lower frequencies of Cs at positions −1 and +3. Two of these conserved nucleotides, the −2 C and +2 G, are particularly interesting given that they are complementary bases equidistant from the TSD. When this analysis was extended to include 50 nucleotides flanking either side of the *Mu* insertion, additional positions with significant deviations from the expected were not identified at rates higher than would be expected by chance (data not shown).

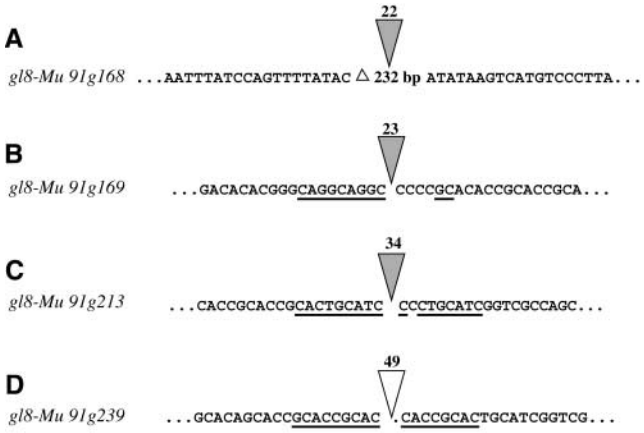

FIGURE 5.—Insertion-site sequences of four noncanonical *Mu* insertions. Shaded triangles indicate *MuDR* transposons and the open triangle represents a *Mu4* or *Mu4*-like transposon. Numbers above triangles indicate allele numbers as shown in Figure 1 and Table 1. Underlined text indicates aberrant TSDs. *gl8-Mu 91g168* contains a 232-bp deletion from the *gl8* 5′ region with no TSD. *gl8-Mu 91g169* contains a seven-nucleotide deletion from the 3′ TSD followed by an insertion of four C's, leaving only a 2-bp TSD. *gl8-Mu 91g213* has a mismatch in the 9-bp TSD. The 3′ duplication has an A-to-C transversion. *gl8-Mu 91g239* has a single-nucleotide deletion from the 3′ TSD, resulting in an 8-bp TSD.

Analysis of *RescueMu* insertion sites: To determine if the nonrandom nucleotide composition observed among *gl8* insertion sites is a common feature of *Mu* insertion sites across the genome, a large collection of *RescueMu* insertion sites was analyzed as described above for the *gl8-Mu* insertion sites. A total of 369 independent *RescueMu* insertion-site sequences generated by the Walbot laboratory were recovered from GenBank by identifying forward and reverse sequences from individual *RescueMu* events. Eighteen of these *RescueMu* insertion sites have TSD lengths other than 9 bp (Table 4). An additional nine *RescueMu* insertion sites were recovered that contained mismatched TSDs (data not shown). Therefore, 315 *RescueMu* sequences remained after removal of these noncanonical *RescueMu* insertion sites and insertion sites that did not have at least 60 bp of sequence flanking each side of the TSD. These insertion sites (consisting of 60 bp flanking each side of the TSD plus the 9-bp TSD) were aligned using the genetic algorithm described previously. Table 5 contains the nucleotide composition and total chi-square values for the 10 positions flanking the TSD and the 9-bp TSD for the optimal *RescueMu* alignment. Within the TSD, six of the nine positions have nucleotides that differ from the expected nucleotide frequency at the 99% confidence interval and two of the remaining three positions have nucleotides that differ from the expected nucleotide frequency at the 95% confidence interval. The one remaining position (position 4) has a significantly lower-than-expected frequency of A's. Position 4 was therefore designated as B, in accordance with the International

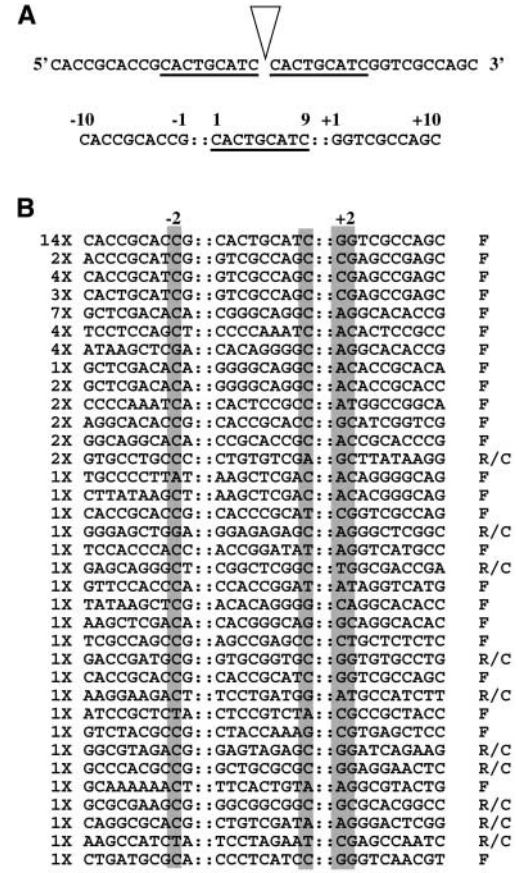

FIGURE 6.—Alignment of the 35 unique insertion sites from Table 1. (A) The top sequence illustrates an example of an insertion site as present in a *gl8-Mu* allele. The underlined sequence indicates the 9-bp TSD and the triangle represents the *Mu* transposon. The bottom sequence illustrates the target site as present in the corresponding progenitor allele. The nine nucleotides in the TSD are designated as positions 1–9. The 9-bp TSD is separated from each set of flanking sequences by a double colon. The 10 nucleotides 5′ to the TSD are designated positions −1 to −10 and the 10 nucleotides 3′ to the TSD are designated as positions +1 to +10. (B) The DNA strand shown for each of the 35 independent insertion sites is the strand that gives the most nonrandom nucleotide composition as determined by the genetic algorithm (see MATERIALS AND METHODS). DNA strands oriented 5′ to 3′ with respect to the *gl8* gene are indicated with an F and sequences in which the opposite strand was selected are indicated by R/C. The four positions identified in Table 2 as having significant deviations from the expected nucleotide frequency at the 99% confidence interval are shaded. The number to the left of each sequence indicates the number of *gl8-Mu* alleles with the identical 29-bp insertion site.

Union of Biochemistry (IUB) ambiguity code, in the *RescueMu* consensus TSD and in the combined consensus TSD. The 9-bp consensus TSD for the *RescueMu* insertion sequences is 5′ CTCBCAGAC 3′, which is strikingly similar to the consensus 9-bp TSD from the *gl8-Mu* insertion sites and the reverse complements of the CRESSE *et al.* (1995) and HANLEY *et al.* (2000) consensus TSDs. The four consensus sequences shown in Table 3 can be combined into a single 9-bp consensus TSD of 5′ C-T-C-B-G/C-A/C-G/A-A/G-C 3′.

TABLE 2

Nucleotide composition of 35 unique insertion sites

	Sequence 5′ of TSD										
Nucleotide	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	Exp.[a]
C	9	11	16	21*	8	16	9	12	30**	*2***	13
G	14	5	11	*3**	15	5	8	7	*2**	15	10
A	7	12	*2**	8	8	13*	12	10	*1**	12	7.5
T	5	7	6	3	4	1	6	6	2	6	4.3
Sum χ^2	2.8	7.4	5.4	10*	4.5	9.8*	5.1	2.6	36**	15**	

	TSD									
Nucleotide	1	2	3	4	5	6	7	8	9	Exp.
C	17	9	19	14	14	12	6	*4**	23**	13
G	10	7	11	11	12	11	13	17*	4	10
A	5	10	4	*2**	6	9	12	8	4	7.5
T	3	9*	1	8	3	3	4	6	4	4.3
Sum χ^2	2.5	8.2*	6.9	7.3	1.1	0.83	7.3	11*	13**	

	Sequence 3′ of TSD										
Nucleotide	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	Exp.
C	9	10	*2***	14	16	15	13	15	13	17	13
G	10	20**	14	15	10	8	9	5	10	13	10
A	15**	*1**	14*	*0**	6	8	8	13*	5	3	7.5
T	1	4	5	6	3	4	5	2	7	2	4.3
Sum χ^2	11*	16**	16**	11*	1.4	0.83	0.28	8.2*	2.5	5.9	

A total of 35 *g18-Mu* insertion alleles from Table 1 have unique insertion sites as defined by the 10 nucleotides 5′ of the 9-bp TSD, the 9-bp TSD, and the 10 nucleotides 3′ of the 9-bp TSD. Chi-square values were calculated for each individual nucleotide at all positions and the sum of those values is listed for each position in the table. The symbols * and ** indicate chi-square values or nucleotides frequencies with chi-square values that are significantly different from expected at the 95 and 99% confidence intervals, respectively. Underlined and boxed text indicate the nucleotides that occur at significantly higher-than-expected rates at the 95 and 99% confidence intervals, respectively, and therefore are consensus nucleotides at that position. Italics indicate nucleotides that are significantly underrepresented at a given position.

[a] Expected nucleotide compositions represent the number of times each nucleotide is expected to appear at each position based on the nucleotide composition of the 35 combined insertion sites.

As was the case for the *gl8-Mu* insertion sites, the nucleotides with the highest deviations from the expected nucleotide composition were observed at the positions directly flanking the TSD. The three positions flanking either side of the TSD had a conserved nucleotide at the 99% confidence interval. These nucleotides are CCT at positions −1, −2, and −3, respectively, and AGG at positions +1, +2, and +3, respectively. Hence, the three positions immediately flanking either side of the TSD are conserved for the paired 3-bp inverted repeats CCT and ATT. Therefore, each of the four conserved nucleotides identified from the *gl8-Mu* insertions is also conserved at its respective position in the *RescueMu* insertion sites. The larger size of the *RescueMu* data set, however, apparently allowed the additional conserved nucleotides to be identified. A plot of the total chi-square at each of the 129 positions reveals that nearly all the positions with significant deviations from the expected occur within a 15-bp region centered on the TSD and that the positions with the highest chi-squares directly flank the 9-bp TSD, particularly the −2 and +2 positions (Figure 7).

Analysis of the *RescueMu* sequence alignment also revealed a significant difference between the nucleotide composition of the sequences to the left and right of the TSD (Figure 8A). GC profiles of the *gl8* and *bz1* genes reveal that the preferred target site for *Mu* insertions also occurs at the interface between regions with low and high GC content (Figure 9). These data suggest that *Mu* transposons may have a preference for inserting into regions where DNA composition is transitioning from low to high GC content.

DISCUSSION

Analysis of a collection of 79 *gl8* mutant alleles derived from *Mu* stocks revealed that the vast majority (at least 75/79) contained *Mu* insertions. Of these 75 alleles, 62

TABLE 3

Similarity among reported 9-bp TSD consensus sequences

Source	N^a	TSD position								
		1	2	3	4	5	6	7	8	9
Cresse[b]	19	C	T/C	C	T	G/C	C	A	A	C
Hanley[b]	450	C>G	T>C	C	T>A	G/C	C/A	A/G	G/A	C
g18-Mu[c]	35	C	T*	C	T	G	A	A	G*	C**
RescueMu[c]	315	C**	T**	C**	B	C*	A*	G**	A**	C**
Combined	819	C	T	C	B	G/C	A/C	G/A	A/G	C

[a] Number of insertion events used to generate the indicated consensus sequence.

[b] These consensus sequences are shown as the reverse complement of the original reports by Cresse *et al.* (1995) and Hanley *et al.* (2000). These consensus sequences were derived from randomly isolated *Mu* transposons and, hence, there is no apparent reason for selection of one orientation over the other. The consensus sequence derived from the *g18* sequences, however, is based on the direction of transcription of the *g18* gene and therefore consensus sequences are oriented with respect to the *g18* gene.

[c] Data from this study. Nucleotides shown in the consensus sequence have the largest chi-square value among the nucleotides that occur at higher-than-expected rates at each position. The symbols * and ** indicate that the chi-square value for that nucleotide is significant at the 95 and 99% confidence intervals, respectively. The IUB ambiguity code, B (for T, C, and G) is used for position 4 in the *RescueMu* and combined consensus sequences.

had *Mu* insertions in the ~140-bp 5′ UTR of the gene. These 62 alleles include numerous independent insertions at exactly the same nucleotide positions of the 5′ UTR. This provides strong evidence for the targeting of *Mu* transposons not only into the 5′ UTR of the *gl8* gene but also into specific positions within the 5′ UTR. Although the genetic screen used to isolate the *gl8-Mu* alleles ensured that each allele arose via an independent transposition event, the independence of alleles that contain *Mu* insertions at the same nucleotide positions was further confirmed by their (1) distinctive transposon identities, (2) transposon orientations, and (3) the wild-type progenitor alleles. For example, of the 15 independent *Mu* insertions at position 67, 4 are *MuDR* and 11 are *Mu1*. Of the 11 *Mu1* insertions, 6 are oriented left to right and 5 are oriented right to left. In addition, these 11 *Mu1*-induced alleles were derived from both Q67 and B79 progenitors. Hence, this collection of *gl8* alleles provides convincing evidence for the preferential insertion of *Mu* transposons into specific regions of a maize gene.

Because each of the *gl8-Mu* alleles in this study was

TABLE 4

Noncanonical TSDs from a collection of 369 *RescueMu* insertion sites

TSD length (bp)	No.	TSD	GenBank no.
6	3	AGAAGG	AZ921440 and AZ921438
		GAGGAG	AZ919395 and AZ919391
		GACGGG	AZ918273 and AZ918270
7	4	CAGAGAA	AZ921717 and AZ921716
		TTTAAAA	AZ919844 and AZ919843
		CTCCGGC	AZ921111 and AZ921110
		CCATGTC	AZ921405 and AZ921404
8	5	GAACAGGA	AZ919590 and AZ919587
		CGTCCCGT	AZ919268 and AZ919267
		TGGTCCGG	AZ919154 and AZ919153
		GTTGGGAG	AZ918009 and AZ918007
		TCCGCGAG	AZ918462 and AZ918461
10	6	CTTGTTTGGT	AZ920795 and AZ920794
		GTCGCCCAGG	AZ919608 and AZ919605
		GCCGCGCGCC	AZ920516 and AZ920515
		CTCATCTGAC	AZ920400 and AZ920398
		GGCCGGTGAG	AZ920273 and AZ920271
		GGCTGGCGCC	AZ918706 and AZ918705

TABLE 5

Nucleotide composition of 315 unique *RescueMu* insertion sites

	Sequence 5′ of TSD										
Nucleotide	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	Exp.[a]
C	123	118	111	104	132	92	103	189**	226**	93	110
G	*50**	61	77	78	66	72	61	*22***	*28***	*30***	73
A	56	70	65	57	45	77	75	47	*24***	57	63
T	86	66	62	76	72	74	76	57	*47**	135**	69
Sum χ^2	13**	3.6	1.4	1.7	9.6	6.8	5.5	98**	190**	88**	0
	TSD										
Nucleotide		1	2	3	4	5	6	7	8	9	Exp.
C		154**	128	150**	103	137*	126	88	*81**	162**	110
G		81	*36***	*41***	88	77	58	110**	78	83	73
A		*37***	*38***	*33***	*42**	48	83*	70	96**	*32***	63
T		*43***	113**	91*	82	53	*48**	47*	60	*38***	69
Sum χ^2		39**	57**	48**	12**	14**	19**	33**	28**	55**	0
	Sequence 3′ of TSD										
Nucleotide	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	Exp.
C	*78***	*81**	*47***	114	137*	109	109	119	109	121	110
G	59	161**	135**	75	55	75	86	73	69	74	73
A	113**	*41**	67	45	54	77	66	55	68	50	63
T	65	*32***	66	81	69	54	54	68	69	70	69
Sum χ^2	54**	140**	91**	6.6	12**	7.4	6.6	1.6	0.77	3.4	0

Composition analysis of 315 *RescueMu* insertion sites as defined by the 10 nucleotides 5′ of the 9-bp TSD, the 9-bp TSD, and the 10 nucleotides 3′ of the 9-bp TSD. Chi-square values were calculated for each individual nucleotide at all positions and the sum of those values is listed for each position in the table. The symbols * and ** indicate chi-square values or nucleotides frequencies with chi-square values that are significantly different from expected at the 95 and 99% confidence intervals, respectively. Underlined and boxed text indicate the nucleotides that occur at significantly higher-than-expected rates at the 95 and 99% confidence intervals, respectively, and therefore are consensus nucleotides at that position. Italics indicate nucleotides that are significantly underrepresented at a given position.

[a] Expected nucleotide compositions represent the number of times each nucleotide is expected to appear at each position based on the nucleotide composition of the 315 combined insertion sites.

selected on the basis of the presence of a mutant phenotype, it is likely that the observed distribution of *Mu* transposons in this collection does not represent a random sampling of insertion events in the *gl8* gene because insertions into introns or the 3′ UTR may not have resulted in a mutant phenotype. However, this phenotypic selection cannot explain the clustering of multiple independent *Mu* insertions in the 5′ UTR because *Mu* insertions elsewhere in the *gl8*-coding region also yield mutant phenotypes but were recovered at much lower frequencies. Analysis of *Mu* insertion alleles from a reverse genetics screen that did not depend upon phenotypic selection would provide a useful comparison. However, *Mu*-based reverse genetic screens typically do not yield a sufficient number of alleles per gene to provide evidence of targeting.

Another advantage of working with a defined set of independent insertion alleles generated from a related *Mu* stock is that alleles are not limited to ones that can be amplified by primers specific to known TIR sequences. If alleles that cannot be amplified with the Mu-TIR primer are present in the collection, insertions in such alleles can still be identified by DNA gel blot hybridization. Of the 79 *gl8-Mu* alleles analyzed in this study, only 1 allele failed to amplify with the Mu-TIR primer and was shown by gel blot analysis to have an insertion. This suggests that most, if not all, of the active *Mu* transposons present in at least the Schnable lab *Mu* stocks have now been identified. Of course, different *Mu* stocks may contain additional active *Mu* transposons.

In addition, because 72 of the 75 *gl8-Mu* alleles in which a *Mu* transposon was identified were derived from related *Mu* stocks maintained by the Schnable lab, the insertion frequency of each *Mu* transposon can be compared. Of the 72 insertions from the Schnable lab *Mu* stock, 41 were generated by insertions of *Mu1*, 18 by *MuDR*, 3 by *Mu8*, 1 by *Mu2*, 1 by *Mu4*, 1 by *Mu10*, 5 by *Mu11*, and 2 by *Mu12*. The overwhelming majority

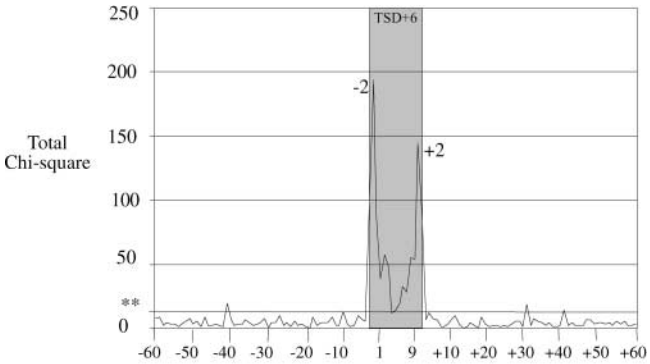

FIGURE 7.—Chi-squares were calculated at each of the 129 positions for the 315 aligned *RescueMu* insertion sequences that consisted of the TSD and 60 bp flanking each side of the TSD. The symbol ** indicates the cutoff at the 99% confidence interval with 3 d.f. The 15-bp target site (TSD +6) is indicated by a shaded box and the −2 and +2 positions that have the highest chi-square values are indicated.

(59/72 or 82%) of these insertions were of the two transposons that are typically most common in *Mu* stocks, *Mu1* and *MuDR*. The typically less common *Mu2*, *Mu4*, and *Mu8* transposons were responsible for relatively few of the *gl8-Mu* alleles. Hence, insertions in the *gl8* gene seem to occur at frequencies proportional to their abundance in typical *Mu* stocks. This is in contrast to observations involving other maize genes that appear to be preferentially targeted by specific *Mu* transposons. For example, all but 2 of the 24 *Mu*-induced *bz1* alleles that have been characterized (TAYLOR *et al.* 1986; BROWN *et al.* 1989; HARDEMAN and CHANDLER 1989, 1993; SCHNABLE *et al.* 1989; BRITT and WALBOT 1991; DOSEFF *et al.* 1991; CHANDLER and HARDEMAN 1992) contain *Mu1* or *Mu1-del* insertions. Analyses of 8 *Mu*-induced *sh1* alleles and 9 *Mu*-induced *kn1* alleles suggest that these genes may be more susceptible to *MuDR* (HARDEMAN and CHANDLER 1993) and *Mu8* (GREENE *et al.* 1994) insertions, respectively. It is of course possible that these apparent differences in susceptibility to insertion of particular classes of *Mu* transposons may be artifacts resulting from the small number of alleles analyzed or may reflect differences among the *Mu* stocks used to isolate these alleles.

The novel *Mu* transposons *Mu10*, *Mu11*, and *Mu12* were found as new insertions from an active *Mu* line and have characteristics typical of *Mu* transposons such as conserved TIRs and the ability to generate 9-bp TSDs upon insertion. The partial TIR sequences from *Mu10*, *Mu11*, and *Mu12* exhibit a high degree of identity to TIRs from known *Mu* transposons but are clearly distinct from each other and all previously identified *Mu* TIR sequences. Of the 39 bp of sequence obtained from these TIRs, 18 nucleotides are conserved among all *Mu* TIRs. Eight additional positions that are identical in the TIRs of *Mu1*, *Mu2*, *Mu3*, *Mu4*, *Mu5*, *Mu7/rcy*, *Mu8*, and *MuDR* are not conserved among the TIRs of *Mu10*, *Mu11*, and *Mu12*. Four of these eight divergent positions are part of the predicted MURA transposase-binding site that extends from nucleotides 25–56 (BENITO and

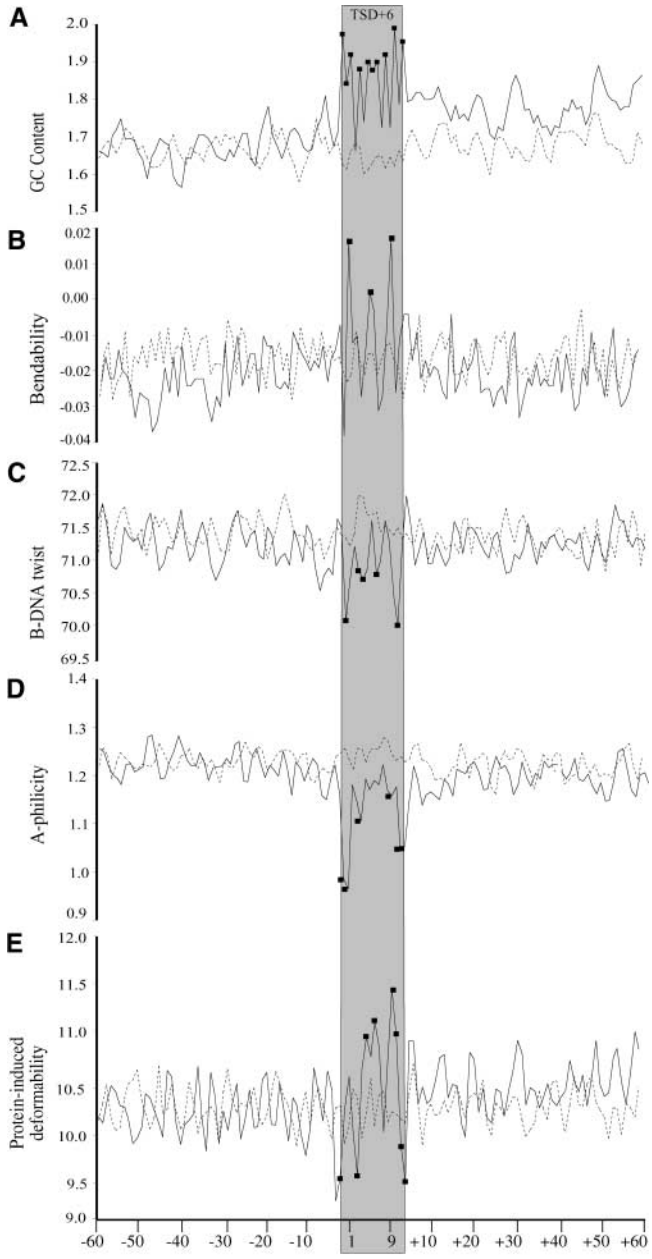

FIGURE 8.—Profiles of GC content (A), bendability (B), B-DNA twist (C), A-philicity (D), and protein-induced deformability (E) of the 315 *RescueMu* insertion sites. The average values of GC content and bendability were calculated using a window size of 3 bp and a shift increment of one; A-philicity, protein-induced deformability, and B-DNA twist were calculated using a window size of 2 bp and a shift increment of one as described at http://www.fruitfly.org/~guochun/pins.html. The solid line represents data from the *RescueMu* insertion sequences. The dashed line represents the average of 315 129-bp cDNA fragments that were randomly selected from maize ESTs. Solid squares indicate positions within the 15-bp target site that are significantly different from random DNA as determined by *t*-test analysis with $P < 0.05$.

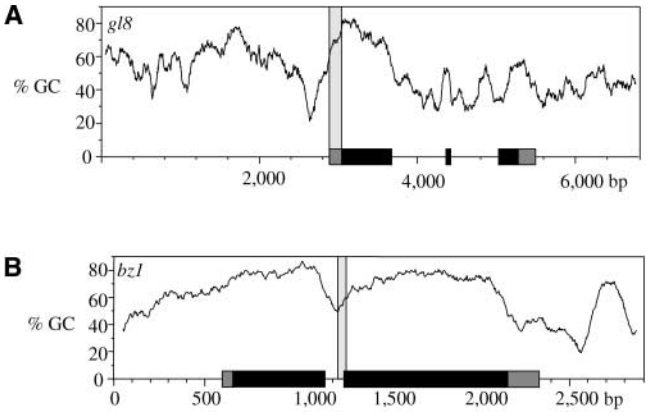

FIGURE 9.—GC content profiles from *gl8* and *bz1* suggest that GC content may influence *Mu* insertion preferences. GC profiles were generated with the WINDOW algorithm of the GCG sequence analysis package. A window size of 100 bp and a shift increment of three were used. Each GC profile is overlaid with a schematic of the appropriate gene structure. The regions of each gene that have experienced high rates of *Mu* insertion are shaded. Of the 75 *gl8* alleles analyzed in the study, 62 contain insertions in the shaded region. *Mu* transposons from each of the four *bz1* alleles that have been positioned by sequencing inserted in the shaded region shown in the *bz1* gene. Genomic sequence of *bz1* was obtained from GenBank accession no. X135000 and 5′ and 3′ UTRs were determined from the *bz1* cDNA (GenBank accession no. X13502).

WALBOT 1997). Due to the placement of the *Mu*-TIR primer used in this experiment only a small part (nucleotides 25–39) of the predicted MURA-binding site could be recovered and analyzed. A high degree of conservation in this predicted binding site is thought to be required for MURA binding because MURA fails to bind *in vitro* to the divergent left TIR of the apparently inactive *Mu5* transposon (BENITO and WALBOT 1997). However, the identification of *Mu10*, *Mu11*, and *Mu12* as new insertions from a *Mu* stock suggests that MURA can bind to the divergent TIRs of these *Mu* transposons. Isolation of the complete TIR sequences and the internal sequences of *Mu10*, *Mu11*, and *Mu12* may significantly enhance our understanding of MURA-binding specificity and *Mu* activity in general.

The identification of only three previously uncharacterized *Mu* transposons in this study suggests that the number of active *Mu* transposon family members is relatively small. Previous studies using DNA gel blotting to compare hybridization patterns obtained with *Mu* TIR specific probes to patterns obtained from probes specific to the internal regions of the cloned *Mu* transposons suggested that only 50–60% of the *Mu* TIRs in the genome are associated with known *Mu* transposons (HARDEMAN and CHANDLER 1989). If *Mu* transposons insert at rates proportional to their abundance in *Mu* stocks, as the data presented in this study suggest, the three novel *Mu* transposons identified in this study would not account for the 40–50% of *Mu* TIRs not associated with previously cloned *Mu* transposons. Hence, there are probably a large number of *Mu* transposons that are not actively inserting into genes. *Mu5* may be an example of such an inactive transposon because no *Mu5* transposons have ever been found to have caused a mutation by inserting into a gene (TALBERT *et al.* 1989). Earlier reports suggested that *Mu4* might also be inactive (TALBERT *et al.* 1989). The identification of a putative *Mu4* insertion in *gl8-Mu 91g239* suggests that *Mu4* may indeed be active, although the possibility that a novel *Mu* transposon with the same TIR sequences as *Mu4* inserted into *gl8-Mu 91g239* cannot be excluded.

Small insertions, deletions, and single-nucleotide substitutions in TSDs, such as the ones associated with these *gl8-Mu* alleles, have been identified previously in revertant alleles resulting from transposon excision (reviewed by SAEDLER and NEVERS 1985) and in somatic excision products of *Mu* transposons (BRITT and WALBOT 1991; DOSEFF *et al.* 1991). However, the aberrant insertion sites of several *gl8* alleles (Figure 5), to our knowledge, are the first reports of aberrant insertion sites (excluding deletions such as that of *gl8-Mu 91g168*) flanking resident *Mu* transposons. In addition to the noncanonical TSDs in the *gl8-Mu* insertion collection, 18 *RescueMu* insertion sites with TSDs other than 9-bp were identified from the 369 *RescueMu* insertion sites obtained from GenBank sequences (Table 4). The 6-, 7-, and 8-bp TSDs are likely the result of 1–3 nucleotide deletions from one of the original 9-bp TSDs. These insertion sites are therefore similar to the insertion site of *gl8-Mu 91g239* that has an 8-bp TSD that appears to have resulted from a deletion of a G from the right TSD. The 10-bp TSDs could result from the addition of a single nucleotide at one of the TSD/TIR borders that, by chance, matched the corresponding nucleotide in the opposite flanking sequence. Alternatively, the 10-bp TSD could result from a 10-bp duplication instead of the typical 9-bp duplication. These two possibilities cannot be distinguished because the progenitor sequences are not available for the *RescueMu* insertions. In addition, sequencing errors cannot be ruled out as the cause of the noncanonical TSD associated with the *RescueMu* insertion sites. However, it is unlikely that all of the variation in *RescueMu* TSD length is a consequence of sequencing errors. This can be inferred from a higher frequency of recovery of TSDs with variations in length (18/369) than the frequency of recovery of insertion sites with mismatched TSDs (9/369), which provides an upper bound for sequencing errors. The recovery of noncanonical TSDs flanking both resident *Mu* and *RescueMu* transposons suggests that such TSDs can result from aberrant insertion events, although aborted transposition events cannot be ruled out. Variation in TSD length has been reported from *Mu*-like transposons in Arabidopsis (YU *et al.* 2000) and rice (TURCOTTE *et al.* 2001) but has not been previously reported for *Mu* transposons.

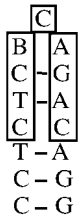

Figure 10.—Potential hairpin loop formed by base pairing within the 15-bp target site centered on position 5 of the 9-bp TSD. The nine positions that represent the TSD are boxed.

The tendency for *Mu* transposons to preferentially insert into genic regions has now become fairly well established (Hanley *et al.* 2000; Raizada *et al.* 2001) yet the mechanism for this preference is not clear. Prior analyses of the 9-bp TSDs that occur at the site of *Mu* insertions have produced similar consensus sequences (Cresse *et al.* 1995; Hanley *et al.* 2000). This study reports two additional consensus sequences from independent collections of *Mu* insertion sites that are strikingly similar to those reported previously. Consistencies among these consensus sequences derived from independent collections of *Mu* insertion events suggest some biological significance to the consensus sequences. The most obvious characteristic of these consensus sequences is that they are GC rich. It is not yet clear, however, whether this GC richness is a cause or an effect of the tendency of *Mu* transposons to insert into genes, which are themselves typically GC rich (Salinas *et al.* 1988; Carels *et al.* 1998).

The composition of the sequences flanking TSDs has not been reported in previous studies involving *Mu* target sites. Statistical analyses performed in this study, however, identified several highly conserved nucleotides in the positions directly flanking the 9-bp TSDs of both the *gl8-Mu* and *RescueMu* insertion sites. Each of the three positions flanking both sides of the TSD from the *RescueMu* insertion sequences had a consensus nucleotide that differed from its expected nucleotide frequency at the 99% confidence interval. In particular, the conservation of the complementary C and G at positions −2 and +2 is very common in these data sets. For example, each of the three *gl8* target sites that experienced the highest numbers of independent *Mu* insertion events (nucleotide positions 67, 77, and 33, which had 15, 9, and 7 insertions, respectively) has a −2 C and a +2 G. In addition, the two other reported sites of multiple independent *Mu* insertions, three *bz1* alleles (*bz1-rcy*, *bz1-mu1*, and *bz1-A58*) and two *Kn1* alleles (*Kn1-mum7* and *Kn1-mum1*), also have the conserved −2 C and +2 G. This conservation of C and G at the −2 and +2 position appears to hold true for other collections of *Mu* insertions as well. For example, reanalysis of the insertion sites of the 23 *Mu* insertion alleles reviewed by Chandler and Hardeman (1992) revealed that 12 have both a C and a G at positions −2 and +2 and all but 2 of these insertion sites have either a C at −2 or a G at +2. In addition, of the 13 randomly isolated *Mu* insertion alleles reported by Cresse *et al.* (1995), in which flanking sequences were reported for both sides of the *Mu* insertion, 6 had a C and a G at positions −2 and +2 and all 13 had at least a −2 C or a +2 G.

Combining the three conserved nucleotides flanking the TSD with the consensus 9-bp TSD (from the *RescueMu* sequences) generates a 15-bp consensus target site for *Mu* insertions of 5′ CCT::CTCBCAGAC::AGG 3′. This 15-bp target site includes five pairs of complementary nucleotides. Although it is tempting to speculate that this target site could allow for the formation of a hairpin loop that could serve as a recognition site for *Mu* insertion (Figure 10), individual insertion sites from the *gl8-Mu* and *RescueMu* data rarely include inverted repeat structures that would allow such hairpins to form. However, to determine whether DNA secondary structure may play some role in *Mu* insertion preferences, measurements of physical properties of DNA such as bendability (Brukner *et al.* 1995), A-philicity (Lu *et al.* 2000), B-DNA twist (Gorin *et al.* 1995), and protein-induced deformability (Olson *et al.* 1998) were determined. Analysis of these physical properties in DNA flanking 467 *P*-element insertion sites revealed significant differences in each of the properties compared to random DNA (Liao *et al.* 2000). Although this type of analysis is not appropriate for the collection of *gl8* insertion sites because each allele, although independent, is from the same gene, a similar analysis of the data from the *RescueMu* insertion sites was performed. As with the *P*-element insertion sites, *RescueMu* insertion sites showed significant differences in each of the physical properties at insertion sites compared to random DNA (Figure 8). An increase in bendability of DNA is associated with susceptibility to DNA cleavage. For example, regions of DNA that are prone to cleavage by DNaseI have been shown to have increased bendability (Brukner *et al.* 1995). In addition, bent DNA has been shown to create favorable sites for retroviral integration (Muller and Varmus 1994) as well as for *mariner* DNA transposon insertion (Lampe *et al.* 1998). The propensity for DNA to be in the A-DNA *vs.* B-DNA form also influences DNA-protein interactions (Otwinowski *et al.* 1988). A-DNA is more relaxed than the more common B-DNA form and therefore the sizes of the major and minor grooves differ between A- and B-DNA. In addition, A-DNA is thought to allow increased hydrogen bonding while B-DNA may allow interactions with larger exocyclic groups such as NH_2-NH_2 (Gorin *et al.* 1995). B-DNA twist and A-philicity are independent methods to determine the propensity of DNA to form A-DNA. The low B-DNA twist and low A-philicity at the *RescueMu* insertion sites indicate that hydrogen bonding may not be significant in binding of proteins involved in *Mu* transposition to the target site. Protein-induced deformability measures the change in DNA conformation in response to protein binding. The significant differences in protein-induced deformability at the *RescueMu* insertion sites are an indication that the conformation of

insertion sites has the potential to change in response to protein binding. In summary, these results suggest that, like *P* elements, *Mu* transposons may rely, at least to some extent, on DNA secondary structure to select insertion sites. However, although *RescueMu* transposons appear to mimic the behavior of natural *Mu* transposons in all aspects thus far studied (Raizada *et al.* 2001; and this study), because *RescueMu* transposons are not natural *Mu* transposons, conclusions about *Mu* insertion preferences cannot be made solely on the basis of sequences recovered from *RescueMu*.

The identification of conserved nucleotides at positions directly flanking the 9-bp TSD in the collection of *gl8-Mu* insertion sites and confirmed by analysis of *RescueMu* insertion sites does not offer an explanation for the dramatic targeting of *Mu* transposons to the 5′ UTR of *gl8*. For example, the *gl8* 5′ UTR does not contain a perfect consensus site that could explain the targeting. The only obvious feature of the *gl8* sequence that appears to be well correlated to increased *Mu* insertion frequency is the presence of tandem repeats of a CACNG motif. Although the correlation between the occurrence of the CACNG motifs and high *Mu* insertion activity in the *gl8* gene is quite high, the challenge is to determine if the correlation is merely by coincidence or if there is a biological significance to the correlation. If the correlation has biological significance, then sequences associated with other regions that experience preferential *Mu* insertion would also be expected to have a high frequency of CACNG motifs. Unfortunately, only three additional regions have been identified that may have potential insertion preference for *Mu* transposons. These regions occur in the *kn1*, *bz1*, and *adh1* genes (reviewed by Bennetzen *et al.* 1993). CACNG motifs were not found in tandem repeats in any of these genes and, in addition, CACNG motifs were not found at increased frequencies among the *RescueMu* insertion-site sequences. Hence, the correlation between preferential *Mu* insertion and CACNG motifs in the *gl8* 5′ UTR appears to be coincidental.

GC profiles were also compared among *gl8* and the three other genes that have possible preferential sites for *Mu* insertion. The preferential *Mu* insertion sites for *gl8* and *bz1* occur in a transition region between low and high GC content (Figure 9). A similar pattern in the GC profiles in the preferred insertion region of *kn1* or *adh1* was not observed. However, the fact that the *kn1* and *adh1* alleles were selected on the basis of their ability to condition dominant and partial inactivation phenotypes, respectively, may have been the driving force for the observed targeting in these genes (reviewed by Bennetzen 1996). The collections of *gl8* and *bz1* alleles are also biased by their ability to condition a mutant phenotype; however, although insertions can be recovered throughout the coding regions, there is an apparent bias in both genes for *Mu* insertion into specific regions of the genes that are transitions between regions of low and high GC content. The average GC profile of the *RescueMu* insertion sites also shows a statistically significant difference in GC content of the left and right flanking sequences, indicating that *RescueMu* transposons also tend to insert into transition regions in GC content. The large difference in GC content between the AT-rich 5′ region of *gl8* and the exceptionally GC-rich region of exon 1 may offer an explanation for the dramatic targeting of *Mu* insertions into the 5′ UTR of *gl8*.

This suggests that *Mu* preferential insertion may be a common phenomenon within the maize genome that has not been previously detected due to the manner in which *Mu* insertion data are obtained and analyzed. For example, a recent analysis of 80 *Mu*-flanking sequences selected from a collection of 761 random *Mu* insertions did not find a preference for *Mu* insertion in coding *vs.* noncoding regions within transcribed sequences (Hanley *et al.* 2000). One possible explanation for this discrepancy is that the 80 *Mu*-flanking sequences used in the analysis by Hanley *et al.* (2000) were selected on the basis of their sequence similarity to ESTs. Because EST collections are typically enriched for 3′ sequences, *Mu* insertions in 5′ regions may have been systematically excluded from the Hanley *et al.* (2000) study. This may be especially significant as large transitions in GC content may be more common in the 5′ regions of genes as a result of the tendency for 5′ regions to be AT rich as opposed to the GC richness of the first exon. In addition, only unique insertion sites were selected for the Hanley *et al.* (2000) analysis. Therefore, multiple insertion events, such as those observed for the *gl8* gene, would not have been included in the analysis. Another possibility is that while insertion hot spots exist in some genes, they may not always occur in the same regions of all genes. This would explain why collections of randomly isolated *Mu* insertions associated with many different genes do not show insertion-site preferences, while collections of alleles isolated from at least some single genes show insertion-site preferences. Hence a more frequent occurrence of *Mu* insertion hotspots cannot be ruled out on the basis of the available collections of *Mu* insertion events. However, it also cannot be excluded that the targeting observed is unique to *gl8* and is a consequence of properties inherent in *gl8*, such as its map position, expression pattern, or chromatin structure.

We thank Manish Raizada (University of Guelph) for discussions regarding *RescueMu* data. This research was funded by grants from the National Science Foundation (IBN-9808559 and DBI-9975868) to P. S. Schnable, B. J. Nikolau, and Schnable *et al.* (P. S. Schnable, D. Ashlock, G. Churchill, X. Gu, K. Lamkey, M. Lee, and G. Naylor). M. Packila's (Bemidji State University, Bemidji, MN) participation in a summer internship in the Schnable laboratory was partially supported by the National Science Foundation-Research Experience for Undergraduates (NSF-REU) site grant DBI 9732256 to D. Oliver and P. Chitnis and by journal paper no. J-19221 of the Iowa Agriculture and Home Economic Experiment Station, Ames, Iowa, project no. 3409, and by the Hatch Act and State of Iowa.

Note added in proof: Data regarding the function of the GL8 protein are provided in Xu *et al.* (2002).

LITERATURE CITED

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

Barker, R. F., D. V. Thompson, D. R. Talbot, J. Swanson and J. L. Bennetzen, 1984 Nucleotide sequence of the maize transposable element *Mu1*. Nucleic Acids Res. **12:** 5955–5967.

Benito, M. I., and V. Walbot, 1997 Characterization of the maize *Mutator* transposable element *MURA* transposase as a DNA-binding protein. Mol. Cell. Biol. **17:** 5165–5175.

Bennetzen, J. L., 1996 The *Mutator* transposable element system of maize. Curr. Top. Microbiol. Immunol. **204:** 195–229.

Bennetzen, J. L., R. P. Fracasso, D. W. Morris and D. S. Robertson, 1987 Concomitant regulation of *Mu1* transposition and *Mutator* activity in maize. Mol. Gen. Genet. **208:** 57–62.

Bennetzen, J. L., P. S. Springer, A. D. Cresse and M. Hendrickx, 1993 Specificity and regulation of the *Mutator* transposable element system in maize. Crit. Rev. Plant Sci. **12:** 57–95.

Berg, D. E., and M. L. Howe, 1989 *Mobile DNA*. American Society of Microbiology, Washington, DC.

Britt, A. B., and V. Walbot, 1991 Germinal and somatic products of *Mu1* excision from the *Bronze-1* gene of *Zea mays*. Mol. Gen. Genet. **227:** 267–276.

Brown, W. E., D. S. Robertson and J. L. Bennetzen, 1989 Molecular analysis of multiple *Mutator*-derived alleles of the *Bronze* locus of maize. Genetics **122:** 439–445.

Brukner, I., R. Sanchez, D. Suck and S. Pongor, 1995 Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. EMBO J. **14:** 1812–1818.

Bureau, T. E., and S. R. Wessler, 1992 *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell **4:** 1283–1294.

Bureau, T. E., and S. R. Wessler, 1994 *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell **6:** 907–916.

Capy, P., C. Bazin, D. Higuet and T. Langin, 1998 *Dynamics and Evolution of Transposable Elements*. Chapman & Hall, New York.

Carels, N., P. Hatey, K. Jabbari and G. Bernardi, 1998 Compositional properties of homologous coding sequences from plants. J. Mol. Evol. **46:** 45–53.

Chandler, V. L., and K. J. Hardeman, 1992 The *Mu* elements of *Zea mays*. Adv. Genet. **30:** 77–122.

Chen, C. H., K. K. Oishi, B. Kloeckener-Gruissem and M. Freeling, 1987 Organ-specific expression of maize *Adh1* is altered after a *Mu* transposon insertion. Genetics **116:** 469–477.

Chomet, P., D. Lisch, K. J. Hardeman, V. L. Chandler and M. Freeling, 1991 Identification of a regulatory transposon that controls the *Mutator* transposable element system in maize. Genetics **129:** 261–270.

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

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

Dellaporta, S. L., J. Wood and J. B. Hicks, 1983 A plant DNA minipreparation: version II. Plant Mol. Biol. Rep. **1:** 19–22.

Doseff, A., R. Martienssen and V. Sundaresan, 1991 Somatic excision of the *Mu1* transposable element of maize. Nucleic Acids Res. **19:** 579–584.

Emerson, R. A., G. W. Beadle and A. C. Fraser, 1935 A summary of linkage studies in maize. Cornell Univ. Agric. Exp. Stn. Memoir **180:** 1–83.

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

Goldberg, D. E., 1989 *{G}enetic {A}lgorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA.

Gorin, A. A., V. B. Zhurkin and W. K. Olson, 1995 B-DNA twisting correlates with base-pair morphology. J. Mol. Biol. **247:** 34–48.

Greene, B., R. Walko and S. Hake, 1994 *Mutator* insertions in an intron of the maize *knotted1* gene result in dominant suppressible mutations. Genetics **138:** 1275–1285.

Hanley, S., D. Edwards, D. Stevenson, S. Haines, M. Hegarty *et al.*, 2000 Identification of transposon-tagged genes by the random sequencing of *Mutator*-tagged DNA fragments from *Zea mays*. Plant J. **22:** 557–566.

Hardeman, K. J., and V. L. Chandler, 1989 Characterization of *bz1* mutants isolated from *Mutator* stocks with high and low numbers of *Mu1* elements. Dev. Genet. **10:** 460–472.

Hardeman, K. J., and V. L. Chandler, 1993 Two maize genes are each targeted predominantly by distinct classes of *Mu* elements. Genetics **135:** 1141–1150.

Hershberger, R. J., C. A. Warren and V. Walbot, 1991 *Mutator* activity in maize correlates with the presence and expression of the *Mu* transposable element *Mu9*. Proc. Natl. Acad. Sci. USA **88:** 10198–10202.

Hsia, A. P., and P. S. Schnable, 1996 DNA sequence analyses support the role of interrupted gap repair in the origin of internal deletions of the maize transposon, *MuDR*. Genetics **142:** 603–618.

Lampe, D. J., T. E. Grant and H. M. Robertson, 1998 Factors affecting transposition of the Himar1 mariner transposon *in vitro*. Genetics **149:** 179–187.

Levy, A. A., and V. Walbot, 1991 Molecular analysis of the loss of somatic instability in the *bz2::mu1* allele of maize. Mol. Gen. Genet. **229:** 147–151.

Liao, G. C., E. J. Rehm and G. M. Rubin, 2000 Insertion site preferences of the P transposable element in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **28:** 3347–3351.

Lu, X. J., Z. Shakked and W. K. Olson, 2000 A-form conformational motifs in ligand-bound DNA structures. J. Mol. Biol. **300:** 819–840.

Maniatis, T., E. F. Fritsch and J. Sambrook, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Muller, H. P., and H. E. Varmus, 1994 DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:** 4704–4714.

Nevers, P., N. S. Shepherd and H. Saedler, 1985 Plant transposable elements. Adv. Bot. Res. **12:** 102–103.

Olson, W. K., A. A. Gorin, X. J. Lu, L. M. Hock and V. B. Zhurkin, 1998 DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. USA **95:** 11163–11168.

Otwinowski, Z., R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak *et al.*, 1988 Crystal structure of trp repressor/operator complex at atomic resolution. Nature **335:** 321–329.

Peterson, P. A., 1988 The mobile element systems in maize, pp. 43–68 in *Plant Transposable Elements*, edited by O. Nelson. Plenum Press, New York.

Qin, M. M., D. S. Robertson and A. H. Ellingboe, 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.

Raizada, M. N., and V. Walbot, 2000 The late developmental pattern of *Mu* transposon excision is conferred by a cauliflower mosaic virus 35S-driven MURA cDNA in transgenic maize. Plant Cell **12:** 5–21.

Raizada, M. N., G. L. Nan and V. Walbot, 2001 Somatic and germinal mobility of the *RescueMu* transposon in transgenic maize. Plant Cell **13:** 1587–1608.

Robertson, D. S., 1978 Characterization of a *Mutator* system in maize. Mutat. Res. **51:** 21–28.

Rogers, S. O., and A. J. Bendich, 1985 Extraction of DNA from milligram amounts of fresh herbarium and mummified plant tissues. Plant Mol. Biol. **5:** 69–76.

Rowland, L. J., and J. N. Strommer, 1985 Insertion of an unstable element in an intervening sequence of maize *Adh1* affects transcription but not processing. Proc. Natl. Acad. Sci. USA **82:** 2875–2879.

Saedler, H., and P. Nevers, 1985 Transposition in plants: a molecular model. EMBO J. **4:** 585–590.

Saghai-Maroof, M. A., K. M. Sloiman, R. A. Jorgensen and R. W. Allard, 1984 Ribosomal DNA spacer-length polymorphisms

in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA **81:** 8014–8018.

Salinas, J., G. Matassi, L. M. Montero and G. Bernardi, 1988 Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. Nucleic Acids Res. **16:** 4269–4285.

Schnable, P. S., and P. A. Peterson, 1986 Distribution of genetically active Cy transposable elements among diverse maize lines. Maydica **31:** 59–81.

Schnable, P. S., P. A. Peterson and H. Saedler, 1989 The *bz-rcy* allele of the *Cy* transposable element system of *Zea mays* contains a *Mu*-like element insertion. Mol. Gen. Genet. **217:** 459–463.

Schnable, P. S., P. S. Stinard, T. J. Wen, S. Heinen, D. Weber *et al.* 1994 The genetics of cuticular wax biosynthesis. Maydica **39:** 279–287.

Shepherd, N. S., W. F. Sheridan, M. G. Mattes and G. Deno, 1988 The use of *Mutator* for gene-tagging: cross-referencing between transposable element systems, pp. 137–147 in *Plant Transposable Elements*, edited by O. Nelson. Plenum Press, New York.

Spradling, A. C., D. M. Stern, I. Kiss, J. Roote, T. Laverty *et al.*, 1995 Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc. Natl. Acad. Sci. USA **92:** 10824–10830.

Stinard, P. S., and P. S. Schnable, 1993 Four-point linkage data for *ae pr lw2 gl8* on 5L. Maize Genet. Coop. Newsl. **67:** 8–9.

Stinard, P. S., D. S. Robertson and P. S. Schnable, 1993 Genetic isolation, cloning, and analysis of the *Mutator*-induced, dominant antimorph of the maize amylose extender1 locus. Plant Cell **5:** 1555–1566.

Talbert, L. E., G. I. Patterson and V. L. Chandler, 1989 *Mu* transposable elements are structurally diverse and distributed throughout the genus *Zea*. J. Mol. Evol. **29:** 28–39.

Taylor, L. P., and V. Walbot, 1987 Isolation and characterization of a 1.7-kb transposable element from a Mutator line of maize. Genetics **117:** 297–307.

Taylor, L. P., V. L. Chandler and V. Walbot, 1986 Insertion of 1.4 KB and 1.7 KB *Mu* elements into the *bronze1* gene of *zea mays* L. Maydica **31:** 31–45.

Turcotte, K., S. Srinivasan and T. Bureau, 2001 Survey of transposable elements from rice genomic sequences. Plant J. **25:** 169–179.

Walbot, V., 1991 The *Mutator* transposable element family of maize. Genet. Eng. **13:** 1–37.

Walbot, V., 1992 Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. Annu. Rev. Plant Physiol. **43:** 49–82.

Xu, X., C. R. Dietrich, M. Delledonne, Y. Xia, T. J. Wen *et al.*, 1997 Sequence analysis of the cloned *glossy8* gene of maize suggests that it may code for a beta-ketoacyl reductase required for the biosynthesis of cuticular waxes. Plant Physiol. **115:** 501–510.

Xu, X., C. R. Dietrich, R. Lessire, B. J. Nikolau, P. S. Schnable, 2002 The endoplasmic reticulum-associated maize GL8 protein is one of the components of the acyl-CoA elongase complex involved in the production of cuticular waxes. Plant Physiol. (in press).

Yu, Z., S. I. Wright and T. E. Bureau, 2000 *Mutator*-like elements in *Arabidopsis thaliana*: structure, diversity and evolution. Genetics **156:** 2019–2031.

Zhang, Q., J. Arbuckle and S. R. Wessler, 2000 Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. Proc. Natl. Acad. Sci. USA **97:** 1160–1165.

Communicating editor: J. A. Birchler