MOLECULAR AND CELLULAR BIOLOGY, Feb. 1988, p. 737–746
0270-7306/88/020737-10$02.00/0

Vol. 8, No. 2

Insertion and Excision of *Caenorhabditis elegans* Transposable Element Tc1

DAVID EIDE† AND PHILIP ANDERSON*

Department of Genetics, University of Wisconsin, 445 Henry Mall, Madison, Wisconsin 53706

Received 8 September 1987/Accepted 10 November 1987

The transposable element Tc1 is responsible for most spontaneous mutations that occur in *Caenorhabditis elegans* variety Bergerac. We investigated the genetic and molecular properties of Tc1 transposition and excision. We show that Tc1 insertion into the *unc-54* myosin heavy-chain gene was strongly site specific. The DNA sequences of independent Tc1 insertion sites were similar to each other, and we present a consensus sequence for Tc1 insertion that describes these similarities. We show that Tc1 excision was usually imprecise. Tc1 excision was imprecise in both germ line and somatic cells. Imprecise excision generated novel *unc-54* alleles that had amino acid substitutions, amino acid insertions, and, in certain cases, probably altered mRNA splicing. The DNA sequences remaining after Tc1 somatic excision were the same as those remaining after germ line excision, but the frequency of somatic excision was at least 1,000-fold higher than that of germ line excision. The genetic properties of Tc1 excision, combined with the DNA sequences of the resulting *unc-54* alleles, demonstrated that excision was dependent on Tc1 transposition functions in both germ line and somatic cells. Somatic excision was not regulated in the same strain-specific manner as germ-line excision was. In a genetic background where Tc1 transposition and excision in the germ line was not detectable, Tc1 excision in the soma still occurred at high frequency.

Transposable elements are responsible for a variety of genetic phenomena in many organisms (for a review, see reference 62). They are a major source of spontaneous mutations because of their ability to cause insertions, deletions, and other DNA rearrangements. In addition to eliminating gene function, insertion of a transposable element within or near a gene can alter the level, tissue specificity, or developmental timing of gene expression (24, 41, 52). Excision of transposable elements can also dramatically affect gene activity. Excision is imprecise for most transposons, and this results in sequence alterations remaining at the sites of excision (2, 5, 9, 12, 53, 58, 61, 67, 69). Imprecise excision of transposable elements can lead to proteins that have altered molecular weights (63), enzymatic activities (14, 69), and developmental patterns of gene expression (10, 40). The insertion and excision of transposable elements, therefore, might be a significant source of genetic variability in nature (for a review, see reference 68).

The genetic activity of transposable elements is often tightly controlled. The frequencies of transposition and excision can be regulated in a strain- and tissue-specific manner. For example, transposition and excision of certain maize transposons depend on the presence of other elements within the genome (for a review, see reference 25). *Drosophila* P factors transpose and excise only in the germ line (20) and only after certain interstrain crosses (4; for a review, see reference 21). Eucaryotic transposons usually encode proteins required for their own transposition and excision. Such activities are generally referred to as transposases, although their roles in the transposition process are unknown. The regulation of transposase activity is central to the regulation of transposable element activity. For example, the strain and tissue specificities of P factor transposition and excision are due to regulation of transposase activity (22, 36).

Transposition of the element Tc1, contained in the nematode *Caenorhabditis elegans*, is regulated in a strain- and tissue-specific manner. Multiple copies of Tc1 are present in the genomes of all wild isolates of *C. elegans* (19, 37), yet Tc1 transposes at a detectable frequency in only certain strains (15, 16, 43, 45). In the *C. elegans* wild-type variety Bergerac, transposition of Tc1 is responsible for most spontaneous mutations. In other wild-type varieties (e.g., Bristol and DH424), transposition of Tc1 is rare or not detectable.

Excision of Tc1 is also regulated in a strain- and tissue-specific manner. Tc1-induced mutations often revert toward a wild-type phenotype after excision of the element (15, 43). Reversion of such mutations in the germ line occurs only in strains that are active for Tc1 transposition (43, 45). Thus, germ line excision of Tc1 appears to be transposase dependent. The frequency of Tc1 excision in the soma is much higher than that in the germ line and can be detected by both genetic (15) and biochemical (18) techniques.

We reported previously the isolation and genetic properties of Tc1 insertions affecting the *C. elegans unc-54* gene (15), which encodes one of two myosin heavy-chain isozymes expressed in body wall muscle cells (23, 39). In this paper, we consider in more detail the genetic and molecular properties of Tc1 transposition and excision. We determined the DNA sequences at the insertion sites of 11 independent *unc-54::Tc1* mutations. We sequenced 20 *unc-54* alleles generated either by germ line or by somatic excision of Tc1. We investigated the genetic properties of *unc-54::Tc1* mutations in both germ line and somatic cells and interpret these genetic properties in light of the DNA sequences of Tc1 insertions and excisions. Finally, we investigated the regulation of Tc1 transposition and excision in different tissues and for different genetic backgrounds.

MATERIALS AND METHODS

Genetic procedures. The conditions for growth of *C. elegans* and methods for its genetic manipulation have been described previously (6). *unc-54::Tc1* mutants are descen-

* Corresponding author.
† Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

TABLE 1. Frequencies of germ line and somatic reversion of *unc-54::Tc1* mutants

Strain[a]	Genotype	Frequency of *unc-54*$^+$ revertants	
		Germ line[b]	Somatic
TR452	*unc-54(323::Tc1)*	3×10^{-5}	7×10^{-3}
TR454	*unc-54(r328::Tc1)*	1×10^{-5}	6×10^{-3}
TR670	*unc-54(r362::Tc1)*	4×10^{-5}	3×10^{-2}
TR672	*unc-54(r366::Tc1)*	1×10^{-5}	3×10^{-3}
TR669	*unc-54(r387::Tc1)*	9×10^{-6}	1×10^{-2}
TR668	*unc-54(r388::Tc1)*	6×10^{-6}	2×10^{-3}
TR667	*unc-54(r390::Tc1)*	2×10^{-5}	7×10^{-3}
TR451	*unc-54(r322::Tc1)*	$<6 \times 10^{-7}$	9×10^{-4}
TR666	*unc-54(r361::Tc1)*	$<6 \times 10^{-7}$	6×10^{-4}
TR671	*unc-54(r327::Tc1)*	$<6 \times 10^{-7}$	$<2 \times 10^{-4}$
TR656	*unc-54(r360::Tc1)*	$<6 \times 10^{-7}$	$<2 \times 10^{-4}$

[a] Strains having insertions at an identical position in *unc-54* are grouped together.

[b] Germ line reversion frequencies were estimated by Poisson analysis of the distribution of revertants among independent cultures (see Materials and Methods).

dants of strain EM1002 (18), our canonical *C. elegans* variety Bergerac stock. *unc-54::Tc1* mutants are selected as partial revertants of the strain *unc-105(n490) II* (15, 50). We removed the *unc-105* mutation from the primary isolate of each insertion by crossing each strain with wild-type Bergerac males; the *unc-54* insertions were then segregated from *unc-105*. For three insertions (*r387*, *r388*, and *r390*), *unc-105* was removed by first crossing the primary isolates with wild-type Bristol males. The resulting *unc-54* single mutants were then crossed six times with Bergerac males. In this way, we isolated each *unc-54* mutation in Bergerac genetic backgrounds that are essentially isogenic.

Germ line reversion frequencies were estimated by Poisson analysis of the distribution of wild-type revertants among 20 to 40 independent cultures. The proportion of cultures that did not contain revertants (P_0) was used to calculate each frequency. ($P_0 = e^{-an}$; a = reversion frequency; n = number of animals per culture.) This method corrects for revertants that occur early during growth of a culture and for multiple revertants that occur within a single culture. When no revertants were obtained in any culture, we concluded that the frequencies were less than the reciprocal of the total number of worms screened (Tables 1 and 2). Because of sampling error, however, such estimates are only approximate (the confidence intervals are about 63%). The upper limits for the 95% confidence intervals are approximately threefold higher than the frequencies shown in Tables 1 and 2. Data for frequency calculations are as follows: n = 40,000 for all cultures; P_0(TR451) = 40/40; P_0(TR452) = 6/20; P_0(TR454) = 13/20; P_0(TR623) = 3/13; P_0(TR656) = 40/40; P_0(TR657) = 37/40; P_0(TR666) = 40/40; P_0(TR667) = 8/20; P_0(TR668) = 16/20; P_0(TR669) = 14/20; P_0(TR670) = 4/20; P_0(TR671) = 40/40; P_0(TR672) = 12/20; P_0(TR759) = 50/50; and P_0(TR1299)= 46/47.

Somatic revertants have a characteristic phenotype (15), and their frequency was estimated by counting the number of egg-laying, partial revertants among total gravid hermaphrodites. The somatic reversion frequencies shown in Tables 1 and 2 are based on 6 to 19 independent events. Somatic revertants were not observed for strains TR671 and TR656. We concluded that their frequencies were less than the reciprocal of the total number of worms screened (Table 1). As discussed above, these estimates are approximate.

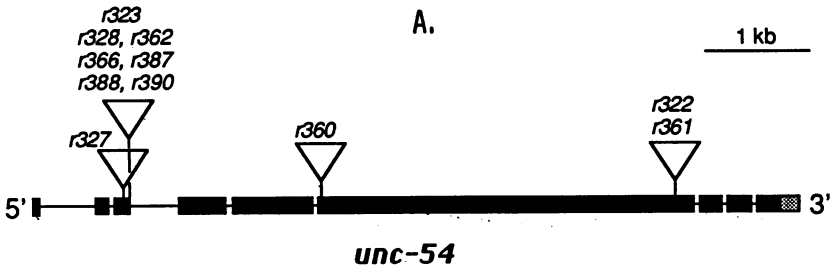
To substitute the *unc-54(r323::Tc1)* mutation in the DH424 genetic background, we crossed the Bergerac strain TR452 [genotype *unc-54(r323::Tc1)*] with wild-type DH424 males. Cross-progeny hermaphrodites were picked, and homozygous *unc-54* mutants were isolated in the next generation, yielding strain TR623. This procedure was repeated a second and third time, yielding strains TR657 and TR759, respectively. We continued the backcrossing series for a total of 10 crosses with DH424. The Bristol mutation *lev-11(x12)* was introduced after backcross 5, and a *lev-11 unc-54* recombinant was identified. (*lev-11* is approximately 1 map unit left of *unc-54* on LGI.) A crossover that separated *lev-11* from *unc-54* was identified after backcross 6, and the resulting *unc-54* single mutant was crossed four additional times with DH424, yielding strain TR1299. This procedure isolated the Bergerac *unc-54(r323::Tc1)* region from the remainder of chromosome I by two recombination events. *unc-54* and terminal regions of LGI in TR1299 are derived from Bergerac, but the remainder of the genome is DH424.

Biochemical methods. Our methods for growing nematodes, extracting DNA, and cloning *unc-54::Tc1* alleles into lambda vectors have been described previously (16, 65). Germ line revertant alleles were cloned in a similar manner, except that *Sal*I complete digestion products were ligated to *Sal*I-digested EMBL3 lambda DNA (27). For cloning somatic excision alleles, we prepared nematode DNA from cultures that were enriched for somatic DNA and empty sites. Somatic excision alleles were cloned by ligating *Xba*I-*Sal*I double-digestion products to *Xba*I-digested lambda 2149 DNA (32). We obtained somatic excision clones TR#25 through TR#28 from a culture of strain TR451 [genotype *unc-54(r322::Tc1)*] that had been grown to starvation and incubated for 14 days without food. Starvation prevents reproduction, and somatic excisions accumulate in the population (17). We obtained somatic excision clones TR#29 through TR#33 from a culture of strain TR814 [genotype *lin-27(b151) unc-54(r323::Tc1)*]. The mutation *lin-27(b151)* causes sterility and greatly reduced numbers of germ cells (J. Kimble, personal communication). A culture of strain TR814 was grown to starvation at 20°C, refed, and shifted to 25°C for 4 days. We then harvested the animals and extracted their DNA. Southern blots indicated that as much as 40% of the *unc-54* gene copies in TR814 can be empty sites. We visually screened TR451 and TR814 cultures before harvesting to ensure the absence of phenotypic revertants. Somatic excision clones were isolated randomly from the somatic

TABLE 2. Effects of genetic background on germ line and somatic reversion of *unc-54(r323::Tc1)*

Strain[a]	Genotype	No. of backcrosses with DH424	Frequency of *unc-54*$^+$ revertants	
			Germ line	Somatic
TR452	*unc-54(r323::Tc1)*	0	3×10^{-5}	7×10^{-3}
TR623	*unc-54(r323::Tc1)*	1	4×10^{-5}	6×10^{-3}
TR657	*unc-54(r323::Tc1)*	2	2×10^{-6}	7×10^{-3}
TR759	*unc-54(r323::Tc1)*	3	$<5 \times 10^{-7}$	3×10^{-3}
TR1299	*unc-54(r323::Tc1)*	10	5×10^{-7}	2×10^{-3}

[a] Strains were derived by repeatedly crossing the Bergerac mutation *unc-54(r323::Tc1)* with wild-type variety DH424 (see Materials and Methods). After each backcross, an increasing fraction of the genome is DH424 material. The frequencies of germ line reversion were estimated by Poisson analysis of the distribution of revertants among independent cultures (see Materials and Methods).



B.

unc-54$^+$	1806-GTCCTCCATAACTTGCGT-1823
r327	GTCCTCCATA<CAGTGCTGG...Tc1...CCAGCACTG>TAACTTGCGT
unc-54$^+$	1840-TGCTCATCTACGTAAGTT-1857
r323, r328, r362, r366, r387, r388, r390	TGCTCATCTA<CAGTGCTGG...Tc1...CCAGCACTG>TACGTAAGTT
unc-54$^+$	3705-CCACTCGGTATTATCTCC-3722
r360	CCACTCGGTA<CAGTGCTGG...Tc1...CCAGCACTG>TATTATCTCC
unc-54$^+$	7084-TGAAGAGATACCAAGAGC-7101
r322	TGAAGAGATA<CAGTGCTGG...Tc1...CCAGCACTG>TACCAAGAGC
r361	TGAAGAGATA<CACTCCTGG...Tc1...CCAGCACTG>TACCAAGAGC

FIG. 1. Tc1 insertion sites in *unc-54*. (A) Tc1 insertion sites in *unc-54* shown relative to the protein-coding regions of the gene. The limits of *unc-54* shown here are the AUG translational initiation codon and the AAUAAA polyadenylation signal (33). ▬, Exons; ———, introns; ▩, untranslated 3′ region. (B) DNA sequences of Tc1 insertional junctions shown below the corresponding wild-type DNA sequences. Sequences corresponding to the Tc1 element are indicated (< . . . Tc1 . . . >). The TA dinucleotides that flank each insertion are underlined. The nucleotide positions in *unc-54* are described by Karn et al. (33).

DNA of $>10^6$ animals. Therefore, they almost certainly represented independent excision events.

M13 cloning procedure. We subcloned Tc1 insertional junction fragments into the sequencing vector M13mp19 (42) and sequenced them by the method of Sanger et al. (60). For most insertions, appropriate restriction fragments containing each Tc1 insert and flanking *unc-54* DNA were gel purified (70) from lambda clones, digested with *Bal*I, ligated to appropriately digested M13mp19 DNA, and transfected into *Escherichia coli* JM101. Two clones were retained for each insertion allele. Each clone extends from a Tc1 *Bal*I site, located 9 base pairs from each terminus of Tc1, to a restriction site in adjacent *unc-54* DNA; one clone covers the insertional junction nearest the 5′ end of *unc-54*, and the other clone covers the insertional junction nearest the 3′ end of *unc-54*. For insertions *r322* and *r361*, an analogous method was used, except that subclones extended from a Tc1 *Eco*RV site, located 17 base pairs from each terminus of Tc1, to a restriction site in adjacent *unc-54* DNA. Germ line and somatic excision alleles were subcloned in a similar manner.

Derivation of an insertion site consensus sequence. The DNA sequences surrounding the points of insertion are known for 24 Tc1 elements (54, 56, 57; I. Mori, G. M. Benian, D. G. Moerman, and R. H. Waterston, Proc. Natl. Acad. Sci. USA, in press) (Fig. 1). These insertions are located at 16 different sites on various chromosomes. We considered each site only once and compared the two DNA strands of each site to all other sites. We considered 16 nucleotides for each strand and aligned them at the TA dinucleotide into which Tc1 inserts. We wrote a computer program that compares all possible DNA consensus sequences with all possible combinations of 16 DNA strands. For each consensus sequence and each combination of strands, we totaled the number of nucleotides that match the consensus. We accepted as significant any base or combination of bases in the consensus sequence whose probability of occurrence was less than 5%. We used the binomial expansion and the base composition of all 32 DNA strands to evaluate the probability of occurrence. For example, the G residue at position −5 is significant ($P = 0.006$) because 8 of the 16 strands shown in Fig. 2 contain a G at this position and G residues constitute only 19.7% of total bases. The combination of DNA strands and the consensus sequence that yielded the largest number of significant matching bases (114

A.

Insertion Site	5'-Sequence-3'	Reference
unc-54(r323::Tc1)	GCT CATCTACGTA AGT	this paper
unc-54(r322::Tc1)	GAA GAGATACCAA GAG	"
unc-54(r327::Tc1)	CGC AAGTTATGGA GGA	"
unc-54(r360::Tc1)	GAG ATAATACCGA GTG	"
unc-22(st136::Tc1)	AAG GATGTACATT GAA	Mori et al., in press
unc-22(st137::Tc1)	CTT GATGTACCAG GAA	"
unc-22(st139::Tc1)	ACC AATGTACCAT CTT	"
unc-22(st140::Tc1)	CCA GTGGTAGCTT CTC	"
unc-22(st141::Tc1)	AAC GACATATCCC AAA	"
unc-22(st185::Tc1)	CGG CATCTATGTC GCC	"
lin-12(e1979::Tc1)	GCA TGTATATGTA AAC	"
unc-15(r408::Tc1)	GAT GAGATATGTG TGT	"
pCe(Be)T1	GCA CATATATTTG AAA	(57)
pCe2001	ACA TATTTATGTA CTT	(58)
stP1	AGA GATATAGGTT T--	(46)
Tc1(Hin)	TCA GTCATAACTA ACG	(55)

B.

Number of Occurrences Among 16 DNA Strands	Position Relative To Insertion Site															
	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8
G	6	3	3	**8**	1	**4**	**4**	0	0	2	**7**	2	3	6	3	3
A	5	5	7	3	**12**	1	**8**	0	**16**	1	1	3	7	5	6	5
T	1	1	3	2	3	**9**	2	**16**	0	**7**	1	**10**	4	2	4	4
C	4	7	3	3	0	2	2	0	0	**6**	**7**	1	2	3	2	3
Consensus:					G	A	G/T	A/G	T	A	T/C	G/C	T			

FIG. 2. Consensus sequence for insertion of Tc1. (A) DNA sequences of 16 Tc1 insertion sites. For each site, the strand shown is the one that maximizes the score of the derived consensus sequence (see Materials and Methods). (B) Site matrix that describes the DNA strands shown in panel A. The consensus sequence was derived by accepting as significant any base or combination of bases whose probability of occurrence is less than 5% (see Materials and Methods). Underlined values indicate significant bases.

out of 144 total possible matches) are shown in Fig. 2. This process has the effect of choosing for each insertion site the DNA strand that maximizes the score of the derived consensus.

RESULTS

Strong site-specificity of Tc1 insertion. We determined the insertion sites of 11 Tc1-induced *unc-54* mutations. We cloned Tc1-induced *unc-54* alleles and determined the DNA sequences at both termini of each insert. By comparing the sequences of the wild-type *unc-54* gene (33) with those of the insertional junctions, we identified the site of each insertion. Our genetic methods ensure that each insertion represents an independent mutational event (15).

Fig. 1A shows the sites of Tc1 insertion within *unc-54*, and Fig. 1B shows the DNA sequences of each insertional junction fragment. Tc1 displays a striking target site preference. Of 11 Tc1 insertions, 7 are located at a single site within the gene; this site represents a "hotspot" for insertion of Tc1 in *unc-54*. Two insertions, *unc-54(r322::Tc1)* and *unc-54(r361::Tc1)*, are both located at a second site (site 2) within the gene. The remaining alleles, *unc-54(r327::Tc1)* and *unc-54(r360::Tc1)*, represent insertions at sites 3 and 4. Each insertion is located within an *unc-54* exon.

Each of the 11 Tc1 insertions is flanked by TA dinucleotides; only one copy of the TA dinucleotide is found at the corresponding site in the wild-type sequence. These TA dinucleotides might represent a 2-base-pair target site duplication that is formed during insertion. Duplications of the target site are characteristic of transposable genetic elements (7, 28). However, since TA is itself an inverted repeat, these nucleotides might be part of Tc1. In this case, no target site duplication is formed upon insertion (see discussion of this point in reference 55). All Tc1 insertions sequenced to date (a total of 24) are flanked by TA dinucleotides (54, 56, 57; Mori et al., in press) (Fig. 1).

We observed one DNA sequence polymorphism within the inverted repeats of Tc1. Tc1 elements have 54-base-pair inverted repeat termini, and we sequenced the outermost 9 or 19 nucleotides of each repeat (see Materials and Methods). With one exception, the regions of Tc1 that we sequenced were identical to each other and to those of the canonical Tc1 element pCe(Be)T1 (56). In the exceptional case, one inverted repeat of the *unc-54(r361::Tc1)* insertion contained 2-base-pair differences compared with all other sequenced termini, including the other terminal inverted repeat of *r361*.

Consensus sequence for Tc1 insertion. We compared the nucleotides surrounding the *unc-54::Tc1* insertion sites to those in other loci. The DNA sequences of 12 additional Tc1 insertion sites are available for comparison (54, 56, 57; Mori et al., in press). Mori et al. have described (in press) a consensus sequence for Tc1 insertion that is based on 11 of these sites. We compared the 16 sites and derived a consensus sequence that best describes the similarities among sites. Our methods for deriving the consensus are described more fully in Materials and Methods, and our results are shown in Fig. 2. The consensus sequence 5'-GA$^{GA}_{TG}$TA$^{TG}_{CC}$T-3' best describes the similarities among all 16 insertion sites. This consensus sequence is similar to that derived by Mori et al. (in press). The DNA sequence similarities of independent insertion sites suggest that the specificity of Tc1 insertion is

due, at least in part, to a preference for certain primary DNA sequences (see Discussion).

Frequencies of germ line and somatic reversion. Many Tc1-induced mutations are unstable; they revert to a wild-type phenotype in both germ line and somatic cells. We determined the frequencies of germ line and somatic reversion for each of our 11 *unc-54::Tc1* alleles; our results are shown in Table 1. Since genetic background can significantly affect reversion (43, 45; see below), we measured reversion frequencies in Bergerac genetic backgrounds that are essentially isogenic (see Materials and Methods).

Germ line revertants of *unc-54::Tc1* mutants are detected as wild-type (non-Unc) animals that occur spontaneously in cultures of the paralyzed mutants. These revertants breed true, and their *unc-54*$^+$ gene is a stably inherited genetic trait. We observed germ line revertants for 8 of our 11 *unc-54::Tc1* mutants. Each of the seven hotspot insertions exhibited similar frequencies of reversion (Table 1); the average for the seven insertions was 1.8×10^{-5} per gamete. We obtained revertants for one other insertion, *unc-54(r322::Tc1)*. The frequency of *r322* reversion, however, was quite low (less than 6×10^{-7} per gamete; Table 1), and we were unable to accurately measure its reversion frequency. We have not observed revertants for the remaining three insertions; their reversion frequencies are each less than 6×10^{-7} per gamete.

Somatic revertants of *unc-54::Tc1* mutants are detected as egg-laying-proficient, partially motile animals. *unc-54* mutants are egg-laying defective (Egl), because the sex muscles (a set of 16 cells required for egg laying) express *unc-54* myosin heavy chains. Reversion of *unc-54::Tc1* mutations in ancestors or certain descendants of the postembryonic mesoblast restores egg-laying ability because the sex muscles are revertant. Such animals are partially motile because all or part of the postembryonic muscle cells are revertant as well. The reversion event in these animals is confined to the soma; their germ lines are exclusively mutant (15). Thus, we can determine the frequencies of somatic reversion in the mesoblast cell lineage by measuring the frequencies of non-Egl, non-Unc, non-wild-type revertants.

We observed somatic revertants for 9 of our 11 *unc-54::Tc1* mutants. Each of the seven hotspot insertions exhibited similar frequencies of somatic reversion (Table 1); the average for the seven insertions was nearly 1% per animal. Both *r322* and *r361*, which are insertions at site 2 (Fig. 1), reverted at a frequency of approximately 0.08% per animal. We have not observed somatic revertants for the remaining insertions; their frequencies were less than 0.02% per animal.

The very high frequency of somatic reversion for insertions at the hotspot reflects the very high frequency of Tc1 somatic excision (18). It is not possible to compare directly the frequency of germ line reversion with the frequency of somatic reversion. The frequencies of germ line reversion represent revertants per gamete among a population of gametes. The frequencies of somatic reversion represent revertants per animal among a population of animals. Individual animals, however, contain many somatic cells, and reversion in any one of a number of cells would be detected as a somatic revertant. We estimated that at most there are 20 cells during development in which we would detect somatic reversion. This estimate is based on the known lineages of the muscle cells in which we detect somatic reversion (66). Somatic excision occurs in most, if not all, somatic cells (17). Assuming that Tc1 excision in the approximately 20 cells in which we can detect reversion is representative of its behavior in all somatic cells, we calculated that the frequency of somatic reversion in Bergerac, when expressed as revertants per cell, is approximately 1,000-fold higher than the frequency of germ line reversion. This estimate is consistent with the accumulated level of empty sites observed in Bergerac genomic DNA (18). Empty sites are products of somatic excision; they are the chromosomal sequences that remain after Tc1 excision in somatic cells.

The sites of Tc1 insertion correlated strongly with the frequencies of germ line and somatic reversion. Insertions at the hotspot reverted at high frequencies in both tissues. Alleles *r322* and *r361*, which are insertions at site 2, reverted in the soma at lower, but detectable frequencies and in the germ line at or below our level of detection. Alleles *r327* and *r369*, which are located at sites 3 and 4, failed to revert at detectable frequencies in either tissue.

These data indicate that both the germ line and somatic reversion properties of *unc-54::Tc1* mutants are determined by the site of insertion and not by the nature of the inserted element. This could be due to direct effects of the insertion sites on the frequency or nature of excision or to constraints imposed by our genetic methods. Since revertants were selected phenotypically, only excisions that restored gene function were detected. The differences between sites, then, reflect the ability of the *unc-54* myosin heavy-chain protein to tolerate the products of imprecise excision. Results presented below demonstrate that the latter explanation is correct.

DNA sequences of germ line revertants. Germ line revertants of *unc-54::Tc1* mutants result from excision of the element (15). To determine what types of germ line excision are responsible for reversion, we sequenced the excision sites of 11 independent *unc-54*$^+$ revertants. We sequenced five wild-type revertants derived from the hotspot insertions *r323* and *r328*, two partial revertants derived from the hotspot insertion *r323*, and four wild-type revertants derived from the non-hotspot insertion *r322*. The wild-type revertants are phenotypically indistinguishable from EM1002, our canonical *C. elegans* variety Bergerac wild-type strain. The partial revertants are more motile than the parental Tc1 insertion mutant but are slower than the wild type. The DNA sequences of these 11 revertants are presented in Fig. 3.

Each of these 11 revertants resulted from imprecise excision of Tc1. The absence of precise excisions in this set indicates that such events are rare. Each revertant resulted from simple deletion of Tc1 sequences and, occasionally, adjacent target site DNA. All breakpoints occurred no farther than 3 base pairs from the termini of Tc1.

The wild-type revertants *unc-54(r847)* through *unc-54(r850)* contain 3- or 6-base-pair insertions relative to the wild-type gene (Fig. 3A). Since a multiple of 3 base pairs was inserted, the translational reading frame of *unc-54* mRNA was maintained. These alleles encode proteins that contain insertions of one (Asn) or two (Met-Tyr) amino acids at the excision site.

The wild-type revertants *unc-54(r851)* through *unc-54(r855)* contain identical 4-base-pair insertions (TATGTA versus TA for the wild-type gene [Fig. 3B]). The translational reading frame of *unc-54* mRNA is not maintained in these revertants. Why are they phenotypically wild type? The site of insertion-excision in these alleles is very close to a 5′ splice site in *unc-54* mRNA. We propose that these alleles are phenotypically wild type because mRNA splicing is altered, such that the 4-nucleotide insertion is removed from the mRNA (see Discussion). The partial revertant allele *r837* also contains a 4-base-pair insertion relative to the

A.	
unc-54$^+$	5'-AGATACCA-3'
unc-54(r322::Tc1)	AGATA<CAG...Tc1...CTG>TACCA
WT revertants	
unc-54(r847)	AGATACA---ACCA
unc-54(r848)	AGATACATGTACCA
unc-54(r849)	AGATACATGTACCA
unc-54(r850)	AGATACATGTACCA
Somatic excision clones	
TR#25	AGATA--TGTACCA
TR#26	AGATA--TGTACCA
TR#27	AGATA--TGTACCA
TR#28	AGATACA--TACCA
B.	
unc-54$^+$	5'-ATCTACGT-3'
unc-54(r323::Tc1) and *unc-54(r328::Tc1)*	ATCTA<CAG...Tc1...CTG>TACGT
WT revertants	
unc-54(r851)	ATCTA--TGTACGT
unc-54(r852)	ATCTA--TGTACGT
unc-54(r853)	ATCTA--TGTACGT
unc-54(r854)	ATCTA--TGTACGT
unc-54(r855)	ATCTA--TGTACGT
Partial revertants	
unc-54(r837)	ATCT--CTGTACGT
unc-54(r661)	AT------GTACGT
Somatic excision clones	
TR#29	ATCTA---GTACGT
TR#30	ATCTACA--TACGT
TR#31	ATCTA--TGTACGT
TR#32	ATCTA--TGTACGT
TR#33	ATCTA------CGT

FIG. 3. DNA sequences resulting from Tc1 excision. The DNA sequences of germ line revertants and somatic excision clones are shown relative to the sequence of the wild-type gene and the Tc1 insertion mutation from which they were derived. Nucleotides of the Tc1 element are set apart in brackets (< . . . Tc1 . . . >). The TA dinucleotides that flank each insertion are underlined. (A) Germ line revertants and somatic excision clones derived from *unc-54(r322::Tc1)*; (B) germ line revertants and somatic excision clones derived from *unc-54(r323::Tc1)* or *unc-54(r328::Tc1)*. Both *r323* and *r328* are Tc1 insertions at the hotspot but are inserted in opposite orientations.

wild-type gene (TCTGTA versus TA). Again, we believe that altered mRNA splicing corrects the translational reading frame of this revertant. The partial revertant *r661* contains a single base substitution relative to the wild-type gene, but this revertant may have altered mRNA splicing as well (see Discussion).

When considering the nature of Tc1 excision, it is important to consider the relative frequencies with which these different revertants occur. Insertions at the hotspot reverted to wild type in the germ line at frequencies that were at least 30-fold higher than those of insertions at any other site (Table 1). We estimate that partial revertants of hotspot insertions occur at frequencies that are 100-fold lower than those of wild-type revertants. (Two partial revertants were isolated in an experiment in which 403 wild-type revertants were obtained.) We sequenced five wild-type revertants of hotspot insertions; all of them contained identical imprecise excisions. Therefore, this sequence was by far the most frequent germ line excision that we have detected. Germ line excisions were selected as revertants, however, and excisions that did not restore function were not isolated.

DNA sequences of Tc1 somatic excisions. Tc1 excision in somatic cells is sufficiently frequent that empty sites can be detected in total genomic DNA with Southern hybridizations (18). In strains that contain very few germ cells, empty-site fragments accumulated to as much as 40% of total *unc-54* DNA (data not shown; see Materials and Methods). To determine the nature of somatic excision and to compare it to germ line excision, we sequenced nine independent clones of empty-site *unc-54* fragments. We sequenced five clones derived from hotspot insertion *r323* and four clones derived from non-hotspot insertion *r322*. Since empty sites were cloned at random and since their isolation was not based on functional reversion, their sequences likely represent the most frequent somatic excisions. The DNA sequences remaining after somatic excision are shown in Fig. 3.

Eight of the nine empty sites resulted from imprecise excision; only one clone resulted from precise excision, in which the wild-type sequence was restored. A 4-base-pair insertion (TATGTA versus the wild-type sequence TA) was the most frequent somatic excision for both the hotspot and *r322* insertion sites. This sequence, when generated at the hotspot, restored wild-type *unc-54* gene function (see above). This sequence, when generated at the *r322* site, would not restore wild-type *unc-54* gene function, because of a frameshift in the translational reading frame. These observations explain why hotspot insertions revert in germ line and somatic cells at such high frequencies relative to those at all other sites. The hotspot is fortuitously tolerant of the preferred type of Tc1 imprecise excision.

The second most frequent somatic excision (TACATA versus the wild-type TA) was a closely related event. In this case, the TA dinucleotides plus 2 base pairs at the opposite end of Tc1 remained after excision. This event is symmetrical to the excisions described above, and a single mechanism may be responsible for both products. For example, imprecise excision might involve single-strand breaks staggered by 2 base pairs at the ends of Tc1. Ligation and mismatch repair of the resulting DNA termini could generate both of these empty sites.

Tc1-induced mutations revert to wild type in the germ line only in strains that are active for Tc1 transposition (43, 45; see below). Germ line excision, therefore, is transposase dependent. The most frequent excision that we observed in the soma had the same DNA sequence as the most frequent excision that we observed in the germ line (Fig. 3). We conclude, therefore, that somatic excision is also transposase dependent.

Effects of genetic background on germ line and somatic reversion. Our ability to detect both germ line and somatic reversion of *unc-54* hotspot insertions allowed us to determine the effects of genetic background on each of these events. *C. elegans* wild-type variety DH424 contains a copy number of Tc1 similar to that of Bergerac, but these elements are quiescent for germ line transposition (15). We crossed the hotspot insertion *unc-54(r323::Tc1)* repeatedly with wild-type DH424 males. *r323* homozygotes were isolated after 1, 2, 3, and 10 backcrosses (see Materials and Methods), and we measured the frequencies of germ line and somatic reversion at each stage. Our results are shown in Table 2. Reversion of *unc-54(r323::Tc1)* in the germ line of DH424 was barely detectable, about 100-fold lower than the

frequency of reversion in Bergerac. Reversion of *unc-54(r323::Tc1)* in the soma of DH424, however, was only slightly less than that of Bergerac. We conclude that activity of Tc1 in the soma can be regulated independently of the germ line.

DISCUSSION

Tc1 is responsible for most spontaneous mutations that occur in *C. elegans* variety Bergerac. We have examined the insertion and excision of Tc1 in the *unc-54* myosin heavy-chain gene. We have shown that (i) Tc1 insertion is strongly site specific, (ii) Tc1 excision is usually imprecise, (iii) the DNA sequences remaining after Tc1 excision are essentially the same in germ line and somatic cells, (iv) the frequency of somatic excision in *C. elegans* variety Bergerac is approximately 1,000-fold higher than the frequency of germ line excision, and (v) somatic excision is not regulated in the same strain-specific manner as germ line excision is.

The site specificity of Tc1 insertion is not surprising. Most transposable elements, including both procaryotic and eucaryotic elements, exhibit a preference for particular target sequences (3, 13, 26, 29–31, 47, 71; Mori et al., in press). Perhaps the most extreme example of this is the bacterial transposon Tn*7*, which inserts at a single site on the *E. coli* chromosome (38). Can the consensus sequence that describes similarities between different Tc1 insertion sites (Fig. 2) explain adequately the site specificity of Tc1 insertion? We think not. If the consensus sequence alone were responsible for site specificity, then the *unc-54* insertional hotspot should match the consensus most closely. This was not the case. The hotspot matches at seven of nine consensus positions, but *unc-54* contains seven other sites that match at eight of nine positions. Thus, the consensus sequence alone does not identify the hotspot as being particularly favorable.

Insertion of Tc1 is also strongly gene specific. We estimate that Tc1 insertion into the Bergerac *unc-22* gene is about 100-fold more frequent than insertion into the Bergerac *unc-54* gene (11, 15). Does the *unc-22* gene contain 100-fold more insertion site consensus sequences? Although this is possible, it is unlikely. By extrapolating from a partial DNA sequence of the *unc-22* gene, Mori et al. (in press) estimate that *unc-22* contains sevenfold more occurrences of a Tc1 insertion site consensus sequence than does the *unc-54* gene. (The consensus sequence described by Mori et al. [in press] is closely related to that shown in Fig. 2.) Tc1 clearly prefers certain DNA sequences, but factors in addition to primary DNA sequence must influence Tc1 target site specificity.

Tc1 excises imprecisely in both germ line and somatic cells. Of 20 independent excisions that we sequenced, only 1 was precise. Imprecise excision is a common feature of eucaryotic transposable elements. Such events have been described for yeast cells (53), *D. melanogaster* (2, 9, 61, 69), maize (63, 67), mice (12), and *Antirrhinum majus* (5, 10). Germ line and somatic excisions are closely related events. The most frequent germ line excision has a characteristic sequence (TATGTA versus TA for the wild-type site). The most frequent somatic excision has this same DNA sequence. Since germ line excision is transposase dependent (43, 45; Table 2), we conclude that somatic excision is also transposase dependent.

The excision of most eucaryotic transposable elements and perhaps some procaryotic elements is transposase dependent. The mechanism of excision in such cases is probably related to the mechanism of transposition. For example, excision of Tn*10* in *E. coli* (1, 46), *Ac* in maize (for a review, see reference 25), and T-DNA in *Agrobacterium tumefaciens* (34, 64) appears to be either an intermediate in or a consequence of the transposition process. The relationship of Tc1 excision to transposition is unknown; perhaps excision is a required intermediate in transposition. Extrachromosomal linear and circular copies of Tc1 have been previously described (55, 59). These molecules are presumably generated by excision, but they could be transposition intermediates as well. The extrachromosomal circular molecules are especially interesting. Retroviral proviruses may integrate via a circular intermediate (49), and examples of site-specific recombination involving circular molecules are numerous (for example, see reference 8). It is intriguing that the inferred DNA sequence at the point of Tc1 circularization (CTGTACAGT [circularization point underlined] [59]) has the potential to base pair with the Tc1 insertion site consensus sequence at five of nine positions; the five positions are contiguous and span the point of insertion. Perhaps such base pairing is partially responsible for insertion site specificity.

The location of a Tc1 insert relative to the protein-coding region of *unc-54* strongly influences the apparent frequency of reversion. Excision of Tc1 is imprecise, and individual sites are able to tolerate different types of imprecise excision. For example, the *r322* insertion is located within an exon, and only excisions that maintained the translational reading frame and did not disrupt myosin function were detected. *r322*, therefore, reverted to wild type at a much lower frequency than did insertions at the hotspot, which was tolerant of a wider variety of imprecise excisions. We detected neither germ line nor somatic revertants for insertions *r327* and *r360*. We have detected somatic excision on Southern blots, however, for both of these alleles (15). Since we detected excision biochemically but not genetically, we conclude that these two insertion sites do not tolerate the sequence alterations remaining after excision.

Several revertants of hotspot insertions contain 4-base-pair inserts relative to the normal gene. Why are these alleles functional? We believe that altered mRNA splicing compensates for the 4-base-pair insert. The Tc1 insertional hotspot is located 1 base pair away from the 5′ splice site of *unc-54* intron 3 (33). We propose that insert-containing revertants have a new 5′ splice site that is displaced 4 bases upstream of the normal splice site. We compared the DNA sequences of hotspot revertants with those of the eucaryotic 5′ splice site consensus (Fig. 4). *C. elegans* 5′ splice sites are closely related to those of all eucaryotes (48; unpublished nematode information compiled by T. Blumenthal). Wild-type revertants *r851* through *r855* contain a region, displaced from the normal splice site by 4 bases, that aligns favorably with the consensus 5′ splice signal (Fig. 4B). Splicing at the upstream site would remove the 4-base insertion from the mRNA; the resulting protein would have a wild-type amino acid sequence. A similar situation applies to the partial revertant *r837* (Fig. 4C). In this case, however, the resulting protein would contain a single amino acid substitution (Tyr-114 → Ser-114); we believe that the *r837* phenotypic defect is due to this amino acid substitution. *r837* and *r851* through *r855* contain normal or near-normal amounts of *unc-54* protein (A. Bejsovec and P. Anderson, unpublished results), indicating that if splicing occurs at the displaced 5′ site, it is efficient. Splicing at the upstream site might be favored, either because of its location directly 5′ to the normal splice site (35) or because the splicing machinery has sequence or structural bias that we do not recognize (51).

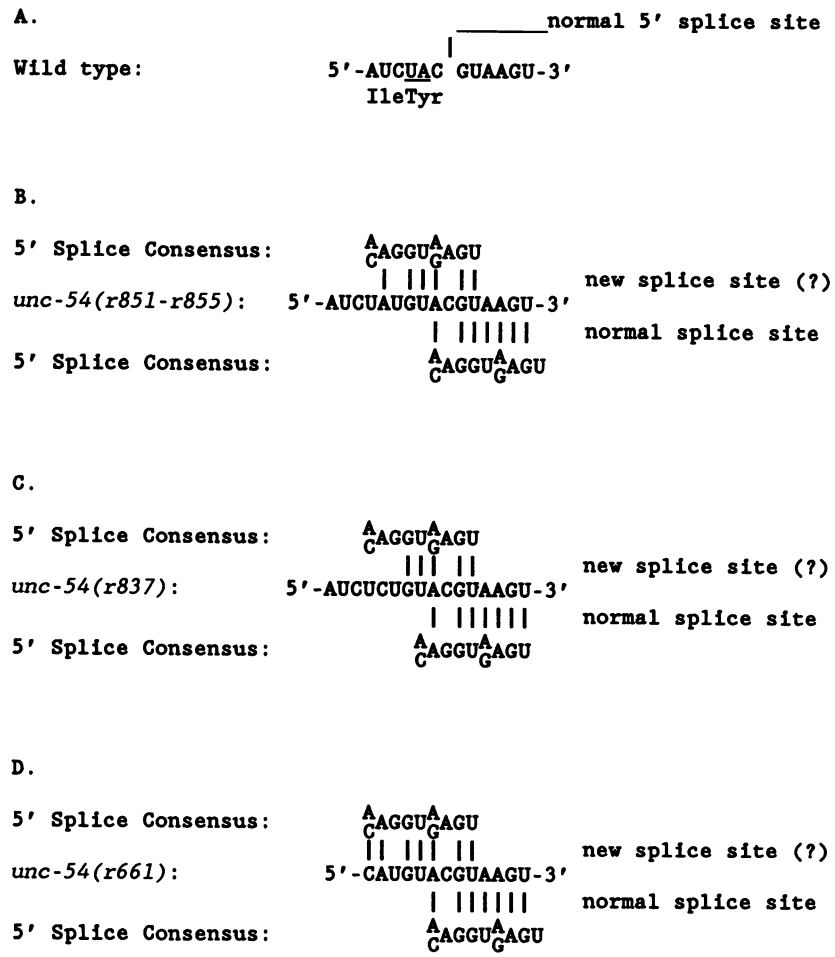


FIG. 4. Model for altered mRNA splicing in revertants of hotspot insertions. (A) mRNA and amino acid sequences of wild-type *unc-54* in the vicinity of the insertional hotspot. The site of Tc1 insertion is underlined. (B) mRNA sequences of wild-type revertants *unc-54(r851)* through *unc-54(r855)* compared with the eucaryotic 5′ splice site consensus sequence (48). Positions of identity are indicated by vertical lines. *r851* through *r855* contain a 4-nucleotide insert compared to the wild-type gene. These revertants have two potential 5′ splice sites. Use of the upstream splice site results in mRNA that encoded a fully wild-type protein. Use of the normal splice site results in mRNA that contained a translational frameshift. (C) mRNA sequence of the partial revertant *unc-54(r837)* compared with the eucaryotic 5′ splice site consensus sequence. *r837* contains a 4-nucleotide insert relative to the wild-type gene and has two potential 5′ splice sites. Use of the upstream splice site results in mRNA that encodes a protein having a tyrosine → serine amino acid substitution. Use of the normal splice site results in mRNA that contains a translational frameshift. (D) mRNA sequence of *unc-54(r661)* compared with the eucaryotic 5′ splice site consensus sequence. *r661* contains two potential 5′ splice sites. Use of the upstream splice site results in mRNA that contains a translation frameshift. Use the normal splice site results in mRNA that encodes a protein having an isoleucine → methionine acid substitution.

The partial revertant allele *r661* contains a single C-to-G transversion relative to the wild type, resulting in replacement of isoleucine 113 with methionine. The phenotypic defect of *r661* likely results, at least in part, from this amino acid substitution. *unc-54(r661)* accumulates approximately 20% of the normal amount of *unc-54* protein (A. Bejsovec and P. Anderson, unpublished results). Thus, the mutant protein may be subject to degradation. An alternative explanation for the low level of *unc-54* protein in *r661* is possible. The single base substitution in *r661* created a potential new 5′ splice site which, like the alleles discussed above, is displaced 4 base pairs upstream of the normal site (Fig. 4D). Use of the displaced splice site would delete four bases from the mRNA, and the resulting translational frameshift would lead to a nonfunctional product. By this model, the low level of *unc-54* protein is due to a low level of functional mRNA, which is obtained by inefficient utilization of the normal 5′ splice site.

The amino acid substitutions of *r661* and *r837* are located within the myosin globular head. Ile-113 and Tyr-114 are located near the ATP-binding site. The amino acid substitutions of *r661* and *r837* may affect ATP binding or ATPase activity. Moerman et al. (44) have described a class of *unc-54* mutations that cause amino acid substitutions in this region. Such alleles are dominant suppressors of the twitching phenotype of *unc-22* loss-of-function mutations. The alleles *r661* and *r837* have behavioral phenotypes similar to those described by Moerman et al. (44), and we have tested them for suppression of *unc-22* twitching. *r661*, when heterozygous to *unc-54*$^+$, suppressed the twitching phenotype of *unc-22(s12)*. This result supports our conclusion that the *r661* protein has an altered amino acid sequence, and it suggests that the *r661* phenotypic defect is due at least in part to that alteration. *r837*, when heterozygous to *unc-54*$^+$, did not suppress the twitching of *unc-22(s12)*.

Transposition and excision of Tc1 in the germ line occur

only in certain genetic backgrounds. For example, Tc1 is active in the germ line of *C. elegans* variety Bergerac (15, 43, 45) but does not transpose or excise in the germ line of *C. elegans* variety DH424 (15; Table 2). The soma of strain DH424, however, behaves unlike the germ line. Hotspot insertion mutants reverted to wild type in the soma at high frequency in the DH424 genetic background (Table 2). Since somatic excision is transposase dependent (see above), we conclude that activity of Tc1 in the soma of variety DH424 is regulated independently of the germ line. This confirms previous work concerning somatic activity of Tc1 in *C. elegans* variety Bergerac (17, 18). Germ line activity of Tc1, furthermore, can be mutated without affecting the soma. Collins et al. (11) describe the isolation of mutants in which the frequency of Tc1 transposition and excision in the germ line, but not the soma, is dramatically increased.

ACKNOWLEDGMENTS

We thank B. Engels for advice concerning the quantitative aspects of our data and our colleagues Beth Spangler, Suzanne Sprunger, Judith Kimble, and John Collins for critically reading the manuscript. We are grateful to I. Mori, G. Benian, D. Moerman, and R. Waterston for sharing the DNA sequences of their Tc1 insertions before publication.

This work was supported by Public Health Service individual research grant GM30132 from the National Institutes of Health and by the University of Wisconsin Training Grant in Cell and Molecular Biology.

LITERATURE CITED

1. **Bender, J., and N. Kleckner.** 1986. Genetic evidence that Tn*10* transposes by a nonreplicative mechanism. Cell **45**:801–815.
2. **Bender, W., M. Akam, F. Karch, P. A. Beachy, M. Peifer, P. Spierer, E. B. Lewis, and D. S. Hogness.** 1983. Molecular genetics of the *bithorax* complex in *Drosophila melanogaster*. Science **221**:23–29.
3. **Berg, D. E., M. A. Schmandt, and J. B. Lowe.** 1983. Specificity of transposon Tn*5* insertion. Genetics **105**:813–828.
4. **Bingham, P. M., M. G. Kidwell, and G. M. Rubin.** 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P strain-specific transposon family. Cell **29**:995–1004.
5. **Bonas, U., H. Sommer, and H. Saedler.** 1984. The 17 kb *Tam1* element of *Antirrhinum majus* induces a 3 bp duplication upon integration into the chalcone synthase gene. EMBO J. **3**:1015–1019.
6. **Brenner, S.** 1974. The genetics of *Caenorhabditis elegans*. Genetics **77**:71–94.
7. **Calos, M. P., L. Johnsrud, and J. H. Miller.** 1978. DNA sequence at the integration sites of the insertion element IS1. Cell **13**:411–418.
8. **Campbell, A.** 1983. Bacteriophage lambda, p. 65–103. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
9. **Carbonare, B. D., and W. J. Gehring.** 1985. Excision of *copia* element in a revertant of the white-apricot mutation of *Drosophila melanogaster* leaves behind one long-terminal repeat. Mol. Gen. Genet. **199**:1–6.
10. **Coen, E. S., R. Carpenter, and C. Martin.** 1986. Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. Cell **47**:285–296.
11. **Collins, J., B. Saari, and P. Anderson.** 1987. Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. Nature (London) **328**:726–728.
12. **Copeland, N. G., K. W. Hutchinson, and N. A. Jenkins.** 1983. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving viral LTRs. Cell **33**:379–387.
13. **Daniels, G. R., and P. L. Deininger.** 1985. Integration site preferences of the *Alu* family and similar repetitive DNA sequences. Nucleic Acids Res. **13**:8939–8954.
14. **Dooner, H. K., and O. E. Nelson.** 1977. Controlling element-induced alterations in UDP glucose: flavonoid glucosyltransferase, the enzyme specified by the *bronze* locus in maize. Proc. Natl. Acad. Sci. USA **74**:5623–5627.
15. **Eide, D., and P. Anderson.** 1985. Transposition of Tc1 in the nematode *C. elegans*. Proc. Natl. Acad. Sci. USA **82**:1756–1760.
16. **Eide, D., and P. Anderson.** 1985. The gene structures of spontaneous mutations affecting a *Caenorhabditis elegans* myosin heavy chain gene. Genetics **109**:67–79.
17. **Emmons, S. W., S. Roberts, and K. S. Ruan.** 1986. Evidence in a nematode for regulation of transposon excision by tissue-specific factors. Mol. Gen. Genet. **202**:410–415.
18. **Emmons, S. W., and L. Yesner.** 1984. High-frequency excision of transposable element Tc1 in the nematode *Caenorhabditis elegans* is limited to somatic cells. Cell **36**:599–605.
19. **Emmons, S. W., L. Yesner, K. Ruan, and D. Katzenberg.** 1983. Evidence for a transposon in *Caenorhabditis elegans*. Cell **32**:55–65.
20. **Engels, W. R.** 1979. Extrachromosomal control of mutability in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79**:4570–4574.
21. **Engels, W. R.** 1983. The P family of transposable elements in *Drosophila*. Annu. Rev. Genet. **17**:315–344.
22. **Engels, W. R.** 1984. A *trans*-acting product needed for P factor transposition in *Drosophila*. Science **226**:1194–1196.
23. **Epstein, H. F., R. H. Waterston, and S. Brenner.** 1974. A mutant affecting the heavy chain of myosin in *Caenorhabditis elegans*. J. Mol. Biol. **90**:291–300.
24. **Errede, B. T., T. Cardillo, F. Sherman, E. Dubois, J. Deschamps, and J. Waime.** 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. Cell **25**:427–436.
25. **Federoff, N. V.** 1983. Controlling elements in maize, p. 1–63. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
26. **Freund, R., and M. Meselson.** 1984. Long terminal repeat nucleotide sequence and specific insertion of the *gypsy* transposon. Proc. Natl. Acad. Sci. USA **81**:4462–4464.
27. **Frischauf, A.-M., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.
28. **Grindley, N. D.** 1978. Insertion generates duplication of a nine base pair sequence at its target site. Cell **13**:419–426.
29. **Halling, L. M., and N. Kleckner.** 1982. A symmetrical six-basepair target site sequence determines Tn*10* insertion specificity. Cell **28**:155–163.
30. **Ikenaga, H., and K. Saigo.** 1982. Insertion of a movable genetic element, *297*, into the T-A-T-A box for the H3 histone gene of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79**:4143–4147.
31. **Inouye, S., S. Yuki, and K. Saigo.** 1984. Sequence-specific insertion of the *Drosophila* transposable genetic element *17.6*. Nature (London) **310**:332–333.
32. **Karn, J., S. Brenner, and L. Barnett.** 1983. New bacteriophage lambda vectors with positive selection for cloned inserts. Methods Enzymol. **101**:3–20.
33. **Karn, J., S. Brenner, and L. Barnett.** 1983. Protein structural domains in the *Caenorhabditis elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. USA **80**:4253–4257.
34. **Koukolikova-Nicola, Z. Schillito, R. D. Hohn, B. Wang, K. Van Montagu, and M. Zambryski.** 1985. Involvement of circular intermediates in the transfer of T-DNA from *Agrobacterium tumefaciens* to plant cells. Nature (London) **313**:191–196.
35. **Lang, K. M., and R. A. Spritz.** 1983. RNA splice site selection: evidence for a 5′-3′ scanning model. Science **220**:1351–1355.
36. **Laski, F. A., D. C. Rio, and G. M. Rubin.** 1986. Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. Cell **44**:7–19.
37. **Liao, L. W., B. Rosenzweig, and D. Hirsh.** 1983. Analysis of a transposable element in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **80**:3585–3589.

38. **Lichtenstein, C., and S. Brenner.** 1982. Unique insertion site of Tn*7* in the *E. coli* chromosome. Nature (London) **297**:601–603.
39. **Mackenzie, J. M., Jr., F. Schachat, and H. F. Epstein.** 1978. Immunocytochemical localization of two myosins within the same muscle cells in *Caenorhabditis elegans*. Cell **15**:413–419.
40. **McClintock, B.** 1965. The control of gene action in maize. Brookhaven Symp. Biol. **18**:162–184.
41. **McGinnis, W., A. W. Shermoen, and S. K. Beckendorf.** 1983. A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. Cell **34**:75–84.
42. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
43. **Moerman, D. G., G. M. Benian, and R. H. Waterston.** 1986. Molecular cloning of the muscle gene *unc-22* in *Caenorhabditis elegans* by Tc1 transposon tagging. Proc. Natl. Acad. Sci. USA **83**:2579–2583.
44. **Moerman, D. G., S. Plurad, R. H. Waterston, and D. L. Baillie.** 1982. Mutations in the *unc-54* myosin heavy chain gene of *C. elegans* that alter contractility but not muscle structure. Cell **29**:773–781.
45. **Moerman, D. G., and R. H. Waterston.** 1984. Spontaneous *unc-22* IV mutations in *C. elegans* var. Bergerac. Genetics **108**:859–877.
46. **Morisato, D., and N. Kleckner.** 1984. Transposase promotes double strand breaks and single strand joints at Tn*10* termini *in vivo*. Cell **39**:181–190.
47. **O'Hare, K., and G. M. Rubin.** 1983. Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**:25–35.
48. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
49. **Panganiban, A. T., and H. M. Temin.** 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.
50. **Park, E.-C., and H. R. Horvitz.** 1986. *C. elegans unc-105* mutations affect muscle and are suppressed by other mutations that affect muscle. Genetics **113**:853–867.
51. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46**:681–690.
52. **Reynolds, A., J. Felton, and A. Wright.** 1981. Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. Nature (London) **293**:625–629.
53. **Roeder, G. S., and G. R. Fink.** 1980. DNA rearrangements associated with a transposable element in yeast. Cell **21**:239–249.
54. **Rose, A. M., L. J. Harris, N. R. Mawji, and W. J. Morris.** 1985. Tc*1*(*Hin*): a form of the transposable element Tc1 in *C. elegans*. Can. J. Biochem. Cell Biol. **63**:752–756.
55. **Rose, A. M., and T. P. Snutch.** 1984. Isolation of the closed circular form of the transposable element Tc1 in *Caenorhabditis elegans*. Nature (London) **311**:485–486.
56. **Rosenzweig, B., L. W. Liao, and D. Hirsh.** 1983. Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res. **11**:4201–4209.
57. **Rosenzweig, B., L. W. Liao, and D. Hirsh.** 1983. Target sequences for the *C. elegans* transposable element Tc1. Nucleic Acids Res. **11**:7137–7140.
58. **Ross, D. G., J. Swan, and N. Kleckner.** 1979. Nearly precise excision: a new type of DNA alteration associated with the translocatable element Tn*10*. Cell **167**:733–738.
59. **Ruan, K., and S. W. Emmons.** 1984. Extrachromosomal copies of transposon Tc1 in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **81**:4018–4022.
60. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
61. **Searles, L. L., A. L. Greenleaf, W. E. Kemp, and R. A. Voelker.** 1986. Sites of P element insertion and structures of P element deletions in the 5′ region of *Drosophila melanogaster RpII215*. Mol. Cell. Biol. **6**:3312–3319.
62. **Shapiro, J. A. (ed.).** 1983. Mobile genetic elements. Academic Press, Inc., New York.
63. **Shure, M., S. Wessler, and N. Federoff.** 1983. Molecular identification and isolation of the *Waxy* locus in maize. Cell **35**:225–233.
64. **Stachel, S. E., B. Timmerman, and P. Zambryski.** 1986. Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens* to plant cells. Nature (London) **322**:706–712.
65. **Sulston, J. E., and S. Brenner.** 1974. The DNA of *C. elegans*. Genetics **77**:95–104.
66. **Sulston, J. E., and H. R. Horvitz.** 1977. Postembryonic cell lineages of the nematode *Caenorhabditis elegans*. Dev. Biol. **56**:110–156.
67. **Sutton, W. D., W. L. Gerlach, D. Schwartz, and W. J. Peacock.** 1984. Molecular analysis of *Ds* controlling element mutations at the *Adh1* locus of maize. Science **223**:1265–1268.
68. **Syvanen, M.** 1984. The evolutionary implications of mobile genetic elements. Annu. Rev. Genet. **18**:271–293.
69. **Tsubota, S., and P. Schedl.** 1986. Hybrid dysgenesis-induced revertants of insertions at the 5′ end of the *rudimentary* gene in *Drosophila melanogaster*: transposon-induced control mutations. Genetics **114**:165–182.
70. **Weislander, L.** 1979. A simple method to recover intact high molecular weight RNA and DNA after electrophoresis in low gelling temperature agarose gels. Anal. Biochem. **98**:305–309.
71. **Zerbib, D., P. Gamas, M. Chandler, P. Prentki, S. Bass, and D. Galas.** 1985. Specificity of insertion of IS1. J. Mol. Biol. **185**:517–524.