S Elements: A Family of *Tc1*-Like Transposons in the Genome of *Drosophila melanogaster*

Peter J. Merriman, Craig D. Grimes, Jaroslaw Ambroziak, David A. Hackett, Pamela Skinner and Michael J. Simmons

Department of Genetics and Cell Biology, BioScience Center, University of Minnesota, St. Paul, Minnesota 55108-1095

Manuscript received May 25, 1995
Accepted for publication August 25, 1995

ABSTRACT

The *S* elements form a diverse family of long-inverted-repeat transposons within the genome of *Drosophila melanogaster*. These elements vary in size and sequence, the longest consisting of 1736 bp with 234-bp inverted terminal repeats. The longest open reading frame in an intact *S* element could encode a 345-amino acid polypeptide. This polypeptide is homologous to the transposases of the *mariner-Tc1* superfamily of transposable elements. *S* elements are ubiquitous in *D. melanogaster* populations and also appear to be present in the genomes of two sibling species; however, they seem to be absent from 17 other Drosophila species that were examined. Within *D. melanogaster* strains, there are, on average, 37.4 cytologically detectable *S* elements per diploid genome. These elements are scattered throughout the chromosomes, but several sites in both the euchromatin and β heterochromatin are consistently occupied. The discovery of an *S*-element-insertion mutation and a reversion of this mutation indicates that *S* elements are at least occasionally mobile in the *D. melanogaster* genome. These elements seem to insert at an AT dinucleotide within a short palindrome and apparently duplicate that dinucleotide upon insertion.

TRANSPOSABLE elements have been found in the genomes of many organisms from diverse taxa, including plants, animals, bacteria and fungi (BERG and HOWE 1989). Their widespread distribution indicates that they have been highly successful as genetic parasites, propagating and transposing within genomes. Genetic studies have demonstrated that transposable elements are a primary cause of mutations and chromosome rearrangements (LIM and SIMMONS 1994). It is therefore likely that they have played an important role in the evolution of many species. Molecular studies have shown that transposable elements are structurally and functionally diverse. Some transpose by an excision/insertion mechanism, whereas others transpose through an RNA molecule that is reverse-transcribed into DNA and then integrated into the genome. The enzymes used in these activities are often encoded by the elements themselves.

Some of the most detailed studies of transposable elements have been carried out with *Drosophila melanogaster*, where as much as 10–15% of the DNA is mobile. Altogether, >40 distinct families of transposable elements have been identified in this organism (LINDSLEY and ZIMM 1992). The best-understood are the ≈3-kb *P* and *hobo* elements, which have short inverted nucleotide repeats at their termini and which encode *trans*-acting transposases (ENGELS 1989; CALVI *et al.* 1991). These elements have been used extensively in genetic analysis, both as insertional mutagens and as transformation vectors. Other elements, such as the *mariner* transposon from *D. mauritiana* and *D. simulans*, are currently being developed for these purposes (LIDHOLM *et al.* 1993; LOHE *et al.* 1995).

Evolutionary studies have indicated that some families of transposons are present in distantly related taxa. The *mariner* family is a good example. *Mariner*-like elements have been found in several orders of insects, in flatworms and roundworms, and also in human beings (MACLEOD and ROBERTSON 1993; ROBERTSON 1993; H. M. ROBERTSON, personal communication). These elements are 1.2–1.3 kb long and are bounded by short inverted repeats. Genetic studies have shown that a single long open reading frame (ORF) in the 1286-bp *Mos1 mariner* element from *D. mauritiana* encodes a transposase (MEDHORA *et al.* 1991). The broad taxonomic distribution of these elements suggests that the *mariner* family is very ancient, possibly tracing back to the origin of the metazoan lineage. However, within taxa *mariner*-like elements have a patchy distribution. For example, although *D. mauritiana* and *D. simulans* contain *mariner* elements in their genomes, the closely related *D. melanogaster* does not (CAPY *et al.* 1992). This indicates that *mariner*-like elements have been lost from some branches of the evolutionary tree. There is also strong evidence that *mariner*-like elements have occasionally been transferred across species boundaries (MARUYAMA and HARTL 1991; ROBERTSON 1993; ROB-

Corresponding author: Michael J. Simmons, Department of Genetics and Cell Biology, 250 BioScience Center, University of Minnesota, 1445 Gortner Ave., St. Paul, MN 55108-1095.
E-mail: simmo004@maroon.tc.umn.edu

ERTSON and MACLEOD 1993). The evolution of this transposon family therefore seems to involve both horizontal and vertical dimensions.

DNA sequencing studies have revealed that the *mariner*-like elements are related to another group of transposons defined by the *Tc1* element of the nematode *Caenorhabditis elegans* (DOAK *et al.* 1994; ROBERTSON 1995). These transposons are larger than the *mariner*-like elements (*Tc1* is 1.6 kb) and some have longer inverted terminal repeats (54 bp in *Tc1 vs.* 28 bp in *mariner*). Genetic and molecular analyses have shown that *Tc1* is transpositionally active and that it encodes a transposase (MOERMAN and WATERSTON 1989; VOS *et al.* 1993). Other *Tc1*-like elements have been found in fish, fungi and insects (BREZINSKY *et al.* 1990; DABOUSSI *et al.* 1992; RADICE *et al.* 1994). The putative transposases of these elements are homologous to the *mariner* transposase, so it seems that all these elements belong to a very widespread transposon superfamily.

In *D. melanogaster,* two *Tc1*-like elements, *HB1* and *Bari-1,* have been characterized (BRIERLY and POTTER 1985; HENIKOFF and PLASTERK 1988; CAIZZI *et al.* 1993); both have 27-bp inverted terminal repeats, but neither has been shown to be transpositionally active. Cytological and molecular analysis indicates that the *Bari-1* elements are concentrated in a single tandem array near the *Responder* (*Rsp*) locus in the alpha heterochromatin of chromosome 2R. A few *Bari-1* elements have also been found at scattered sites in the euchromatin. The cytological distribution of the *HB1* element is currently unknown. Here we report the discovery of another *Tc1*-like transposon in the *D. melanogaster* genome. This element, called *S,* is ubiquitous in *D. melanogaster* populations. Unlike *Bari-1,* it is found at many sites in the euchromatin and appears to be transpositionally active.

MATERIALS AND METHODS

Drosophila strains: Genetically marked stocks of *D. melanogaster* were obtained from diverse sources; the chromosomes and markers in these stocks are described in LINDSLEY and ZIMM (1992). *D. melanogaster* stocks derived from natural populations came mainly from collections made between 1978 and 1987 in the central and eastern United States (KOCUR *et al.* 1986) and from collections from many countries assembled by MARGARET KIDWELL (KIDWELL *et al.* 1983); a few stocks were obtained from other investigators. *D. simulans* stocks came from our collections and from stock centers. Stocks of all other Drosophila species came from the National Drosophila Species Resource Center, Bowling Green, Ohio.

Genomic DNA libraries, cloning and restriction mapping: DNA libraries were prepared by ligating *Eco*RI-digested genomic DNA from *D. melanogaster* adults into the *Eco*RI-digested lambdaZAPII bacteriophage vector (Stratagene), which had been treated with alkaline phosphatase to minimize self-ligation. Recombinant molecules were packaged *in vitro* using Gigapack (Stratagene), and the resulting phage were plated on XL1-Blue *Escherichia coli* cells for screening by standard plaque-lift methods. Hybond-N^+ (Amersham) was used as the DNA-binding membrane. Purified phage clones were converted into single-stranded phagemid clones by following the *in vivo* excision protocol provided by Stratagene. These phagemid clones were then converted into double-stranded pBluescript plasmid clones by isolating the phagemid DNA from infected cells and transforming it into cells that were free of helper phage. Standard procedures were used to isolate Drosophila, phagemid and plasmid DNA. Transformation of *E. coli* cells was accomplished using the procedures of CHUNG *et al.* (1989), and restriction enzyme digestions were performed according to the supplier's instructions.

Polymerase chain reactions (PCR): DNA amplification reactions were performed in volumes of 25–100 μl overlaid with 50–100 μl paraffin oil. Each reaction contained 0.2 mM of each of the four deoxyribonucleotides, 75 ng of one or 30 ng of each of two oligonucleotide primers, a buffer (15 mM Tris pH 8.8, 60 mM KCl, 2.8 mM $MgCl_2$ or 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.0% Triton X-100, 1.15 mM $MgCl_2$), *Taq* DNA polymerase (supplied either by Perkin Elmer Cetus or Promega) and template DNA. DNA templates were obtained from plasmid clones, purified Drosophila genomic DNA, previous PCR products or crude genomic DNA extracts from individual flies (GLOOR and ENGELS 1992). The temperature regimes for the amplification reactions are described with the results.

DNA primers: Oligonucleotide primers for the *su(s)* gene were obtained from R. A. VOELKER. Primers for the *S* element were purchased from Oligos, Etc. The primer that was used in PCR to screen Drosophila strains for *S* elements was 5′CACTTTTGAGACTGTCAAGAAACTC3′, denoted S-IR, and spanned nucleotides 202–226 in the left inverted repeat and 1533–1509 in the right inverted repeat of the element cloned from the *su(s)* gene. This sequence differs by 2 bp from the corresponding sequence in the element shown in Figure 4 below.

DNA sequencing: Nested deletion subclones of one of the *S*-element clones (pS2) were constructed according to the procedures of HENIKOFF (1987) and then sequenced by the dideoxy method using USB's Sequenase kit and ^{35}S-labeled dATP. Other *S*-element clones were sequenced using oligonucleotide primers derived from the sequence of pS2. Except for the terminal repeats and immediately adjacent regions, both strands were sequenced in each of these clones. Selected PCR products were sequenced on one strand by the dideoxy method using ^{32}P-end-labeled primers and *Taq* DNA polymerase. The DNA sequence data were analyzed using the GCG software developed at the University of Wisconsin. All DNA sequences have been deposited in the GenBank data base (accession numbers U33461-U33470).

Southern blotting: DNA was fractionated in agarose gels and transferred to Hybond-N^+ membranes by capillary blotting using 0.4 N NaOH/0.6 M NaCl as the transfer solution. After air drying, the blots were prepared for hybridization by washing in 0.1% SDS/0.1× SSC for 15 min at 65° and shaking in hyb-solution [5× SSCP (0.75 M NaCl, 0.75 M sodium citrate, 0.005 M K_2PO_4 pH 6.8), 35% (reduced stringency) or 50% (high stringency) deionized formamide, 50 mM Tris (pH 7.5), 1× Denhardt's solution (0.02% Ficoll-400, 0.02% polyvinylpyrrolidone, 0.02% nuclease-free BSA), 1% SDS, 5% dextran sulfate and 100 μg/ml heat-denatured salmon sperm DNA] for at least 6 hr at 42°. Radioactive probes, prepared by random primer labeling of DNA with ^{32}P-dCTP, were hybridized with the blots by shaking overnight in hyb-solution at 42°. The hybridized blots were then washed in 0.1% SDS/0.1× SSC, first for 30 min at 42°, and then two more times, each for 20 min, either at 42° (reduced stringency) or at 65° (high stringency). The washed blots were exposed to X-ray film at −70° with two intensifying screens.

***In situ* hybridization and cytological analysis:** Polytene chro-

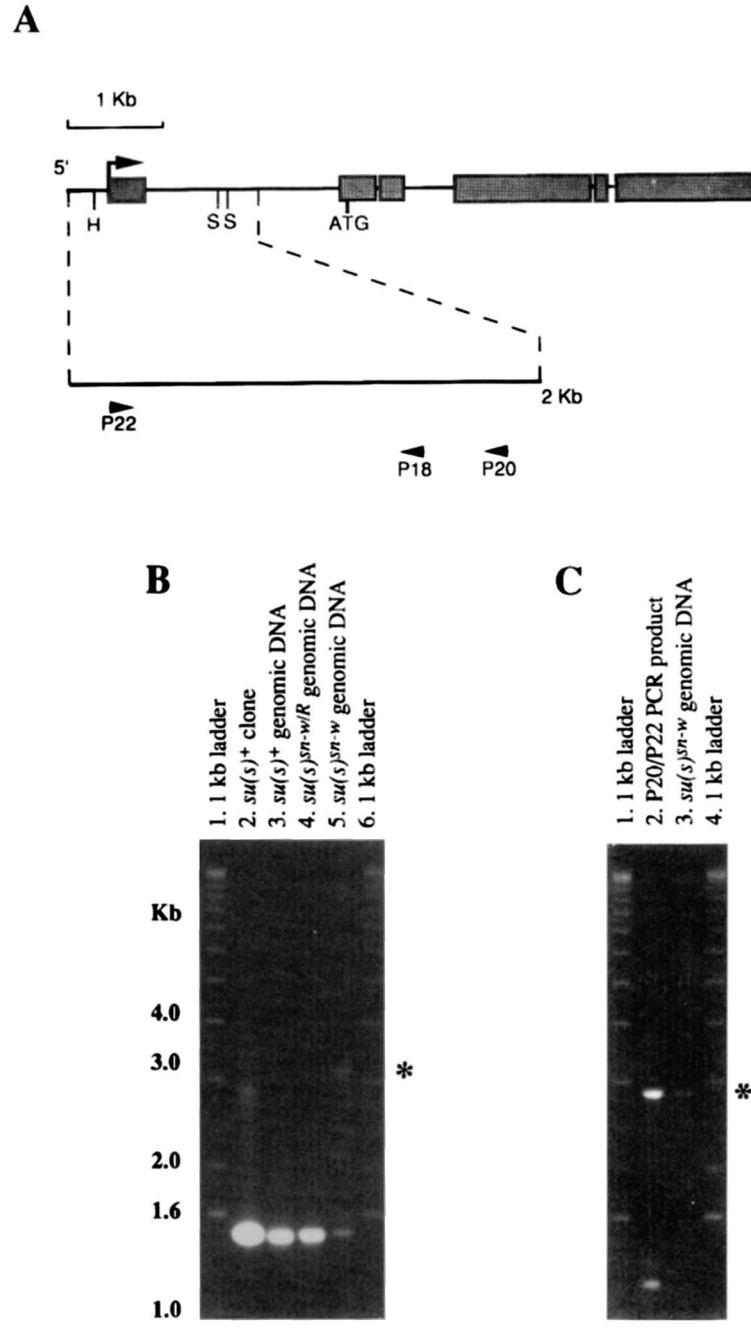


FIGURE 1.—Characterization of an insertion in the $su(s)^{sn\text{-}w}$ mutation. (A) Map of the *su(s)* gene showing exons (rectangles) and introns. Nucleotides are numbered from left to right according to the coordinate system of VOELKER *et al.* (1990). The transcription initiation site (bent arrow) is at coordinate 415, and the putative translation initiation codon (ATG) begins at coordinate 2975. The cleavage sites for the restriction enzymes *Hin*dIII (H) and *Sal*I (S) are indicated. The 5′ end of the gene is enlarged to show the positions of oligonucleotides that were used in PCR analysis; arrows indicate the direction in which these oligonucleotides would prime DNA synthesis. (B) Products from PCR amplification of *su(s)* DNA using primers P22 and P20. Amplifications were conducted in 100-μl volumes with 30 ng of each primer, 2.5 units AmpliTaq DNA polymerase (Perkin Elmer Cetus) and 1–2.5 μl template DNA. The $su(s)^{+}$ clone that was amplified was p4.1 (CHANG *et al.* 1986); it was linearized by digestion with *Eco*RI, and diluted to 1 ng/μl. Genomic DNA for the other three reactions was extracted from 20–50 adult males and resuspended in a volume of 20 μl. The profile for these reactions was 1.5 min at 94°, followed by 30 cycles of 1 min at 94°, 1 min at 55° and 3 min at 70°, followed by 5 min at 70°. *, the PCR product with the 1.7-kb insertion in the $su(s)^{sn\text{-}w}$ mutation. A template-free control made from the same reaction mix failed to yield any detectable product (not shown). (C) Products from PCR amplification of *su(s)* DNA with primers P22 and P18. Amplifications were conducted as in B, except that the profile was 1.5 min at 94°, followed by 30 cycles of 1 min at 94° and 3 min at 70°, followed by 5 min at 70°. The template for the reaction in lane 2 was a 1-μl sample from a gel slice containing the 3.2-kb PCR product shown in B, lane 5. This sample was expected to contain smaller PCR products as contaminants. The template for the reaction in lane 3 was 1 μl $su(s)^{sn\text{-}w}$ genomic DNA from the same extract that yielded the products shown in B, lane 5. *, the PCR product with the 1.7-kb insertion in the $su(s)^{sn\text{-}w}$ mutation. A template-free control made from the same reaction mix failed to yield any detectable product (not shown).

Clone	Left Flank	Right Flank
pSsu(s)	C G C A C A T A T A/S/T A T A T A A A T C	
pSa	G G C A T A C A T A/S/T A T A T G G G G A	
pSb	C A A A G A A A T A/S/T A T A A C A A T T	
pS1, pS2	A T T A C A T A T A/S/T A T A T G T T T C	
pS3	T T G C C A A A T A/S/T A T T T G G C A A	
pSc	T A T A C A T A T A/S-interrupted	
pSd	T A G A T A T A T A/S-interrupted	
pS4	C A C A T A C A T A/S/T A T A T G G A A A	
pS5	T G C A T A G A T A/S-interrupted	
Consensus	N N N A Y A N A T A/S/T A T A T R N N N N 5'	3'

FIGURE 2.—Sequences flanking cloned *S* elements. Palindromic sequences are underlined.

mosome preparations from the salivary glands of female third instar larvae were hybridized with a biotinylated probe made from the plasmid clone pSsu(s) according to published methods (LIM 1993); the labeled chromosomes were analyzed with phase-contrast optics at 500×.

RESULTS

A novel insertion mutation in the *suppressor of sable* gene: A spontaneous mutation of the *X*-linked *suppressor of sable* gene led to the discovery of the *S* family of transposable elements. This mutation, denoted $su(s)^{sn\text{-}w}$, was identified because it suppressed the phenotype of *singed-weak* (sn^w), a *P*-element-insertion mutation of the *X*-linked *singed* (*sn*) bristle gene; see SIMMONS *et al.* (1987) for a brief account of the discovery of the suppressor mutation. Males with the sn^w mutation have moderately singed bristles, but $su(s)^{sn\text{-}w}$ sn^w males have wild-type bristles. Recombination tests indicated that the suppressor mutation was tightly linked to the *yellow* (*y*) gene near the tip of the *X* chromosome, and complementation tests with *Y* chromosomes carrying terminal segments of the *X* chromosome established that the mutation was an allele of *su(s)*, one of the classical suppressor loci known to lie in this region. In addition, a revertant of the suppressor, denoted $su(s)^{sn\text{-}w/R}$, was discovered in a homozygous *y* $su(s)^{sn\text{-}w}$ sn^w stock. Males carrying this reversion had weak-singed rather than wild-type bristles.

Initially, the $su(s)^{sn\text{-}w}$ mutation and its revertant were characterized by genomic Southern blots hybridized with probes made from *su(s)* clones. These blots indicated that the $su(s)^{sn\text{-}w}$ mutation was associated with a 1.7-kb insertion between the *Hin*dIII and *Sal*I cleavage sites at the 5′ end of the *su(s)* gene (Figure 1A), and that the revertant had apparently lost this insertion. For more detailed analysis, DNA was amplified from $su(s)^{sn\text{-}w}$ and $su(s)^{sn\text{-}w/R}$ flies by PCR using two primers, P22 and P20, flanking the *Hin*dIII and *Sal*I sites in the *su(s)* gene (Figure 1A). Genomic $su(s)^{sn\text{-}w}$ DNA generated four products, 3.2, 2.2, 1.6 and 1.5 kb long (Figure 1B). The largest of these presumably corresponded to the 1.7-kb insertion plus 1.5 kb of flanking DNA, and the smallest apparently corresponded to wild-type $su(s)^+$ DNA. This smallest product was also seen when genomic $su(s)^{sn\text{-}w/R}$ DNA was amplified with the P22 and P20 primers, a result consistent with the Southern data that showed that the revertant allele had lost the 1.7-kb insertion. The 1.6-kb product, just slightly larger than the wild-type product, appeared to represent a nearly complete loss of the insertion in the $su(s)^{sn\text{-}w}$ DNA. Similar results were obtained when other *su(s)* primers were substituted for P20 in amplifications of $su(s)^{sn\text{-}w}$ DNA, except that no product corresponding to the 2.2-kb band was seen. This indicated that the 2.2-kb band was an artifact peculiar to the P20 primer.

The 3.2-, 1.6- and 1.5-kb bands generated in the initial amplification of the $su(s)^{sn\text{-}w}$ DNA suggested that this DNA was heterogeneous in structure, or that it behaved anomalously in PCR. To distinguish between these possibilities, we amplified the same genomic DNA sample in a PCR with a higher annealing temperature (70°), which might be expected to suppress anomalous products, and with primer P18 substituted for P20, to elimi-

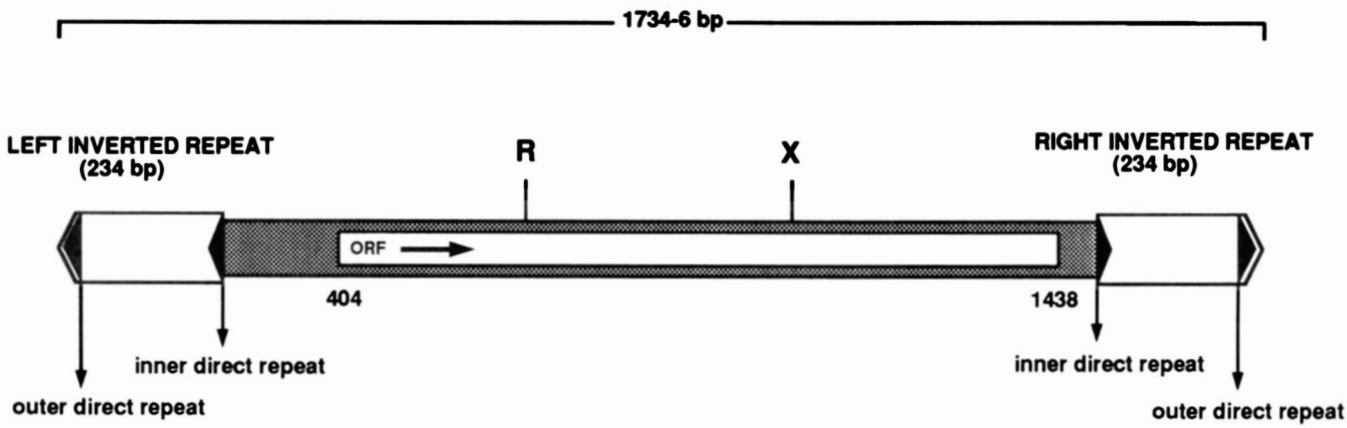


FIGURE 3.—Structure of the *S* element showing the long terminal inverted repeats, the short direct repeats within the inverted repeats, and the long internal ORF. Restriction enzyme recognition sites: R, *Eco*RV; X, *Xba*I.

TABLE 1

Variation in the sequences of the terminal inverted repeats of five 1.7-kb *S* elements

Position[a]	pS1	pSsu(s)	pSb	pS3	pSa
15	AA	AA	**GA**	AA	AA
17	TT	TT	**GT**	TT	TT
25	CC	CC	CC	CC	**CA**
28	TT	TT	TT	CC	CC
30	TT	TT	TT	**CT**	TT
31	**AG**	**AG**	**GA**	AA	GG
40	TT	TT	TT	AA	AA
43	AA	**CA**	**AC**	AA	AA
49	GG	**AG**	**GA**	GG	GG
59	CC	CC	CC	GG	GG
60	TT	TT	**GT**	**CT**	TT
61	AA	AA	AA	CC	CC
67	AA	AA	AA	TT	TT
74	**GT**	TT	TT	TT	TT
78	TT	TT	TT	**TC**	TT
92	CC	CC	CC	CC	**AC**
93	**AG**	AA	AA	AA	AA
104	AA	AA	AA	GG	GG
111	AA	TT	TT	AA	AA
135	AA	AA	AA	AA	**AG**
141	CC	CC	CC	TT	**TC**
154	AA	AA	AA	CC	**CA**
157	AA	AA	AA	TT	**TA**
162	TT	TT	TT	CC	**CT**
166	TT	TT	TT	TT	**CT**
191	CC	CC	CC	TT	**TC**
193	AA	AA	AA	**TA**	AA
194	AA	AA	AA	**TA**	AA
213	CC	CC	CC	GG	**GC**
216	TT	TT	TT	**CT**	TT

[a] The numbers indicate the positions of variable nucleotides in the inverted repeats of the cloned *S* elements, counting inward from the terminus. Nucleotides in the left inverted repeat are given in the first column under each clone. The corresponding nucleotides from the opposite strand of the right inverted repeat are given in the second column. Mismatches between the left and right inverted repeats are shown in boldface.

nate the 2.2-kb artifact. This amplification generated a single 2.9-kb product corresponding to the 1.7-kb insertion plus the expected amount of flanking *su(s)* DNA (Figure 1C). The smaller PCR products seen in the initial amplification were therefore suppressed, suggesting that they were caused by an amplification anomaly such as strand slippage or mispairing, and not by heterogeneity in the template DNA.

Cloning the *su(s)$^{sn-w}$* insertion and sequences homologous to it: Probes made from the 3.2- and 2.9-kb products from these PCR amplifications were used to screen two genomic DNA libraries (S.1, from a *y su(s)$^{sn-w}$* strain, and 6C4, from an unrelated *su(s)*$^+$ strain) for sequences homologous to the insertion in the *su(s)$^{sn-w}$* mutation. Five clones (pS1, pS2, pS3, pS4 and pS5) were isolated from the S.1 library, and four (pSa, PSb, pSc and pSd) were isolated from the 6C4 library. In addition, the 1.5-kb PCR product from *su(s)*$^+$ DNA was used to screen a genomic DNA library (S.2, from a *y su(s)$^{sn-w}$ cv v car* strain) for a clone containing the *su(s)$^{sn-w}$* mutation. A single such clone, designated pSsu(s), was identified.

Each of these clones was mapped by digestion with an array of restriction enzymes, and the regions that were homologous with the *su(s)$^{sn-w}$* insertion were identified by Southern hybridization with probes made from *su(s)$^{sn-w}$* PCR products. A pattern formed by the recognition sites for two restriction enzymes (*Eco*RV and *Xba*I) was found in six of the clones, including pSsu(s), suggesting a diagnostic motif for the type of element inserted in the *su(s)$^{sn-w}$* allele. Nested deletion subclones of pS2, which contained this motif, were constructed and sequenced. The results showed that pS2 contained a 1.7-kb element with long inverted repeats at each end. Because this putative transposon had been isolated from the S.1 library, we named it the *S* element. Oligonucleotides made from the sequence of this element were then used to determine the sequences of the elements in all the other clones. In five of these [pSsu(s), pSa, pSb, pS1 and pS3], a 1.7-kb element with the same basic organization as the element in pS2 was identified. In the others, an element with an incomplete sequence was found.

***S* element insertion sites:** Figure 2 shows the sequences around the insertion site of each cloned *S* element, including three that were interrupted, either by the insertion of a different transposable element or by enzymatic cleavage during cloning. The combined sequence data do not unambiguously define the ends of these elements because a six-base palindrome, ATA/TAT, was present at each insertion site. However, by comparing the sequence of the pSsu(s) clone with that of the wild-type *su(s)* gene (VOELKER *et al.* 1990), we know that no more than two of these six nucleotides could be a part of the element; moreover, by analogy with other transposons, these two nucleotides could be the result of a target-site duplication. If we assume that such a duplication was created by the insertion of each element, then the cloned 1.7-kb *S* elements were either 1734 bp long [pSu(s), pSa, pSb, pS1 and pS2] or 1736 bp long (pS3). The other cloned elements were shorter, either because they were interrupted or had internal deletions.

Various lengths of flanking sequence were determined for each of the cloned *S* elements. Among these, two (pS1 and pS2) had identical flanking sequences, indicating that they represented the same insertion; however, there were nucleotide differences between the *S* elements within these clones, suggesting that the stock from which they came was polymorphic for two slightly different elements at this insertion site. Inspection of the sequences flanking the 10 cloned *S* elements revealed a consensus sequence, AYANATA/TATATRN, which is a quasi-palindrome. Although we cannot be sure about the exact insertion site, one possibility is that

```
1      cagtttgtca agaaactgttta cacaccgcaaaa taagtagaattt ttgactttaaag gccaaaattaag   70

71   ggtttttttgctt aattaaacgcaa ttttttttatgaa atataattaaac aatatttatttt acttataaatta  142

143  aaaaacaaattc aatatatcaaat atacaagaaaat aaacaacaaatt tcttgtttacac acttttgagagt  214

215  gccaagaaactc tttacacagttt tgggttcctact ttgttttgctct ttttcttagaaa caatctcatttt  286

287  tccgttatttttt gtcttatgcatt ccttttttacaac gcttctattgca atttttttcactt tgcttgtgaaat  358

359  tttgttgatcta acgtgcttaaag cgaattattaaa tttaatgaaATG CCTGGAAAGAGA TTGGCTTTTGAA  430
                                                  M   P   G   K   R   L   A   F   E
431  GTGACCCAGCTA ATATACTATAAC CACCAGTTGGGA AAATCTATTCCT GAATTAGTAGAA ATATTTTCCGTA  502
      V   T   Q   L   I   Y   Y   N   H   Q   L   G   K   S   I   P   E   L   V   E   I   F   S   V
503  TCCCGTAAAACC GTCTATAATATT TTAAATCGTGCG GAAAAAGAGGGC AGGCTTGAACCT AAGAGTGGTGGT  574
      S   R   K   T   V   Y   N   I   L   N   R   A   E   K   E   G   R   L   E   P   K   S   G   G
575  GGGTGTAAAACG AAAATTAACAAG CGAGTAGACCGC CTTATTATGCGA AAAGCGATTGCG AACCCCCGAATC  646
      G   C   K   T   K   I   N   K   R   V   D   R   L   I   M   R   K   A   I   A   N   P   R   I
647  TCGGTCAGATCA CTTGCTCAGGAT ATCAGGGAAGAA TGTCACCTAACT GTATCACACGAA ACTGTGCGCCAA  718
      S   V   R   S   L   A   Q   D   I   R   E   E   C   H   L   T   V   S   H   E   T   V   R   Q
719  GTCATCCTACGC CATAGGTACTCT TCAAGAGTTGCA AGAAAAAAGCCT TTGCTATCAGAG ATCAATATTGAA  790
      V   I   L   R   H   R   Y   S   S   R   V   A   R   K   K   P   L   L   S   E   I   N   I   E
791  AAGCGTCATTCA TTCGCTGTGAGC ATGATGGATCAT GCGGAAGAGTAC TGGGATGACGTC ATATTTTGTGAC  862
      K   R   H   S   F   A   V   S   M   M   D   H   A   E   E   Y   W   D   D   V   I   F   C   D
863  GAAACAAAAATG ATGCTCTTTTAT AACGATGGGCCA AGCAGAGTATGG CGCAAACCGTTG AGTGCGCTAGAA  934
      E   T   K   M   M   L   F   Y   N   D   G   P   S   R   V   W   R   K   P   L   S   A   L   E
935  ACACAAAATATA ATTCCAACAATC AAATTTGGAAAA TTGTCAGTGATG ATTTGGGGCTGT ATTTCCAGCCAT 1006
      T   Q   N   I   I   P   T   I   K   F   G   K   L   S   V   M   I   W   G   C   I   S   S   H
1007 GGAGTGGGCAAA CTAGCCTTTATT GAAAGCACTATG AATGCCGTGCAA TATCTAGATATT TTAAAAACAAAT 1078
      G   V   G   K   L   A   F   I   E   S   T   M   N   A   V   Q   Y   L   D   I   L   K   T   N
1079 TTGAAGGCCAGT GCAGAAAAATTT GGTTTGTTTAGC AACAACAAGCCA AATTTTAAGTTT TATCAGGACAAT 1150
      L   K   A   S   A   E   K   F   G   L   F   S   N   N   K   P   N   F   K   F   Y   Q   D   N
1151 GATCCCAAACAT AAAGAGTACAAT GTACGCAACTGG CTACTCTATAAC TGTGGCAAGGTG ATCGATACGCCC 1222
      D   P   K   H   K   E   Y   N   V   R   N   W   L   L   Y   N   C   G   K   V   I   D   T   P
1223 CCTCAGAGTCCT GATCTAAACCCC ATTGAAAATTTG TGGGCCTACTTA AAGAAGAAGGTT GCAAAAAGGGGC 1294
      P   Q   S   P   D   L   N   P   I   E   N   L   W   A   Y   L   K   K   K   V   A   K   R   G
1295 CCCAAAACTCGA CAACAACTCATG GCTGCGATAATC GAAGAGTGGGAA AAGATCCCGCTT GAATATGACCTA 1366
      P   K   T   R   Q   Q   L   M   A   A   I   I   E   E   W   E   K   I   P   L   E   Y   D   L
1367 AAAAAACTCATA CATTCCATGAAA AAAAGGCTTCAA CTTGTAGCCAAA GCCAATGGGGGT CATACTAAATAC 1438
      K   K   L   I   H   S   M   K   K   R   L   Q   L   V   A   K   A   N   G   G   H   T   K   Y
1439 taaaacttttca aatattatcaaa ataattaaaaaa tttaggattaaa cttaggtttagt gtttTGTGTAAA 1510

1511 gagtttcttgac actctcaaaagt gtgtaaacttga aatttgttgttt attttcttgtat atttgatatatt 1582

1583 gaatttgttttt taatttataagt aaaataaatatt gtttaattatat ttcataaaaaaa ttgcgtttaatt 1654

1655 aagcgaaaaacc cttaattttgac ctttaaagtcaa aaattctactta ttttacggtgtg taaacagtttct 1726

1727 tgacaaactg                                                                  1736
```

FIGURE 4.—Sequence of the 1736-bp *S* element in the clone pS3 (GenBank accession number U33463). The long inverted terminal repeats are in boldface, the short direct repeats within them are underlined, and the long ORF and its putative polypeptide are shown in upper case. The underlined italicized sequences are restriction enzyme recognition sites (GATATC, *Eco*RV, and TCTAGA, *Xba*I).

each element was inserted into the target DNA after staggered cleavage around a central TA dinucleotide within a palindrome. Cleavage 5′ to the T on each DNA strand would create a gap into which the element could be inserted, and repair of this gap would then generate a 2-bp target-site duplication. On this hypothesis, three of the uninterrupted *S* elements had been inserted into the palindrome AT/AT, one into the palindrome ATAT/ATAT, one into the palindrome ACATAT/ATATGT and one into the palindrome TTGCCAAAT/ATTTGGCAA. To investigate the nature of *S* element excisions, we sequenced a PCR product from the revertant allele $su(s)^{sn\text{-}w/R}$ and found that all but 2 bp from the *S* element's right end had been lost; however, all the flanking nucleotides, including the presumptive target-site duplication, had been retained. This revertant allele was therefore due to an "imprecise" excision of the inserted *S* element.

The inverted terminal repeats of *S* elements: Figure 3 shows the overall structure of the six 1.7-kb *S* elements that were sequenced. Each element had 234-bp inverted terminal repeats that, in turn, contained two almost perfect direct repeats of 21 bp. The outer direct repeat was indented 5 bp from the end of the element and the inner direct repeat was located at the inside margin of the inverted repeat. Direct repeats within inverted

TABLE 2

Similarity matrix comparing the sequences of 1.7-kb *S* elements[a]

	pS1	pS2	pSsu(s)	pSb	pSa	pS3
pS1	1.0000	0.9977	0.9844	0.9810	0.9602	0.9054
pS2		1.0000	0.9850	0.9815	0.9608	0.9060
pSsu(s)			1.0000	0.9931	0.9573	0.9048
pSb				1.0000	0.9550	0.9031
pSa					1.0000	0.9262
pS3						1.0000

[a] Similarities were computed using the GCG program Pileup.

terminal repeats have also been found in the *Minos* element isolated from *D. hydei* (FRANZ and SAVAKIS 1991), in the *Tdr1* element from the zebrafish, *Danio rerio* (IZSVAK *et al.* 1995), and in the *Tc3* element from *C. elegans* (P. ANDERSON, personal communication). The inverted terminal repeats of *S* elements are very A:T-rich (77% in pS3). Within a particular element, the left and right repeats are nearly identical, differing in fewer than 10 nucleotides. Greater differences are seen when the left repeats from two elements are compared with each other (as many as 21 nucleotide differences), or when the right repeats are compared with each other (as many as 16 differences). This indicates that the inverted repeats are more similar within than between elements.

Of the 468 nucleotides within the two inverted repeats, 46 were variable among the six 1.7-kb elements that were sequenced (Table 1). An interesting feature of this variation is that nucleotide substitutions in the left repeat were often accompanied by the same substitutions at corresponding positions in the right repeat. For example, the left inverted repeats of the elements in clones pSsu(s) and pS3 differed in 20 positions and the right inverted repeats differed in 15 positions; however, 13 of these 15 positions corresponded to nucleotide differences in the left repeats. This suggests that nucleotide substitutions in the two inverted repeats occur in a concerted manner.

Another feature of the sequence variation within the inverted repeats is that in some elements the left and right repeats appear to possess segments derived from different elements. The element in the clone pSa is an example. The outer halves of the repeats in this element are reasonably well matched (only three mismatches in the first 134 nucleotides), but the inner halves are not (seven mismatches in the next 100 nucleotides). The reason seems to be that a portion of the left repeat beyond nucleotide 135 is similar to the left repeat of the element in clone pS3, but the corresponding portion of the right repeat is similar to the right repeats of the elements in clones pS1, pSsu(s) and pSb. It therefore appears that segments of the left and right inverted repeats in the pSa element were derived from two different kinds of *S* elements.

The coding sequences of *S* elements: Among the six 1.7-kb elements that were analyzed, the 1736-bp-long element in the clone pS3 (Figure 4) had the longest ORF. This ORF, from bp 404 to 1438, could encode a polypeptide of 345 amino acids (aa). In the other 1.7-kb *S* elements, this ORF was interrupted by stop codons and frameshifts. The putative polypeptide of the long ORF in pS3 has a predicted molecular weight of 40 kD and an estimated isoelectric point of 10.53. Leucines at positions 6, 13, 20 and 27 in this basic polypeptide form a leucine zipper motif that could play a role in protein-protein and/or protein-DNA interactions.

Sequence variation among 1.7-kb *S* elements: The 1.7-kb *S* elements that were sequenced differed in as much as 9.4% of their nucleotides. Table 2 gives the similarity matrix for these six elements. The pS1 and pS2 elements, which represent the same insertion, differed only slightly from each other. These two elements and the pSsu(s) and pSb elements form a closely related group. The pSa element is somewhat removed from this group, and the pS3 element, which was the only element with a long ORF, is even more removed. These distance data clearly show that the *S* family has undergone mutational diversification.

***S*-element homology with other transposons:** Computer analysis (Table 3) of the long ORF within the pS3 clone indicates that *S* elements belong to the *mariner-Tc1* superfamily of transposons (DOAK *et al.* 1994; ROBERTSON 1995). This superfamily includes the *Tc1* element from *C. elegans* (ROSENZWEIG *et al.* 1983; SCHUKKINK and PLASTERK 1990), the *Uhu* element from *D. heteroneura* and other Hawaiian Drosophila (BREZINSKY *et al.* 1990, 1992), the *Minos* element from *D. hydei* (FRANZ and SAVAKIS 1991; FRANZ *et al.* 1994), the *HB1* (BRIERLEY and POTTER 1985) and *Bari-1* (CAIZZI *et al.* 1993) elements from *D. melanogaster,* and the *mariner* element from *D. mauritiana* and *D. simulans* (JACOBSON *et al.* 1986; MEDHORA *et al.* 1991). Recent analyses have identified other members of this transposon superfamily, including a large number of *mariner*-like elements in many different arthropods (ROBERTSON 1993; ROBERTSON and MACLEOD 1993), a *mariner*-like element in the fungus *Fusarium oxysporum* (DABOUSSI *et al.* 1992) and several *Tc1*-like elements in the genomes of differ-

TABLE 3

Comparison of putative polypeptides encoded by transposons related to the *S* element

Element	Accesion No.	Species	Transposon length (bp)	Polypeptide length (aa)	Similar[a]	Identical[b]
S (pS3)	U33463	*D. melanogaster*	1736	345	1.000	1.000
Bari-1	X67681	*D. melanogaster*	1726	340	0.494	0.311
HB1	X01748	*D. melanogaster*	1643–49	148[c]	0.476	0.279
mariner	M14653	*D. mauritana*	1286	346	0.428	0.194
Minos2	Z29098	*D. hydei*	1773	361[d]	0.494	0.325
Uhu	X17356	*D. heteroneura*	1647	192[c]	0.562	0.365
Tc1	X01005	*C. elegans*	1611[e]	343[f]	0.504	0.319

[a] Fraction of amino acids similar to those in the putative *S* polypeptide.
[b] Fraction of amino acids identical to those in the putative *S* polypeptide.
[c] This polypeptide lacks a region comparable to the C-terminal amino acids of the putative *S* polypeptide.
[d] Compared to the putative *S* polypeptide, this polypeptide has 22 extra amino acids at its N-terminus.
[e] SCHUKKINK and PLASTERK (1990) suggest that the element is 1 bp longer than reported in the original sequence (ROSENZWEIG *et al.* 1983).
[f] Based on the cDNA identified by VOS *et al.* (1993).

ent fish (WILSON *et al.* 1990; HIERHORST *et al.* 1992; RADICE *et al.* 1994; IZSVAKS *et al.* 1995). These transposons range from 1.2 to 1.8 kb in size and have inverted terminal repeats, albeit of quite different lengths. The long ORFs in intact members of the *mariner-Tc1* superfamily evidently encode transposases.

The inverted terminal repeats of *S* elements also provide evidence for membership in the *mariner-Tc1* transposon superfamily. Figure 5 compares the first 26 nucleotides of the left inverted repeat, *i.e.*, the terminal nucleotides plus the 21-nucleotide direct repeat within the inverted repeat, with sequences in the inverted terminal repeats of other transposons, including *P* and *hobo* that do not belong to the *mariner-Tc1* superfamily.

Element	ITR(bp)		Sequence	DR (bp)
S	234		CAGTT-T-G-T-C-AAGAAACTGTT-T-ACACA...	21
Bari-1	27		CAGTcaT-GgT-C-A--AAAtTaTT-Tt-CACAa	
HB1	27		CAGcTgT-GtT-C--AGAAA-aaTag---CAgtgc	
Uhu	46		CAGT----G-T-CttAGA-gCT...	
Minos	254	...	CAGT-gc-G---C-AA-----TGTT-TaACACA...	18
Tc1	54		CAGTgcT-Gg--C$_T$AAGtAACTtTT-Tt-Ct (caaa) atatccactttggtttttttgtgtg	
Tc3	462		CAGT----G-T-g--gGAAA--GTTcT-AtAcga...	32
Tdr1	211		CAGTTgaaG-T-C-g-GAA---GTT-T-ACAtAc...	12+
mariner	28	c	CAGg--T-G-TaC-AAGt-AggGaa-TgtCggtt	
hobo	12		CAG----------AGAA-CTG--------CA	
P	31		CA----T-G----At-GAA...	

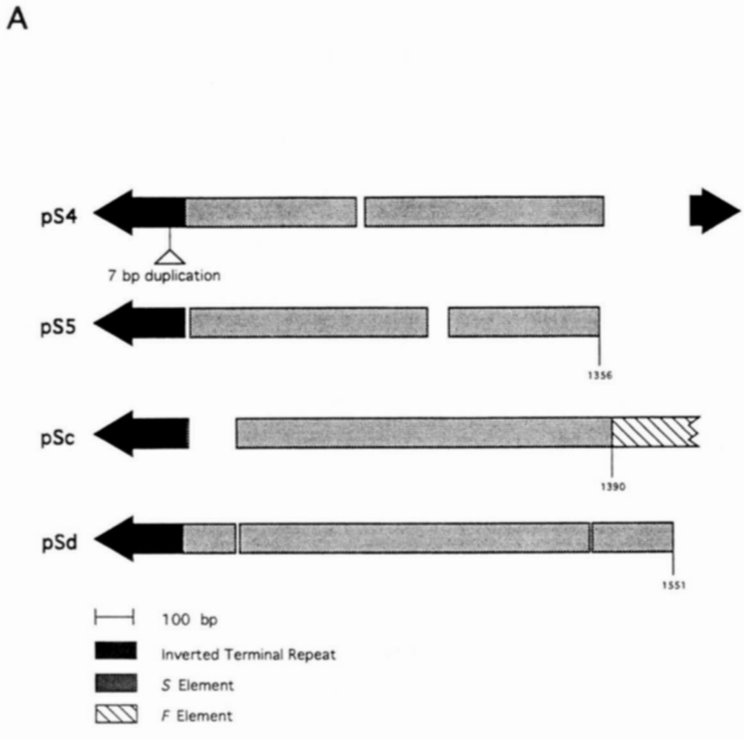
FIGURE 5.—Comparison of the terminal nucleotides of *S* to the inverted terminal repeats (ITR) of several transposons. Nucleotide identities are shown in upper case, differences in lower case. Gaps to bring the sequences into alignment are indicated by hyphens, and nucleotides that belong to direct repeats (DR) within the ITRs are underlined. In the *Tdr1* element, only the perfect direct repeats noted by IZSVAK *et al.* (1995) are underlined; however, these may be extended outward by several nucleotides if imperfect matches are allowed.

In several cases, the ends of these transposons (and therefore the lengths of their repeats) are not precisely defined. Nevertheless, it is still possible to align the sequences and identify similarities. The presumptive ends of six of the 11 elements (*S, Bari-1, Uhu, Tc1, Tc3* and *Tdr1*) are demarcated by the sequence CAGT, which has previously been recognized as a characteristic of the *Tc1* transposon family (HENIKOFF 1992); this motif is also present in the *Minos* element, but in *Minos* it is indented 43 bp from the presumptive end. *HB1, mariner* and *hobo* have the trinucleotide CAG at or near their ends and *P* has the dinucleotide CA at its ends. Proceeding rightward from the terminus, we can recognize other regions of similarity between *S* and the other transposons; however, the functional significance of these similarities is not known.

Four of the elements listed in Figure 5 have direct repeats within their inverted terminal repeats. In *Tc3* these have been shown to be binding sites for the element's transposase (COLLOMS *et al.* 1994). It therefore seems plausible that the direct repeats within the terminal repeats of *S* elements serve a similar function. Curiously, the direct repeats of the *Minos* element are found outside the CAGT tetranucleotide that is characteristic of the ends of several *mariner-Tc1* elements.

Incomplete *S* elements: Four of the 10 cloned *S* elements were incomplete (Figure 6A), including two (pSd and pS5) that were truncated during cloning and one (pSc) that was interrupted by the insertion of another transposon. All four of these incomplete elements had deletions of internal sequences, and one of them (pS4) had a 7-bp duplication in the left inverted terminal repeat. These clones demonstrate that structurally "defective" *S* elements are present in the *D. melanogaster* genome.

In screening data banks for sequences similar to *S* elements, we discovered that *S*-element fragments are present in two clusters of heat shock response genes on



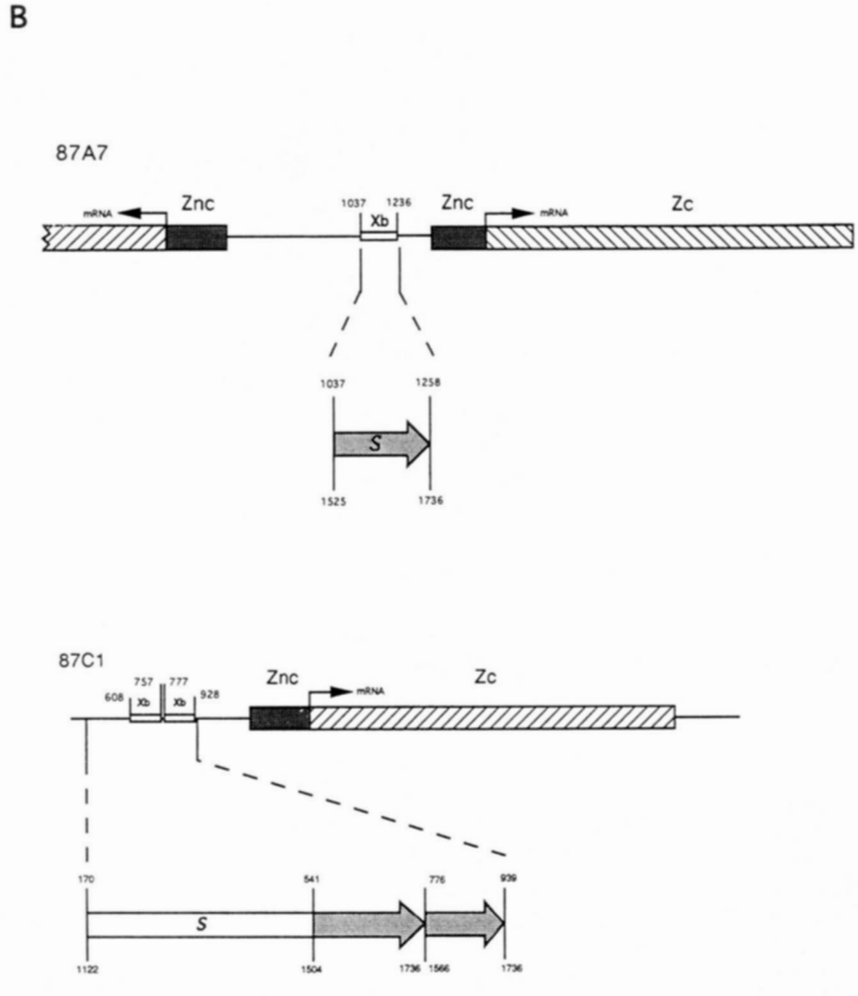


FIGURE 6.—(A) Structure of incomplete *S* elements. The coordinates indicating the truncation points are based on the sequence in Figure 4. (B) Location of *S*-element fragments in the *hsp70* loci from cytological regions 87A7 and 87C1. Zc, coding region of the *hsp70* genes; Znc, noncoding promoter. The coordinates of the Xb homology sequences are given above the maps and those of the *S*-element fragments are given below.

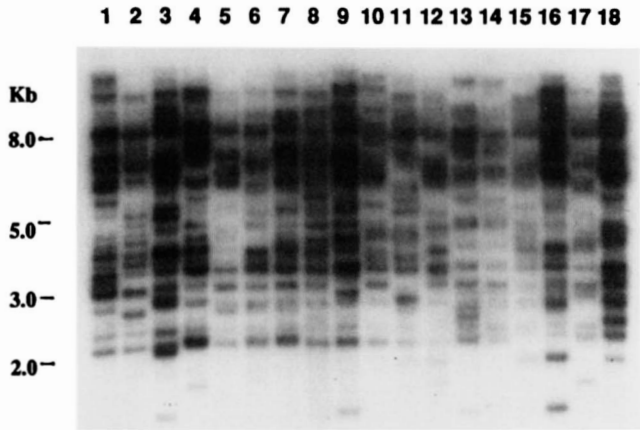


FIGURE 7.—Southern hybridization of an *S*-element probe with genomic DNA from 18 wild-type strains of *D. melanogaster*. The probe, made by random primer labeling of a PCR product generated from the *su(s) S* element with the S-IR primer, was hybridized with the blot under conditions of high stringency. Strains: 1, Oregon-R B; 2, Gaiano, Italy; 3, Sexi, Spain; 4, Samarkand, Uzbekistan; 5, Winnepesaukee, New Hampshire, 74i; 6, Winnepesaukee, New Hampshire, 76i; 7, Raleigh, North Carolina, NC38; 8, Raleigh, North Carolina, NC44; 9, St. Paul, Minnesota, HS4; 10, St. Paul, Minnesota, HS5; 11, Iquitos, Peru; 12, Surinam, 78; 13, Bujumbura, Burundi, II; 14, Bujumbura, Burundi, I; 15, Israel, QA-B81; 16, Hachijojima, Japan, 77; 17, Tottori, Japan; 18, Canberra, Australia, #70. Strains 1, 2 and 4 were from S. BECKENDORF, strains 9 and 10 were from F. SHEEN and all the others were from M. KIDWELL.

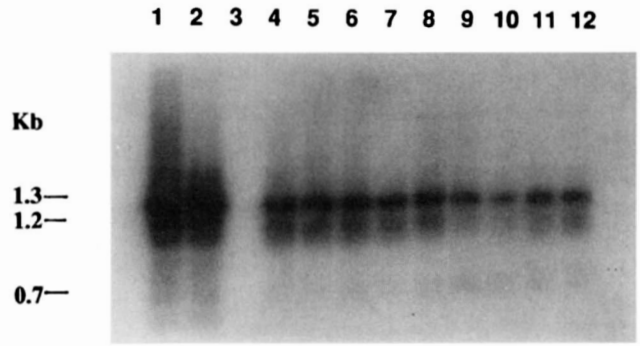


FIGURE 8.—PCR amplification of *S* elements in various *D. melanogaster* genomes. Amplifications were conducted in 25 μl volumes with 75 ng S-IR primer, 0.65 units Taq DNA polymerase (Promega) and 2 μl template DNA. Genomic DNA templates were prepared according to GLOOR and ENGELS (1992). The reaction profile was 2 min 15 sec at 92°, followed by 30 cycles of 45 sec at 92°, 2 min at 55° and 2 min at 72°, followed by 5 min at 72°. The products were fractionated in a 0.8% agarose gel, blotted to a nylon membrane and hybridized with a ^{32}P-labeled internal *S*-element probe spanning bp 274–1424. Amplification templates: 1, gel-purified 3.2-kb PCR product from Figure 1B, lane 5; 2, pSb clone, diluted to 1 ng/μl; 3, no template; 4, *y sn^3 v car;* 5, *C(2)EN, b cn bw;* 6, *ci ey^R;* 7, *cn bw;* 8, Oregon-R B; 9, Amherst, Massachusetts, 88–6; 10, Winnepesaukee, New Hampshire, 76i; 11, Israel, QA-B81; 12, Hachijojima, Japan, 77. Templates 4–8 are from laboratory strains and templates 9–12 are from strains recently derived from natural populations.

chromosome *3*. These fragments are situated upstream of the three *hsp70* genes in the 87C1 cluster and between the two divergently transcribed *hsp70* genes in the 87A7 cluster (Figure 6B). Sequence analysis shows that the *S*-element fragments in these clusters correspond to segments previously denoted as Xb homology sequences (MIRKOVITCH *et al.* 1984). The 87A7 cluster contains only one *S*-element fragment and it was derived from an inverted terminal repeat. The 87C1 cluster contains two *S*-element fragments inserted next to each other, one from an inverted terminal repeat and the other from the right half of an *S* element. Previous studies have indicated that the region around these fragments is an attachment site for the nuclear scaffold (MIRKOVITCH *et al.* 1984), raising the possibility that the AT-rich *S*-element termini play a role in chromatin organization.

Taxonomic distribution of *S* elements: Genomic Southern blotting and PCR were used to screen strains of *D. melanogaster* for *S* elements. For the Southern analysis, DNA was extracted from various strains and digested with *Eco*RI, which does not cut within 1.7-kb *S* elements. The autoradiogram in Figure 7 shows the results for 18 wild-type strains derived from populations all over the world. Many fragments from each genome hybridized with an internal *S*-element probe, indicating that *S* sequences are moderately repetitive. Moreover, some of the same hybridizing fragments appeared to be present in several different genomes, suggesting a limited conservation of *S*-element position.

A more extensive investigation of the distribution of *S* elements in *D. melanogaster* populations was conducted using PCR with a primer spanning 25 nucleotides near the inner border of the inverted repeat sequence. This primer, denoted S-IR, was used to amplify genomic DNA obtained from single flies. Altogether 114 strains were screened, including 15 laboratory stocks and 99 stocks derived from natural populations; most of the latter were from North America, but there were representatives from the Mediterranean Basin, central Africa, South America, central Asia, Japan and Australia. In every amplification of genomic DNA, three PCR products were clearly observed and each hybridized with an internal *S*-element probe; see Figure 8 for examples. The largest of these products, 1.3 kb long, corresponded to the sequence between the primer sites in the inverted repeats. The other PCR products, 1.2 and 0.7 kb long, were apparently artifactual because they were generated from a cloned *S* element as well as from elements present in genomic DNA.

Other Drosophila species were also examined for the presence of *S* elements. Figure 9 shows the results of reduced stringency hybridization between an *S*-element probe and genomic DNA from seven stocks of *D. simulans* and one stock of *D. mauritiana*. In each case, three bands were detected, including one corresponding to high molecular weight DNA. However, these bands

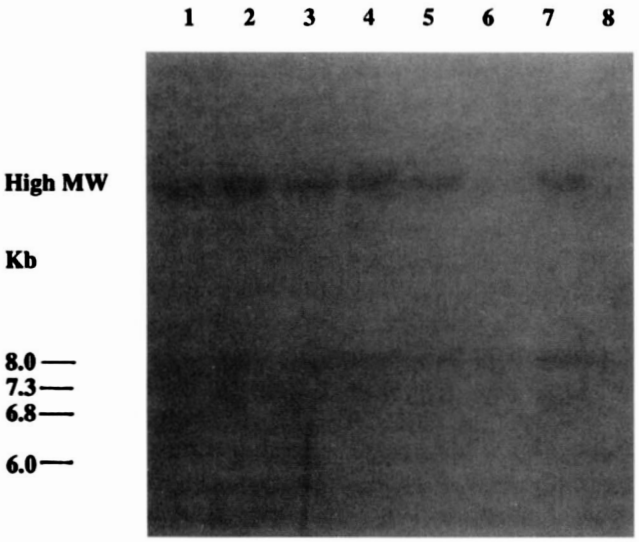


FIGURE 9.—Southern hybridization of an *S*-element probe with genomic DNA from *D. simulans* and *D. mauritiana*. The probe, the same one used in Figure 7, was hybridized under conditions of moderate stringency. Strains: 1, *D. simulans*, St. Anthony Park, Minnesota, F-3; 2, *D. simulans*, St. Anthony Park, Minnesota, F-19; 3, *D. simulans*, Roseville, Minnesota, R-1; 4, *D. simulans*, Northeast Minneapolis, Minnesota, N-1; 5, *D. simulans*, Northeast Minneapolis, Minnesota, N-2; 6, *D. simulans, v*, Cal Tech Stock Center; 7, *D. simulans*, National Drosophila Species Resource Center, BG-0251.0; 8, *D. mauritiana*, National Drosophila Species Resource Center, BG-0241.0.

were fainter than many of the bands seen with *D. melanogaster* DNA (not shown), suggesting that the *simulans* and *mauritiana S*-like sequences were only partially complementary to the *melanogaster* probe. PCR amplifications of *D. simulans* and *D. mauritiana* genomic DNA using the S-IR primer failed to generate any detectable products. Reduced stringency hybridization was also used to screen for *S* sequences in 17 other Drosophila species (*affinis, ananassae, austrosaltans, cardini, dunni dunni, equinoxialis, funebris, hydei, immigrans, kepuluana, nebulosa, paulistorum, pseudoobscura, tropicalis, virilis, willistoni* and *yakuba*); however, in no case were any bands observed (data not shown).

Cytological distribution of *S* elements on *D. melanogaster* chromosomes: Ten different strains were analyzed by *in situ* hybridization of an *S*-element probe to polytene chromosomes. The strains included wild-type and marked laboratory stocks, as well as stocks derived recently from natural populations. The hybidizing probe was made from the plasmid pSsu(s), which contains unique sequences from the *su(s)* locus as well as a 1.7-kb *S* element. The unique *su(s)* sequences served as a positive control, labeling cytological position 1B12–13. This analysis was limited to the chromosomes of a single female from each strain and the labeled sites were localized only within lettered subdivisions on the cytological map.

Table 4 summarizes the data. Altogether, 383 labeled sites were detected, including the positive control site, which was labeled in all 10 of the strains; however, probably only one of them, L, which was homozygous for the $su(s)^{sn\text{-}w}$ mutation, actually contained an *S* element at this site. After adjusting for the control site, the average number of cytologically detectable *S* elements in a strain was 37.4. However, this average is inflated by the data from two strains, L and Canton S, which had 63 and 90 cytologically detectable *S* elements, respectively. All the other strains had between 23 and 32 labeled sites. The abundance of *S* elements in the L and Canton S strains suggests that the *S* family may have been unusually active in them. The occurrence of the *S*-induced $su(s)^{sn\text{-}w}$ mutation in strain L is consistent with this speculation.

Among the 10 strains analyzed, the average number of labeled sites per chromosome arm ranged from 6.2 on 2L and 3L to 8.7 on the *X*. Labeled sites were concentrated in the basal euchromatin and β-heterochromatin of each chromosome arm, but they were also scattered throughout the euchromatin. Only one telomere in one strain was labeled, indicating that *S* elements are not associated with this specialized chromosome structure. Labeled sites were also seen in the highly condensed heterochromatin abutting the chromocenter.

Neglecting the control site in 1B, 13 sites were consistently labeled in all ten strains: on the *X* chromosome, region 20; on 2L, 26B, 39E, and region 40; on 2R, 41A-E, 41F, 42C, and 44D; on 3L, region 80; on 3R, 82C, 82D, 82F and 87C. The last of these sites includes the *hsp70* loci, which sequencing analysis has shown to contain *S*-element fragments inserted upstream of one of the two genes (see above). The very small *S* fragment that was found between the two *hsp70* loci in 87A was not detected by *in situ* hybridization. In addition to the 13 sites that were consistently labeled by the *S*-element probe, seven sites were labeled in at least five of the strains: on the *X*, 16D; on 2L, 38C; on 2R, 46C; on 3L, 67B, 71D and 75B; on 3R, 98F. One of these, 67B, contains a set of heat shock genes (LINDSLEY and ZIMM 1992), suggesting a further association between the *S*-element and heat shock-inducible loci. On average, 20/37.4 = 53% of the sites labeled by the *S*-element probe were common to a majority of the strains examined.

DISCUSSION

S elements belong to the widespread *mariner-Tc1* superfamily of transposable elements. The members of this superfamily have inverted terminal repeats of different lengths and encode a transposase about 340 aa long. Studies with two members, *Tc1* and *Tc3*, suggest that this transposase binds to sequences near the ends of the inverted repeats (VOS *et al.* 1993; COLLOMS *et al.* 1994; VOS and PLASTERK 1994), and that it makes staggered cuts when an element is excised from or inserted into chromosomal DNA (VAN LUENEN *et al.* 1994). All the evidence suggests that these diverse transposons are

TABLE 4

The number of cytological sites in the chromosome arms of 10 different strains of *D. melanogaster* that were labeled by an *S*-element probe derived from the clone pSsu(s)

Strain	*X*	2L	2R	3L	3R	Total
Oregon-R B (USA)	8	6	4	5	5	28
Canton S (USA)	30	17	16	14	14	91
w m f	4	5	9	3	6	27
HS4 (USA)	9	4	6	4	7	30
L [$su(s)^{sm\text{-}w}$]	21	6	13	11	12	63
bw; st	4	5	7	6	7	29
Gaiano (Italy)	4	4	4	5	7	24
Canberra (Australia)	8	6	4	6	9	33
Iquitos (Peru)	6	6	8	2	12	34
Samarkand (Uzbekistan)	2	3	6	6	7	24
Total	96	62	77	62	86	383

quite ancient, and that during the course of their evolution, some have been transferred across species boundaries (MACLEOD and ROBERTSON 1993; ROBERTSON 1993, 1995).

A subgroup of elements within the *mariner-Tc1* superfamily (*S, Tc3, Minos, Tdr1* and related transposons from various fish species) is unusual in having rather long inverted terminal repeats with short direct repeats embedded within them. One copy of the direct repeat is located near the outer end of the inverted repeat, where the transposase presumably binds, and the other is located at or near the inner end (*S, Minos* and *Tdr1*), or in the middle (*Tc3*) of the inverted repeat. The distance between the two direct repeats ranges from 142 (in *Tc3*) to 201 bp (in *Minos*), and the direct repeats of three of the elements (*S, Tc3* and *Tdr1*) are similar in sequence. These conserved features suggest that the direct repeats are functionally significant. Indeed, physical studies have demonstrated that the transposase of *Tc3* binds to both copies of its direct repeats (COLLOMS *et al.* 1994). One possibility is that this double binding helps to form a secondary structure through intrastrand pairing between the two long inverted repeats. Such a secondary structure might be necessary for the physical and chemical interactions that occur during transposition. There is, by the way, indirect evidence that the two inverted repeats of *S* elements become associated with each other during DNA replication, since base changes in one repeat tend to be correlated with identical base changes in the other (see Table 1). This correlation suggests that the two inverted repeats pair, and that one is used as a template for the correction of the other by a process of gene conversion.

In addition to a role in transposition, the inverted repeats of these elements may have an influence on chromosome organization. Fragments of *S* inverted repeats in two clusters of heat shock genes are close to, and possibly coincident with, sequences that bind to components of the nuclear scaffold (MIRKOVITCH *et al.* 1984). The AT-rich *S* inverted repeats may therefore be contact points between the chromosomes and the nuclear scaffold.

It is interesting that *Bari-1,* another *Tc1*-like element in the *D. melanogaster* genome, is concentrated in a tandem array in the alpha heterochromatin (CAIZZI *et al.* 1993). This may also be the location of the *S*-like elements in *D. simulans* and *D. mauritiana,* since only three *Eco*RI fragments from the genomes of these species hybridize with an internal *S*-element probe, and one of these fragments represents high molecular weight DNA. However, the true genomic location of these *S*-like elements will only become known after the elements are cloned and used as probes for *in situ* hybridization with *D. simulans* and *D. mauritiana* chromosomes.

The members of the *mariner-Tc1* superfamily seem to insert at TA dinucleotides, which they duplicate in the process (VAN LUENEN *et al.* 1994). *S* elements follow this rule, but with a slight modification: they insert preferentially into short palindromes centered on the TA dinucleotide. This behavior might provide a mechanism for increasing the length of inverted terminal repeats in some members of the *mariner-Tc1* superfamily. An insertion into a palindrome would, in effect, create a longer inverted repeat that might become functionally incorporated into the element.

Studies of *Tc3* have suggested a possible mechanism for transposition of the members of the *mariner-Tc1* superfamily (VAN LUENEN *et al.* 1994). The *Tc3* elements are 2.3 kb long with 462-bp inverted terminal repeats. When these elements excise from chromosomal DNA, the transposase makes staggered cuts at and near the ends of the element. The effect of this staggered cleavage is to release an element lacking a CA dinucleotide from the 5′ end of each of its DNA strands. This dinucleotide remains in the chromosomal DNA at the site from which the element was excised. If it is not removed by exonuclease activity, it will form a "footprint" in the chromosomal DNA when the gap caused by the excision of the element is repaired. The most commonly ob-

served footprint is a TG dinucleotide from the element's right end, plus a duplication of the TA dinucleotide at the insertion site. This is exactly the structure we found in the revertant allele of $su(s)^{sn\text{-}w}$, suggesting that *S*-element excision is mechanistically similar to that of *Tc3*.

Studies of the transposase of the *Tc1* element have identified two functional domains, one apparently facilitating generalized binding to DNA and the other facilitating specific binding to 24 nucleotides just inside the ends of the inverted repeats of *Tc1* elements (VOS *et al.* 1993). Like the putative *S* transposase, the *Tc1* transposase, termed Tc1A by VOS *et al.* (1993), is a basic protein; however, unlike the putative *S* transposase, Tc1A does not have a leucine zipper near its amino terminus. Moreover, the Tc1A protein is encoded by two exons separated by a 41-bp intron, whereas the *S* coding sequence seems to be devoid of introns. The transposases of *Tc3* and *Minos* are also encoded by two exons separated by a short intron (COLLOMS *et al.* 1994; FRANZ *et al.* 1994).

The transpositional activity of *S* elements appears to be quite limited. As far as is known, only a single *S*-insertion mutation has been identified. The *mariner*, *Tc1* and *Tc3* elements are also relatively inactive, especially in the germline, unless they are stimulated by certain genetic factors; for example, *mariner* transposition is stimulated by a factor called *Mos1*, which is actually a transposase-producing copy of the *mariner* element (MEDHORA *et al.* 1991). It is not understood why this copy of the *mariner* element is so effective in mobilizing other *mariner* elements. Nor is it understood what stimulates the movement of the *Tc1* and *Tc3* elements in the so-called mutator strains of *C. elegans*. However, transposase-producing elements are probably involved (VOS *et al.* 1993).

The nature of the factors that stimulate transposition will have to be ascertained to develop the members of the *mariner-Tc1* superfamily as transformation vectors. The prospect of using *mariner-Tc1* elements in germline transformation experiments has been attractive because these elements are taxonomically widespread and may therefore be useful as vectors in a wide range of species. However, attempts to obtain transformants with these elements have been rather disappointing. *mariner*-based transformation vectors have succeeded in getting transgenes inserted into the *D. melanogaster* genome, but once inserted, these transgenes are essentially immobile (LIDHOLM *et al.* 1993; LOHE *et al.* 1995). Efforts to use *Tc1* as a transformation vector in *C. elegans* have also met with very limited success (J. SHAW, personal communication). Although there are many possible explanations for these disappointing results, one that deserves special attention is the possiblity that the transposition of this class of elements may be adversely affected by increasing the element's size. There is every reason to believe that the inverted terminal repeats of the *mariner-Tc1* elements play an important role in transposition. If these repeats are separated by an abnormally large segment of DNA, as would be the case in a transgene construct, they might not be able to come together to initiate or complete a critical aspect of the transposition process. If this is so, the members of the *mariner-Tc1* superfamily may not prove to be useful as general transformation vectors.

We thank ROBERT VOELKER for *su(s)* clones and primers and for hosting M.J.S. during a sabbatical leave, KENNETH TINDALL for expert advice on PCR and PCR-based DNA sequencing, JOHNG LIM for performing the *in situ* hybridizations, JOAN GRAVES and GREGG SIMONSON for technical advice, MARGARET KIDWELL and CHRISTINE PRESTON for providing many Drosophila stocks, JOHN RAYMOND for technical help, ZSUZSANNA ISVAK and ZOLTAN IVICS for helpful discussions, PHILIP ANDERSON for sharing unpublished data, and LINDA DAGHESTANI for discovering the *S*-element homology in *D. simulans*. This work was supported by National Institutes of Health grant GM-40263 and by a sabbatical appointment for M.J.S. at the National Institute of Environmental Health Sciences.

LITERATURE CITED

BERG, D. E., and M. M. HOWE, 1989 *Mobile DNA*. American Society for Microbiology, Washington, D.C.

BREZINSKY, L., G. V. L. WANG, T. HUMPHREYS and J. HUNT, 1990 The transposable element Uhu from Hawaiian *Drosophila*—member of the widely dispersed class of Tc1-like transposons. Nucleic Acids Res. **18:** 2053–2059.

BREZINSKY, L., T. D. HUMPHREYS and J. A. HUNT, 1992 Evolution of the transposable element *Uhu* in five species of Hawaiian *Drosophila*. Genetica **86:** 21–35.

BRIERLEY, H. L., and S. S. POTTER, 1985 Distinct characteristics of loop sequences of two *Drosophila* foldback transposable elements. Nucleic Acids Res. **13:** 485–500.

CAIZZI, R., C. CAGGESE and S. PIMPINELLI, 1993 *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics **133:** 335–345.

CALVI, B. R., T. J. HONG, S. D. FINDLEY and W. M. GELBART, 1991 Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, Activator and Tam3. Cell **66:** 465–471.

CAPY, P., J. R. DAVID and D. L. HARTL, 1992 Evolution of the transposable element mariner in the *Drosophila melanogaster* species group. Genetica **86:** 37–46.

CHUNG, C. T., S. L. NIEMELA and R. H. MILLER, 1989 One-step preparation of competent *Escherichia coli:* transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA **86:** 2172–2175.

COLLOMS, S. D., H. G. A. M. VAN LUENEN and R. H. A. PLASTERK, 1994 DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase. Nucleic Acids Res. **22:** 5548–5554.

DABOUSSI, M. J., T. LANGIN and Y. BRYGOO, 1992 Fot1, a new family of fungal transposable elements. Mol. Gen. Genet. **232:** 12–16.

DOAK, T. G., F. P. DOERDER, C. L. JAHN and G. HERRICK, 1994 A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E' motif. Proc. Natl. Acad. Sci. USA **91:** 942–946.

ENGELS, W. R., 1989 P elements in *Drosophila melanogaster*, pp. 437–484 in *Mobile DNA*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, D.C.

FRANZ, G., and C. SAVAKIS, 1991 *Minos*, a new transposable element from *Drosophila hydei*, is a member of the Tc1-like family of transposons. Nucleic Acids Res. **19:** 6646.

FRANZ, G., T. B. LOUKENS, G. DIALEKTAKI, C. R. L. THOMPSON and C. SAVAKIS, 1994 Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. Proc. Natl. Acad. Sci. USA **91:** 4746–4750.

GLOOR, G., and W. ENGELS, 1992 Single fly preps for PCR. Drosoph. Inform. Service **71:** 148–149.

HENIKOFF, S., 1987 Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155:** 156–165.

HENIKOFF, S., 1992 Detection of *Caenorhabditis* transposon homologs in diverse organisms. New Biol. **4:** 382–388.

HENIKOFF, S., and R. H. A. PLASTERK, 1988 Related transposons in *C. elegans* and *D. melanogaster*. Nucleic Acids Res. **16:** 6234–6235.

HIERHORST, J., K. LEDERIS and D. RICHTER, 1992 Presence of a member of the Tc*1*-like transposon family from nematodes and Drosophila within the vasotocin gene of a primitive vertebrate, the Pacific hagfish *Eptatretus stouti*. Proc. Natl. Acad. Sci. USA **89:** 6798–6802.

IZSVAK, Z., Z. IVICS and P. B. HACKETT, 1995 Characterization of a Tc1-like transposable element in zebrafish (*Danio rerio*). Mol. Gen. Genet. **247:** 312–322.

JACOBSON, J. W., M. M. MEDHORA and D. L. HARTL, 1986 Molecular structure of a somatically unstable transposable element in *Drosophila*. Proc. Natl. Acad. Sci. USA **83:** 8684–8688.

KIDWELL, M. G., T. FRYDRYK and J. B. NOVY, 1983 The hybrid dysgenesis potential of *Drosophila melanogaster* strains of diverse temporal and geographic origin. Dros. Inf. Serv. **59:** 63–69.

KOCUR, G. J., E. A. DRIER and M. J. SIMMONS, 1986 Sterility and hypermutability in the *P-M* system of hybrid dysgenesis in *Drosophila melanogaster*. Genetics **114:** 1147–1163.

LIDHOLM, D.-A., A. R. LOHE and D. L. HARTL, 1993 The transposable element *mariner* mediates germline transformation in *Drosophila melanogaster*. Genetics **134:** 859–868.

LIM, J. K., 1993 *In situ* hybridization with biotinylated DNA. Drosoph. Inform. Serv. **72:** 73–77.

LIM, J. K., and M. J. SIMMONS, 1994 Gross chromosome rearrangements mediated by transposable elements. BioEssays **16:** 269–275.

LINDSLEY, D. L., and G. ZIMM, 1992 The genome of *Drosophila melanogaster*. Academic Press, New York.

LOHE, A. R., D.-A. LIDHOLM and D. L. HARTL, 1995 Genotypic effects, maternal effects and grand-maternal effects of immobilized derivatives of the transposable element *mariner*. Genetics **140:** 183–192.

MARUYAMA, K., and D. L. HARTL, 1991 Evidence for interspecific transfer of the transposable element mariner between *Drosophila* and *Zaprionus*. J. Mol. Evol. **33:** 514–524.

MEDHORA, M., K. MARUYAMA and D. L. HARTL, 1991 Molecular and functional analysis of the *mariner* mutator element *Mos1* in Drosophila. Genetics **128:** 311–318.

MIRKOVITCH, J., M. MIRAULT and U. K. LAEMMLI, 1984 Organization of higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell **39:** 223–232.

MOERMAN, D. G., and R. H. WATERSTON, 1989 Mobile elements in Caenorhabiditis elegans and other nematodes, pp. 537–556 in *Mobile DNA*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, D.C.

RADICE, A. D., B. BUGAJ, D. H. A. FITCH and S. W. EMMONS, 1994 Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. Mol. Gen. Genet. **244:** 606–612.

ROBERTSON, H. M., 1993 The *mariner* element is widespread in insects. Nature **356:** 241–245.

ROBERTSON, H. M., 1995 The *Tc1-mariner* superfamily of transposons in animals. J. Insect Physiol. **41:** 99–105.

ROBERTSON, H. M., and E. G. MACLEOD, 1993 Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. **2:** 125–139.

ROSENZWEIG, B., L. W. LIAO and D. HIRSH, 1983 Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res. **11:** 4201–4209.

SCHUKKINK, R. F., and R. H. A. PLASTERK, 1990 TcA, the putative transposase of the *C. elegans* Tc1 transposon, has an N-terminal DNA binding domain. Nucleic Acids Res. **18:** 895–900.

SIMMONS, M. J., J. D. RAYMOND, M. J. BOEDIGHEIMER and J. R. ZUNT, 1987 The influence of nonautonomous *P* elements on hybrid dysgenesis in *Drosophila melanogaster*. Genetics **117:** 671–685.

VAN LUENEN, H. G. A. M., S. D. COLLOMS and R. H. A. PLASTERK, 1994 The mechanism of transposition of Tc3 in C. elegans. Cell **79:** 293–301.

VOELKER, R. A., J. GRAVES, W. GIBSON and M. EISENBERG, 1990 Mobile element insertions causing mutations in the Drosophila *suppressor of sable* locus occur in DNaseI hypersensitive non-translated sequences. Genetics **126:** 1071–1082.

VOS, J. C., and R. H. A. PLASTERK 1994 Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. EMBO J. **13:** 6125–6132.

VOS, J. C., H. G. A. M. VAN LUENEN and R. H. A. PLASTERK, 1993 Characterization of the *Caenorhabiditis elegans* Tc*1* transposase in vivo and in vitro. Genes Dev. **7:** 1244–1253.

WILSON, M. R., A. MARCUZ, F. VAN GINKEL, N. W. MILLER, L. W. CLEM *et al.*, 1990 The immunoglobulin M heavy chain constant region gene of the channel catfish, *Ictalurus punctatus:* an unusual mRNA splice pattern produces the membrane form of the molecule. Nucleic Acids Res. **18:** 5227–5233.

Communicating editor: R. S. HAWLEY