*Genet. Res. Camb.* (1994), **64**, *pp.* 87–97 *With 7 text-figures*  

# The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family

WILLIAM D. WARREN[1]*†, PETER W. ATKINSON[2] AND DAVID A. O'BROCHTA[1]
[1] *Center for Agricultural Biotechnology, Maryland Biotechnology Institute, College Park, MD 20742-3351, USA*
[2] *CSIRO Division of Entomology, Black Mountain, Canberra, ACT 2601, Australia*

(*Received 19 May 1994 and in revised form*)

**Summary**

The genome of the house fly, *Musca domestica*, contains an active transposable element system, called *Hermes*. Using PCR and inverse PCR we amplified and sequenced overlapping segments of several *Hermes* elements and from these data we have constructed a 2749 bp consensus *Hermes* DNA sequence. *Hermes* termini are composed of 17 bp imperfect inverted repeats that are almost identical to the inverted terminal repeats of the *hobo* element of *Drosophila melanogaster*. Full length *Hermes* elements contain a single long ORF capable of encoding a protein of 612 amino acids which is 55% identical to the amino acid sequence of the *hobo* transposase. Comparison of the ends of the *Hermes* and *hobo* elements to those of the *Ac* element of *Zea mays*, and the *Tam3* element of *Antirrhinum majus*, as well as several other plant and insect elements, revealed a conserved terminal sequence motif. Thus *Hermes* is clearly a member of the *hobo*, *Ac* and *Tam3* (*hAT*) transposable element family, other members of which include the *Tag1* element from *Arabidopsis thaliana* and the *Bg* element from *Zea mays*. The evolution of this class of transposable elements and the potential utility of *Hermes* as a genetic tool in *M. domestica* and related species are discussed.

## 1. Introduction

In the twenty years or so since eukaryotic transposable elements were first identified and characterized at the molecular level, much has been discovered about their structure, genetic properties, phylogenetic distribution, population dynamics and the mechanisms of their movement and its regulation. However, much has still to be learnt. In particular, little is known about the origins of these mobile genetic elements, and in most cases the details of their evolution have yet to be elucidated.

The *Ac*/*Ds* controlling element system of *Zea mays* represents one of the better genetically and molecularly characterized transposable element systems in higher eukaryotes. *Ac* elements are autonomous, capable of inducing their own excision and transposition, whereas the non-autonomous *Ds* elements are capable of movement only in the presence of *Ac* (McClintock, 1956). Studies on the distribution of *Ac* and *Ds* sequences in *Z. mays* have shown that all lines examined contain sequences which hybridize to *Ac* and *Ds* (Geister *et al.* 1982; Fedoroff, Wessler & Shure, 1983; Behrens *et al.* 1984) and DNA hybridization studies have shown that *Ac*-like sequences are present in representatives of all taxa in the genus *Zea* (Johns, 1990). Sequences hybridizing to *Ds* have been observed in wheat, rye and barley genomes (Vershinin *et al.* 1987) and an *Ac*-like element has been cloned from *Pennisetum glaucum* (pearl millet) (MacRae *et al.* 1990; MacRae & Clegg, 1992). These observations suggest that *Ac*- and *Ds*-related elements are widespread amongst the Poaceae (grasses) and support the hypothesis that the *Ac*-family of elements is of ancient origin, being present in the progenitor of maize and millet (prior to 25 million years ago) (MacRae & Clegg, 1992).

Recent studies by Calvi *et al.* (1991), Feldmar & Kunze (1991), Hehl *et al.* (1991) and Atkinson, Warren & O'Brochta (1993) suggest that elements related to the *Ac*-element family are widely distributed amongst both plant and animal species. Calvi *et al.* (1991) and Feldmar & Kunze (1991) have shown that the *hobo* element from *Drosophila melanogaster* (a fly), the *Tam3* element from *Antirrhinum majus* (a dicot) and *Ac* from *Z. mays* (a monocot) exhibit low levels of coding sequence similarity that span several hundred

* To whom reprint requests should be sent.
† Present address: CAMBIA, G.P.O. Box 3200, Canberra ACT 2601, Australia

codons. All three elements generate 8 bp target site duplications upon insertion and leave similar empty excision sites that are distinctly different from those observed for other transposable elements (Pohlman, Fendorff & Messing, 1984; Coen, Carpenter & Martin, 1986; Atkinson, Warren & O'Brochta, 1993). These data suggest that *hobo*, *Ac* and *Tam3* belong to a family of related transposable elements, which we have termed the *hAT* element family (Atkinson, Warren & O'Brochta, 1993). Since the *Ac* and *Tam3* sequences appear more closely related to one another than either is to *hobo*, it has been proposed that these three elements have evolved from an ancestral element present prior to the divergence of plants and animals (Calvi *et al.* 1991).

Functional *hobo* elements are 2·9 kb in length and have 12 bp inverted terminal repeats (Streck, MacGaffey & Beckendorf, 1986; Blackman *et al.* 1989; Calvi *et al.* 1991). Like the *Ac* and *Tam3* systems in plants, not all *D. melanogaster* strains contain active *hobo* elements and numerous copies of internally deleted defective elements can be found (McGinnis, Shermeon & Beckendorf, 1983; Streck, MacGaffey & Beckendorf, 1986; Boussy & Daniels, 1991; Pascual & Periquet, 1991). In contrast to the broad distribution observed for *Ac*-family elements in the Poaceae, the distribution of *hobo* sequences within the family Drosophilidae, based on DNA hybridization studies, appears to be very limited. *Hobo* sequences are found only in the melanogaster and montium subgroups of the melanogaster-species group, displaying one of the narrowest distributions of all transposable elements from *D. melanogaster* (Daniels, Chovnick & Boussy, 1990). The *hobo* elements found in the melanogaster species complex, including those from *D. melanogaster*, *D. simulans* and *D. mauritiana*, are almost identical at the nucleotide level (Simmons, 1992). These observations are inconsistent with the idea of an ancient origin of *hobo*-like elements within this family and are more consistent with at least one instance of horizontal transmission, although other evolutionary scenarios could account for their discontinuous distribution (Capy, Anxolabéhère & Langin, 1994). The identification of related, functional elements from other insect species would allow us to understand better how *hobo* arose in the *Drosophila* genome and to understand better the evolutionary history of the *hobo*, *Ac* and *Tam3* family of elements.

As part of a study of the mobility properties of *hobo* in the house fly, *Musca domestica*, we discovered an endogenous *hobo* transposase-like activity in embryos of this species which suggested that *hobo*-like elements might be present (Atkinson, Warren & O'Brochta, 1993). Here we report the full sequence and structural characteristics of a *M. domestica* short inverted repeat-type transposable element belonging to the *hobo*, *Ac* and *Tam3* (*hAT*) element family. We present a detailed comparison of the structure and coding sequences of this element, which we have called *Hermes*, to those of the other *hAT* elements and present evidence suggesting that the *Bg* element from *Zea mays* and the *Tag1* element from *Arabidopsis thaliana* are also members of this family. In addition, we report that various plant, insect and vertebrate transposable elements share a common sequence motif within their inverted terminal repeats. Our data not only show that *hobo*-related elements are present in non-drosophilid species but also generate interesting insights into the structural constraints and evolutionary relationships within this family of transposable elements.

## 2. Materials and methods

### (i) M. domestica *genomic DNA*

The Maryland and Mullinix strains were obtained from the USDA Livestock Insect Laboratory, Beltsville, Maryland and the Florida strain was obtained from the USDA South Atlantic Area Medical and Veterinary Entomology Laboratory. All other strains were obtained from the Department of Entomology, Texas A&M University. All strains examined were derived from wild caught flies with the exception of *sbo*, which is a multiply marked laboratory strain carrying the *stubby*, *brown body* and *ocra eye* mutations. Genomic DNA was prepared either from embryos as described by Miklos (1984) or from single adults by a modification of the method of Lifton (Bender, Spierer & Hogness, 1983).

### (ii) *PCR analysis of strains*

The oligonucleotide primers, 5′-TTGTTGTTACTC-AGTCGC-3′ and 5′-GTTTGATGTTAAGATC-ACC-3′ were used to amplify *Hermes* sequences from genomic DNA prepared from single adult *M. domestica* of various strains. Each PCR reaction contained 50 mM-KCl, 10 mM Tris-HCl (pH 8·3), 1·8 mM-$MgCl_2$, 0·125 $\mu$M dNTPs, 0·2 $\mu$M of each primer, 12% sucrose, 0·2 mM cresol red, 200 ng of template DNA and 2·2 units of *Taq* polymerase (AmpliTaq, Perkin-Elmer) in a total volume of 50 $\mu$l. PCR amplification was performed in a Perkin-Elmer 9600 thermocycler programmed for 95 °C (3 min) then 35 cycles of 50 °C (30 s), 72 °C (2·5 min) and 95 °C (20 s) followed by 10 min at 72 °C. Amplified products were then size fractionated by electrophoresis in 1·2% agarose and visualized by fluorescence in ethidium bromide.

### (iii) *Inverse PCR*

20 $\mu$g of genomic DNA, purified from embryos of the Maryland strain of *M. domestica*, was digested to completion with *Eco*RI or *Bam*HI (New England Biolabs) in accordance with the supplier's recommendations. After electrophoresis through 0·8% agarose, the 2 to 3 kb and 3 to 5 kb fractions of each

digest were excised from the gel and purified by NaI/glass-milk (Gene-Clean, BIO101 Inc). Circularization of the DNA in each fraction was performed at a DNA concentration of 0·5 $\mu$g/ml in 30 mM TrisHCl pH 7·8, 10 mM-$MgCl_2$, 10 mM dithiothreitol, 5 mM ATP and 1600 units/ml T4 DNA ligase (New England Biolabs) for 18 h at 12 °C. Ligated DNA was recovered by NaI/glass-milk purification and digested with *Hin*dIII (New England Biolabs) to linearize *Hermes* containing molecules. Approximately 150 ng of this DNA was used per 100 $\mu$l PCR. PCR amplifications contained 50 mM-KCl, 10 mM Tris-HCl (pH 8·3), 2 mM-$MgCl_2$, 200 $\mu$M dNTPs, one wax bead (AmpliWax, Perkin-Elmer), template DNA and 0·1 $\mu$M of each primer. Amplifications of *Eco*RI ligated molecules were performed with Primer 1 (5′-CTGTGGATCTTTCAAAATCTAAGG-3′) and Primer 4 (5′-GAGACACCTGAACTCAATGT-GC-3′) whereas in those performed on *Bam*HI ligated molecules used Primer 4 and Primer 5 (5′-CGCAGTTCATTGTTTTCATGG-3′). Amplifications were initiated by the addition of 3 units *Taq* polymerase and performed as follows: two cycles of 95 °C (60 s), 58 °C (15 s), 72 °C (5 min), 35 cycles of 95 °C (15 s), 58 °C (15 s), 72 °C (3 min) followed by 10 min at 72 °C. Amplified products were size fractionated by electrophoresis in 1% agarose, gel purified and cloned as blunt-ended fragments into the pBCKS(+) vector (Stratagene).

### (iv) *DNA sequence determination and analysis*

DNA sequencing was performed by the chain termination method of Sanger, Nicklen & Coulsen (1977) using modified T7 DNA polymerase (Sequenase, United States Biochemical) under conditions recommended by the supplier. Full DNA sequence of the cloned PCR products was obtained with the aid of exonuclease III generated deletions (Henikoff, 1987) as well as with *Hermes* specific oligonucleotide sequencing primers. DNA sequence analyses were performed using version 7·1 of the GCG package of programs (Devereaux, Haeberli & Smithies, 1984) and nucleic acid sequence database searches were performed using the BLAST algorithm (Altschul *et al.* 1990) using the electronic-mail search facility provided by the National Center for Biotechnology Information (NCBI).

## 3. Results

### (i) Hermes *sequence and structure*

Inverse PCR, a variation of the polymerase chain reaction that permits the amplification of regions of unknown sequence that flank a known sequence (Ochman, Gerber & Hartl, 1988; Triglia, Peterson & Kemp, 1988), was used to amplify and isolate overlapping segments of several *Hermes* elements from the Maryland strain of *M. domestica*. Initially, genomic DNA was digested with *Eco*RI, the resulting fragments circularized and those containing *Hermes* sequences amplified using PCR with oligonucleotide primers based upon the sequence data presented in Atkinson, Warren & O'Brochta (1993). Prior to circularization the template DNA was size-fractionated to bias the PCR towards amplification of full length *Hermes* sequences. In this way DNAs from three independent *Hermes* elements (denoted E1, E2 and E3) were isolated, cloned and sequenced. These clones contain sequences between an internal *Eco*RI site located within the *Hermes* transposase encoding region and the next *Eco*RI site in the upstream flanking genomic DNA (Fig. 1). The left terminus of *Hermes* was identified by aligning these three elements, which are collinear until approximately 1·4 kb upstream of the *Eco*RI site, at which point their identity abruptly ends (Fig. 2). The 3′ *Hermes* sequences, including the C-terminal transposase encoding region

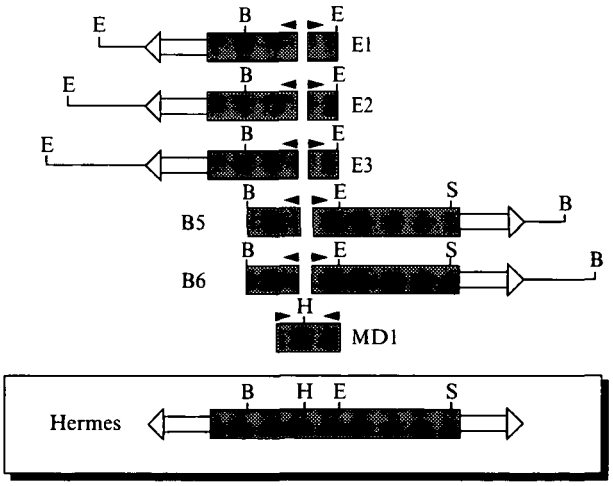


Fig. 1. Summary of independent *Hermes* sequences obtained from PCR and inverse PCR and used in the construction of a consensus full length *Hermes* element. Clones labelled E1, E2 and E3 were obtained from inverse PCR amplification of *Eco*RI circles, B5 and B6 from inverse PCR amplification of *Bam*HI circles and MD1 from degenerate PCR as described in Atkinson *et al.* (1993). Shaded boxes indicate transposase coding sequences, open boxes flanking *Hermes* sequences, open triangles inverted terminal repeats and thin horizontal lines flanking *M. domestica* genomic DNA sequences. The location of PCR primers are indicated by arrowheads. Restriction sites: E, *Eco*RI, *Bam*HI; H, HindIII; S, *Sal*I.

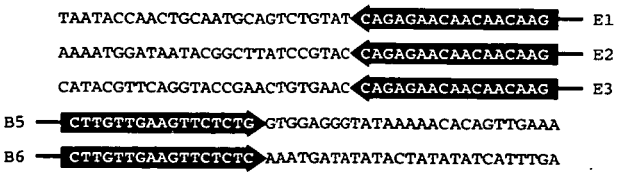


Fig. 2. The termini of *Hermes* were identified by the alignment of sequences obtained from independent elements. DNA sequences of the three left termini (E1, E2 and E3) and two right termini (B5 and B6), including 27 bp of flanking genomic sequences, are given in the 5′ to 3′ orientation. The 17 bp inverted terminal repeats of each *Hermes* element are indicated by solid arrows.

https://doi.org/10.1017/S0016672300032699 Published online by Cambridge University Press

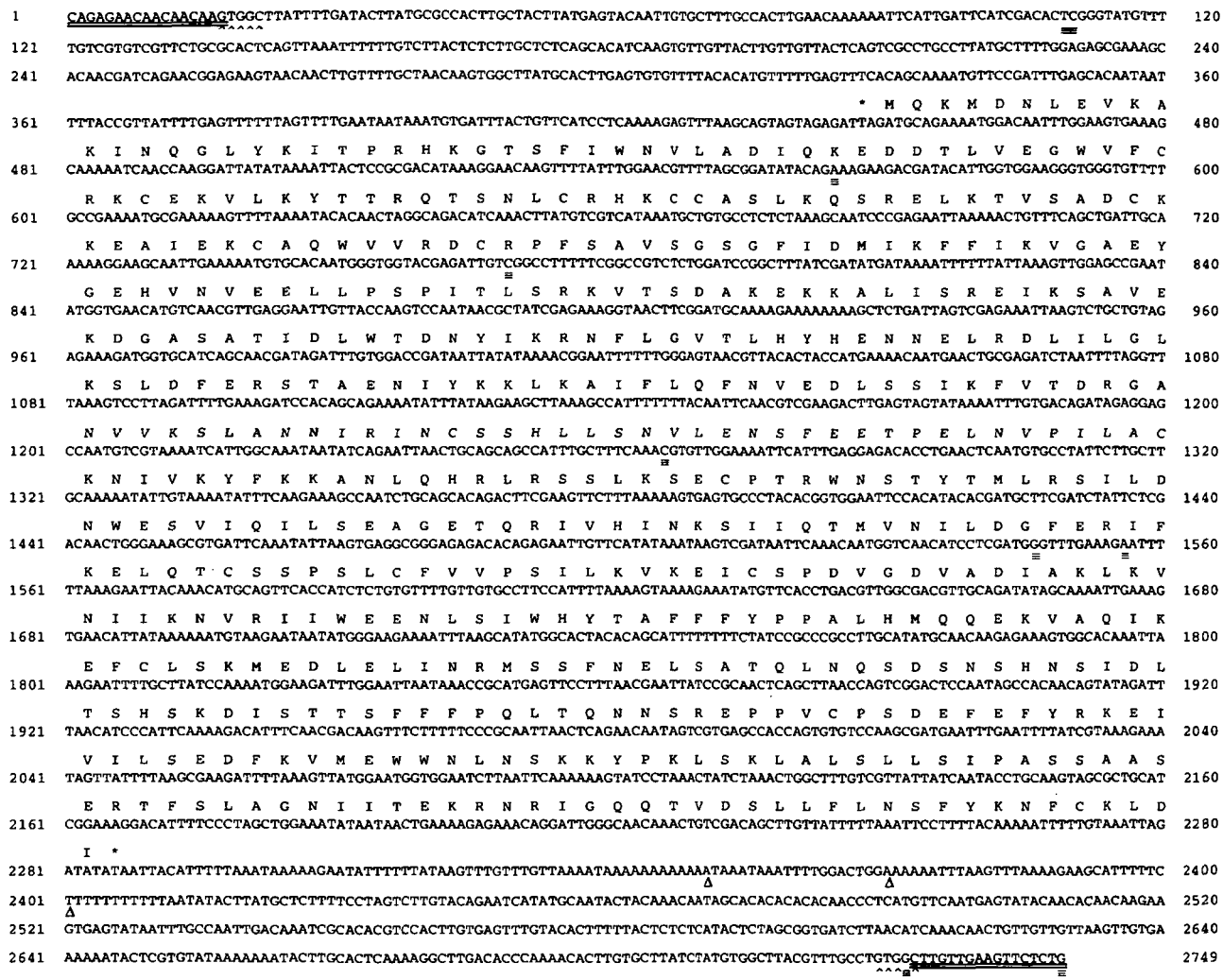

Fig. 3. Consensus *Hermes* nucleotide sequence, including translation of the transposase coding region. Terminal inverted repeat sequences are double underlined, nucleotides differing between independent *Hermes* elements are triple underlined, nucleotide deletions indicated by capital deltas, stop codons by an asterisk (*) and the subterminal pentanucleotide motif conserved between *Hermes* and *hobo* marked with carets (^). Amino acid residues encoded by ORF1 are shown above the DNA sequence. These data have been entered into the Genbank database (accession L34807).

and the sequences that comprise the right end of *Hermes*, were isolated by inverse PCR from *Bam*HI digested genomic DNA. DNAs from two independent *Hermes* elements (denoted B5 and B6), spanning the region from an internal *Bam*HI site within the N-terminal region of the *Hermes* transposase to the next *Bam*HI site in the downstream flanking genomic DNA (Fig. 1), were amplified, cloned and sequenced. The right terminus of *Hermes* was identified by aligning the B5 and B6 sequences and observing the point where their identities end (Fig. 2).

Alignment of the overlapping regions from the E1, E2, E3, B5 and B6 sequences yielded a full length consensus *Hermes* element sequence of 2749 bp (Fig. 3). These data were generated by compiling the sequences of several independent recombinants of each inverse PCR generated product, as well as three additional recombinants (denoted MD1 in Fig. 1) carrying the PCR product whose sequence was reported in Atkinson, Warren & O'Brochta (1993). In this way sequence variation introduced during amplification by *Taq* polymerase was distinguished from naturally occurring sequence variation between elements. In general, *Hermes* elements are quite homogeneous in sequence. Very low levels of nucleotide polymorphism were found between the different *Hermes* elements sequenced and although several single base deletions were observed there were no large DNA insertions or deletions (Fig. 3).

### (ii) Hermes *sequence variation between* M. domestica *strains*

Oligonucleotide primers specific to subterminal *Hermes* sequences were used in a PCR reaction to investigate the sequence length heterogeneity of *Hermes* elements amongst *M. domestica* strains. These oligonucleotides, whose 5′ bases end at positions 195 and 2618 (Fig. 3), were used to amplify internal *Hermes* sequences from genomic DNA extracted from single flies of various strains. As shown in Fig. 4, the

https://doi.org/10.1017/S0016672300032699 Published online by Cambridge University Press

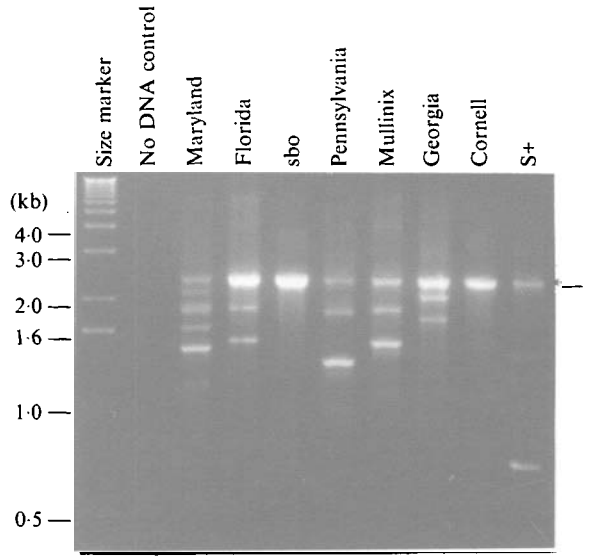


Fig. 4. Ethidium bromide stained agarose gel containing *Hermes* sequences amplified by PCR from genomic DNA extracted from single *M. domestica* of various strains. Size standards are shown leftmost and the arrow at right indicates the 2·4 kb PCR product derived from full-length *Hermes* elements.

largest band amplified with these primers is 2·4 kb. The size of this amplification product is consistent with the data presented in Fig. 3, which predict that these primers would generate a 2424 bp product from a full-length *Hermes* element. All strains examined contain a 2·4 kb band, indicating that all contain at least one full-length, or near full-length element. In the case of *sbo* and Cornell samples, the only *Hermes* element sequences in the genome that contain priming sites for both PCR primers appear to be full-length elements. All other samples contain between 1 and 5 different-sized elements, with the Maryland strain exhibiting the greatest diversity. Variation in the size and number of internally deleted *Hermes* elements was also observed amongst individuals within some strains (data not shown). This pattern of size variation is similar to that observed for other active transposable element systems, including P, *hobo*, *Tam3* and *Ac* (Berg & Howe, 1989). Unfortunately the PCR band intensities seen in Fig. 3 cannot be correlated to element copy number due to preferential amplification of shorter PCR products and the non-linear nature of the PCR amplification conditions used in this experiment.

### (iii) *The* Hermes *transposase*

Examination of the DNA sequence shown in Fig. 3 revealed the presence of a single long open reading frame (ORF1) that begins at nucleotide 450 and ends at 2285. Sequences resembling consensus eukaryotic TATA and CAAT sequences can be found around nucleotides 396 and 354 respectively and four potential polyadenylation signals can be found within 100 bp of the end of ORF1. The locations of the putative CAAT and TATA sequences do not coincide with those proposed for the *hobo* element (Streck, MacGaffey & Beckendorf, 1986; Calvi *et al.* 1991); however, a potential polyadenylation signal sequence is present exactly 14 bp beyond the end of ORF1 in both *hobo* and *Hermes*. Conceptual translation of ORF1 yields a protein sequence comprising 612 amino acids that displays 55% identity and 71% similarity to the *hobo*-transposase (Fig. 5). Thus ORF1 appears to encode part, if not all, of the *Hermes* transposase protein.

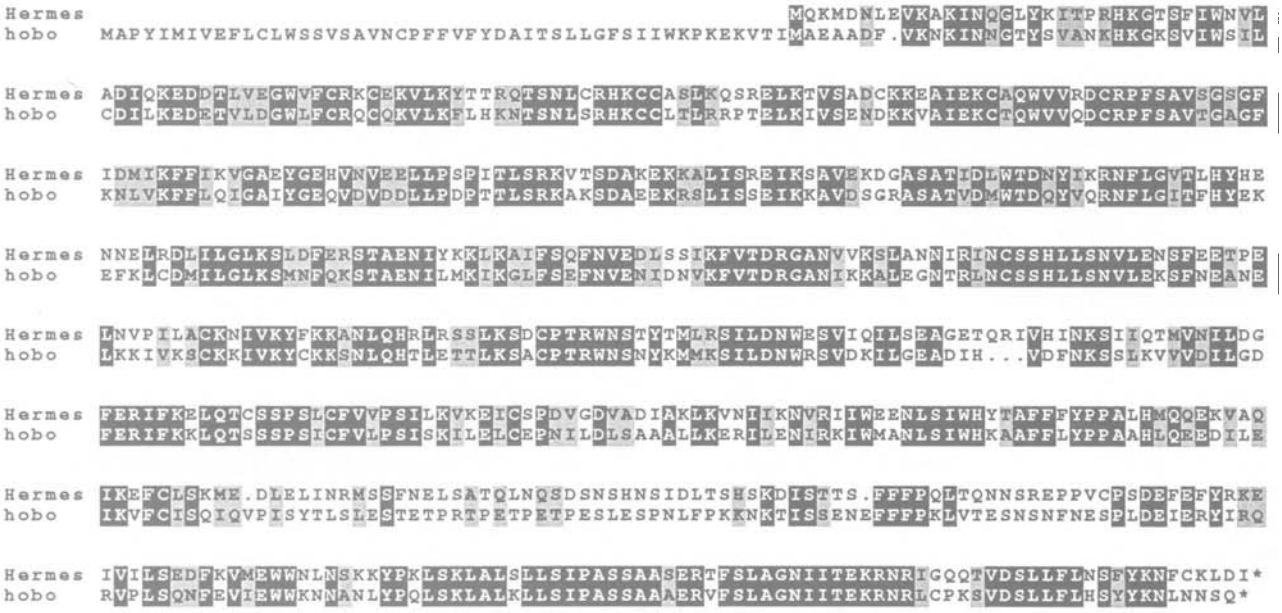


Fig. 5. Amino acid sequence alignment of the *Hermes* and *hobo* transposase coding sequences (translations of GenBank database entries, L34807 and M69216). Identical residues are indicated by white letters on a black background, chemically similar residues as black letters on a shaded background, stop codons by an asterisk (*) and gaps introduced to optimize the alignment by dots (...).

https://doi.org/10.1017/S0016672300032699 Published online by Cambridge University Press

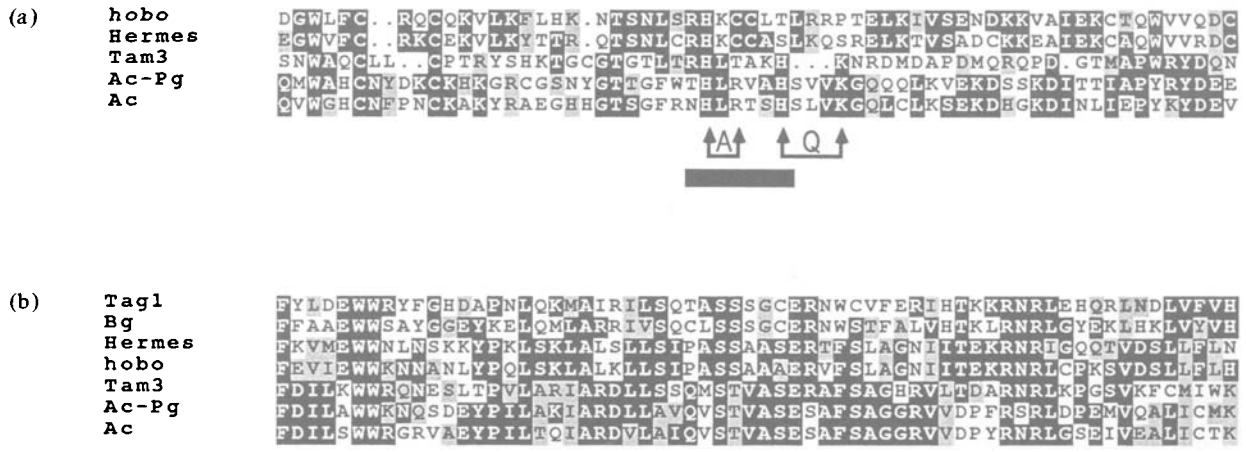


Fig. 6. Amino acid sequence alignments of the putative DNA binding regions (*a*) and highly conserved C-terminal sequences (*b*) of various *hAT* element transposases. Identical residues are indicated by white letters on a black background and chemically similar residues as black letters on a shaded background. The solid black rectangle indicates residues found to be essential for *Ac* transposase DNA binding and the U-shaped arrows indicate the pairs of charged residues that when substituted for alanine (A), or for glutamine (Q), abolish DNA binding (Feldmar & Kunze, 1991).

Comparisons of the *Hermes* sequence with those of the *hobo*, *Ac*, *Tam3* and the *Ac*-like element from *P. glaucum* clearly show that the *Hermes* transposase protein sequence is most similar to that of *hobo*. In addition to the similarities reported by Calvi *et al.* (1991), Feldmar & Kunze (1991) and Hehl *et al.* (1991), we find that all five transposase proteins are indeed alignable over their entire length (data not shown). Two regions are particularly noteworthy. The first, shown in Fig. 6*a*, is that which encompasses the DNA binding region of the *Ac* transposase protein. Although this region of the *Ac* transposase contains no known sequence motifs characteristic of DNA binding domains, it was shown by Feldmar & Kunze (1991) to bind specifically to multiple AAACGG motifs that are clustered subterminally at both ends of the *Ac* element. Further analyses revealed that the deletion or substitution of certain basic amino acids in this region abolishes its DNA binding capability (Feldmar & Kunze, 1991). As seen in Fig. 6*a*, this region of the *Ac* transposase shows good sequence similarity to *Tam3* and the *Ac*-like element from *P. glaucum*, and somewhat lower similarity to the corresponding regions of the *Hermes* and *hobo* transposases. A histidine residue in this region, corresponding to $His^{191}$ of the *Ac* transposase which is known to be important for DNA binding, is conserved in all five proteins. This suggests that this region in general, and the conserved histidine in particular, plays a crucial role in the DNA binding of all five transposases. The second region, shown in Fig. 6*b*, exhibits the highest levels of sequence conservation amongst all *hAT* element transposases and is located near their C-termini. When sequences from this region were used to search translations of the GenBank/EMBL/DDBJ nucleotide sequence databases, sequence similarity to two other transposable elements, *Tag1* from *Arabidopsis thaliana* (Tsay *et al.* 1993) and *Bg* from *Z. mays* (Hartings *et al.* 1991), was discovered. Further analysis of these elements showed that they share several other short regions of coding sequence similarity in various reading frames; however, in the absence of transcriptional information the significance of these other regions is uncertain.

### (iv) Hermes *terminal and subterminal sequences*

Comparison of the left and right terminal sequences of *Hermes* reveals that they are composed of 17 bp imperfect inverted repeats (Fig. 2). The left terminal inverted repeat of *Hermes* differs from that of *hobo* by two bases, while the right terminus of *Hermes* differs from the corresponding region of *hobo* by only a single nucleotide. When the inverted terminal sequences of other members of the *hAT* element family are aligned with those of *Hermes*, an interesting pattern emerges. Although several *hAT* elements have imperfect inverted repeats, all share a conserved A and G at positions 2 and 5 respectively, in their left inverted terminal repeats and a complementary C and T in their right terminal sequences (Fig. 7). This A2G5 pattern is not universal to all short inverted repeat-type elements. Other elements that conform to this pattern (apart from the *hAT* family members discussed above), include the *Ispr* element from *Pisum sativum* (pea) (Bhattacharyya *et al.* 1990), *Tpc1* from *Petroselinum crispum* (Parsley) (Herrmann, Schulz & Hahlbrock, 1988), *1723* from *Xenopus laevis* (Kay & Dawid, 1983), and *TECth1* from *Chironomus thummi* (Wobus *et al.* 1990). Interestingly, all these elements generate 8 bp target site duplications upon transposition. These observations suggest that the *Ispr*,

https://doi.org/10.1017/S0016672300032699 Published online by Cambridge University Press

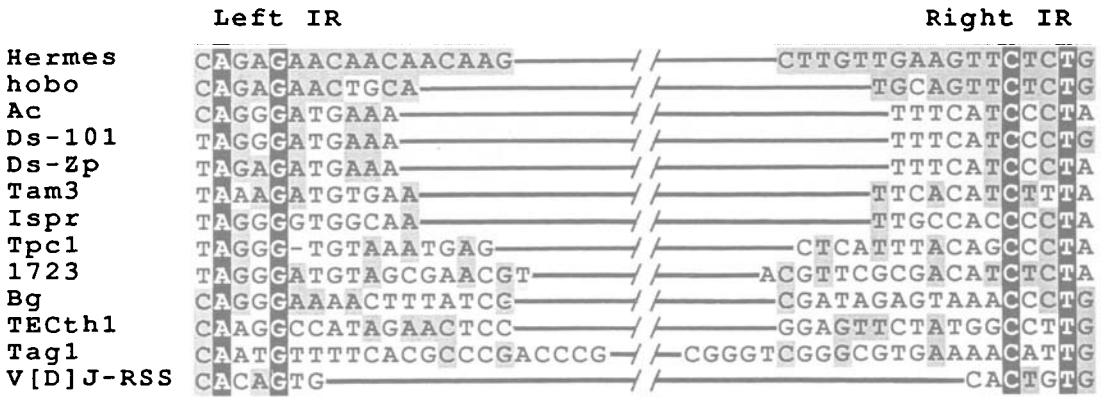


Fig. 7. Nucleotide sequence of the left and right terminal inverted repeats (IRs) of members the *hAT* element family and other related elements. Bases identical in all elements are shown as white letters on a black background and those identical to the *Hermes* element are shown by black letters on a shaded background. A gap in the *Tpc1* sequence was introduced only to preserve the inverted repeat. Data presented here were obtained from the following GenBank database entries or papers: *Hermes* (*M. domestica*) this paper; *hobo* (*D. melanogaster*), M69216; *Ac* (*Z. mays*), X05424 and X05425; *Ds*-101 (*Z. mays*), X07147; *Ds*-Zp (*Z. perennis*), X54710; *Tam3* (*A. majus*), X55078; *Ipsr* (*P. sativum*), Bhattacharyya *et al.* (1990); *Tpc1* (*P. crispum*), Herrmann, Schulz & Hahlbrock (1988); *1723* (*X. laevis*), X00079 and X00077; *Bg* (*Z. mays*), X56877; *Tag1* (*A. thaliana*), L12220; *TECth1* (*C. thummi*), X17627; V[D]J-RSS (Hesse *et al.* 1989).

*Tpc1*, *1723* and *TECth1* elements may excise and transpose using an enzymology similar to that employed by *hAT* elements and may perhaps be evolutionarily related. In addition, the recombination signal sequences (RSS) required for DNA rearrangements of the V(D)J segments of vertebrate immunoglobin and T-cell receptor genes (Tonegawa, 1983; Hesse *et al.* 1989) also conform to this A2G5 pattern (Fig. 7).

With the exception of the inverted terminal repeats, *Hermes* sequences outside of ORF1 show little similarity to those of *hobo* or any of the other members of the *hAT* element family. In the *Ac* element, subterminally located AAACGG repeats are bound by the *Ac* transposase, yet similar repeated motifs are not found in *Tam3*, *hobo* or *Hermes*. The pentanucleotide GTGGC does appear within 20 bp of the left and right termini in both the *hobo* and *Hermes* sequences (Fig. 3) although the significance of this sequence has not been established.

## 4. Discussion

### (i) Hermes/hobo *similarities*

The discovery of a *hobo* transposase-like activity in *M. domestica* embryos (Atkinson, Warren & O'Brochta, 1993) suggested the existence of a related transposable element family in this species. Using PCR and inverse PCR we amplified and sequenced segments of several members of a family of transposable elements, which we have called *Hermes*. *Hermes* elements have 17 bp imperfect inverted repeats that are nearly identical to the inverted repeats of *hobo*. Analysis of the internal sequences of *Hermes* revealed a single long open reading frame whose translation product displays a remarkable level of amino acid sequence similarity to that of the *hobo* transposase gene. These data indicate that *Hermes* is a short inverted repeat-type transposable element belonging to the *hobo*, *Ac* and *Tam3* (*hAT*) element family and are consistent with *Hermes* being the source of the endogenous transposase activity observed in *M. domestica* embryos (Atkinson, Warren & O'Brochta, 1993).

Although we have no direct evidence for ORF1 transcription in *Hermes*, the presence of putative CAAT, TATA and polyadenylation signal sequences, the low level of sequence polymorphism between elements, and the colinearity of its putative protein product with the transposase proteins of *hobo* and the other *hAT* elements, all suggest that *Hermes* is capable of generating a functional transposase protein. Several other pieces of evidence also indicate that *Hermes* is active within the *M. domestica* genome. Genomic Southern blot data show copy number and insertion site variation of *Hermes* elements between two *M. domestica* strains of different geographic origins (Atkinson, Warren & O'Brochta, 1993). In addition, PCR amplification of *Hermes* sequences from these strains plus four others indicates that some contain exclusively full-length elements whereas others also harbour one or more internally deleted derivatives, the sizes of which vary both amongst and between strains. Although the evolutionary relationships of these strains are not known, these data are consistent with *Hermes* being active in some if not all strains examined. Preliminary plasmid-based *Hermes* excision and transposition experiments using embryos of the Maryland strain support this conclusion (unpublished data).

Conceptual translations of the sequences upstream of ORF1 in *Hermes* fail to identify sequences that correspond to the 49 N-terminal residues of the *hobo*

transposase protein sequence proposed by Calvi *et al.* (1991). As the largest PCR products observed in Fig. 4 are consistent with the sequence data presented in Fig. 3, it seems unlikely that the lack of these amino terminal residues in *Hermes* is the result of a deletion. At present, the *hobo* transposase protein sequence, like that of *Hermes*, is based on the conceptual translation of the element's major ORF and a comparison of these sequences to those of *Ac* and *Tam3*. Although these 49 N-terminal residues may perform a special function unique to the *hobo* transposase, it is also possible that they are not present in the mature *hobo* protein, and that translation of the *hobo* transposase may begin at the third AUG encoded by the *hobo* ORF1.

The level of similarity observed between the transposase proteins of *hobo* and *Hermes* shows that these elements are almost certainly evolutionarily related, and consequently it is likely that the transposase of *Hermes* acts on a similar substrate and recognizes a similar DNA sequence to that of the *hobo* transposase. We recently showed that the endogenous transposase activity in *M. domestica* embryos, although capable of imprecisely excising *hobo*, is incapable of mediating *hobo* transposition (O'Brochta *et al.* 1994). One possible explanation for this observation is that although there is sufficient similarity between the two systems to allow the *Hermes* transposase to recognize *hobo* sequences and initiate excision/transposition, normal excision/transposition cannot proceed due to incorrect recognition of essential *cis*-acting sequences or perhaps a lack of appropriate host factors. We are presently investigating excision and transposition of *Hermes* and *hobo* in *M. domestica* and *D. melanogaster* and by interchanging various transposase domains and *cis*-acting sequences between these elements we aim to gain further insight into the mechanism of *hAT* element movement.

### (ii) hAT *element transposase similarities*

In comparing the *Hermes* and *hobo* transposase protein sequences to the transposases of the *Ac*, *Tam3* and the *P. glaucum Ac*-like element, we found, in addition to the sequence similarities previously reported (Calvi *et al.* 1991; Feldmar & Kunze, 1991; Hehl *et al.* 1991; MacRae & Clegg, 1992), a region displaying sequence similarity to the DNA binding domain of *Ac* (Fig. 6*a*). Although the level of sequence similarity in this region is quite low, the overall collinearity of all five transposase sequences and the presence of higher levels of similarity flanking this region indicate that this alignment is appropriate. The DNA binding sequences of *Ac* are very similar to those found in the *P. glaucum Ac*-like element transposase, indicating that these proteins bind the same or very similar DNA sequences. The corresponding regions in the transposase proteins of *hobo* and *Hermes* are also very similar to one another, suggesting that they also have similar DNA binding characteristics. However, apart from a conserved histidine residue known to be important for DNA binding of the *Ac* transposase protein, and a conserved threonine nearby, the level of similarity shared between the plant and insect sequences in this region is quite low. Given the overall similarities between the various *hAT* transposase proteins, it seems reasonable to suggest that they encode enzymically analogous proteins and that the transposase proteins encoded by *hobo* and *Hermes* elements have different DNA recognition specificities than those of the other *hAT* elements.

Upon searching translations of the nucleic acid databases, we discovered that two other known plant transposable elements, *Bg* and *Tag1*, also share sequence similarities to the most conserved region of amino acid similarity between *hAT* elements. These results suggest that the *Bg* and *Tag1* elements, which are known to generate 8 bp target site duplications and have short inverted terminal repeats, are related to one another and to the *hAT* family of elements. *Tag1* was isolated as an insertion into the *chl1* gene during an *Ac* mutagenesis screen (Tsay *et al.* 1993), suggesting that *Ac* may be able to mobilize *Tag1*. The *Ac/Ds* and *Bg/rbg* systems, however, appear to be incapable of crossmobilizing one another (Peterson, 1986).

### (iii) hAT *element inverted terminal repeats*

A comparison of the terminal inverted repeats of *Hermes* with those of the other members of the *hAT* family, including the *Bg* and *Tag1* elements, revealed a previously undocumented sequence similarity (Fig. 7). These elements, although having inverted repeats of various lengths and sequence compositions, all have an A at position 2 and a G at position 5 of their left termini, and complementary bases at the corresponding positions in their right termini. This observation suggests that these nucleotides play a central role in the biochemistry of recombination in this family of elements. A survey of the termini of other short inverted repeat-type elements revealed that several others, including *Ispr*, *Tpc1*, *TECth1*, and *1723* also follow this pattern. We feel that this A2G5 motif is unlikely to be fortuitous. All these elements share the common feature of generating an 8 bp insertion site duplication and, in the cases where excision has been studied, cause the addition of supernumerary nucleotides that form short palindromes at the site of rejoining (Pohlman, Fendorff & Messing, 1984; Coen, Carpenter & Martin, 1986; Atkinson, Warren & O'Brochta, 1993). The proposed mechanisms of breakage and rejoining following *hAT* element excision and the recombination of the variable (V), diversity (D) and joining (J) regions in lymphocytes are both thought to involve

the formation and resolution of hairpin structures (Coen, Carpenter & Martin, 1986; Lieber, 1991; Roth *et al.* 1992). Examination of the recombination signal sequence heptamer repeats that delimit the various V(D)J sequences removed during lymphoid cell development revealed that they also follow the A2G5 pattern. This similarity in the structure of the recombination reaction-products and the substrates for recombination suggest that V(D)J recombination and *hAT* element excision may share common enzymologies.

### (iv) *Distribution and evolution of* hAT *elements*

The identification and characterization of *Hermes* in *M. domestica* indicates that *hobo*-like elements are not as restricted in their distribution in insects as previously thought. Although we have not undertaken a large-scale search for related elements in other species, we have identified additional members of this family from a number of non-drosophilid insects, including *Musca vetustissima* (Muscidae), *Lucilia cuprina* (Calliphoridae) and *Bactrocera tryoni* (Tephritidae) (unpublished data). All presently available data are consistent with this family of elements being of ancient origin as thus far all insect *hAT* element sequences appear more similar to each other than they are to the plant *hAT* elements. With the exception of the *Tag1* and *Bg* elements, for which full transposase protein sequences have yet to be determined, *hAT* element phylogeny inferred from transposase coding sequence data appears consistent with the accepted phylogenetic relationships of the species from which the elements were isolated.

The *hobo* element is capable of transposition in *M. domestica* embryos as well as in several other Dipteran species (O'Brochta *et al.* 1994). In plants both the *Ac* and *Tam3* elements have been shown to function in a broad range of non-host cells (Baker *et al.* 1986; Van Sluys, Tempe & Fendoroff, 1987; Knapp *et al.* 1988; Yoder *et al.* 1988; Martin *et al.* 1989; Chuck *et al.* 1993). Thus it would appear that these elements either transpose using a mechanism that requires general cellular factors that are highly conserved among species or alternatively they do not require any *trans*-acting factors other than their transposase. The ability of these elements to function in heterologous cellular environments would appear to be a general characteristic of *hAT* elements and is of significance for two reasons. Firstly, this characteristic makes this class of elements ideal tools for studying and manipulating genes and genomes in a variety of genera. Because of their abilities to excise and insert into various genomic locations, transposable elements have been used as gene vectors and gene tagging agents in a variety of organisms. Although this technology has reached high levels of sophistication in *Drosophila*, gene vectors and tagging agents are currently unavailable for non-drosophilid insects. *Hermes* therefore is an ideal candidate for a non-drosophilid insect gene vector for use in *M. domestica* and related insect species. Secondly, the ability of these elements to function in other species without modification implies that, if given the opportunity, they have the capability of establishing themselves in other species. This potential for horizontal transmission is important to consider when trying to understand the distribution and evolution of this class of elements.

To account for the discontinuous distribution of *hobo* sequences in the drosophilids it has been proposed that *hobo* elements have recently invaded *D. melanogaster* and closely related species (Daniels, Chovnick & Boussy, 1990; Boussy & Daniels, 1991; Pascual & Periquet, 1991; Simmons, 1992). Although this is a reasonable hypothesis, given the propensity of this class of elements to function in heterologous cellular environments, other possibilities could also account for these observations (Capy, Anxolabéhère & Langin, 1994). As several investigators have observed weak hybridization of *hobo* sequences to genomic DNA from many drosophilid species (Streck, MacGaffey & Beckendorf, 1986; Daniels, Chovnick & Boussy, 1990), it may well be that *hAT* elements are widely distributed within the melanogaster species group and that previous studies lacked the sensitivity to detect them. The identification of conserved sequences in both *hobo* and *Hermes* will now permit a thorough PCR based survey of drosophilid and non-drosophilid species for the presence or absence of *hAT* elements. Only through the isolation and characterization of further members of this family of mobile genetic elements from both insects and plants will their origins and evolution be elucidated.

We thank T. Friedlander and K. Saville for helpful discussions and for critiquing of the many drafts of this manuscript. This work was supported in part by a grant (GM48102) from the National Institutes of Health.

## References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Atkinson, P. W., Warren,W. D. & O'Brochta, D. A. (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proceedings of the National Academy of Sciences, U.S.A.* **90**, 9693–9697.

Baker, B., Schell, J., Lörz, H. & Fedoroff, N. (1986). Transposition of the maize controlling element '*Activator*' in tobacco. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 4844–4848.

Behrens, B., Fedoroff, N., Laird, A., Müller-Neumann, M., Starlinger, P. & Yoder, J. (1984). Cloning of the *Zea mays* controlling element *Ac* from the *wx-m7* allele. *Molecular and General Genetics* **194**, 346–347.

Bender, W., Spierer, P. & Hogness, D. S. (1983). Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the bithorax complex in *Drosophila melanogaster*. *Journal of Molecular Biology* **168**, 17–33.

Berg, D. E. & Howe, M. M. (1989). *Mobile DNA*. Washington DC: American Society for Microbiology Publications.

Bhattacharyya, M. K., Smith, A. M., Ellis, T. H. N., Hedley, C. & Martin, C. (1990). The wrinkled-seed character of the pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**, 115–122.

Blackman, R. K., Macy, M., Koehler, D., Grimaila, R. & Gelbart, W. M. (1989). Identification of a fully-functional *hobo* transposable element and its use for germ-like transformation. *EMBO Journal* **8**, 211–217.

Boussy, I. A. & Daniels, S. B. (1991). *hobo* elements in *Drosophila melanogaster* and *D. simulans*. *Genetical Research* **58**, 27–34.

Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. (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., Anxolabéhère, D. & Langin, T. (1994). The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends in Genetics* **10**, 7–12.

Chuck, G., Robbins, T., Nijjar, C., Ralston, E., Courtney-Gutterson, N. & Dooner, H. K. (1993). Tagging and cloning of a Petunia flower color gene with the maize transposable element *Activator*. *The Plant Cell* **5**, 371–378.

Coen, E. S., Carpenter, R. & Martin, C. (1986). Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47**, 285–296.

Daniels, S. B., Chovnick, A. & Boussy, I. A. (1990). Distribution of the *hobo* transposable element in the genus *Drosophila*. *Molecular Biology and Evolution* **7**, 589–606.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

Fedoroff, N., Wessler, S. & Shure, M. (1983). Isolation of the transposable maize controlling elements *Ac* and *Ds*. *Cell* **35**, 235–242.

Feldmar, S. & Kunze, R. (1991). The ORFa protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain. *EMBO Journal* **10**, 4003–4010.

Geiser, M., Weck, E., Döring, H.-P., Werr, W., Courage-Tebbe, U., Tillmann, E. & Starlinger, P. (1982). Genomic clones of a wild-type allele and a transposable element-induced mutant allele of the sucrose synthase gene of *Zea mays* L. *EMBO Journal* **1**, 1455–1460.

Hartings, H., Spilmont, C., Lazzaroni, N., Rossi, V., Salamini, F., Thompson, R. D. & Motto, M. (1991). Molecular analysis of the *Bg-rbg* transposable element system of *Zea mays* L. *Molecular and General Genetics* **227**, 91–96.

Hehl, R., Nacken, W. K. F., Krause, A., Saedler, H. & Sommer, H. (1991). Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveals homologies to the *Ac* element of maize. *Plant Molecular Biology* **16**, 369–371.

Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods in Enzymology* **155**, 156–165.

Herrmann, A., Schulz, W. & Hahlbrock, K. (1988). Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. *Molecular and General Genetics* **212**, 93–98.

Hesse, J. E., Lieber, M. R., Mizuuchi, K. & Gellert, M. (1989). V(D)J recombination: a functional definition of the joining signals. *Genes in Development* **3**, 1053–1061.

Johns, M. A. (1990). Sequences related to the maize transposable element *Ac* in the genus *Zea*. *Journal of Molecular Evolution* **30**, 493–499.

Kay, B. K. & Dawid, I. B. (1983). The 1723 element: a long, homogenous, highly repeated DNA unit interspersed in the genome of *Xenopus laevis*. *Journal of Molecular Evolution* **170**, 583–596.

Knapp, S., Coupland, G., Uhrig, H., Starlinger, P. & Salamini, F. (1988). Transposition of the maize transposable element *Ac* in *Solanum tuberosum*. *Molecular and General Genetics* **213**, 285–290.

Lieber, M. R. (1991). Site-specific recombination in the immune system. *FASEB Journal* **5**, 2934–2944.

MacRae, A. F., Learn, G. H., Karjala, M. & Clegg, M. T. (1990). Presence of an *Activator* (*Ac*)-like sequence in *Pennisetum glaucum* (pearl millet). *Plant Molecular Biology* **15**, 177–179.

MacRae, A. F. & Clegg, M. T. (1992). Evolution of *Ac* and *Ds1* elements in select grasses (Poaceae). *Genetica* **86**, 55–66.

Martin, C., Prescott, A., Lister, C. & MacKay, S. (1989). Activity of the transposon *Tam3* in *Antirrhinum* and tobacco: possible role of DNA methylation. *EMBO Journal* **8**, 977–1004.

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology* **21**, 197–216.

McGinnis, W., Shermoen, A. W. & Beckendorf, S. K. (1983). A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* **34**, 75–84.

Miklos, G. L. G. (1984). The isolation of high molecular weight DNA from adult heads of *D. melanogaster*. *Drosophila Information Service* **60**, 222–223.

O'Brochta, D. A., Warren, W. D., Saville, K. J. & Atkinson, P. W. (1994). Interplasmic transposition of *Drosophila hobo* elements in non-drosophilid insects. *Molecular and General Genetics* **244**, 9–14.

Ochman, H., Gerber, A. S. & Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623.

Pascual, L. & Periquet, G. (1991). Distribution of *hobo* transposable elements in natural populations of *Drosophila melanogaster*. *Molecular Biology and Evolution* **8**, 282–296.

Peterson, P. A. (1986). Mobile genetic elements in maize: a force in evolutionary and plant breeding processes. In *Genetics, Development and Evolution* (ed. J. B. Gustafson, G. L. Stebbins and F. J. Ayala), pp. 47–78. New York: Plenum.

Pohlman, R. F., Fendoroff, N. V. & Messing, J. (1984). The nucleotide sequence of the maize controlling element *Activator*. *Cell* **37**, 635–643.

Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J. & Gellert, M. (1992). V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* **70**, 983–991.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

Simmons, G. M. (1992). Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. *Molecular Biology and Evolution* **9**, 1050–1060.

Streck, R. D., MacGaffey, J. E. & Beckendorf, S. K. (1986). The structure of *hobo* transposable elements and their insertion sites. *EMBO Journal* **5**, 3615–3623.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**, 575–581.

Triglia, T., Peterson, M. G. & Kemp, D. J. (1988). A

procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Research* **16**, 8186.

Tsay, Y.-F., Frank, M. J., Page, T., Dean, C. & Crawford, N. M. (1993). Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. *Science* **260**, 342–344.

Van Sluys, M. A., Tempe, J. & Fedoroff, N. (1987). Studies on the introduction and mobility of the maize *I* element in *Arabidopsis thaliana* and *Daucus carota*. *EMBO Journal* **6**, 3881–3889.

Vershinin, A., Salina, E. A., Svitashev, S. K. & Shumny, V. K. (1987). The occurrence of *Ds*-like sequences in cereal genomes. *Theoretical and Applied Genetics* **73**, 428–432.

Wobus, U., Baumlein, H., Bogachev, S. S., Bosevich, I. V., Panitz, R. & Kolesnikov, N. N. (1990). A new transposable element in *Chironomus thummi*. *Molecular and General Genetics* **222**, 311–316.

Yoder, J. I., Palys, J., Alpert, K. & Lassner, M. (1988). *Ac* transposition in transgenic tobacco plants. *Molecular and General Genetics* **213**, 291–296.