4048–4054 *Nucleic Acids Research, 1997, Vol. 25, No. 20*

Tc7, a Tc1-hitch hiking transposon in *Caenorhabditis elegans*

René Rezsohazy[+], Henri G. A. M van Luenen, Richard M. Durbin[1] and Ronald H. A. Plasterk*

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands and [1]The Sanger Center, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, UK

Received July 7, 1997; Revised and Accepted September 4, 1997

ABSTRACT

We have found a novel transposon in the genome of *Caenorhabditis elegans*. Tc7 is a 921 bp element, made up of two 345 bp inverted repeats separated by a unique, internal sequence. Tc7 does not contain an open reading frame. The outer 38 bp of the inverted repeat show 36 matches with the outer 38 bp of Tc1. This region of Tc1 contains the Tc1-transposase binding site. Furthermore, Tc7 is flanked by TA dinucleotides, just like Tc1, which presumably correspond to the target duplication generated upon integration. Since Tc7 does not encode its own transposase but contains the Tc1-transposase binding site at its extremities, we tested the ability of Tc7 to jump upon forced expression of Tc1 transposase in somatic cells. Under these conditions Tc7 jumps at a frequency similar to Tc1. The target site choice of Tc7 is identical to that of Tc1. These data suggest that Tc7 shares with Tc1 all the sequences minimally required to parasitize upon the Tc1 transposition machinery. The genomic distribution of Tc7 shows a striking clustering on the X chromosome where two thirds of the elements (20 out of 33) are located. Related transposons in *C.elegans* do not show this asymmetric distribution.

INTRODUCTION

Of the DNA or class II transposons (1), which transpose by excision and reintegration into the genome without an RNA intermediate, the Tc1/mariner family is the most widespread (2–4). Members of this family have been found in fungi, insects, nematodes, vertebrates and recently, Tc1/mariner related sequences have even been identified in the human genome (5–8). Genome sequencing projects reveal new groups of repetitive sequences of which many display the features of transposable elements of different classes (9–13). However, most of these cases seem to correspond to non-autonomous, deleted transposons. Thus the mobility of the elements identified through the genome sequencing projects remains hypothetical.

In *Caenorhabditis elegans*, six groups of DNA transposons have been identified to date (14), Tc1 and Tc3 being the best characterized. Tc1 and Tc3 are members of the Tc1/mariner family (2–4). Each Tc1/mariner transposon encodes a transposase which shares 35% identity with the transposases encoded by the other members of the family. The elements are delimited by inverted repeat sequences which are unrelated except for the last four nucleotides (5′-CAGT) which are conserved within the entire Tc1/mariner family. To mediate the transposition reaction, the transposase recognizes and binds to the terminal 30 bp of its cognate element (15). Thus, Tc1 and Tc3 transposases do not activate each others transposon (15–18). Tc1 and Tc3 insert into TA dinucleotides which are duplicated upon integration (19). However, within a given genomic region, only a subset of TA dinucleotides are chosen as insertion sites, and the insertion patterns differ between Tc1 and Tc3 (20).

In the wild-type *C.elegans* strain Bristol N2, Tc1 transposition is detectable only in the somatic cells, whereas Tc3 transposition is undetectable in both soma and germline (21). However, in the strain Bergerac BO, as well as in other mutator strains, Tc1 transposition also occurs in the germline (21–25). Some mutations, like *mut-2 (r759)* and *mut-7 (pk204)* also activate germline transposition of Tc3, Tc4 and Tc5 (26,27) (R.F. Ketting and R.H.A.P., unpublished observations). Recent *in vivo* and *in vitro* studies have shown that both Tc1 and Tc3 jump via a cut-and-paste process (19,28). After their excision, a double strand break is left at the donor site which is sealed by the host DNA repair machinery. This repair process often leaves characteristic footprints at the donor site (29). Interestingly, no transposition-proficient deletion derivative of Tc1 or Tc3 has been identified until now, whereas deleted versions of the P element in *Drosophila* or Ac in Maize, which presumably also transpose via a similar cut-and-paste mechanism, are widespread within the genomes of their hosts (30,31).

Recently, a new repetitive element consisting of two large inverted repeat sequences separated by a short unique sequence has been identified (13). The termini of the inverted repeats show strong similarity to the ends of Tc1 (36 out of 38 nt), suggesting

*To whom correspondence should be addressed. Tel: +31 20 512 2081; Fax: +31 20 669 1383; Email: rplas@nki.nl

[+]Present address: Catholic University of Louvain, Unit of Genetics, 5(bte3) Place Croix du Sud, 1348 Louvain-La-Neuve, Belgium

that transposition of this novel element can be mediated by the Tc1 transposase. In this paper, we characterize this new repetitive element. We have tested its ability to transpose its target site choice and its polymorphic distribution among different *C.elegans* strains. Since we find that it is fully mobile in the germ-line, it rightly deserves to be categorized as a transposon; we call it Tc7. The data suggest that Tc7 shares with Tc1 all the sequence requirements to make use of the Tc1 transposition machinery.

MATERIALS AND METHODS

Analysis of Tc7 genomic distribution

Analysis of the Tc7 distribution in the genome of different *C.elegans* strains was performed by Southern hybridization as described in Sambrook *et al.* (32). Genomic DNAs extracted from Bristol N2 and RW7097 strains were digested with *Sau*96I, and run on an agarose gel. The samples were blotted on a nitrocellulose filter. An internal fragment of Tc7 was PCR-amplified from Bristol N2 genomic DNA using primers 8RR14 (5′-atgtagctcgtgatcaggcc-3′) and 8RR15 (5′-gtgtagagtaatcttgagc-3′). The PCR product was cut out of an agarose gel. The agarose plug was placed in a perforated tube containing a glasswool filter and this was again put in a tube to elute and recover the DNA by centrifugation. The eluate was phenol extracted twice, chloroform extracted once, precipitated and used to make a radiolabelled probe by random primed labelling (33).

Detection of Tc7 transposition under Tc1-transposase expression

The stable transgenic Bristol N2 line NL818 (28), harbouring the Tc1-transposase gene under the control of a heat-shock promoter (pRP465) (15), was heat shocked for 2 or 4 h at 33°C. Genomic DNA was isolated as previously described (20) after a recovery at 18°C for 12 h. Somatic transposition of Tc7 in a 1 kb region of the *gpa-2* gene was scored by nested PCRs (20) using primers specific for Tc7 (8RR12, 5′-gccgctttatcacttgccatg-3′ and 8RR13, 5-acataggcctgatcacgagc-3′) and primers specific for the *gpa-2* target (AB3550 and AB5623) (20). The PCR products were analyzed on 1% agarose gels using the 1 kb DNA ladder (Gibco BRL) as a size marker. The PCR products were sequenced using the ABI PRISM™. Dye terminator cycle sequencing kit (Perkin Elmer) following the manufacturer instructions. Sequencing products were run and analyzed on an ABI automatic sequenator.

Database searches and computer analysis

Similarity searches through the *C.elegans* genome database (ACeDB, release of October 10, 1996) (34) and through GenBank/EMBL databases were performed using BLAST (35). Further sequence editing and analysis were made using the GCG package (University of Wisconsin, Madison).

RESULTS

A new repetitive element in the *C.elegans* genome

Using the 54 bp inverted repeat sequence of Tc1 for a similarity search of the *C.elegans* genome database we found, in addition to several Tc1 elements, 10 hits which define a new class of repetitive elements (Table 1, Tc7-1 to Tc7-10). Two of these hits (Tc7-1 and Tc7-2 in Table 1) correspond to Tc1-related sequences recently reported by Oosumi *et al.* (13). These elements are 921–923 bp sequences, made of two 345–347 bp inverted repeats and a conserved middle section that lacks a large ORF (Fig. 1A). Of the terminal 38 bp of their inverted repeats, 36 bp were identical to the ends of Tc1 and like the transposons of the Tc1/mariner family they were flanked by TA dinucleotides (Fig. 1B). These TA dinucleotides could be the result of a sequence duplication during integration as has been shown for Tc1 and Tc3 (20,28). Taken together these sequence features suggest that this element corresponds to a new transposon, which we call Tc7. No other sequence similarities were found between Tc7 and Tc1, nor between Tc7 and any other known repetitive element.

These 10 Tc7 elements were identical, with the exception of a 1 bp addition within the middle region of four sequences and a separate single base pair addition in one of the inverted repeats of two sequences (Fig. 1A). Furthermore, two elements showed a small deletion of 6 and 44 bp at different locations. The copy number of Tc7 can not be estimated by extrapolation of the number of elements found in the database which contains ~50% of the entire genome because the Tc7 elements are not equally spread over the genome (see below).

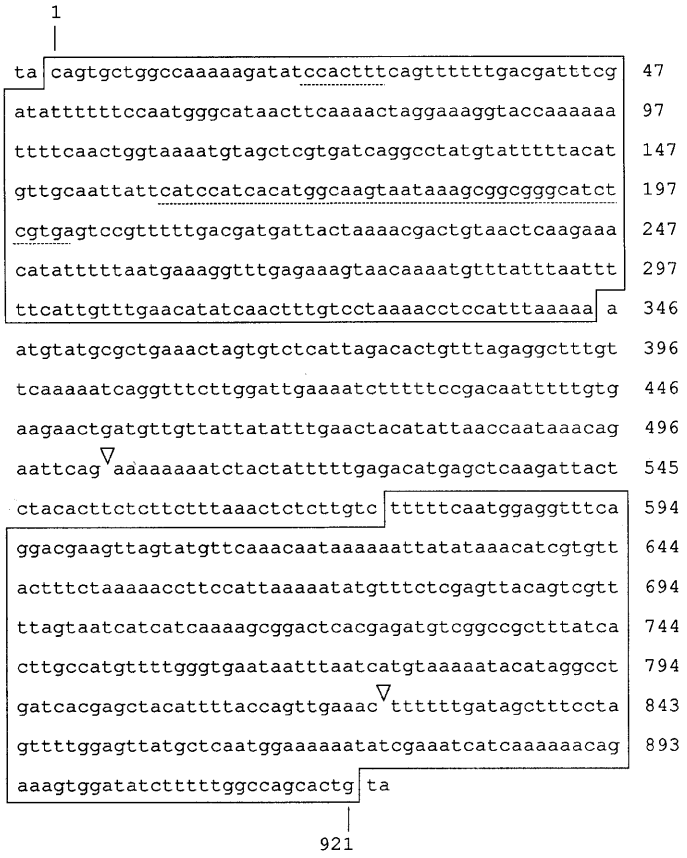
Similarity searches through the GenBank/EMBL databases using the Tc7 sequence as a probe did not reveal related sequences in other species. However, 23 additional sequences were hit in the *C.elegans* genome database. These sequences define a more heterogeneous group of elements. Seven of them were delimited by the canonical 5′-CAGT nucleotides common to the Tc1/mariner family, and they were flanked by TA dinucleotides (Tc7-d1 to Tc7-d7, Table 1). These 808–931 bp long sequences shared 62–83% sequence identity with Tc7. They also had 0.35 kb inverted repeats at their extremities, but in contrast to Tc7 their terminal sequences showed more sequence divergence with the ends of Tc1. Finally, the last 16 hits shared sequence similarities with only one Tc7 end and were presumed to correspond to degenerate, incomplete Tc7 derivatives (Tc7-d8 to Tc7-d23, Table 1). However, these Tc7 derivatives had the canonical 5′-CAGT sequence next to a TA dinucleotide at their extremity. The 23 Tc7-related elements do not represent a distinct sub-group of Tc7 elements but they seem to have independently diverged over time.

Twenty of the 33 Tc7 and Tc7-related elements (61%) were located on the X chromosome whereas only 13 elements were found on the autosomes (Fig. 2). Only 25% of the *C.elegans* sequence in the data base used for the searches is derived from the X chromosome. This means that Tc7 is not equally spread over the entire genome; the elements are clustered on the X chromosome.

Tc7 is active in the germline of mutator strains

In mutator lines Tc1 transposition is activated in the germline and new inheritable insertions (or excisions) can be detected in the genome of originally isogenic strains. We analyzed the Tc7 content of eight *mut-6* lines (25) derived from the same parental strain (RW7097) (25), which were cultivated in parallel for three months. DNA from these lines was digested with *Sau*96I, for which the Tc7 sequence does not contain a recognition site. The DNA was analyzed on a Southern blot using a Tc7 probe (Fig. 3). Interestingly, the genomes of the *mut-6* lines and Bristol N2 contained a similar number of Tc7 copies, whereas the number of Tc1 elements is about 10 times higher in this mutator strain than

A



B

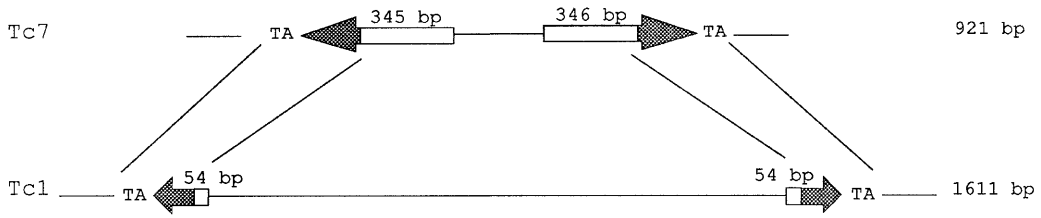


Figure 1. (**A**) Nucleotide sequence of Tc7-1 (see Table 1). The inverted-repeat sequences are boxed. The dotted lines underline the 6 bp and 44 bp sequences missing in Tc7-8 and Tc7-9, respectively. Small triangles indicate the single-base pair insertions found in Tc7-3, Tc7-4, Tc7-5, Tc7-6, Tc7-7, Tc7-9 and Tc7-10 (between nt 503 and 504 in Tc7-1) and in Tc7-3 and Tc7-4 (between nt 824 and 825 in Tc7-1). (**B**) Schematic representation of Tc7-1 and Tc1. Open arrows represent the inverted repeats. Shaded areas correspond to the regions of similarity between Tc1 and Tc7 inverted repeats.

in Bristol N2 (H.G.A.M.V.L. and R.H.A.P., unpublished observations). Between the lines maintained in parallel, most of the Tc7 bands were conserved. However, in several lines a few bands had been lost or gained, showing that Tc7 is mobile in the germ-line. Similar experiments using originally isogenic *mut-7* lines (R.F. Ketting and R.H.A.P., unpublished observations) showed that Tc7, like Tc1, is also active in the germ-line of this mutator strain (data not shown). Based on these results we conclude that Tc7 is a transposable element.

Tc7 transposition upon Tc1-transposase expression

Tc7 is an active transposon in the germline of several mutator strains, however we have not yet identified a Tc7 element with a gene coding for transposase. Since the binding site of the Tc1 transposase within Tc1 inverted repeats (17) is fully contained within the 38 bp sequences shared with the Tc7 extremities, we tested whether Tc7 could transpose when Tc1 transposase was expressed. Transgenic lines which harbour the Tc1 transposase

Table 1. Tc7 and Tc7-related elements

Tc7 variant	cosmid	coordinates	chromosome	size (bp)*	% identity†	gaps†
Tc7-1	ZK899	29218-30142	X	925	100	0
Tc7-2	C07D10	12383-13307	II	925	100	0
Tc7-3	D1009	997-71	X	927	100	2
Tc7-4	M02D8	35345-36271	X	927	100	2
Tc7-5	F02D10	7329-8254	X	926	100	1
Tc7-6	F57F5	23079-24004	V	926	100	1
Tc7-7	F23A7	6849-7774	X	926	100	1
Tc7-8	F10D7	30229-29311	X	919	100	1
Tc7-9	F19H6	10092-10973	X	882	100	2
Tc7-10	B0365	21017-21942	V	926	100	1
Tc7-d1	C30G4	5129-5936	X	808	83	10
Tc7-d2	R06C1	31287-32210	I	924	66	9
Tc7-d3	T08B2	20785-21708	I	924	66	13
Tc7-d4	F48B9	23972-24904	X	933	66	10
Tc7-d5	F52D2	19835-20765	X	931	65	9
Tc7-d6	E03G2	20069-19144	X	926	64	5
Tc7-d7	ZC47	26931-27858	I	928	62	13
Tc7-d8	C13E3	6850-6721	X	130	100	1
Tc7-d9	F45H7	36362-35684	III	679	93	1
Tc7-d10	R06C1	20136-20352	I	217	82	1
Tc7-d11	W02B8	20270-20534	II	265	82	5
Tc7-d12	C33D3	17462-17062	X	401	80	1
Tc7-d13	F52D2	35550-35890	X	341	80	8
Tc7-d14	T12D12	22756-22934	IV	179	79	3
Tc7-d15	F48B9	29798-29840	X	43	76	0
Tc7-d16	T05A12	15567-16129	IV	563	75	7
Tc7-d17	C05D9	5567-5324	X	247	74	6
Tc7-d18	K08E4	6623-6852	IV	230	68	1
Tc7-d19	T19D7	27640-28369	X	730	67	7
Tc7-d20	F07G6	7519-7656	X	138	66	2
Tc7-d21	F23D12	32609-33448	X	840	64	6
Tc7-d22	C05D12	38241-37511	II	731	63	2
Tc7-d23	C14E2	9579-9261	X	319	62	6

Based on the sequence data available in the ACeDB release of October 10th 1996.
*Including the flanking TA dinucleotides.
†Indentity scores are derived from a pairwise comparison with the reference sequence Tc7-1 using the Bestfit program (GCG, University of Wisconsin, Madison, USA).

gene under the control of a heat-shock promoter were generated. Transposition in the somatic cells was scored by PCR, using transposon specific primers and primers specific for a genomic target [the *gpa-2* gene (36)]. After heat-shock, several Tc7 insertions were detected in *gpa-2*, whereas without heat-shock, Tc7 transposition was much less frequent (35 versus 3 insertions; Fig. 4). Under the same conditions Tc1 transposition was also detectable. A comparison of the number of *gpa-2* insertions scored for Tc1 verses Tc7 using the same transgenic line and identical DNA inputs in the PCR showed that Tc7 is only four times less active than Tc1 (data not shown).

Tc7 target choice

Tc1 always inserts into TA dinucleotides, but uses only a small subset of all TA dinucleotides within the genome. Among those dinucleotides, a few are very often chosen and define 'hotspots' for Tc1 insertion (20). Since Tc7 can use Tc1 transposase for its mobility, we determined the distribution of Tc7 insertion sites within *gpa-2* and compared it to the distribution of Tc1 insertion sites. Forty two Tc7 insertion sites were sequenced (Fig. 5). As expected, Tc7 always inserted into TA dinucleotides. Furthermore, the distribution of the Tc7 insertions along the *gpa-2* sequence

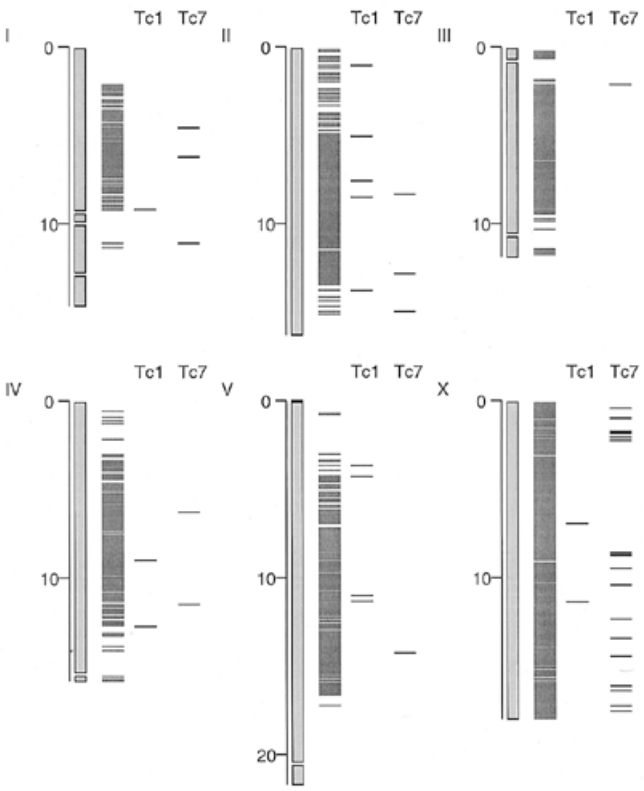


Figure 2. Distribution of Tc1, Tc7 and Tc7-related (see Table 1) elements mapped to the *C.elegans* genome. In light shade are shown the contigs on the physical map and the sequenced regions are shown dark shade. Horizontal lines indicate the locations of the Tc1 and Tc7 elements mapped in Bristol N2. This distribution is based on the *C.elegans* genomic sequence data released up to October 10, 1996.

perfectly fits the Tc1 insertion pattern. Most of the sequenced Tc7 insertions reside at TA dinucleotides that are also hit by Tc1, and the hottest insertion sites for Tc7 are the same as for Tc1.

DISCUSSION

Searches in the *C.elegans* genome database have revealed putative transposons or transposon fossils, including heterogeneous groups of elements presumably derived from the transposable elements Tc2 and Tc5 (5,13). The element analyzed here, which we call Tc7, and for which two variants have been described recently by Oosumi *et al.* (13), has all the hallmarks of a transposable element. Tc7 contains 345–347 bp inverted repeats of which the terminal 38 bp are nearly identical to the ends of Tc1 and is flanked by TA dinucleotides. Furthermore we have shown that Tc7 is mobile in the germline of independent mutator lines. However, Tc7 does not contain an ORF and is, therefore, presumably unable to make its own transposase. The 38 bp sequence shared with Tc1 encompasses the Tc1-transposase binding site and we have found that Tc1 transposase can promote Tc7 transposition.

Tc1/mariner DNA transposons always jump into TA dinucleotides. It has been shown for Tc1 and Tc3 that only a small fraction of the available TA dinucleotides is actually chosen for insertion. Tc1 and Tc3 each have a distinct preference for certain insertion sites (20). The pattern of Tc1 insertions in the *gpa-2* gene obtained

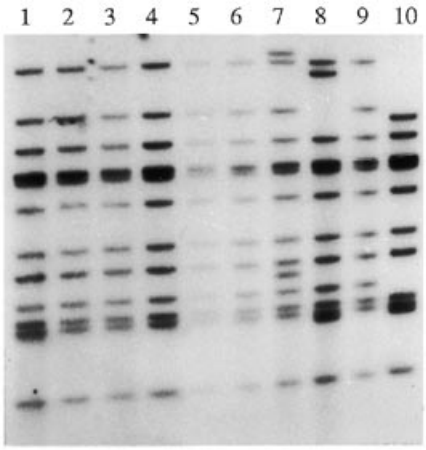


Figure 3. Southern blot hybridization of Bristol N2 (lane 10) and RW7097 DNAs (lanes 1–9) digested with *Sau*96I. Samples loaded in lanes 2–9 correspond to eight originally isogenic RW7097 lines which were cultivated in parallel for three months. The sample loaded in lane 1 corresponds to the parental RW7097 line. The Southern blot was probed with a Tc7 fragment containing the unique sequence and a large part of one of the inverted repeats (the end of the inverted repeat which is identical to Tc1 is not included in the probe).

in vitro with Tc1 transposase purified from *Escherichia coli* and a plasmid with the *gpa-2* gene as target is identical to that obtained *in vivo*, suggesting that the target choice is only determined by the transposase (28). Since we showed that the Tc1 transposase can cause Tc7 jumping, we expected that Tc1 transposase driven transposition of Tc7 would lead to the same target site specificity; we found this indeed to be the case.

No similarities between Tc7 and any known transposon were found in addition to the terminal 38 bp. Using the Tc7 sequence as a query to search through the *C.elegans* genome database, we found 23 related sequences of which 16 corresponded to single Tc7-related ends. The seven other hits, like Tc7 itself, were delimited by long inverted repeats and flanked by TA dinucleotides. However they showed more sequence heterogeneity, and in particular they showed sequence divergence within the terminal 38 bp. These sequence changes affect nucleotides supposed to be necessary for the binding of the Tc1 transposase based on methylation-interference experiments (28) and this would imply that these elements are not mobile. Alteratively these elements may depend for their mobility on other enzymes not yet identified.

The Tc7 elements and Tc7-related sequences are not evenly distributed within the *C.elegans* genome: two thirds of the Tc7 and Tc7-related sequences are located on the X chromosome (Fig. 2). In contrast, the Tc1 elements and other repetitive sequences are randomly distributed among the six chromosomes (37). The X chromosome has several unique features (38); e.g. the more uniform distribution of genes, repetitive sequences and recombination events along the X chromosome, the X chromosome contains more G and C runs and there is dosage compensation of the expression level of genes located on X. The clustering of Tc7 on the X chromosome might be another unique feature of this chromosome. Alternatively, it could also reflect a recent activation of an X chromosome located Tc7 element and a preference for local jumps. Yet another explanation could be that the parental strain of Bristol N2 originally did not contain Tc7

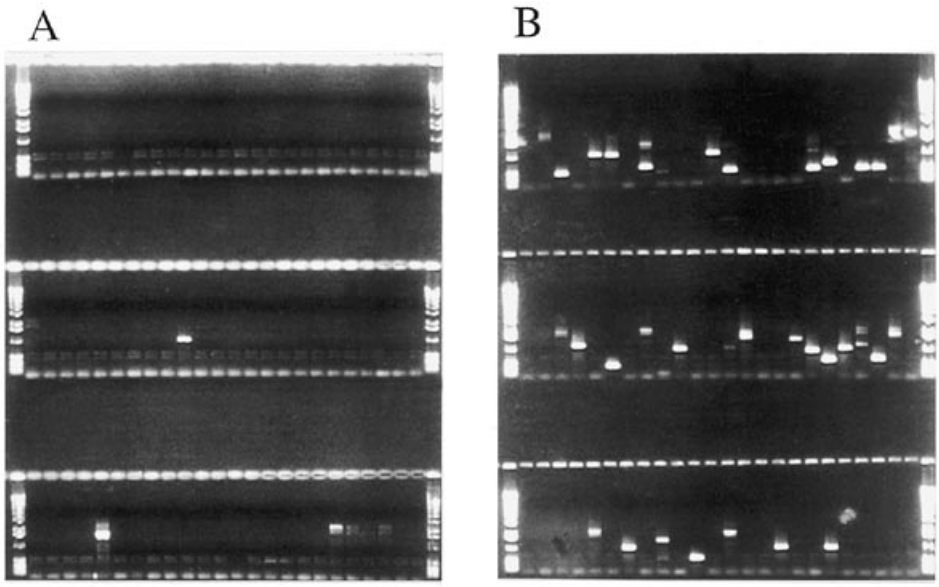


Figure 4. PCR detection of Tc7 somatic insertions in *gpa-2* without heat-shock (**A**) or after heat-shock (**B**) of the NL818 strain (for details see Materials and Methods). The lower band present in all lanes in (A) is due to a PCR artefact.

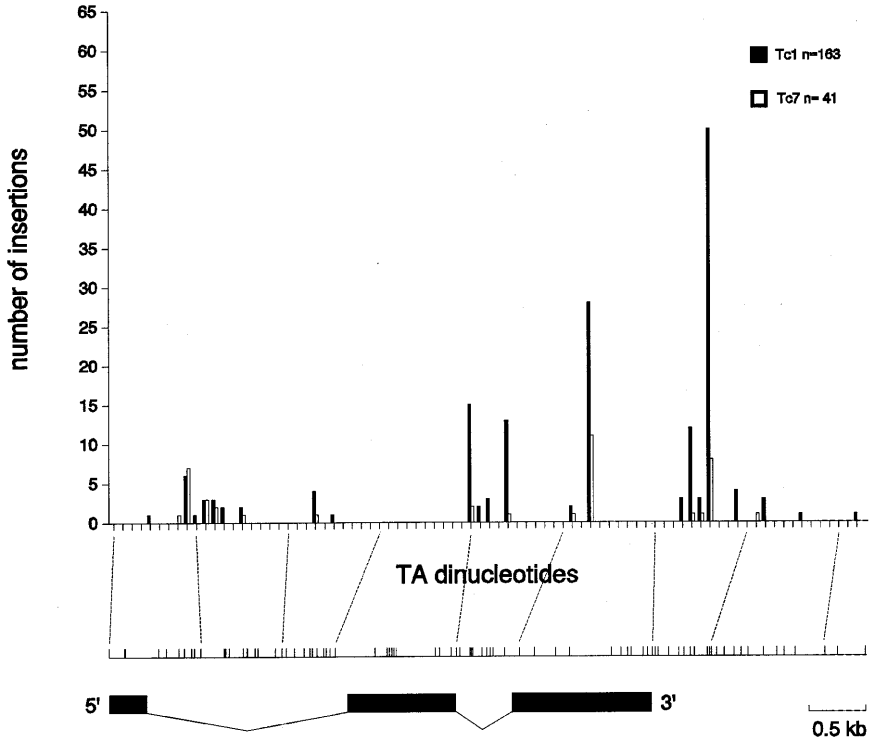


Figure 5. Comparison of the distribution of Tc1 insertions (20) and Tc7 insertions within the *gpa-2* target. Every mark on the x-axis represents a TA dinucleotide. The first and the last TA dinucleotides represented on the x-axis are located at positions 2058 and 3098 respectively. The intron/exon structure of the *gpa-2* gene and the actual position of the TA dinucleotides within its sequence are depicted underneath the x-axis.

elements and has only recently acquired an X chromosome through a cross with a strain that carried Tc7 elements.

The structure of Tc7 and its relationship with Tc1 is reminiscent of the Ac/Ds and the mutator families of transposable elements in maize (30,31,39–41). The Ac transposons are autonomous elements encoding their own transposase and they can mobilize various deleted derivatives (Ds elements) which still contain the terminal inverted repeats required for binding of the transposase. A variety of mutator-like elements of maize also show different levels of sequence diversity, while all having terminal sequences

in common (31,41). However, Tc7 is not a deletion derivative of Tc1; it has undergone a substitution of the unique internal region and most of the terminal inverted repeats. Alternatively, the Tc7 elements could have originated from a Tc1-like autonomous element which has not been identified yet or has been lost in the present-day strains. Thus far, Tc7 is the only group of Tc1-derived elements found in the *C.elegans* genome.

The analysis of the Tc7 content of originally isogenic mutator strains showing germline transposition of Tc1 suggests that Tc7 is also active in these strains. The copy number of Tc7 in the genome of Bristol N2 is comparable to that of Tc1. However, unlike the copy number of Tc1, the copy number of Tc7 has not increased in mutator strains. The transposition efficiency of Tc7 is of the same order of magnitude as that of Tc1. It is not clear why the Tc7 copy number has not increased in mutator strains and why thus far no Tc7 alleles have been obtained in a forward genetic screen (21–23,25).

ACKNOWLEDGEMENTS

We thank P.Borst, R.Ketting, G.Verlaan and S.Wicks for critical reading of the manuscript. This work was supported by a BIOTECH fellowship (B102CT94-8167) from the Commission of the European Communities to R.R. and a grant (5 R01 RR10082-02) from the NIH-NCRR to RHAP. RD is supported by the Wellcome Trust.

REFERENCES

1 Finnegan,D.J. (1992) *Curr. Opin. Genet. Dev.*, **2**, 861–867.
2 Henikoff,S. (1992) *New Biol.*, **4**, 382–388.
3 Robertson,H.M. (1995) *J. Insect Phys.*, **41**, 99–105.
4 Plasterk,R.H.A. (1996) In Saedler,H. and Gierl,A. (eds.), *Current Topics in Microbiology and Immunology, Transposable elements*. Springer-Verlag, Berlin Heidelberg, Germany, Vol. 204, pp. 125–143.
5 Morgan,G.T. (1995) *J. Mol. Biol.*, **254**, 1–5.
6 Auge-Gouillou,C., Bigot,Y., Pollet,N., Hamelin,M., Meunier-Rotival,M. and Periquet,G. (1995) *FEBS Lett.*, **368**, 541–546.
7 Oosumi,T., Belknap,W.R. and Garlick,B. (1995) *Nature*, **378**, 672.
8 Reiter,L.T., Murakami,T., Koeuth,T., Pentao,L., Muzny,D.M., Gibbs,R.A. and Lupski,J.R. (1996) *Nature Genet.*, **12**, 288–297.
9 Oosumi,T., Garlick,B. and Belknap,W.R. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 8886–8890.
10 Hodgkin,J., Plasterk,R.H.A. and Waterston,R.H. (1995) *Science*, **270**, 410–414.
11 Waterston,R. and Sulston,J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 10836–10840.
12 Smit,A.F.A. and Riggs,A.D. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 1443–1448.
13 Oosumi,T., Garlick,B. and Belknap,W.R. (1996) *J. Mol. Evol.*, **43**, 11–18.
14 Plasterk,R.H.A. and Van Luenen,H.G.A.M. (1997) In Riddle,D., Blumenthal,T., Meyer,B.J. and Pries,J. (eds.), *The Nematode Caenorhabditis elegans, II*. Cold Spring Harbor Laboratory Press, NY, USA, pp. 97–116.
15 Vos,J.C., Van Luenen,H.G.A.M. and Plasterk,R.H.A. (1993) *Genes Dev.*, **7**, 1244–1253.
16 Van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1993) *EMBO J.*, **12**, 2513–2520.
17 Vos,J.C. and Plasterk,R.H.A. (1994) *EMBO J.*, **13**, 6125–6132.
18 Colloms,S.D., Van Luenen,H.G.A.M. and Plasterk,R.H.A. (1994) *Nucleic Acids Res.*, **22**, 5548–5554.
19 Van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1994) *Cell*, **79**, 293–301.
20 Van Luenen,H.G.A.M. and Plasterk,R.H.A. (1994) *Nucleic Acids Res.*, **22**, 262–269.
21 Emmons,S.W. and Yesner,L. (1984) *Cell*, **36**, 599–605.
22 Emmons,S.W., Yesner,L., Ruan,K. and Katzenberg,D. (1983) *Cell*, **32**, 55–65.
23 Moerman,D.G. and Waterston,R.H. (1984) *Genetics*, **108**, 859–877.
24 Collins,J., Saari,B. and Anderson,P. (1987) *Nature*, **328**, 726–728.
25 Mori,I., Moerman,D.G. and Waterston,R.H. (1988) *Genetics*, **120**, 397–407.
26 Collins,J., Forbes,E. and Anderson,P. (1989) *Genetics*, **121**, 47–55.
27 Collins,J. and Anderson,P. (1994) *Genetics*, **137**, 771–781.
28 Vos,J.C., de Baere,I. and Plasterk,R.H.A. (1996) *Genes Dev.*, **10**, 755–761.
29 Plasterk,R.H.A. (1991) *EMBO J.*, **10**, 1919–1925.
30 Engels,W.R. (1989) In Berg,D.E. and Howe,M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, USA, pp. 437–484.
31 Fedoroff,N.V. (1989) In Berg,D.E. and Howe,M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, USA, pp. 375–411.
32 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
33 Feinberger,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
34 Durbin,R. and Thierry-Mieg,J. (1994) The ACEDB genome database. In Suhai,S. (ed.) *Computational Methods in Genome Research*. Plenum, NY, USA.
35 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
36 Fino Silva,I. and Plasterk,R.H.A. (1990) *J. Mol. Biol.*, **215**, 483–487.
37 Naclerio,G., Cangiano,G., Coulson,A., Levitt,A., Ruvolo,V. and La Volpe,A. (1992) *J. Mol. Biol.*, **226**, 159–168.
38 Waterston,R.H., Sulston,J.E. and Coulson,A.R. (1997) In Riddle,D.L., Blumenthal,T., Meyer,B.J. and Priess,J.R. (eds) *Caenorhabditis elegans II*. Cold Spring Harbor Laboratory Press, NY, USA, pp. 23–45.
39 Peacock,W.J., Dennis,E.S., Gerlach,W.L., Sachs,M.M. and Schwartz,D. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 347–354.
40 Sutton,W.D., Gerlach,W.L., Schwartz,D. and Peacock,W.J. (1984) *Science*, **223**, 1265–1268.
41 Chomet,P., Lisch,D., Hardeman,K.J., Chandler,V.L. and Freeling,M. (1991) *Genetics*, **129**, 261–270.