

---

**Sequence identity between an inverted repeat family of transposable elements in *Drosophila* and *Caenorhabditis***

---

L.J.Harris, D.L.Baillie<sup>1</sup> and A.M.Rose

---

Department of Medical Genetics, University of British Columbia, Vancouver, BC V6T 1W5 and  
<sup>1</sup>Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

---

Received February 26, 1988; Revised and Accepted May 31, 1988

---

**ABSTRACT**

The Tc1-like transposable elements, originally described in *Caenorhabditis elegans*<sup>1,2</sup>, have a much wider phylogenetic distribution than previously thought. In this paper, we demonstrate that Tc1 shares sequence identity in its open reading frame and terminal repeats with a new transposable element Barney (also known as TCb1 - Transposon *Caenorhabditis briggsae* 1). Barney was detected and isolated by Tc1 hybridization from the closely related nematode species, *Caenorhabditis briggsae*. The conserved open reading frames of Tc1 and Barney share identity with a structurally similar family of elements named HB found in *Drosophila melanogaster*, after the introduction of 3 small centrally located deletions in HB1. These reading frames would code for proteins with 30% amino acid identity (42% when conservative changes are included). Tc1, Barney and HB1 contain highly conserved blocks of amino acids which are likely to be in the functional domains of the putative transposase.

**INTRODUCTION**

Tc1 belongs to a class of eukaryotic transposable elements which have short terminal inverted repeats. Tc1 is 1610 bp in length, has 54 bp perfect inverted repeats, and contains an ORF with the potential to code for a 273 amino acid protein<sup>3</sup>. A high level of transposition is exhibited by Tc1 in *C. elegans* var. Bergerac (strain B0)<sup>4</sup>.

We wanted to know whether Tc1-like sequences were restricted to the *C. elegans* genome. Using Tc1 as a probe, we have found another member of the class of inverted repeat transposable elements in a second species of nematode, *Caenorhabditis briggsae* (L. Harris, Ph.D. Thesis, University of British Columbia). The *C. briggsae* transposable element Barney and the *C. elegans* element Tc1 show high sequence identity within their ORFs.

HB1 is an inverted repeat transposable element family in *Drosophila melanogaster*<sup>5,6</sup>. An HB element was first discovered inserted into the foldback element FB4 and was initially thought to be an integral part of foldback<sup>5</sup>. It was later shown that this sequence, HB1, was present in only one foldback element and other HB homologous sequences were present dispersed

throughout the genome with a copy number of approximately twenty<sup>6</sup>. HB1 is 1654 bp in length and has 30 bp inverted terminal repeats, similar to the structure of Tc1. When HB is hybridized to a number of *Drosophila melanogaster* strains, the banding pattern is almost identical among strains. Brierley and Potter thus suggested that this element is not active in *Drosophila melanogaster*<sup>6</sup>. We show that there is coding region sequence identity between Tc1, Barney, and HB1, which suggests an evolutionary relationship between these three transposable elements hosted by Arthropoda and Nematoda species.

#### MATERIALS AND METHODS

##### *C. briggsae* Strains

*C. briggsae* strain G16 was provided by A. Fodor who isolated this strain in India. Strain Z was isolated in the United States by B. Zuckerman. Strain BO is derived from a Z strain received from B. Zuckerman by D. Hirsh in 1978<sup>7</sup> and subsequently sent to the Caenorhabditis Genetics Center in 1980. Strains Z and BO were obtained by us from the Caenorhabditis Genetics Center, Columbia, Mo.

##### Cloning of the Barney Sequence

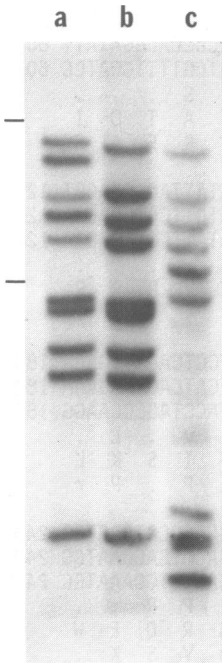
An EcoRV Tc1 fragment probe (containing the central 1572 bp of Tc1) was used to screen a Charon 4 *C. briggsae* strain G16 library (constructed by T. Snutch, Simon Fraser University). A Tc1-hybridizing 6.2 kb EcoRI fragment of the purified phage VR#2 was subcloned into pUC19. A 2.3 kb SalI/EcoRI fragment of the above 6.2 kb sequence was further subcloned and named pCbh11. The pCbh11 insert was cloned into the Bluescript (-) vector in both orientations to facilitate sequencing (plasmids pCbh12 and pCbh17).

##### Genomic Hybridizations

Nematode strains were grown as described<sup>8</sup> and DNA prepared by a method modified from Emmons et al.<sup>7</sup>. EcoRI-digested genomic DNA (4 ug) for each strain was electrophoresed and transferred to Nytran<sup>9</sup>. Hybridization conditions were a 20 hr incubation at 62°C in 5XSSPE [0.9 M NaCl, 0.05 M disodium hydrogen phosphate, 5 mM EDTA (pH 7.0)]; 0.3%SDS followed by two washes at 68°C in 0.2XSSPE;0.1%SDS.

##### DNA Sequencing and Database Search

Serial deletions of pCbh12 and pCbh17 were carried out using the Exonuclease III/S1 nuclease method of Henikoff<sup>10</sup> and dideoxy sequencing<sup>11</sup> of the double-stranded templates was accomplished<sup>12</sup>. Both strands of the 2.3 kb insert were sequenced. DNA sequences were analyzed using the Delaney sequence



**Figure 1.** Hybridization of Barney to *C. briggsae* genomic DNA isolated from three strains of *C. briggsae*. The probe pCbh11, containing 2.3 kb of TcI-hybridizing sequences, was nick-translated and hybridized at high stringency; lane a - strain B0; lane b - strain Z; and lane c - strain G16. Size markers (-) are 20.3 and 9.4 kb.

program and overlapping sequences were aligned using the DB system<sup>13</sup>. Sequence alignment and formatting was done using the DNA sequence editor, ESEE provided by E. Cabot, Simon Fraser University.

A search for related coding regions was carried out by using the expected amino acid sequence coded for by the major open reading frame of TcI and by searching for matching codons in the entire EMBL database.

## RESULTS AND DISCUSSION

Genomic blot hybridizations to DNA from the *C. briggsae* genome demonstrated the presence of TcI-hybridizing sequences. To determine whether these sequences represented a related transposable element family, sequences were isolated from a *C. briggsae* genomic Charon 4 library using a TcI probe. One of the *C. briggsae* TcI-hybridizing sequences was restricted to a 2.3 kb SalI/EcoRI fragment in pCbh11. When pCbh11 was hybridized to the genomic DNAs of three *C. briggsae* strains, different banding patterns for each of the three strains was observed (Figure 1). We conclude from these results that the sequence in pCbh11 is mobile and has been named a representative of the Barney family of transposable elements.

Nucleic Acids Research

Barney ATGGATCGGAACATTTTGCGCATGCCGAGAAGATCCTAGACGCACATCTACGGATATT 60  
Tc1 ATGGATCGCAACATCTCCGATCAGCAAGAGAAGATCCGCATAGGACCGCCACGGATATT 60  
HB1 ACCACAGATATAGAGGATCGACGCATTGTTTCTACAGCAAAGTCTATCGTTTTGCATCC 60  
Barney pro . . . . . A c . . . . . R . . . . . S . . . .  
Tc1 pro M D R N I L R S A R E D P H R T A T D I  
HB1 pro T T D I E D . R I V S Y S K V Y R F A S

Barney CAACTTTCTGTGACGTCTCCAAATGAACCGGTACCATCGAGAAGAATATTAGAAGACGT 120  
Tc1 CAAATGATTATAAGTTCTCCAAATGAACCTGTACCAAGTAAACGAACTGTTGCGTCGACGT 120  
HB1 TTATGGGACATAAAGTCTGAGCTGAACCTGGGAATCAGCGACGTTACTATTTCGTAGACGA 120  
Barney pro . l S v t . . . . . r . . . . . t . . . . .  
Tc1 pro Q M I I S S P N E P V P S K R T V R R R  
HB1 pro F R D . K . E L N L G I . D V . i . . .

Barney TTGCAAGTTGCTGGACTGCACGGACGAAGACCAGTCAAAAAACCACTCGTCAAGTTTGAAA 180  
Tc1 TTACAGCAAGCAGGACTACACGGACGAAAGCCAGTCAAGAAACCGTTTCATCAGTAAGAAA 180  
HB1 CTACTGAATCAAAATTTAGTGCAGGAGTCCACGAAAGTTCCCTACCTAGCCCAAGG 180  
Barney pro . . . V . . . . . r . . . . . L V . L .  
Tc1 pro L Q Q A G L H G R K P V K K P F I S K K  
HB1 pro . L n Q N F S a . S . R . V . L P . P r

Barney AACCGAAAAGTCGCGTTGAATGGGCTAAACAGCACTTGTCTGGGGCCCCGCGTAGTGG 240  
Tc1 AATCGCATGGTCGAGTTGCGTGGGCAAAAGCGCATCTTCGTTGGGACGTCAGGAATGG 240  
HB1 CATATTAAGGCAAGGTTAAGCTTAGCTAAAACCTACCTAAACTGGCCAGTCTCCAATGG 240  
Barney pro . . K . . . . E . . . . Q . . . . S . . . . P R . .  
Tc1 pro N R M A R V A W A K A H L R W G R Q E W  
HB1 pro H I K . . l S L . . T Y . N . p V S K .

Barney GCAAATCACATCTGGAGCGATGAATCGAAGTTCAATATGTTCCGAACTGATGGTATTTCAG 300  
Tc1 GCTAAACACATCTGGTCTGACGAAAGCAAGTTCAATTTGTTCCGGAGTGATGGAAATTC 300  
HB1 CGTAATATCCTTTGGACTGATGGGTCAAAAATCATGCTATTTGGTGGAACTGGTCTACTA 300  
Barney pro . N . . . . . m . . . . . t . . . . . I Q  
Tc1 pro A K H I W S D E S K F N L F G S D G N S  
HB1 pro R N I l . t . G . . I M . . . . . G T . S L

Barney TGGATT---CGACGTCCATTGGCTCCAGGTATGCTCCACAGTACCAATGTCCAACCTGTG 357  
Tc1 TGGGTA---CGTCGCTCTGTTGGCTCTAGGTAATCTCCAAAGTATCAATGCCAACCGTT 357  
HB1 CAGTATATCTGACGACCTCCAAACACGGAGTATCACCCAAAACACCCAGTGAAGACTTTC 360  
Barney pro . i . . . . . i . . . . . A . . . . . Q . . . . .  
Tc1 pro W V . R R P V G S R Y S P K Y Q C P T V  
HB1 pro Q Y I \* . . P N t E . H . . . . . H P V K . F

Barney AAACATGGAGGTGGATCTGTGATGGTTGGGGATGCTTCTCCGACACTTCTATG-GGACC 416  
Tc1 AAGCATGGAGTGGGAGCGTCATGGTGTGGGGGTGCTTACCAGCACTTCCATG-GGCC 416  
HB1 AATCACGGTGGACCTAAAATCATGGTATGGGCTGTTTTTTTATAATGGTATGAGTCAT 420  
Barney pro . . . . . s D . . . . .  
Tc1 pro K H G G G S V M V W G C F T S T S M G P  
HB1 pro N . . . . p K i . . . . a . . . . F Y N G . V M

Barney ATTGAAGAGAATC---GTTGGAACCATGGATCGATATGTGTACGAAGACATCCTGGAGAA 473  
Tc1 ACTAAGGAGAATC---CAAAGCATTATGGATCGTTTTCAATACGAAAACATCTTTGAAAC 473  
HB1 GCTATGGATTATGATTTATGGTATTATAGACCAAAACGCAATGTAAATATACTTAGTGA 480  
Barney pro . k . . . . . V G T . . . . . y V . . . . . D . . . . . L . N  
Tc1 pro L R R I Q S I M D R F Q Y E N I F E T  
HB1 pro . W I m I Y G . i . Q N A . V . . . . . L S D

Barney	CACAATGAGACCATGGGCAAGAGCAAATTTGGGCCGATCGTGGGTGTTCCAACAGGACAA	533
Tc1	TACAATGCGACCCTGGGCACCTTCAAAATGTGGCCGTGGCTTCGTGTTTCAGCAGGATAA	533
Hb1	TGTCTTATTGTCATATTCTGAATAAAAATATACCCTTAAATGGACATTTCCAACAGGATAA	540
Barney pro	. . . . . R A . l . . S w . . . . .	
Tc1 pro	T M R P W A L Q N V G R G F V F Q Q D N	
Hb1 pro	V l L S y S E * . i p L K w T . . . . .	
Barney	TGACCCGAAGCATACTTCGGGTACATGTCGCCAATTGGTTCAGACGTCGCCGTGTGAACCT	593
Tc1	CGATCCTAAGCATACTTCTCTTCATGTGCGTTTCATGGTTTCAACGTCGTCATGTGCATTT	593
Hb1	TGATCAGAAAACGCAGATGTAATCGGCTAAGAATAGGTTACCCAAAAATAGAATAGATGC	600
Barney pro	. . . . . G . A N . . R . . R . N .	
Tc1 pro	D P K H T S L H V R S W F Q R R H V H L	
Hb1 pro	. Q . R R C K S A k N R . T Q N R i D A	
Barney	CCTAGAATGGCCAAGTCAATCTCCAGACTTGAATCCCATCGAGCATATGTGGGAGGAGCT	653
Tc1	GCTCGATTGGCCAAGTCAGTCTCCGGACTTGAATCCAATAGAGCATTGTGGGAAGAGTT	653
Hb1	AATGCCGTGGCAAGCACCTTCCCATTAAACCGATTGAAAACCTGTATGGGGACAT	660
Barney pro	. e . . . . . m . . . . .	
Tc1 pro	L D W P S Q S P D L N P I E H L W E E L	
Hb1 pro	m P . Q A P P S H . . . . . N . y G d i	
Barney	GGAACGCCCTCAAAGGAGTCAGAGCATCCAATGCCAATCAAAGTTTGCTCAACTCGA	713
Tc1	GGAAGAGCGTCTTGGAGGTATTCCGGCTTCAAATGCAGATGCCAATTCACCCAGTTGGA	713
Hb1	TAAACAGTTTGTCGAAGAAGTCCCCGACGCTAAGACTCAGATTTGGCAAGTTGTGCA	720
Barney pro	. . . . . K . v . . . N Q . . A . . .	
Tc1 pro	E R R L G G I R A S N A D A K F N Q L E	
Hb1 pro	K Q F v S K K S p t S K T Q I w q V v Q	
Barney	AGCTGCTTGAAGAGTATCCCGATGACGTTGGTTCAGACGCTCCTGGAGTCGATGCCACG	773
Tc1	AAACGCTTGAAGAGTATCCCATGTGAGTTATTCACAAGCTGATCGACTCGATGCCACG	773
Hb1	GGATACATGGGCAAAAATCTCCCAAACCTTGCTAGGACTTGGTGGACTTCATGCCCG	780
Barney pro	A . . . . S . . . . t l v Q T . l e . . . .	
Tc1 pro	N A W K A I P M S V I H K L I D S M P R	
Hb1 pro	D T . A K . . P K P C * D . v . F . . . .	
Barney	TAGATGCAAGGCTGTTATCGATGCGAAAGGATACCCAACGAAATATTA	822
Tc1	TCGTTGTCAAGCTGTTATTGATGCAAACGGATACGCGACAAAGTATTA	822
Hb1	TGGGTGTAAGGCTGTGCTGGCTAACAAAGGCTATCCAGCCAAGTATTG	829
Barney pro	. . K . . . . K . . . p . . . . *	
Tc1 pro	R C Q A V I D A N G Y A T K Y *	
Hb1 pro	G . K . . l A N K . . p A . . *	

**Figure 2.** DNA and amino acid sequence of Barney and Hb1 compared with Tc1. Hb1 and Barney amino acid sequence are each compared to Tc1 amino acid sequence. "." indicates amino acid conservation, a lower case letter indicates a similar amino acid change<sup>8</sup>, and an upper case letter indicates a significant amino acid change. Tc1 sequence from Rosenzweig et al.<sup>3</sup> and Hb1 sequence taken from Brierley and Potter<sup>7</sup>.

The Barney elements have been characterized and the insertion within pCbh11 (Barney.10) sequenced (L. Harris, Ph.D. Thesis, University of British Columbia). Upon comparison with Tc1, the sequence identity was found to be primarily restricted to the open reading frames. The ORF of Barney starts and stops at the same position in the sequence as the ORF of Tc1 does. Within the

```

HB1   IR                               GCACTGCTATTTTTATGAACACAGCTGTA
                                     |*||| |||*| || || |||||
Tc1   IR  AGAAAAAGTTACACACAAAAAACCAAAAGTGGATATCTTTTGGCCAGCACTGTA
          ||||| ||||| ||||| || || ||||| ||||| ||||| ||||| |||||
Barney IR TTTCAAAAAGTTATCTTCAAAAAACAAGTTGTCGCATTCTTTATGGCCAGTACTGTA
  
```

Figure 3. Comparison of putative terminal inverted repeat sequences. The putative termini of HB1, Tc1, and Barney are shown below, with the end of the elements on the right. Bars bridging DNA sequences mark sites of sequence identity between HB1 and Tc1 as well as Tc1 and Barney. "\*" indicate identical nucleotides between HB1 and Barney not in common with Tc1. Tc1 sequence from Rosenzweig et al.<sup>3</sup> and HB1 sequence taken from Brierley and Potter<sup>7</sup>.

ORF, the amino acid sequence identity is 74% and the overall nucleotide sequence identity is 71%. In order to discover whether Tc1 and Barney contained any significant regions of identity with other transposable elements, we searched the EMBL data bank. A surprising amount of identity to HB1, a member of the HB transposable element family in *Drosophila melanogaster*, was found.

We compared the DNA sequences of Tc1, Barney and HB1. On first inspection, it seemed that the sequence identity of these elements was limited to the last third of the Tc1 ORF. However, when 3 small deletions of 3, 1, and 3 bp were created at positions 304, 408 and 427 in HB1, the rest of the putative coding region could be aligned with that of Tc1 (Figure 2). Once aligned in this way, the stop codon of the Tc1 ORF and the equivalent stop codon of HB1 are at approximately the same position (one bp difference) from the start of the elements. Within the coding regions there are blocks of highly conserved sequences. The HB sequence contains 3 in-frame stop codons, but sequences preceding and following these stop signals show strong conservation. If these stop codons are ignored, HB1 and Tc1 could code for proteins with 30% amino acid identity. When similar amino acid changes (as defined by Doolittle<sup>14</sup>) are taken into account, 42% amino acid identity is obtained. The amino acids in common between the Tc1 and HB1 coding regions are also shared with Barney, with the exception of seven residues (of which four residues are conserved changes in Barney).

No identity was observed in the sequences between the ORFs and the inverted repeat termini. The inverted repeats, however, do show sequence identity between Tc1, Barney and HB1 (Figure 3). A sequence frameshifted one basepair from alignment with the Tc1 sequence is the probable terminal repeat of Barney and has 68% sequence identity with the Tc1 inverted repeat. The absence of any sequence identity between Tc1 and Barney from the ORF to the

inverted repeat of Tc1 suggests that this putative Barney terminus is functionally conserved. Although the HB1 terminal repeats are only 30 bp long, they show greater than 50% identity with the Tc1 and Barney terminal repeats. We propose that the transposable element families of Tc1, Barney, and HB1 share a common ancestry. Most likely HB elements in *D. melanogaster* were mobile in the past, became immobile and have accumulated stop codons and frameshifts.

The class of transposable elements having short terminal repeats is widely distributed and includes Tc1 of *Caenorhabditis elegans*<sup>1,2,3</sup>, Mu of maize<sup>15</sup>, mariner of *Drosophila mauritiana*<sup>16</sup>, and HB, P, and hobo of *Drosophila melanogaster*<sup>6,17,18</sup>. These transposable elements have in common relatively short terminal inverted repeats (12 to 213 bp) which flank an open reading frame. Unlike the retrotransposons which share numerous structural and behavioral characteristics<sup>19</sup>, the inverted repeat elements show considerable variation. HB elements, P and Mu are quite heterogeneous in length<sup>6,17,20,21</sup>. Mu of *Zea mays* has many behavioral characteristics in common with Tc1, including a high level of somatic excision and formation of extrachromosomal forms<sup>22,23,24,25,26</sup>. However Mu, P and mariner showed no similarity to Tc1 with regard to DNA sequences and coding potential. We have shown that that several members of this class do in fact share common sequences in what are likely to be the functionally important parts of their coding regions. It appears therefore that within the inverted repeat elements, a subclass of closely related elements exists. This subclass includes Tc1 and Barney in the Nematoda and HB in Arthropoda. Since these phyla are evolutionarily quite distant it is possible that Tc1-like elements will have a wide distribution within the eukaryotes. Alternately, their relatedness may be to a greater extent due to horizontal transmission between a parasitic nematode and its insect host.

#### ACKNOWLEDGEMENTS

This work was supported by a Medical Research Council of Canada studentship to L.J.H. and operating grants from the Medical Research Council of Canada to A.M.R. and the Natural Sciences and Engineering Council of Canada to D.L.B.

#### REFERENCES

1. Emmons, S.W., Yesner, L., Ruan, K., & Katzenberg, D. (1983) *Cell* 32, 55-65.
2. Liao, L.W., Rosenzweig, B., and Hirsh, D. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3585-3589.
3. Rosenzweig, B., Liao, L.W., and Hirsh, D. (1983) *Nucl. Acids Res.* 11, 4201-4209.

4. Moerman,D.G., and Waterston,R.H. (1984) *Genetics* 108, 859-877.
5. Potter,S.S. (1982) *Nature* 297, 201-204.
6. Brierley,H.L. and Potter,S.S. (1985) *Nucl. Acids Res.* 13, 485-500.
7. Emmons,S.W., Klass,M.R. and Hirsh,D. (1979) *Proc. Natl. Acad. Sci USA* 76, 1333-1337.
8. Brenner,S. (1974) *Genetics* 77, 71-94.
9. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
10. Henikoff,S. (1987) *Methods Enzymol.* in press.
11. Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
12. Hattori,M. & Sakaki,Y. (1986) *Anal. Biochem.* 152, 232-238.
13. Staden,R. (1980) *Nucl. Acids Res.* 8, 3673-3694.
14. Doolittle,R.F. (1979) from *The Proteins*, Vol. IV, 3rd ed., eds. H.Neurath and R.L.Hill (Academic Press Inc.) pp.1-118.
15. Barker,R.F., Thompson,D.V., Talbot,D.R., Swanson,J. and Bennetzen,J.L. (1984) *Nuc. Acids Res.* 12, 5955-5969.
16. Jacobson,J.W., Medhora,M.M. and Hartl,D.L. (1983) *Proc. Natl. Acad. Sci. USA* 83, 8684-8688.
17. O'Hare,K. and Rubin,G.M. (1983) *Cell* 34, 25-35.
18. Streck,R.D., MacGaffey,J.E., and Beckendorf,S.K. (1986) *EMBO J.* 5, 3615-3623.
19. Weiner,A.M., Deininger,P.L. and Efstratiadis,A. (1986) *Ann. Rev.Biochem.* 55, 631-661.
20. Chandler,V., Rivin,C. and Walbot,V. (1986) *Genetics* 114, 1007-1021.
21. Taylor,L.P. and Walbot,V. (1987) *Genetics* 117, 297-307.
22. Emmons,S.W. and Yesner,L. (1984) *Cell* 36, 599-605.
23. Rose,A. and Snutch,T.P. (1984) *Nature* 311, 485-486.
24. Ruan,K. and Emmons,S.W. (1984) *Proc. Natl. Acad. Sci. USA* 81:4018-4022.
25. Robertson,D. (1978) *Mutat. Res.* 51, 21-28.
26. Sundaresan,V. and Freeling,M. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4924-4928.