The transposable element Uhu from Hawaiian *Drosophila* — member of the widely dispersed class of Tc1-like transposons

Laura Brezinsky, Gordon V.L.Wang[1], Tom Humphreys[1] and John Hunt
Cancer Research Center of Hawaii, 1236 Lauhala St. Honolulu, HI 96813 and [1]University of Hawaii, Pacific Biomedical Research Center, Honolulu, HI 96822, USA

Received January 2, 1990; Revised and Accepted March 7, 1990 EMBL accession no. X17356

ABSTRACT

We report the complete nucleotide sequence of the transposable element Uhu from the vicinity of the alcohol dehydrogenase (Adh) gene of *Drosophila heteroneura* (an endemic Hawaiian *Drosophila*). The complete element is about 1650 base-pairs (bp) long, has 46 – 50 base-pair inverse imperfect repeats at it's ends, and contains a large open reading frame potentially encoding a 192 amino acid protein. We demonstrate that Uhu belongs to a class of transposable elements which includes Tc1 from *Caenorhabditis elegans*, Barney from *Caenorhabditis briggsae*, and HB1 from *Drosophila melanogaster*. All of these elements share significant sequence similarity, are approximately 1600 base pairs long, have short inverse terminal repeats (ITRs), contain open reading frames (ORFs) with significant sequence identity, and appear to insert specifically at TA sequences generating target site duplications.

INTRODUCTION

Transposable elements, DNA sequences which move from one location to another within a genome, were first discovered by Barbara McClintock in the 1940's. McClintock demonstrated that, not only were these elements capable of transposition, but that this transposition might affect the function of genes near the sites of insertion (1). Since the original discovery, transposable elements of various structural types have been found in a wide variety of both prokaryotic and eukaryotic organisms (2,3,4,5,6,7,8).

A repetitive element was discovered during the course of restriction mapping the alcohol dehydrogenase gene region of *D. heteroneura* (9,10) which reacted to genomic southern blots of DNA from four additional species of Hawaiian *Drosophila* in the *planitibia* subgroup (*D. silvestris*, *D. planitibia*, *D. differens* and *D. picticornis*) (10). *In situ* chromosomal hybridizations comparing four of these species, *D. heteroneura*, *D. silvestris*, *D. planitibia*, and *D. differens*, indicated extensive differences in chromosomal distribution of the 80–150 copies of this element (10).

In this paper we describe the cloning of four copies of the Uhu element from *D. heteroneura* (Uhu-1, Uhu-2, Uhu-3, and Uhu-4) and present the complete DNA sequence of Uhu-1. It shares structural and sequence similarity with the transposable elements Tc1 from *C. elegans (11)*, Barney from *C. briggsae* (12,13), and the HB family of transposable elements from *D. melanogaster* (14) and therefore represents a newly discovered member of this class of elements.

Tc1, a highly conserved 1610 bp transposable element found in all *C. elegans* strains examined (15), has a genomic copy number ranging from 30–300 (11,16), 54 bp perfect ITRs and contains two ORFs on the same DNA strand in different translational reading frames. The smaller of the two ORF's which potentially encodes a 112 amino acid protein is nested within the larger ORF which potentially encodes a 273 amino acid protein (11). All Tc1 elements characterized are flanked by a TA dinucleotide which may represent a duplicated target site (11,17,19,21). Although the mechanism of transposition is not known, extrachromosomal copies of Tc1 detected in *C. bergerac* somatic cells are presumed to be excision products (22,23). Tc1 has been shown to undergo frequent, spontaneous excision in somatic cells of the *Bergerac* and *Bristol* strains (18,24,25) but germline transposition of Tc1 has been detected only in the *Bergerac* strain (17,20,26,27,28). Mutator activity responsible for germline activity has been mapped to several locations in the *Bergerac* genome (26), and recent evidence suggests that the mutator itself is transposable (28).

Barney, a family of transposable elements in *C. briggsae*, is closely related to Tc1 (12,13). A composite Barney element constructed from 2 deletion mutants, is 1616 bp, including 80 bp imperfect ITR's which end with the TA dinucleotide possibly representing a duplicated target site (13). Barney has a large open reading frame (12) which has 71% nucleotide sequence identity, and 74% amino acid sequence identity with the large Tc1 open reading frame (12).

HB1 is the major representative of the highly variable HB family of transposable elements from *D. melanogaster* (14). HB1 is 1655 bp long with 30 bp imperfect ITR's terminating with the TA dinucleotide. When three small deletions are created in the HB1 sequence, a region aligns with the large Tc1 open reading frame revealing a 30% amino acid sequence identity (12). HB1 banding patterns on genomic southern blots are similar across four strains of *D. melanogaster* indicating that it is not actively transposing in these strains (14).

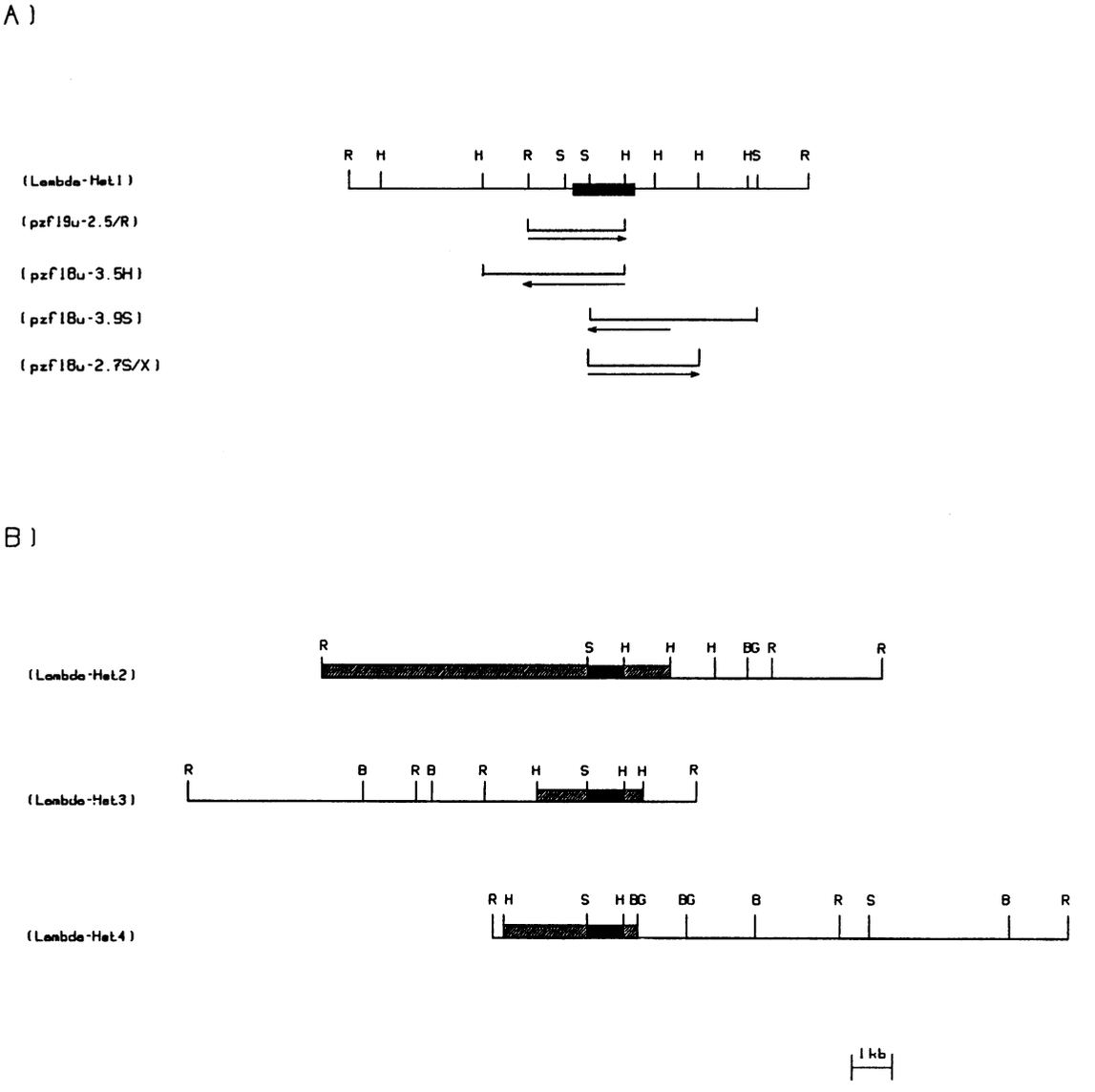


Figure 1. Restriction maps of D. heteroneura DNA containing four independent isolates of Uhu. Restriction enzymes used are;R, EcoRI; H,HindII; S,SalI; BG BglII; B, BamHI. A) Restriction map of the λ-Het1 insert containing Uhu-1. The boxed region is Uhu-1 as determined by DNA sequence data. Restriction fragments subcloned into pzf vectors are indicated below the map. Arrows indicate sequencing strategy. B) Restriction maps of λ-Het2, λ-Het3, and λ-Het4 containing Uhu-2, Uhu- 3, and Uhu-4 respectively. Restriction fragments containing portions of the Uhu elements as determined by southern blot analysis are hatched or blackened. The blackened box represents the conserved 0.9 kb H/S fragments which hybridize strongly to Uhu-1, flanking fragments hybridizing more weakly are hatched.

MATERIALS AND METHODS

Drosophila Stocks

All *Drosophila* stocks were maintained in the lab from wild caught flies (9). The *D. heteroneura* stock used to construct the genomic DNA library was an isofemale line collected on the island of Hawaii in 1972. Genomic DNA libraries were constructed in the λ Charon4 vector from a volume of adult flies from these stocks as described previously (9).

Cloning

Uhu-1 was recognized as a repetitive *D. heteroneura* sequence in an ADH containing λ clone (9). Two subclones (pBR322-3.5H and pBR322-3.9S) containing overlapping restriction fragments yielded repetitive banding patterns (9,10) when reacted to *D. heteroneura* genomic southern blots. Restriction fragments containing the repetitive sequences were subcloned into the pzf plasmid vectors producing subclones pzf18u-3.5H, pzf18u-3.9S, pzf18u-2.7S/X and pzf19u-2.5H/RI which were used in sequencing as indicated (Fig. 1). Three additional clones; λ-Het2, λ-Het3, and λ-Het4 from the D. heteroneura genomic library were isolated by plaque hybridization to pBR322-3.5H, and contain Uhu-2, Uhu-3 and Uhu-4 respectively (Fig. 1). Restriction maps of the transposable elements in these clones were constructed by standard techniques of restriction mapping (29,30) and Southern blot analysis (31) using pBR322-3.5H and pBR322-3.9S as radiolabelled probes.

DNA Amplification

JM101 (32) was used as a host strain for all plasmid and phagemid (phage-like particles) (33) amplifications. Plasmid DNA used for cloning and restriction mapping was purified from overnight liquid cultures using the alkaline lysis method (34). Plasmids were

precipitated in isopropyl alcohol, resuspended in TE bufer, and reprecipitated in 2.5 M NH_4oAc (29). Plasmid DNA used for probes was purified by cesium chloride equilibrium-gradient centrifugation. The pzf plasmids were induced to synthesize phagemids containing single-stranded DNA by superinfection of 250 ml plasmid cultures with M13KO7 helper phage (33). Phagemids were precipitated with polethylene glycol (PEG) and single-stranded DNA purified by phenol extraction (29). In most cases it was necessary to further purify single stranded DNA with Elutip-D columns (Schleicher and Schuell). Samples were loaded onto the Elutip-D columns in 0.5 M NaCl and eluted with 2.0 M NaCl. Bacterial strains K802 and LE392 were used as hosts for λ amplifications. λ phage were amplified as plate lysates, and purified by PEG precipitation followed by banding on cesium chloride equilibrium gradients (29,35). λ DNA was subsequently purified by proteinase K digestion and phenol/chloroform extraction (29).

Restriction Mapping and Southern Blot Analysis

Restriction enzymes were purchased from International Biotechnologies Inc. and used in conjunction with the manufacturer's supplied buffers. Horizontal agarose gel electrophoresis was conducted under standard conditions for restriction mapping (29,30). Southern transfer onto Nytran was carried out per the supplier's instructions (31). Plasmids were labeled by random priming with ^{32}P-dCTP, and passed over Elutip-D columns to eliminate unincorporated nucleotides. Blots were prehybridized for 3 hours at 68°C in 6× SSC, 5× Denhardt's solution, 0.5% SDS and 0.1 mg/ml denatured salmon sperm DNA and hybridized overnight at 68°C in the same solution with the addition of 5×10^5 cpm/ml of probe. Blots were washed at 68°C; once in 5× SSC, twice in 1× SSC and twice in 0.5× SSC.

DNA Sequencing

Both strands of Uhu-1 were sequenced completely using the dideoxy sequencing method of Sanger and Coulson (36). Deletion subclones of pzf18-3.5H, pzf18-3.9S pzf18-2.7S/X and pzf19-2.5H/RI produced using the Dale deletion method (37). Gaps in the resulting sequence were filled using synthetic oligonucleotides as sequencing primers for the parental clones. The sequences of the Uhu-2, Uhu-3, and Uhu-4 termini were obtained from pzf18U subclones containing these elements. Synthetic olionucleotides were used as sequencing primers based on the sequence of Uhu-1.

Data Analysis

DNA sequence data was edited and analysed using BIONET National Computer Resource for Molecular Biology.

RESULTS

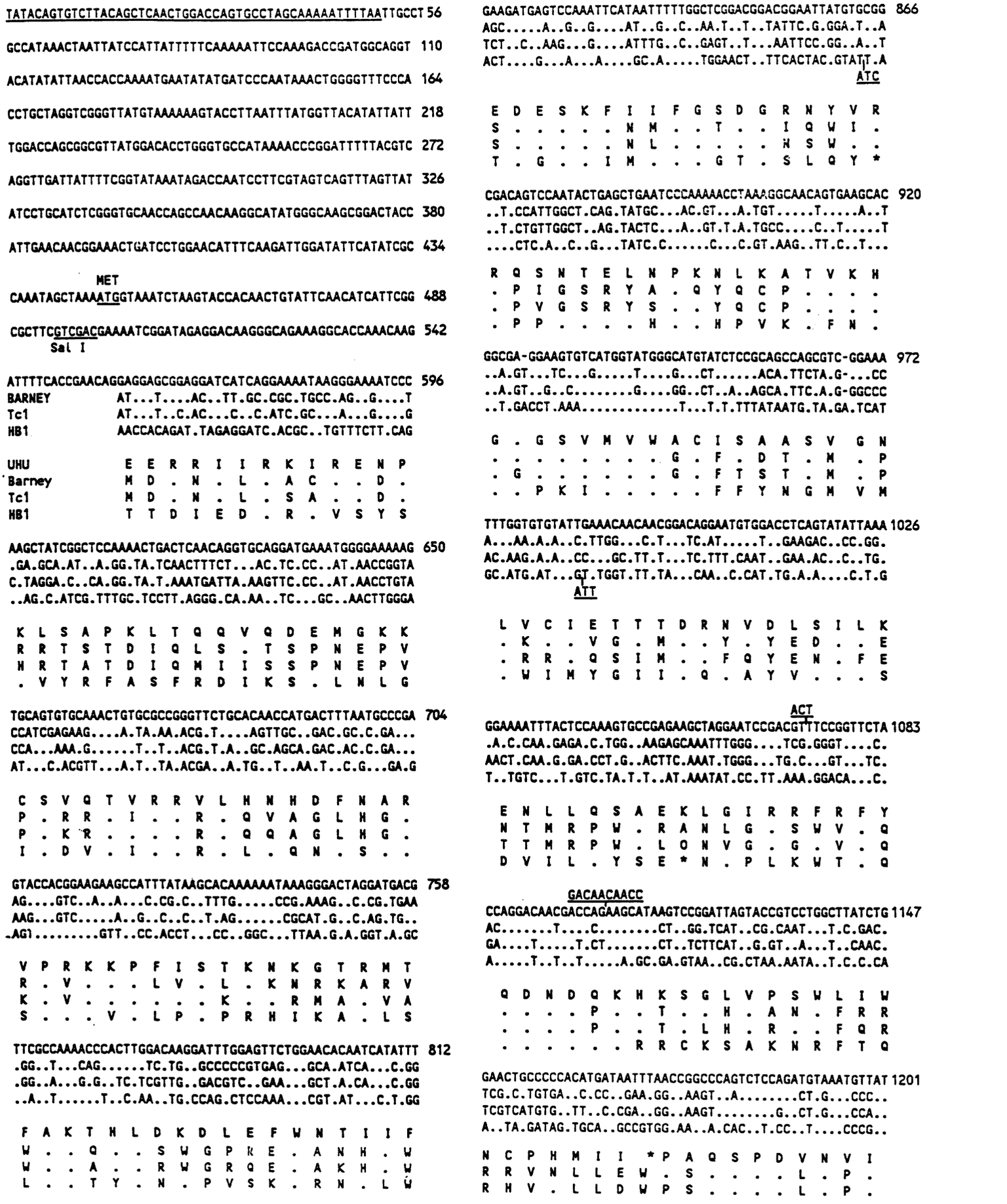
Uhu Sequence

Southern blot analysis showed that a repetitive sequence is contained within a 2.2 kb region of the *D. heteroneura* Adh containing clone λ-Het1 (9,10). Based on this information Uhu was subcloned and sequenced according to the strategy outlined in Fig. 1. A transposable-like element between 1646−1655 bp long with inverse terminal imperfect repeats of 46−50 bp was recognized. The sequences of the termini of three additional Uhu elements (Uhu-2, Uhu-3, and Uhu-4) were compared to the sequence of Uhu-1 to define the ends of Uhu (data not shown). Because the sequence TATA, which is both a direct and an inverse repeat of itself, is found at both ends of Uhu-1, Uhu-2, Uhu-3, and Uhu-4, it is not possible to determine the precise ends of the element without additional data. TATA could represent the original genomic target sites of insertion which were duplicated during transposition in which case Uhu is 1646 bp long with 85% identical 46 bp inverse terminal repeats. Alternatively, the TATA sequences could represent the outside ends of the inverted terminal repeats in which case Uhu is 1655 bp long with 86% identical 50 bp inverse terminal repeats.

Sequence Comparison Between Uhu and Known Transposable Elements

The GenBank DNA database library was searched unsuccessfully. The Uhu sequence was then subjected to a *Drosophila* coding bias analysis and a region approximately 900 bp long was found with protein-encoding codon usage in various reading frames. This region was translated in all three translational reading frames and submitted to the Protein Identification Resource database. The search revealed an 89 amino acid sequence with 33.7% identity between Uhu and Tc1. This sequence is contained in an open reading frame begining at postion 447 and ending at position 1023, which potentially encodes a 192 amino acid protein with 23% amino acid sequence identity to the putative Tc1 protein. By inserting 3 gaps of 1, 3 and 8 nucleotides at positions 926, 1247 and 1261, respectively, and 2 deletions of 3 and 10 bp at positions 1072 and 1100, respectively, in the Uhu sequence, the entire 273 amino acid ORF frame of Tc1 aligns with a 273 amino acid sequence of Uhu (Fig. 2). Aligned in this way, Uhu and Tc1 share a 46% nucleotide sequence and a 40% amino acid sequence identity, over the entire region of the Tc1 ORF. Significant sequence identity between these two elements is restricted to this region. Both sequences have a TAA stop codon at the same relative position corresponding to the end of the Tc1 open reading frame. Putative start codons are not aligned (Fig. 2). The start codon in the Uhu sequence nearest the apparent Tc1 start codon at position 558 is found 111 bp 5′ at position 447. However, by inserting a T at position 560 in the Uhu sequence, Uhu would have a start codon corresponding to the putative Tc1 start codon. In the Uhu sequence, a one bp deletion relative to Tc1 at position 926 results in a frame shift which is corrected by a 10 bp insertion at position 1100. This 174 bp frame shifted sequence encodes 54 amino acids with several long stretches identical to the putative Tc1 protein. An additional frame shift occurs at position 1261 as a result of the 8 bp deletion in the Uhu sequence relative to Tc1. In frame stop codons at positions 1023, 1056, 1170 and 1203 in the Uhu sequence would terminate translation of the Uhu protein before the stop codon corresponding to the end of the Tc1 open reading frame, and eliminate regions of significant sequence identity to Tc1.

The Uhu sequence was also compared with Barney (12) and HB1 (14), transposable elements of known sequence similarity to Tc1 (12,38). Barney was aligned with Uhu by allowing the same gaps and deletions of the Uhu DNA sequence as listed above. When aligned in this way, Barney and Uhu share a 47% nucleotide sequence identity, and a 38% amino acid sequence identity, in the region which aligns with the Tc1 open reading frame. Alignment of HB1 with the modified Uhu sequence required 3 deletions of 3 bp, 1bp and 3 bp at postitions 863, 967, and 986, respectively, in the HB1 sequence. When aligned in this way, Uhu and HB1 share a 42% nucleotide sequence identity, and a 32% amino acid sequence identity, in the region which


TATACAGTGTCTTACAGCTCAACTGGACCAGTGCCTAGCAAAAATTTTAATTGCCT 56
GCCATAAACTAATTATCCATTATTTTTCAAAAATTCCAAAGACCGATGGCAGGT 110
ACATATATTAACCACCAAAATGAATATATGATCCCAATAAACTGGGGTTTCCCA 164
CCTGCTAGGTCGGGTTATGTAAAAAAGTACCTTAATTTATGGTTACATATTATT 218
TGGACCAGCGGCGTTATGGACACCTGGGTGCCATAAAACCCGGATTTTTACGTC 272
AGGTTGATTATTTTCGGTATAAATAGACCAATCCTTCGTAGTCAGTTTAGTTAT 326
ATCCTGCATCTCGGGTGCAACCAGCCAACAAGGCATATGGGCAAGCGGACTACC 380
ATTGAACAACGGAAACTGATCCTGGAACATTTCAAGATTGGATATTCATATCGC 434
MET
CAAATAGCTAAAATGGTAAATCTAAGTACCACAACTGTATTCAACATCATTCGG 488
CGCTTCGTCGACGAAAATCGGATAGAGGACAAGGGCAGAAAGGCACCAAACAAG 542
Sal I
ATTTTCACCGAACAGGAGGAGCGGAGGATCATCAGGAAAATAAGGGAAAATCCC 596
BARNEY AT...T....AC..TT.GC.CGC.TGCC.AG..G....T
Tc1 AT...T..C.AC...C..C.ATC.GC...A...G....G
HB1 AACCACAGAT.TAGAGGATC.ACGC..TGTTTCTT.CAG
UHU E E R R I I R K I R E N P
Barney M D . N . L . A C . . D .
Tc1 M D . N . L . S A . . D .
HB1 T T D I E D . R . V S Y S
AAGCTATCGGCTCCAAAACTGACTCAACAGGTGCAGGATGAAATGGGGAAAAAG 650
.GA.GCA.AT..A.GG.TA.TCAACTTTCT...AC.TC.CC..AT.AACCGGTA
C.TAGGA.C..CA.GG.TA.T.AAATGATTA.AAGTTC.CC..AT.AACCTGTA
..AG.C.ATCG.TTTGC.TCCTT.AGGG.CA.AA..TC...GC..AACTTGGGA
K L S A P K L T Q Q V Q D E M G K K
R R T S T D I Q L S . T S P N E P V
H R T A T D I Q M I I S S P N E P V
. V Y R F A S F R D I K S . L N L G
TGCAGTGTGCAAACTGTGCGCCGGGTTCTGCACAACCATGACTTTAATGCCCGA 704
CCATCGAGAAG....A.TA.AA.ACG.T....AGTTGC..GAC.GC.C.GA...
CCA...AAA.G.....T..T..ACG.T.A..GC.AGCA.GAC.AC.C.GA...
AT...C.ACGTT...A.T..TA.ACGA..A.TG..T..AA.T..C.G...GA.G
C S V Q T V R R V L H N H D F N A R
P . R R . I . . R . Q V A G L H G .
P . K R . . . . R . Q Q A G L H G .
I . D V . I . . R . L . Q N . S . .
GTACCACGGAAGAAGCCATTTATAAGCACAAAAAATAAAGGGACTAGGATGACG 758
AG....GTC..A..A...C.CG.C..TTTG.....CCG.AAAG..C.CG.TGAA
AAG...GTC.....A..G..C..C..T.AG.....CGCAT.G..C.AG.TG..
-AG1.....GTT..CC.ACCT...CC..GGC...TTAA.G.A.GGT.A.GC
V P R K K P F I S T K N K G T R M T
R . V . . . L V . L . K N R K A R V
K . V . . . . . K . . R M A . V A
S . . . V . L P . P R H I K A . L S
TTCGCCAAAACCCACTTGGACAAGGATTTGGAGTTCTGGAACACAATCATATTT 812
.GG..T...CAG.....TC.TG..GCCCCCGTGAG...GCA.ATCA...C.GG
.GG..A...G.G..TC.TCGTTG..GACGTC..GAA...GCT.A.CA...C.GG
..A..T.....T..C.AA..TG.CCAG.CTCCAAA...CGT.AT...C.T.GG
F A K T H L D K D L E F W N T I I F
W . . Q . . S W G P R E . A N H . W
W . . A . . R W G R Q E . A K H . W
L . . T Y . N . P V S K . R N . L W
GAAGATGAGTCCAAATTCATAATTTTTGGCTCGGACGGACGGAATTATGTGCGG 866
AGC.....A..G..G....AT..G..C..AA.T..T..TATTC.G.GGA.T..A
TCT..C..AAG...G....ATTTG..C..GAGT..T...AATTCC.GG..A..T
ACT....G...A...A....GC.A.....TGGAACT..TTCACTAC.GTATT.A
ATC
E D E S K F I I F G S D G R N Y V R
S . . . . . N M . . T . . I Q W I .
S . . . . . N L . . . . . H S W . .
T . G . . I M . . . G T . S L Q Y *
CGACAGTCCAATACTGAGCTGAATCCCAAAAACCTAAAGGCAACAGTGAAGCAC 920
..T.CCATTGGCT.CAG.TATGC...AC.GT...A.TGT.....T.....A..T
..T.CTGTTGGCT..AG.TACTC...A..GT.T.A.TGCC....C..T.....T
....CTC.A..C..G...TATC.C.....C...C.GT.AAG..TT.C..T...
R Q S N T E L N P K N L K A T V K H
. P I G S R Y A . Q Y Q C P . . . .
. P V G S R Y S . . Y Q C P . . . .
. P P . . . . H . . H P V K . F N .
GGCGA-GGAAGTGTCATGGTATGGGCATGTATCTCCGCAGCCAGCGTC-GGAAA 972
..A.GT...TC...G.....T....G...CT.....ACA.TTCTA.G-...CC
..A.GT..G..C.....G....GG..CT..A..AGCA.TTC.A.G-GGCCC
..T.GACCT.AAA.....T...T.T.TTTATAATG.TA.GA.TCAT
G . G S V M V W A C I S A A S V G N
. . . . . G . F . D T . M . P
. G . . . . . G . F T S T . M . P
. . P K I . . . . . F F Y N G M V M
TTTGGTGTGTATTGAAACAACAACGGACAGGAATGTGGACCTCAGTATATTAAA 1026
A...AA.A.A..C.TTGG...C.T...TC.AT.....T..GAAGAC..CC.GG.
AC.AAG.A.A..CC...GC.TT.T...TC.TTT.CAAT..GAA.AC..C..TG.
GC.ATG.AT...GT.TGGT.TT.TA...CAA..C.CAT.TG.A.A....C.T.G
ATT
L V C I E T T T D R N V D L S I L K
. K . . V G . M . . Y . Y E D . . E
. R R . Q S I M . . F Q Y E N . F E
. W I M Y G I I . Q . A Y V . . . S
ACT
GGAAAATTTACTCCAAAGTGCCGAGAAGCTAGGAATCCGACGTTTCCGGTTCTA 1083
.A.C.CAA.GAGA.C.TGG..AAGAGCAAATTTGGG....TCG.GGGT....C.
AACT.CAA.G.GA.CCT.G..ACTTC.AAAT.TGGG...TG.C...GT...TC.
T..TGTC...T.GTC.TA.T.T..AT.AAATAT.CC.TT.AAA.GGACA...C.
E N L L Q S A E K L G I R R F R F Y
N T M R P W . R A N L G . S W V . Q
T T M R P W . L Q N V G . G . V . Q
D V I L . Y S E * N . P L K W T . Q
GACAACAACC
CCAGGACAACGACCAGAAGCATAAGTCCGGATTAGTACCGTCCTGGCTTATCTG 1147
AC.....T....C.....CT..GG.TCAT..CG.CAAT...T.C.GAC.
GA....T.....T.CT.....CT..TCTTCAT..G.GT..A...T..CAAC.
A.....T..T..T.....A.GC.GA.GTAA..CG.CTAA.AATA..T.C.C.CA
Q D N D Q K H K S G L V P S W L I W
. . . . P . . T . . H . A N . F R R
. . . . P . . T . L H . R . . F Q R
. . . . . R R C K S A K N R F T Q
GAACTGCCCCCACATGATAATTTAACCGGCCCAGTCTCCAGATGTAAATGTTAT 1201
TCG.C.TGTGA..C.CC..GAA.GG..AAGT..A.....CT.G...CCC..
TCGTCATGTG..TT..C.CGA..GG..AAGT.....G..CT.G...CCA..
A..TA.GATAG.TGCA..GCCGTGG.AA..A.CAC..T.CC..T....CCCG..
N C P H M I I *P A Q S P D V N V I
R R V N L L E W . S . . . . L . P .
R H V . L L D W P S . . . . L . P .

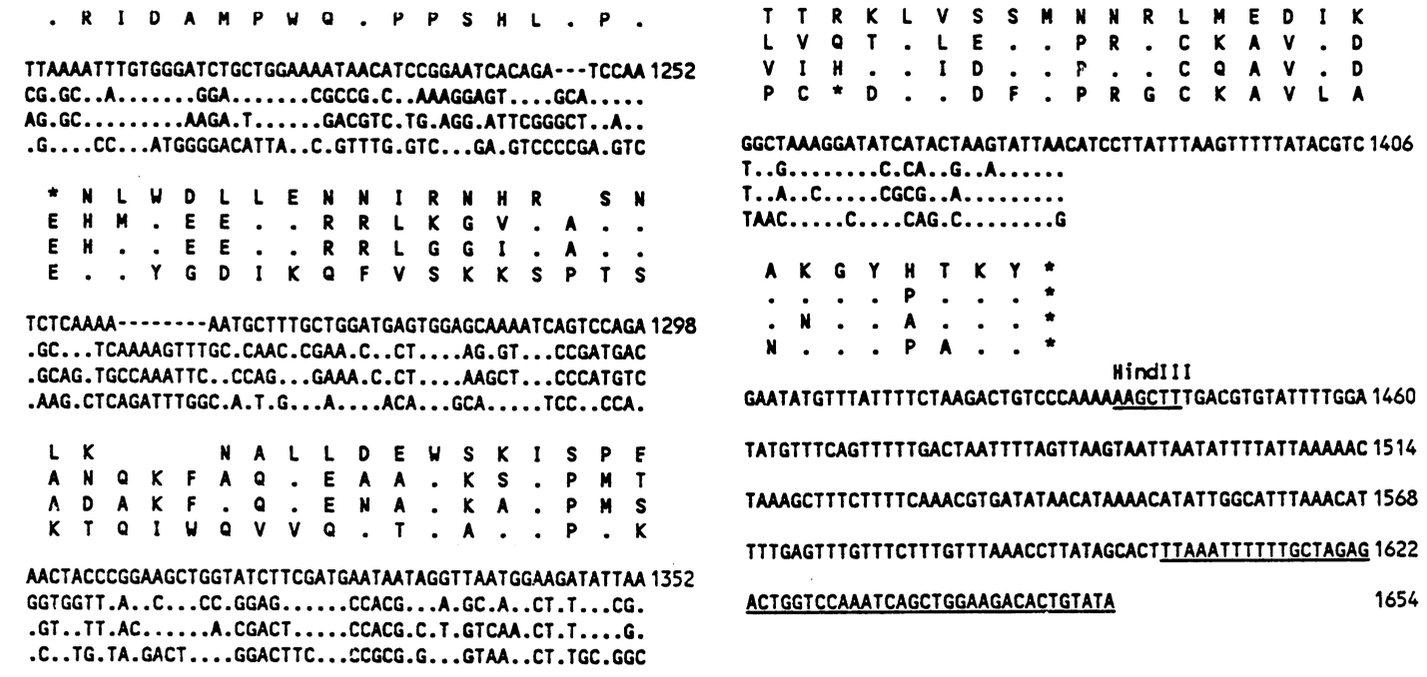


Figure 2. The complete DNA sequence of Uhu aligned with Barney, Tc1, and Hb1. Translations of all 4 elements are shown for the region of identity. Uhu ITR's are bold faced and underlined, a possible Methionine start codon in the Uhu sequence, as well as the HIII and SALI sights, are labeled and bold faced, stop codons are indicated with (*), sequence identity is indicated with (.) Deletions made in order to align these sequences are indicated above the sequences and insertions are indicated as (−).

aligns with the Tc1 open reading frame (Fig. 2). No significant sequence similarity is found between Uhu and either HB1 or Barney outside of the regions which align with the Tc1 open reading frame. When aligned as described, all of these elements have a stop codon at the same position, which corresponds to the end of the Tc1 open reading frame.

The region of identity among all of these elements is positioned similarly relative to their ITR's. The distances between the left ITR and the conserved region in Uhu, Tc1, Barney and HB1 are; 509 bp, 469 bp, 429 bp and, 489 bp respectively. The distances between the right ITR and the common stop codon in Uhu, Tc1, Barney and HB1 are; 225 bp, 213 bp, 185 bp and 272 bp respectively.

Uhu ITR's Are Related To ITR's From Other Transposable Elements

The termini of Uhu are imperfect 46−50 bp inverse repeats with 85% nucleotide sequence identity to each other (Fig. 3). As mentioned above, the sequences found at the ends of the ITR's (TATA) are both direct and inverse repeats of each other. For this reason we can not determine if these TATA tetranucleotides are part of the inverted repeats or, if they represent the original genomic site of insertion which was duplicated during transposition. The 5 bp sequences, CAGTG and CACTG, found at the 5'and 3' ends of the Uhu ITR's, respectively, appear in the termini of Tc1, and at the inside ends of the HB ITR's (Fig. 3). Four of the five terminal bp of the Barney ITR's are also identical to this sequence. Two copies of this 5 bp sequence are found in the 5' and 3' Uhu ITR's (Fig. 3).

Restriction Maps

Restriction maps of Uhu-1, Uhu-2, Uhu-3 and Uhu-4 are presented in Fig. 1. These data, in combination with the known DNA sequence of Uhu-1, indicate that Uhu may be conserved within the *D. heteroneura* genome. The DNA sequence of Uhu-1 confirms that the major portion of the element is contained within a 924 bp *Hind*III/*Sal*I fragment; the 5' end of the element extends 496 bp upstream of the *Sal*I site, and the 3' end of the element extends 234 bp beyond the *Hind*III site (Fig. 2). As shown in Fig. 1, all four isolates of Uhu have a *Hind*III/*Sal*I fragment approximately 900 bp in length containing a significant portion of the sequences reacting to the Uhu-1 probes. Restriction fragments flanking the *Hind*III/*Sal*I fragment of Uhu-2, 3 and 4 also hybridize, although less intensly, to the Uhu-1 probes and thus appear to contain portions of the elements. Relative band intensities on southern blots of Uhu-2, 3 and 4, probed with Uhu-1 probes, are consistent with the sequence data of Uhu-1, indicating that the element may be conserved (data not shown).

DISCUSSION

We have demonstrated that Uhu is a newly discovered member of a class of transposable elements previously identified in *Caenorhabditis elegans*, *Caenorhabditis briggsae*, and *Drosophila melanogaster* (12,38). These elements are approximately 1600 bp long; share significant sequence identity within a region encoding a 273 amino acid sequence; have similar positioning of this region relative to the 5' and 3' ends of the elements; have a stop codon at the same relative position at the end of the region of similarity; have terminal inverted repeats with an identical 4−7 bp sequence at or near the ends; and have the dinucleotide TA at both ends of the element.

The presence of related transposable elements in such distantly related phyla might reflect their presence in common ancestral genomes, horizontal transmission or, convergent evolution. If these elements evolved from a common ancestral sequence maintained in their host genomes during evolution, Uhu would be expected to be more closely related to the *D. melanogaster* sequences than to the *Caenorhabditis* sequences. To the contrary, Uhu is more closely related to the two *Caenorhabditis* elements Tc1 and Barney than to the *Drosophila melanogaster* element HB1. This contrast suggests that the elements may have been

A) UHU ITR'S

```
left    tataCAGTGTCTTACAGCTCAACTGGACCAGTGCCTAGCAAAAATTTTAA
                    *     * **          **           *
right   atatGTCACAGAAGGTCGACTAAACCTGGTCAGAGATCGTTTTTTAAATT
```

B)

UHU	Left ITR	tataCAGTGTCTTACAGCTCAACTGG
		CAGTGCCTAGCAAAAATTTTAA
TC1	Left ITR	taCAGTGCTGGCCAAAAAGATATC
TC3	Left ITR	taCAGTGTGGGAAAGTTCTATA
HB1	Left ITR	TAGCAGTGC}
HB2	Left ITR	TAGCAGTGC}
HB3	Left ITR	TAGCAGTGC}
HB4	Left ITR	TAGCAGTGC}
BARNEY	Left ITR	taCAGTACTGGCCATAAAGAATGC

Figure 3. Comparison of Uhu ITR's with ITR's of other transposons in this class of elements. (A) Left and Right ITR's from Uhu. (*) indicate noncomplimentary positions, sequences which are identical to those found in ITR's from other elements are bold faced and underlined. The tetranucleotides TATA found at the ends of the ITR may represent the duplicated genomic target site,and are indicated in small lettering. (B) Comparison of ITR's from 8 transposable elements. Termini are indicated by (/), inside ends of ITR's are indicated by (}), identical sequences are bold faced and underlined. The precise termini of Uhu have not been determined.

horizontally transmitted between phyla. However, the HB elements may be inactive remnants of an active element as has been suggested (14). If this is true then the HB elements are not constrained by the same selective forces and would diverge more rapidly than active elements in this class. This would account for the high degree of variability among the HB elements as well as the contrasting relationship observed between HB and the other elements. Also, only four elements have been identified in this class and it is possible that there are additional elements which support an alternative hypothesis. Another widely distributed class of transposable elements, the retroposons, found in *D. melanogaster* (6), Yeast (7), Bacteria (2), and *Dictyostelium* (8), are thought to have been transmitted horizontally by viral particles since they resemble retroviral proviruses (39). However, the Tc1-like elements have no apparent sequence or structural similarity to the retroposons (11,12,13,15,28,). Any similarity between Tc1-like elements and other types of viruses, which would suggest a viral mode of transmission, has not been recognized.

Although these elements are structurally related, it is not clear that they are all capable of autonomous transposition. It has been shown that the *Caenorhabditis* element Tc1 is active, but it is thought that the Drosophila HB elements are no longer active. Comparisons of *in-situ* hybridization of Uhu to closely related species of Hawaiian *Drosophila* reveal variability in chromosomal location and genomic copy number indicating that Uhu is actively transposing, or has been active within the 5×10^6 years of speciation of the Hawaiian *Drosophila* (10). Restriction maps of four independent isolates of Uhu indicate that it may be conserved within the *D. heteroneura* genome as would be expected for a functional element, though small differences may not be detected by this method. Comparison of Uhu with Tc1 reveals that, although these elements share significant sequence identity within the Tc1 open reading frame, there are several small differences which would result in vastly different proteins being expressed. All of these differences involve small insertions, small deletions, or point mutations, which do not effect the restriction map in Fig. 1. Perhaps Uhu-1 is a nonautonomous mutant element which depends on expression of transposase by an autonomous element. This autonomous element which would most likely be the parental element of Uhu-1, could still have the same restriction map as Uhu-1 in spite of small sequence differences. It is also possible that Uhu-1 depends on an a distinct autonomous element which does not hybridize with Uhu-1. Other families of transposable elements such as the P elements, harbor nonautonomous mutants which are activated by transposase encoding autonomous elements; perhaps this is also true for the Tc1-like elements. It has been suggested that mutator activity in *Caenorhabditis* is encoded by autonomous Tc1 elements which are capable of activating nonautonmous elements in trans. A Tc1 variant was sequenced with a stop codon within the large open reading frame (40), which may be a nonautonomous element.

Many transposable elements are flanked by short direct repeats which are duplicated genomic target sites (41,42) thought to result from staggered cuts occuring during insertion of the elements. All of the Tc1 like elements have the TA dinucleotide at their termini which are both direct and inverse repeats. Therefore it is difficult to determine if these TA sequences represent duplicated genomic target sites or if they are part of the ITR's. In *Caenorhabditis* the *unc-22* gene which affects muscle development is a favored site for Tc1 mutagenesis and was cloned by Tc1 transposon tagging (43). Comparison of the wild type *unc-22* and unc-54 genes with mutant gene sequences containing Tc1, showed that these elements always insert at TA sequences

(19,17). Either the TA dinucleotide is duplicated during transposition or one copy of the TA dinucleotide is actually part Tc1 and no duplication occurs. Sequences of additional Uhu elements may help to determine the nature of the TATA sequence at the termini of Uhu. If Uhu does not always insert specifically at TATA sequences, and does generate target site duplications, then the sequence of additional elements will show other duplicated nucleotides and allow us to determine the precise ends of Uhu. If all of the Uhu elements sequenced are flanked by the TATA tetranucleotide then we cannot ascertain from sequence data if this tetranucleotide is actually part of the Uhu element which does not generate target site duplications upon insertion or, if Uhu inserts specifically at the sequence TATA which is duplicated during this process.

The structural similarities found among the TC1-like elements: sequence similarity of their ITR's; TA sequences which are likely to be duplicated target sites; a conserved open reading frame, and conserved size, suggest a functional relationship among these elements. It is likely that these elements are activated by identical or related transposase enzymes which recognize genomic TA sequences as target sites, and require the sequences CAGTG in the 5′ ITR and CACTG in the 3′ITR, in order to function. This is supported by the relationship between the *Caenorhabditis* elements Tc1 and Tc3 (a 2.5 kb transposable element from *C. elegans* with ITR's of at least 70 bp (44). These two elements exibit similar transpositional activity although their sequences are apparently unrelated (44) outside of the identical 5 bp terminal sequences (CAGTG at the 5′ends and CACTG at the 3′ ends) (Fig. 3). Both TC1 and TC3 insert specifically at TA sequences which are duplicated during this process (11,17,18,19,21), and are activated in the *Caenorhabditis* 'mutator' strain TR679 but remain inactive in the wild type Bristol strain (44). There is evidence suggesting that ITR's are the sites of transposase activity in other elements (3,45), and the above evidence suggests that this is also true for the Tc1-like elements. It is possible that this transposase is encoded by these elements although no evidence for *in vitro* expression has been reported: (1) All of these elements contain an open reading frame with significant sequence identity which is likely to have been functionally conserved. (2) Codon usage in the Uhu open reading frame is consistent with known *Drosophila* proteins. (3) Possible regulatory signals required for expression have been identified for the Tc1 large open reading frame. (4) A polypeptide from the large open reading frame of TC1 has been expressed in E. coli (46).

ACKNOWLEDGEMENTS

We thank Rob Rowan for helpful advice and discussion. This work was supported by NSF Grant BSR85-21810, and RCMI Grant RR03061.

REFERENCES

1. McClintock, B. (1951) Cold Spring Harbor Symp. Quant. Biol., **16**,13−47.
2. Kleckner, N. (1981) Ann. Rev. Genet., **15**,341−404.
3. Federoff, N. V. (1989) In Berg, D., and Howe, M. (ed) Mobile Genetic Elements. American Society for Microbiology, Washington,D.C.
4. Spradling, A. C., and Rubin, G. M. (1981) Ann Rev. Genet., 15,219−264.
5. Cameron, J. R., Loh, E. Y., and, Davis, R. W. (1979) Cell, **16**,739−751.
6. Mount, S., and Rubin, G. M. (1985) Mol. Cell. Biol., 1630−1638.
7. Boek, J., Styles, C. A., and Fink, G. R. (1985) Cell, **40**, 491−500.
8. Capello, J., Handelsman, K., and Lodish, H (1985) Cell, **43**, 105−115.
9. Bishop, J. G. B. III, and Hunt, J. A. (1988) Mol. Biol. Evol., **5**,415−431.
10. Hunt, J. A., Bishop III, J. G., and Carson, H. L. (1984) Proc. Natl. Acad. Sci. USA, **81**,7146−7150.
11. Rosenzweig, B., Liao, L. W. and Hirsh, D. (1983) Nucl. Acids Res., **11**,4201−4209.
12. Harris, L. J., Baillie, D. L. and Rose, A. M. (1988) Nucl. Acids Res., **16**,5991−5999.
13. Harris, L. J.,Prasad, S. and Rose, A. M.(1989) (submitted to J.Mol.Evol.).
14. Brierly, H. L.,and Potter, S. S. (1985) Nucl. Acids Res., **13**,485−500.
15. Liao, L. W., Rosenzweig, B.,and Hirsh, D. (1983) Proc. Natl. Acad. Sci. USA, **80**,3585−3589.
16. Herman, H.K., and Shaw, J.E. (1987) Trends in Genetics,**3**,222−225.
17. Eide, D., and Anderson,P. (1988) Mol. Cell Biol., **8**,737−746.
18. Eide, D., and Anderson, P. (1985) Genetics, **109**,67−79.
19. Mori, I., Benian, G. M., Moerman, D. G., and Waterston, R. H. (1988) Proc. Natl Acad. Sci. USA, **85**,861−864.
20. Eide, D., and Anderson, P. (1985) Proc. Natl. Acad. Sci. USA, **82**,1756−1760.
21. Ruan, K. and Emmons, S. W. (1987) Nucl. Acids Res,. **15**,6875−6881.
22. Rose, A. M. and Snutch, T. P. (1984) Nature, **311**,485−486.
23. Ruan, K., and Emmons, S. W. (1984) Proc. Natl. Acad. Sci. USA, **81**,4018−4022.
24. Harris, L. J., and Rose, A. M. (1986) Mol. Cell. Biol., **6**,1782−1786.
25. Emmons, S. W. and Yesner, L. (1984) Cell, **36**,599−605.
26. Moerman, D. G., and Waterston, R. H. (1984) Genetics **108**,859−877.
27. Collins, J., Saari, B.and Anderson, P. (1987) Nature, **328**,726−728.
28. Mori, I., Moerman, D. G., Waterston, R. H. (1988) Genetics, **120**,397−407.
29. Maniatis, T., Fritch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbour. N. Y.
30. Southern, E. M. (1980) Meth. Enz., **69**,152.
31. Southern, E. M. (1975) J. Mol. Biol., **98**,503−517.
32. Messing, J. (1983) Meth. Enz., **101**,20−79.
33. Mead, D. A., and Kemp, B. (1986) Chimeric single stranded DNA phage-plasmid cloning vectors. In: A survey of molecular cloning vestors and their uses. Butterworth publishers, Massachusetts.
34. Birnboim, H. C., and Doly, J. (1979) Nuc. Acid. Res. **7**,1513−1523.
35. Zinder, N. D., and Boeke, J. D. (1982) Gene, **19**,1−10.
36. Sanger, F., Nicklen, S.,and Coulson, A. R. (1977) Proc Natl Acad. Sci. USA, **74**,5463−5467.
37. Dale, R. M. K., McClure, B. A.,and Houchins, J. P. (1985) Plasmid, **13**,31−40.
38. Henikoff, S., and Plasterk, R. H. A. (1988) Nucl. Acids Res.,**16**,6234.
39. Flavell, A. J. (1984) Nature, **310**,514−516.
40. Plasterk, R. H. A. (1987) Nucl. Acids Res., **15**,10050.
41. Calos, M. P., Johnsrud, L., and Miller,J. H. (1978) Cell, **13**, 411−418.
42. Grindley, N. D. (1978) Cell, **13**,419−426.
43. Moerman, D. G., Benian, G. M., and Waterston, R. H. (1986) Proc. Natl. Acad. Sci. USA, **83**,2579−2583.
44. Collins, J., Forbes, E., and Anderson, P. (1989) Genetics, 121, 47−55.
45. Engels, W. R. (1989) In Berg, D., and Howe, M. (ed) Mobile Genetic Elements. American Society of Microbiology, Washington, D.C. p 437.
46. Moerman, D. G., and Waterston, R. H. (1989) In Berg, D., and Howe, M. (ed) Mobile Genetic Elements. American Society for Microbiology, Washington, D.C p 537.