Order Number 9107020

Molecular and evolutionary characterization of the transposable element Uhu from Hawaiian *Drosophila*

Brezinsky, Laura, Ph.D.

University of Hawaii, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

MOLECULAR AND EVOLUTIONARY CHARACTERIZATION OF THE TRANSPOSABLE ELEMENT UHU FROM HAWAIIAN DROSOPHILA

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOMEDICAL SCIENCES(GENETICS)

AUGUST 1990

By

Laura Brezinsky

Dissertation Committee

Terrence Lyttle, Chairperson
Tom Humphreys
Rebecca Cann
Stephen R. Palumbi
Ken Kaneshiro
Hampton L. Carson

ABSTRACT

I report the complete nucleotide sequence of three independent isolates of the transposable element Uhu from Drosophila heteroneura (an endemic Hawaiian Drosophila). The complete element is about 1650 base-pairs (bp) long, has 46-50 bp inverse imperfect repeats at its ends, and contains a large open reading frame potentially encoding a 251 amino acid protein. The three randomly selected isolates of Uhu which I sequenced share 93.3% nucleotide sequence identity, indicating that Uhu is well conserved within the D. heteroneura genome. I 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.

Uhu has been found in five species of endemic Hawaiian Drosophila; D. heteroneura, D. silvestris, D. differens, D. planitibia, and D. picticornis. Because the phylogenetic relationship and approximate divergence times for these

species are known, I was able to examine the evolutionary dynamics of Uhu. I have sequenced a 447 bp fragment from a total of 12 Uhu elements from these species. These data provides evidence for the degeneration of Uhu in D. picticornis indicating that it is being lost in this species. The nucleotide substitution rate of Uhu is estimated to be 1.02×10^{-8} substitutions/position/year which is comparable to the estimated rate for structural genes in these flies. Phylogenetic analysis using distance matrix methods confirms previously proposed phylogenetic models of these Drosophila species indicating that Uhu has been distributed among these species by vertical transmission.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Genetic transposition is the movement of DNA sequences from one locus to another. Elements which are capable of transposing within the genome (transposable elements) 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 nearby the sites of insertion (McClintock 1951). Since that time, mobile genetic elements of various structural types have been found in a wide variety of organisms, both prokaryotic and eukaryotic.

A repetitive element has been detected in five species of Hawaiian Drosophila in the planitibia subgroup including; D. heteroneura, D. silvestris, D. planitibia, D. differens, and D. picticornis. In situ chromosomal hybridizations comparing the five species indicates that this element, which has been named Uhu, is transposable and ranges in copy number from 10 to 150 per genome. Uhu is approximately 1650 base-pairs long and is related to the transposable elements Tc1 from Caenorhabditis elegans, Barney from Caenorhabditis briggsae and the HB family of elements found in Drosophila

melanogaster. The basis for this relationship is described in chapter 2 which is a published paper (Brezinsky et al 1990). I am the first author on this paper and was responsible for the writing, all of the laboratory research, and the majority of the analysis. Gordon Wang, the second author, was responsible for the initial computer analysis that established the relationship between Uhu and the other Tc1-like elements. The third author, Tom Humphreys, is my current thesis advisor, and the final author, John Hunt, supported a substantial portion of this project and was previously my thesis advisor.

There has been a great deal of speculation as to the function of transposable elements. It has been shown repeatedly that transposable elements induce mutations, promote genome rearrangements, and, in bacteria, can facilitate the dissemination of genetic information. These processes are obviously important in the evolution of organisms. Rose and Doolittle (1983) propose three mechanisms of speciation affected by transposable elements. In each case, transposable element activity creates mutually incompatible genotypes, and hence provides a selective advantage for assortative mating. The first, termed genetic disease, is exemplified by P element-associated hybrid dysgenesis in Drosophila melanogaster. This model proposes that an element does not ordinarily transpose at a high rate. Transposition activity is greatly increased, however, when the

element is introduced into a new genetic background. This would occur when different populations interbreed, resulting in dysgenic progeny. The second mechanism is termed genome incompatibility. Transposition results in chromosomal alterations which exibit reduced fitness in heterozygotes, e.g., chromosome inversions and translocations. The third mechanism, genome resetting, requires that transposon DNA be functional. For instance, transposons might include regulatory elements which influence developmental pathways. It has been proposed that transposable genetic elements have no evolutionary or phenotypic function other than self perpetuation, thus the term selfish DNA (Orgell and Crick 1980, Doolittle and Sapienza 1980). It has been argued that transposable elements are unlikely to have arisen or been maintained by selection for an evolutionary function for two reasons. First, some of the effects of transposition are destructive and would therefore be selected against, and second, the evolutionary process is not based on maintenance of structures which may be useful in the future, but on selection of phenotypic traits which confer selective advantage in the present. Doolittle and Sapienza (1980) maintain: "When a given DNA or class of DNAs of unproven phenotypic function can be shown to have evolved a strategy (such as transposition) which insures its genomic survival, then no other explanation for its existence is necessary."

Several different classes of transposable elements have been identified from both eukaryotic and prokaryotic organisms. I have compiled a brief description of some of the eukaryotic structural types with an emphasis on Drosophila elements.

Retroposons: This broad class of structurally homologous transposable elements, is represented in Drosophila melanogaster by several different elements including Copia. The retroposons are usually long (5-9 kb), have long terminal (direct) repeats (LTR's) of 200-500 bp, and are found repeated in the genome between 5 and 100 times, (Mount and Rubin 1985). The combined evidence supports the suggestion that retroposons may be evolutionarily related to the vertebrate retroviruses, and transpose via RNA intermediates which reverse transcribe into DNA prior to integration. The complete nucleotide sequence of Copia has been determined (5146 nt) and analyzed with respect to protein coding potential (Mount and Rubin 1985). This analysis revealed homology to several retroviral proteins including reverse transcriptase, nucleic acid binding protein, a protease, and a DNA polymerase. Extrachromosomal DNA forms (linear and circular) of Copia-like elements have been isolated from cultured Drosophila cells and embryos (Flavell and Ish-Horowicz 1981). These molecules contain one or two terminal repeats and resemble unintegrated viral DNA suggesting that they may be transposition intermediates

(Flavell and Ish-Horowicz 1981, Flavell 1984). Virus-like particles (VLPs) with RNA sequences homologous to copia, and having a reverse transcriptase activity, have been detected in cultured Drosophila cells (Shiba and Saigo 1983).

Retroposons have been seen in other organisms such as TY1 of Yeast (Boeke 1985), the TN9 class of bacterial transposons (Kleckner 1981), Dirs-1 in Dictyostelium (Zuker 1981),and IAP in rodents, suggesting that this class of transposable element may be universal.

Foldback (FB): Another class of transposable elements found in D. melanogaster are the foldback (FB) elements. FB elements were named for their tendency to fold back on themselves and form duplexes when denatured and reannealed. Unlike the retroposons, these elements are highly variable with respect to total length, as well as length of the inverted terminal repeats, which consist of a complex structure involving contiguous 31 bp tandem repeats interspersed with a repeated 10 bp sequence. FB elements generate a 9 bp duplication of the target site during insertion into the genome and have a copy number of approximately 30 per haploid genome (Truett et al 1981). Evidence supporting the transposability of the FB elements comes from chromosomal in situ hybridization studies comparing various strains of D. melanogaster (Truett et al 1981).

The HB family of elements: This highly variable family of elements were first discovered as a loop sequences in the foldback elements (Brierley and Potter 1985). HB elements have short inverse terminal repeats and are of variable size due to the presence of large deletions. In-situ hybridizations of HB1 to genomic southern blots of several closely related strains of D. melanogaster yield similar banding patterns indicating that they are no longer actively transposing. It is thought that these elements are inactive remnants of a transposable element which are accumulating mutations thus accounting for the high degree of variability observed.

Tc1: Tc1 is a highly conserved 1610 bp transposable element (Rosenzweig et al 1983) with 54 bp perfect inverse repeats found in all Caenorhabditis elegans strains examined. Tc1 inserts at TA dinucleotide target sites which are duplicated during transposition. This element is active in somatic tissues of most strains of Caenorhabditis, although germline transposition has been observed only in the Bergerac strains (Liao et al 1983). Extrachromosomal forms of Tc1 have been identified in the Bergerac strain of C. elegans which may be transposition intermediates (Rose and Snutch 1984 , Ruan and Emmons 1984) but the mechanism of transposition is not known.

Tc3: This 2.5 kb element with inverse terminal repeats found

in C. elegans is unrelated to Tc1 except for 8 of the 9 terminal nucleotides which are identical (Collins et.al. 1989). Tc3 is activated in C. elegans mutator strains similarly to Tc1 and it has been suggested that these elements are substrates for the same transposase functions.

P elements: Represented by a single known heterogeneous family in the D. melanogaster genome, P elements are found at approximately 30-50 sites per haploid genome. The P elements are associated with the phenomenon of P-M hybrid dysgenesis resulting in elevated rates of sterility, male recombination, mutability, and other germ line abnormalities. These effects are seen in F1 progeny resulting from the cross between P-strain males whose genomes harbour P elements, and M-strain females whose genomes lack P elements. The prototype 2.9 kb element (Rubin et al 1982) codes for its own transposition and complements the non-transposable phenotype of some mutant (defective) P elements (O'Hare and Rubin 1983). Unlike other transposable elements, P elements do not have long terminal repeats; instead, they have short (31bp) inverted terminal repeats. P elements generate an 8 bp duplication of the target site upon insertion (O'Hare and Rubin 83). It is unlikely that the P elements transpose via a RNA intermediate because they do not structurally resemble the retroviruses and no full length transcripts have been found. P elements have also been shown to undergo precise excision from the genome

(Rubin et al 1982). This sort of activity is not indicative of an element which undergoes replicative transformation via a RNA intermediate.

I factors: Another system of hybrid dysgenesis found in D. melanogaster, the I-R system, functions similarly to the P-M system. Progeny of crosses between R-strain (reactive) females and I-strain (inducer) males exhibit hybrid dysgenesis manifested by: elevated mutation rates, nondisjunction, and sterility of female hybrids. The I-R system of hybrid dysgenesis is not associated with male recombination or male sterility resulting in gonadal dysgenesis as in the P-M system. Between 10 and 15 copies of the complete 5.4 kb I factor are present (per haploid genome) in inducer strains (Bucheton, et al 1984). I factors found in reactive strains are inactive.

pDv elements: Recently, a new class of a transposable elements was identified in D. virilis which occupies about 200 sites in the genome, and appears to be mobilized by interspecific hybridization, (Zelentsova et al 1986). Three closely related species were examined and it was noted that the copy number decreases with increased phylogenetic distance. pDv is composed of 36 bp tandemly arranged repeats, flanked by imperfect direct terminal repeats of 80 bp, and 4 bp terminal repeats.

The second part of this thesis involves the evolutionary analysis of Uhu and is found in chapter 3. The endemic Hawaiian Drosophila are an outstanding example of adaptive radiation, comprising approximately 25% of all the identified Drosophila species in the world. It is likely that these species are the descendants of a single ancestral species, and perhaps a single introduction. The current theory is that gravid females were transported from their origin by way of the jet stream, and landed here. Based on polytene chromosome banding patterns, behavioral studies, and studies of internal morphology, it has been suggested that the Hawaiian Drosophila originated in East Asia (Carson and Yoon 1982).

The unique geological history of the Hawaiian islands as well as the extreme isolation of these islands presents an ideal situation for studying the process of evolution. The Hawaiian Islands are the most geographically isolated group of islands in the world, with the nearest continent (the California coast) being 3845 km away. This island chain has been formed as the Pacific tectonic plate migrated over a stationary hot spot (Dalrymple et al 1973). Molten lava was (and still is) spewed from the earth's interior in the form of volcanic eruptions, thereby forming new islands. As the plate moves northwesterly, new islands are formed, and older islands erode below sea level. The resulting chain of islands exhibits a temporal relationship, with the newer islands on the southeasterly end, and the older islands situated to the

north. Presently, the Hawaiian chain is represented by six major islands: Kauai, Oahu, Molokai, Maui, Lanai, and Hawaii. Potassium-argon dating has been used to estimate the ages of these islands (McDougall 1969). The youngest and most southerly island of Hawaii is less than one million years old. Kauai, the northernmost and oldest of the six major islands, is approximately 5.6 million years old. Hawaiian Drosophila are found on all of these islands. It is suggested that the founder(s) of extant species arrived on one of the older islands (Carson and Yoon 1982). Its descendants migrated to the newer islands as they were formed.

Several phylogenetic models for Hawaiian Drosophila have been constructed. One problem which normally arises when constructing such models is that of directionality. It may be possible to group species based on certain characteristics, but it is impossible to determine from the cytological information alone, which group is ancestral. For instance, identification of certain polytene chromosome inversions can be used to construct relationships between species, yet it is impossible to determine which sequence orientation was the original. In the Hawaiian islands, geologic and geographic information makes it possible to confer a reasonable directionality to the scheme. It is suggested that, in most cases, these flies have speciated by islands. This would have occurred by founder events resulting in isolation and divergence of populations. The result is that most of the

islands have a distinct group of species of Drosophila. It becomes possible to date such founder events based on the ages of the islands where the species are found.

Construction of the following phylogenetic relationships was based largely on banding patterns of polytene chromosomes, metaphase chromosome karyotypes, and DNA sequence (Carson and Kaneshiro 1976, Kaneshiro 1976, Sene and Carson 1977, Craddock and Johnson 1979, Speith 1981, Carson and Yoon 1982, Hunt and Carson 1983, DeSalle and Giddings 1986, Rowan and Hunt 1990). D. heteroneura and D. silvestris, found on Hawaii island, are the most recently diverged. Because these species are unique to Hawaii island, they probably diverged within the last 1 million years. Although they differ in appearance, they are capable of hybridizing and producing viable offspring in the laboratory, as well as in the wild (Carson et al 1989). D. planitibia from Maui, and D. differens from Molokai appear to have diverged from a common ancestor within the past 2 million years. Mitochondrial sequence data suggest that D. planitibia and D. differens may have hybridized at some time during their evolutionary history (DeSalle and Giddings 1986). Such an event could have taken place during the Pleistocene when sea level fluctuations resulted in a land bridge between Maui, Molokai, and Lanai. It is suggested that an ancestor common to these two species served as the founder for D. silvestris and D. heteroneura. D. picticornis, found on Kauai, represents the oldest, and most distantly related of these

species. This species has been included as a close relative of the planitibia subgroup based on the presence of four chromosomal inversions which are found in all members of the subgroup. It is suggested that D. picticornis diverged from an ancestor common to the other members of the planitibia subgroup. This divergence would most likely have occurred 4-5 million years ago.

Very little is known about the evolutionary dynamics of transposable elements because most transposable elements are found only in a single species. The discovery of this transposable element in the Hawaiian Drosophila provides a unique opportunity to examine its evolution. It has been suggested that the Tc1-like elements may have been distributed by independent horizontal transmission similarly to the retrotransposons (Harris et al 1989, Harris et al. 1988). However, my data suggest that Uhu has been transmitted vertically among these Hawaiian Drosophila; No evidence for the transmission of Uhu, independent of the host genome, was observed. These data do provide evidence for the degeneracy of Uhu in D. picticornis which is considered to be the oldest species in this study. The Uhu elements which I examined from D. picticornis appear to be diverging rapidly, presumably due to loss of function.

CHAPTER 2

THE TRANSPOSABLE ELEMENT UHU FROM HAWAIIAN DROSOPHILA- A 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

Published in Nucleic Acids Research, volume 18, number 8

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 its 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 ITRs 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 ITRs 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).

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 Charon4 vector from a volume of adult flies from these stocks as described previously (9).

Cloning

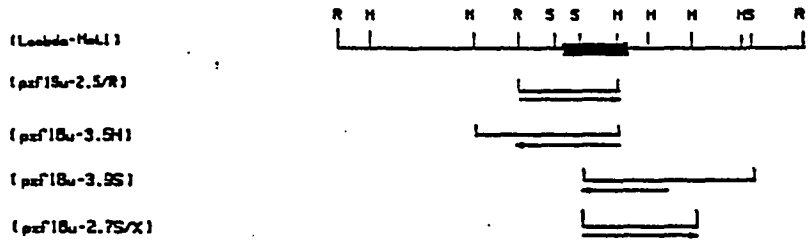
Uhu-1 was recognized as a repetitive D. heteroneura sequence in an ADH containing clone (9,9A). Two subclones (pBR322-3.5H and pBR322-3.9S) containing overlapping restriction fragments yielded repetitive banding patterns (9,9A,10) when reacted to D. heteroneura genomic southern blots. Restriction fragments 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; Lambda-Het2, Lambda-Het3, and Lambda-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 polyethylene 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 amplifications. phage were amplified as plate lysates, and purified by PEG precipitation followed by banding on cesium chloride equilibrium gradients (29,35). DNA was subsequently purified by proteinase K digestion and phenol/chloroform extraction (29).

A)



B)

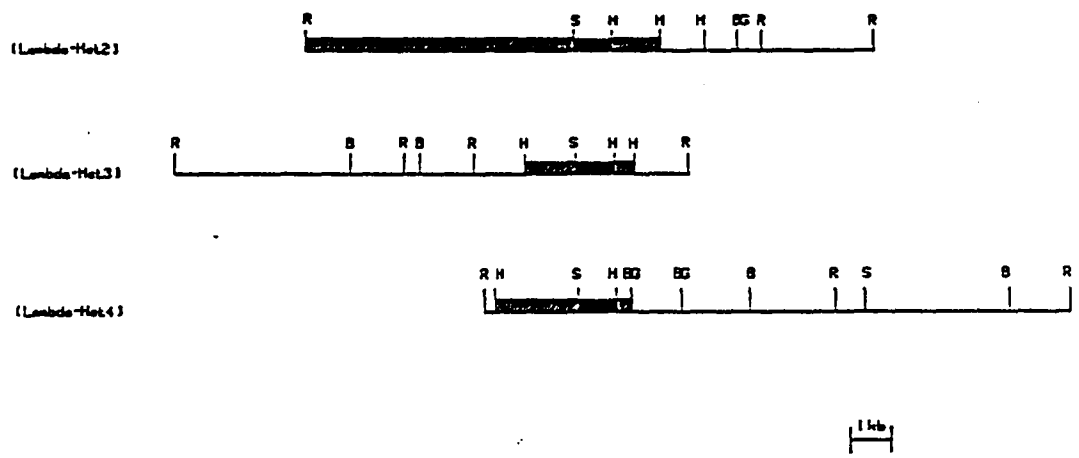


FIGURE 1. Restriction maps of D. heteroneura DNA containing four independent isolates of Uhu. Restriction enzymes used are; R, EcoRI; H,HindII; S,SalI; BG BglII; B, BamHI. A) Restriction map of the Lambda-Het1 insert containing Uhu-1. The boxed region is Uhu-1 as determined by DNA sequence data. Restriction fragments subcloned into pzf vectors are indicated below the map. Arrows indicate sequencing strategy. B) Restriction maps of Lambda-Het2, Lambda-Het3, and Lambda-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.

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 6x SSC, 5x 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 $5x10^5$cpm/ml of probe. Blots were washed at 68°C; once in 5x SSC, twice in 1x SSC and twice in 0.5x 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 Lambda-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 (Figure III and Figure V). 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 2 gaps of 3 and 8 nucleotides at positions 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

Figure II 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 ITRs 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 (-).

```
TATACAGTGTCTTACAGCTCAACTGGACCAGTGCCTAGCAAAAATTTTAATTGCCT  56

GCCATAAACTAATTATCCATTATTTTTCAAAAATTCCAAAGACCGATGGCAGGT  110

ACATATATTAACCACCAAAATGAATATATGATCCCAATAAACTGGGGTTTCCCA  164

CCTGCTAGGTCGGGTTATGTAAAAAAGTACCTTAATTTATGGTTACATATTATT  218

TGGACCAGCGGCGTTATGGACACCTGGGTGCCATAAAACCCGGATTTTTACGTC  272

AGGTTGATTATTTTCGGTATAAATAGACCAATCCTTCGTAGTCAGTTTAGTTAT  326

ATCCTGCATCTCGGGTGCAACCAGCCAACAAGGCATATGGGCAAGCGGACTACC  380

ATTGAACAACGGAAACTGATCCTGGAACATTTCAAGATTGGATATTCATATCGC  434
          MET
CAAATAGCTAAAATGGTAAATCTAAGTACCACAACTGTATTCAACATCATTCGG  488
       SALI
CGCTTCGTCGACGAAAATCGGATAGAGGACAAGGGCAGAAAGGCACCAAACAAG  542

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..
AGT.........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  .  .  .  .  .  N  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  .
```

```
GGCGAGGGAAGTGTCATGGTATGGGCATGTATCTCCGCAGCCAGCGTC-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  O  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  .
  .  R  I  D  A  M  P  W  Q  .  P  P  S  H  L  .  P  .

TTAAAATTTGTGGGATCTGCTGGAAAATAACATCCGGAATCACAGA---TCCAA  1252
CG.GC..A.......GGA.......CGCCG.C..AAAGGAGT....GCA.....
AG.GC.........AAGA.T......GACGTC.TG.AGG.ATTCGGGCT..A..
.G....CC...ATGGGGACATTA..C.GTTTG.GTC...GA.GTCCCCGA.GTC

  *  N  L  W  D  L  L  E  N  N  I  R  N  H  R     S  N
  E  H  M  .  E  E  .  .  R  R  L  K  G  V  .  A  .  .
  E  H  .  .  E  E  .  .  R  R  L  G  G  I  .  A  .  .
  E  .  .  Y  G  D  I  K  Q  F  V  S  K  K  S  P  T  S

TCTCAAAA--------AATGCTTTGCTGGATGAGTGGAGCAAAATCAGTCCAGA  1298
.GC...TCAAAAGTTTGC.CAAC.CGAA.C..CT....AG.GT...CCGATGAC
.GCAG.TGCCAAATTC..CCAG...GAAA.C.CT....AAGCT...CCCATGTC
.AAG.CTCAGATTTGGC.A.T.G...A....ACA...GCA.....TCC..CCA.

  L  K              N  A  L  L  D  E  W  S  K  I  S  P  E
  A  N  Q  K  F  A  Q  .  E  A  A  .  K  S  .  P  M  T
  A  D  A  K  F  .  Q  .  E  N  A  .  K  A  .  P  M  S
  K  T  Q  I  W  Q  V  V  Q  .  T  .  A  .  .  P  .  K

AACTACCCGGAAGCTGGTATCTTCGATGAATAATAGGTTAATGGAAGATATTAA  1352
GGTGGTT.A..C...CC.GGAG......CCACG...A.GC.A..CT.T...CG.
.GT..TT.AC......A.CGACT.....CCACG.C.T.GTCAA.CT.T....G.
.C..TG.TA.GACT....GGACTTC...CCGCG.G...GTAA..CT.TGC.GGC

  T  T  R  K  L  V  S  S  M  N  N  R  L  M  E  D  I  K
  L  V  Q  T  .  L  E  .  .  P  R  .  C  K  A  V  .  D
  V  I  H  .  .  I  D  .  .  P  .  .  C  Q  A  V  .  D
  P  C  *  D  .  .  D  F  .  P  R  G  C  K  A  V  L  A
```

```
GGCTAAAGGATATCATACTAAGTATTAACATCCTTATTTAAGTTTTTATACGTC  1406
T..G.........C.CA..G..A.......
T..A..C.....CGCG..A..........
TAAC.....C....CAG.C.........G

  A  K  G  Y  H  T  K  Y  *
  .  .  .  .  P  .  .  .  *

  .  N  .  .  A  .  .  .  *
  N  .  .  .  P  A  .  .  *
```

HindIII

```
GAATATGTTTATTTTCTAAGACTGTCCCAAAAAAGCTTTGACGTGTATTTTGGA  1460

TATGTTTCAGTTTTTGACTAATTTTAGTTAAGTAATTAATATTTTATTAAAAAC  1514

TAAAGCTTTCTTTTCAAACGTGATATAACATAAAACATATTGGCATTTAAACAT  1568

TTTGAGTTTGTTTCTTTGTTTAAACCTTATAGCACTTTAAATTTTTTGCTAGAG  1622

ACTGGTCCAAATCAGCTGGAAGACACTGTATA                        1654
```

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 10 bp insertion at position 1100 results in a frameshift relative to Tc1. An additional frame shift occurs at position 1261 as a result of the 8 bp deletion in the Uhu sequence relative to Tc1. An frame stop codon at position 1170 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 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 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 ITRs. 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 ITRs Are Related To ITRs 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 ITRs (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 ITRs, respectively, appear in the termini of Tc1,

and at the inside ends of the HB ITRs (Fig. 3). Four of the five terminal bp of the Barney ITRs are also identical to this sequence. Two copies of this 5 bp sequence are found in the 5' and 3' Uhu ITRs (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 HindIII/SalI fragment; the 5' end of the element extends 496 bp upstream of the SalI site, and the 3' end of the element extends 234 bp beyond the HindIII site (Fig. 2). As shown in Fig. 1, all four isolates of Uhu have a HindIII/SalI fragment approximately 900 bp in length containing a significant portion of the sequences reacting to the Uhu-1 probes. Restriction fragments flanking the HindIII/SalI 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).

Figure III Comparison of ITRs from Tc1-like elements. (A) Left and Right ITRs from Uhu-1, Uhu-3, and Uhu-4. (*) indicate noncomplimentary positions. Sequences that are conserved among the Uhu ITRs 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 the left of ITRs from 10 transposable elements. Inside ends of ITRs are indicated by (}), identical sequences are bold faced and underlined.

A) Uhu Inverse Terminal Repeats

UHU-1

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

UHU-3

```
left     tataCAGTGTCTCACAGCTCAACTGGAACAGTGCCTAGCAAAAAATTTAA
                   *                *          * *
right    atatGTCACAGAATGTCGAGTTGACCTTGTCGCGGATCGTTTATAAAATT
```

UHU-4

```
left     tataCAGTGTCTCGCAGCGTATTTGGACCAGTGTCTAGCAAAAAATTTAA
                              *
right    atatGTCACAGAGCGTCGCATAACCCTGGTCACAGATCGTTTTTTAAATT
```

B) Comparison of Left Inverse Terminal Repeats

```
UHU-1    tataCAGTGTCTTACAGCTCAACTGG
           ACCAGTGCCTAGCAAAAATTTTAA
UHU-3    tataCAGTGTCTCACAGCTCAACTGG
           AACAGTGCCTAGCAAAAAATTTAA
UHU-4    tataCAGTGTCTCGCAGCGTATTTGG
           ACCAGTGTCTAGCAAAAAATTTAA
TC1        taCAGTGCTGGCCAAAAAGATATC
TC3        taCAGTGTGGGAAAGTTCTATA
HB1       TAGCAGTGC)
HB2       TAGCAGTGC)
HB3       TAGCAGTGC)
HB4       TAGCAGTGC)
BARNEY     taCAGTACTGGCCATAAAGAATGC
```

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 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 $5x10^6$ 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 ITRs. 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 ITRs; 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 ITRs 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 ITRs 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.

9A. Bishop, J. G. B. III, 1982 Evolution of the alcohol dehydrogenase gene region in the Hawaiian Drosophila. doctoral thesis, University of Hawaii, Department of Genetics

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.

CHAPTER 3

EVOLUTION OF THE WELL CONSERVED TRANSPOSABLE ELEMENT UHU IN FIVE SPECIES OF HAWAIIAN *DROSOPHILA*

ABSTRACT

The complete DNA sequence of three independent isolates of Uhu, a member of the Tc1 like class of transposable elements from *D. heteroneura* (Uhu-1, Uhu-3, and Uhu-4) is presented. These isolates exibit 93.3% nucleotide sequence identity indicating that Uhu is well conserved within this species. A study of the complete DNA sequences of Uhu-1, Uhu-3, and Uhu-4, indicates that the nucleotide substitution rate for Uhu is about 1.02×10^{-8} substitutions/nucleotide/year, which is comparable to the rate for structural genes in these species. Uhu has been identified in four other species of endemic Hawaiian *Drosophila*, *D. silvestris*, *D. differens*, *D. planitibia*, and *D. picticornis*, and 9 Uhu elements were isolated from genomic libraries of these four species. A region of Uhu with well conserved ends and a nucleotide substitution rate representative of the Uhu genome as a whole, was chosen for sequence comparison of elements from different species. Using the polymerase chain reaction, a homologous DNA fragment ranging in size from 264-401 bp, was sequenced from each of these isolates. Analysis of these sequence data

agrees with previously proposed phylogenetic relationships of the host species, indicating that Uhu has been distributed among these Hawaiian Drosophila by vertical transmission. These data also show extensive nucleotide substitutions in Uhu of D. picticornis indicating that it is degenerating and being lost in this species.

INTRODUCTION

Uhu is a repeat sequence found in five species of endemic Hawaiian Drosophila (D. heteroneura, D. silvestris, D. planitibia, D. differens, and D. picticornis) (Hunt, Bishop, and Carson 1984, Bishop and Hunt 1988). I previously demonstrated that Uhu is member of the Tc1-like class of transposable elements (Brezinsky et al. 1990) which includes Tc1 (Liao et al. 1983) and Barney (Harris et al. 1988, Harris et al. 1989,) from Caenorhabditis, and the HB family of elements (Brierly and Potter 1985) from D. melanogaster. The presence of these related elements in such distantly related organisms raises the question of how they were distributed. Evidence has not been available to distinguish wether they have been transmitted horizontally between host genomes similarly to the retroposons, or, if they were distributed from a common ancestral genome and conserved during evolution Harris et al. 1988).

The Hawaiian islands were formed sequentially as the Pacific tectonic plate passed northwesterly across a volcanic hot spot, resulting in a temporally related chain of islands with the oldest island found at the northwest end of the archipelago. Estimated ages of the high islands based on potassium-argon dating are; Kauai (5.6 Myr), Oahu (3.5 Myr), Molokai (1.5 Myr), Maui (1.0 Myr), and Hawaii (0.4 Myr)

(McDougall 1969). The endemic Hawaiian Drosophila represented by about 800 extant species (Carson and Yoon 1982) are thought to have been founded by a single gravid female which arrived on one of the older islands. Its descendants migrated to newly emerged islands and founded new species. The result is a phylogeny which can be given directionality and approximate divergence dates. Phylogenetic models that have been constructed for the Hawaiian Drosophila using morphology, cytology, electrophoresis, and DNA sequence analysis (Kaneshiro 1976, Carson and Kaneshiro 1976, Sene and Carson 1977, Craddock and Johnson 1979, Speith 1981, Carson and Yoon 1982, Hunt and Carson 1983, DeSalle and Giddings 1986, Rowan and Hunt 1990) are in general agreement with the theory that the Hawaiian Drosophila were founded sequentially by island. Very little is known about the evolution of transposable elements due to the fact that most elements are identified in only one species. The presence of Uhu in these five species of Hawaiian Drosophila has allowed me to examine this issue.

METHODS AND MATERIALS

Drosophila stocks

Drosophila stocks were maintained in the lab from wild caught flies as described previously (Bishop and Hunt 1988).

Genomic libraries

Genomic libraries were constructed in the Lambda CharonIV vector from volumes of adult flies as described (Bishop and Hunt 1988).

Cloning

Lambda clones containing independent Uhu elements from *D. heteroneura* (Het1, Het3,and Het4), *D. silvestris* (Silv2 and Silv3), *D. differens* (Diff1 and Diff3), *D. planitibia* (Plan2, Plan3, and Plan4), and *D. picticornis* (Pict1 and Pict4), were isolated from genomic libraries by plaque hybridization to pBR322-3.5H, a clone containing the major portion of Uhu-1 (Bishop and Hunt 1988). Uhu-1 was subcloned and sequenced as described previously (Brezinsky et al 1990). Uhu-3, and Uhu-4 from *D. heteroneura* were subcloned into pzf chimeric plasmid vectors as indicated in figure IV.

DNA Amplification

Lambda phage and plasmid DNA amplification and purification has been described previously (Brezinsky et al 1990).

Polymerase Chain Reaction

Based on the sequences of Uhu-1, Uhu-3, and Uhu-4, I designed synthetic oligonucleotides specific for well conserved regions of Uhu (figure V). These oligonucleotides which define a 447 bp fragment in the three D. heteroneura Uhu elements were used as primers for Polymerase Chain Reaction (PCR) amplifications and subsequent DNA sequencing. DNA from the Lambda clones: Silv2, Silv3, Diff1, Diff3, Plan2, Plan3, Plan4, Pict1, and Pict4 were used as templates for asymetric PCR amplifications which produce predominantly single-stranded DNA (Maniatis et al. 1989) (see appendix). Both strands of each double stranded template were amplified and subsequently sequenced. The GeneAmp kit from Perkin Elmer Cetus was used for all PCR amplifications which were carried out as specified by the supplier; Between 10 and 20 ng of template was used with 100 pmoles of the limiting primer and 1 pmole of the nonlimiting primer, primers were annealed at 50°C, ramp time was 2', and reactions were carried out for 35 cycles followed by an extension at 72°C for 7' to insure that all nacant fragments were completed. The PCR products were purified and concentrated to about 20 ul with centricon 30 filters (Amicon

Corp) and analyzed by agarose gel electrophoresis prior to sequencing (see appendix).

DNA Sequencing

Uhu-1 was sequenced as described previously. Uhu-3 and Uhu-4 were sequenced in both directions using the dideoxy sequencing method (Sanger et al 1977) in conjunction with the Sequenase kit (USB), ^{35}S-dATP was used to label all sequencing reactions. Serial deletions of pzf parental clones were constructed using the Dale deletion method (Dale et al 1985). Gaps in the resulting sequences were filled by using synthetic oligonucleotides as sequencing primers. Single stranded PCR amplification products were also sequenced using the dideoxy sequencing method with the Sequenase kit. Reactions were carried out under standard conditions for single stranded sequencing. Approximately 1/3 of the product from each PCR amplification was used per sequencing reaction. Sequencing reactions were electrophoresed on denaturing 7% polyacrylamide gels and visualized by autoradiography.

Data Analysis

Sequences were aligned using PC Gene. Pairwise analysis was carried out using sequencer 2.0 by Bailey D. Kessing. This program calculates corrected pairwise percent difference based on the algorithms of Wu and Li 1985, and Kimura 1980 which correct for reversals, convergence, and accounts for

transition:transversion bias for DNA sequence data. Phylogenetic trees were constructed using FITCH and KITCH (distance matrix methods) from the PHYLIP 3.2 phylogeny inference package by Joseph Felsenstein 1986, 1989, and PAUP (maximum parsimony analysis) by David L Swafford 1989. Corrected percent nucleotide difference values found in the matrix in table I were used to construct FITCH and KITCH trees. All deletions were considered as single events.

Nucleotide Substitution Rate

Nucleotide substitution rate per year was calculated using the corrected percent difference (K) (Table I) such that the substitution rate k = K/2T where T is the time since divergence (Kimura 1980) and assuming a divergence time between D. melanogaster and the Hawaiian Drosophila of $4x10^7$ years (Beverly and Wilson 1984, Rowan and Hunt 1990).

RESULTS

Sequences of Uhu-1, Uhu-3, and Uhu-4

The DNA sequence of three independent isolates of Uhu (Uhu-1, Uhu-3, and Uhu-4) from D. heteroneura are presented in figure V. The three elements exibit 93.3% nucleotide sequence identity,and are apparently the same size (1656 bp including the terminal TATA tetranucleotides). Previously I found that by inserting 2 gaps and 2 deletions in the Uhu-1 sequence, it could be aligned with the entire region of a putative open reading frame the nematode transposable element, Tc1 (Brezinsky et al. 1990). When aligned in this way, Uhu and Tc1 share a 46% nucleotide sequence identity, and a 40% amino acid sequence identity in this region. The requirement of two frame-shifts in Uhu relative to TC1 indicates that Uhu is mutant and possibly nonfunctional. The Uhu-3 and Uhu-4 sequences require only one of these frame shifts, a 10 bp insertion which is a duplication within the Uhu sequence at position 1096 (figure V). The duplicated sequence extends 11 bp, but since the homologous reading frame is only 10 base pairs shifted from Tc1, it is possible that the duplication event only involved a 10 bp sequence, and the nucleotide found at one end of the repeated sequence was actually part of the original genomic sequence. The following additional differences are seen between Tc1 and the Uhu elements, within

the region aligning with the Tc1 open reading frame: All three isolates of Uhu have a 3 bp insertion at position 1072 and a 3 bp deletion at position 1247. Uhu-3 and Uhu-4 have a 9 bp deletion and Uhu-1 has an 8 bp deletion at position 1261. Uhu-1 has 2 stop codons that are in frame with the TC1 alignment at positions 1170 and 1203, Uhu-3 and Uhu-4 have only one of these stop codons at position 1203 (figure VI).

Nucleotide sequence identity varies in different regions of the Uhu genome. Among Uhu-1, Uhu-3, and Uhu-4, there are two start codons prior to the region of identity with Tc1 that are in frame with the Tc1 alignment at positions 364 and 448 (figure V). I have identified a possible TATA box at position 291 (see below). Based on this information, I divided the Uhu genome into four regions and calculated the nucleotide divergence for each of these regions: the left and right inverse terminal repeats (ITRs) are 18% and 26% divergent respectively, the region between the left ITR and the TATA box (51-291) is 8.3% divergent, the region between the putative TATA box and the end of the region which aligns with the Tc1 open reading frame (292-1381) is 3.7% divergent, and the region between the end of the conserved regio and the right ITR (1382-1605) is 11.6% divergent. Comparison of the left ITR with the right ITR within each element reveals that the ITRs within Uhu-1 are 14% divergent, within Uhu-3 are 8% divergent, and within Uhu-4 are 2% divergent.

Transcriptional signals

A potential TATA box at position 291 (figure V), is separated from the start codon by 73 bp, which is typical of eukaryotic structural genes (Breathrach and Chambon 1981, Nakajima et al 1988). A potential transcriptional initiation site (pyrimidine, A, pyrimidine, pyrimidine, pyrimidine, pyrimidine) at position 328 is separated from the TATA box by 31 bp which is also typical of eukaryotic genes (Breathrach and Chambon 1981, Nakajima et al 1988). A possible polyadenylation site (Proudfoot and Brownlee 1976, Hagenbuchle et al. 1980, Breathrach and Chambon 1981) represented by the sequence ATTAAA is found at position 1508.

Coding potential

All three isolates of Uhu have a well conserved open reading frame which potentially encodes a 251 amino acid protein begining with a start codon at position 364, and ending with an in-frame stop codon at position 1117 (figure V). Overlapping with the open reading frame is a region (position 559 to 1381) that shows significant sequence identity with the entire Tc1 open reading frame. However, the 10 bp insertion within this region, at position 1102, results in a frame-shift relative to Tc1; This frame-shift is followed by the in-frame stop codon at position 1117. The region from the possible TATA box to the end of the region aligning with

the Tc1 open reading frame (292-1381) is highly conserved (96.3%) relative to the rest of the Uhu genome (87%).

Comparison of Uhu from 5 species of Hawaiian Drosophila

Synthetic oligonucleotides that define a 401 bp fragment of Uhu-1 between position 646 and 1046 (figure-V) were used as primers for PCR amplifications to sequence this region DNA from the following additional Uhu elements, Silv2 and Silv3 from D. silvestris, Diff1 and Diff3 from D. differens , Plan2, Plan3, and Plan4 from D. planitibia, and Pict1 and Pict4 from D. picticornis. All sequences were nonambiguous when determined in both directions, indicating that there was no contamination or incorrect priming during the PCR amplifications. These sequences along with the homologous sequences from the three D. heteroneura Uhu elements were aligned and compared as shown in figure VII. A matrix of the corrected percent differences between these homologous sequences is presented in table I. The nucleotide sequence divergence within species is as follows: D. picticornis, 17.6%, D. silvestris, 8.4%, D. heteroneura, 3.7%, D. planitibia, 0.5%, and D. differens, 1%. There are three large deletions (two of which result in frame shifts) in the two D. picticornis elements including, a 21 bp deletion, a 22 bp deletion, and a 181 bp deletion (figure VII). There are two additional deletions among these data, both of which are in frame: a 12 bp deletion in Plan3 and a 6 bp deletion in Silv3.

Both Silv3 and Pict1 have a 4 bp insertion which also results in a frame shift. However, the insertion in Silv3 is a tandem duplication whereas, the insertion in Pict1 is unrelated to the surrounding sequence.

These data indicate that Uhu is well conserved in all of these species except D. picticornis. The D. picticornis sequences exhibit significantly greater average within-species divergence than any of the other species in this study. They also have large deletions which are not found elsewhere.

Maximum Parsimony

The sequence data presented in figure VII were subjected to a maximum parsimony analysis using PAUP; The results are presented in figure VIII and IX. There is one shortest tree consisting of 564 steps (single nucleotide changes) and 13 trees that are one step longer. The cladogram in figure VIII is the strict consensus of these 14 trees which has been constructed with Tc1 indicated as the outgroup. The phylogram which gives relative branch lengths for the same tree is found in figure IX.

The separation of the two Pict sequences (Pict1 and Pict4) and the placement of the Pict4 sequence with to the D. planitibia/D. differens divergence is problematic. Previous data indicate that D. picticornis is a distant relative of D. differens and D. planitibia and probably shares a common ancestor with the other 4 species in this study. However,

there are significant evolutionary rate differences among different lineages as explained below. For taxa where significant rate variation exists among different lineages, trees constructed by cluster analysis with unconstrained branch lengths are more likely to reflect the true phylogeny than maximum parsimony analysis (Felsenstein 1988). On this basis, anomalous placement of pict4 is not surprising since the D. picticornis sequences are the most divergent among this data. Other than the anomolous placement of Pict4, the results are in agreement with previously proposed phylogenetic models for these species.

Having determined that Pict1 is an out-group relative to the other Uhu sequences, I subjected the Uhu sequence data to a maximum parsimony analysis, with Pict1 specified as the out-group. The results of this analysis are presented in figure X. There are 2 shortest trees which consists of 168 steps. Overall, this analysis, in contrast to the previous trees, indicates that SILV3 diverged from a common ancester of the divergence of the D. heteroneura/D. silvestris group and the D. planitibia/D. differens group. Perhaps the common ancestor was an autonomous Uhu element. Interestingly, the 2 trees are different from each other on only one point, the placement of PICT4. One tree places this sequence ancestral to the 4 recently diverged species, which is consistant with the following FITCH analysis of this data; The other tree places PICT4 with the D. planitibia/D. differens group, which is

consistant with the previous maximum parsimony analysis. As explained above, the two D. picticornis sequences are highly divergent and maximum parsimony analysis has difficulty dealing with data where significant rate variation exists.

Distance matrix analysis

Distance matrix methods fit a tree to a matrix of pairwise distances using least squares regression. Corrected distances for the 14 sequences involved in this study are given in table I. Phylogentic trees were constructed using FITCH and KITCH. FITCH fits unrooted trees with unconstrained branch lengths to distance matrices. KITCH fits rooted trees to distance matrices and assumes an evolutionary clock according to which the branch lengths from the root to each tip is the same. Sum of squares which, measures how well the data fit the tree, is calculated for the trees produced by both of these algorithms. Because KITCH assumes a molecular clock and FITCH does not, the residual sum of squares of these trees were compared by F test to determine if there is evolutionary rate variability among the branches of these trees (Sheldon, 1987).

The F ratio was calculated as follows:

$$\frac{(SS[KITCH] - SS[FITCH])/(df[KITCH] - df[FITCH])}{(SS[FITCH])/(df[FITCH])},$$

where SS=sum of squares and df=degrees of freedom.
df=(no. of observations)/2-(no. of branches).
SS[FITCH]=1.1782, SS[KITCH]=3.454,
v_1 =(df[KITCH]-df[FITCH])=12, v_2=df[FITCH]=66.
The calculated F ratio=10.6, the theoretical F=2.5 (p=.01). In other words; there is significant rate variation among these taxa.

The trees constructed with FITCH and KITCH are in agreement with each other. Tc1 was indicated as the outgroup for the FITCH analysis and was defined as the root by the KITCH analysis. The results of this analysis given in figure XI indicate only those branches which are significant as determined by the 95% confidence interval; In most cases, the specific relationships among the D. heteroneura and D. silvestris sequences and among the D. differens and D. planitibia sequences can not be determined from this data because the distances within species are not significantly different from the between species distances as indicated by the 95% confidence intervals (Table I). Pict4 which was anomolously placed with the D. differens/ D. planitibia divergence by maximum parsimony analysis is placed ancestral to the other four species by distance matrix analysis.

Substitution rate

The nucleotide substitution rate of Uhu which was

calculated as described above is 1.02×10^{-8} substitutions nucleotide-1 year-1. The divergence of this 447 bp fragment among Uhu-1, Uhu-3, and Uhu-4 is 5.2% which is not significantly different from the divergence of the Uhu genome as a whole (6.7%). Therefore the substitution rate which was calculated from this data is representative of the whole Uhu genome.

Transition/Transversion.

A rise in the fraction of transversions relative to transitions with divergence time has been observed in the Hawaiian Drosophila for mtDNA (DeSalle 1987) and for the ADH gene (Rowan and Hunt 1990). We observed the same trend for Uhu (figure XII). I have fitted a logarithmic curve to the data points in the graph which illustrates that the percent of transversions is reaching saturation with the most distant taxa which represent about 4×10^{7} years of divergence. The ratio of transition:transversion ranges from 0.633 (Tc1/Hawaiian Drosophila) to 1.23 (D. differens/D. planitibia).

DISCUSSION

I have shown that Uhu is a well conserved transposable element in D. heteroneura. My data indicate that Uhu is also well conserved in D. silvestris, D. planitibia, and D. differens although it is apparently divergent in the oldest species (D. picticornis). Two experiments have suggested that Uhu in D. picticornis was divergent in Uhu in D. heteroneura (Bishop 1982). 1) Signals on genomic southern blots of D. picticornis probed with Uhu were about 10% as intense as were those of the other four species in this study. 2) In-situ hybridizations revealed about 10 sites in D. picticornis as compared with about 150 sites in the other four species. The DNA sequence of homologous fragments from two D. picticornis elements suggests that the previous data were obtained because Uhu is degenerate in D. picticornis. I suggest that Uhu is, or recently has been, active in the four most recently diverged species, but it may have lost function in D. picticornis and is more divergent due to the lack of functional constraints. An analogous effect has been proposed for the highly variable HB elements. These latter elements exhibit similar genomic banding patterns among different strains of D. melanogaster indicating that they are no longer active (Brierly and Potter 1985).

Founder events resulting in genetic bottlenecks have probably occured repeatedly during the divergence of the

Hawaiian Drosophila (Carson and Yoon 1982). I suggest that Uhu has experienced increased activity during these founder events, possibly as a result of stochastic processes associated with genetic bottlenecks. Increased transpositional activity could have resulted from disruption of a balanced polymorphic state existing in the founder population. For instance, if transposition is dependent on autonomous transposable elements, and such elements are maintained at low frequencies in a balanced polymorphic state, then transposition in such a population would occur infrequently. Although these elements would accumulate mutations, complete extinction from the genome may be delayed by their ability to propagate by transposing. If the founder of a new population had an autonomous element in its genome, then the resulting population would have a much higher frequency of these elements than the founder population. The increased gene frequency would be associated with an increase in transpositional activity, which would function to further increase the genomic copy number of functional elements. When such a population reached equilibrium, the frequency of autonomous transposable element would return to its original low level due to the production of deleterious mutations. This would explain why Uhu has a high copy number of highly conserved elements in recently diverged species, and a low copy number of highly divergent elements in the oldest species. This model becomes more likely when one considers

that the founders of new species of Hawaiian Drosophila come from relatively young species populations and are therefore likely to still harbour active elements. A similar mechanism has been invoked for the spontaneous activation of P elements; P strains maintained as vial stocks with small effective population size reverted to M strains at a rate of 5%/stock/year (Bingham et al. 1982). Bingham maintains that transpositional activation was due to stochastic processes associated with bottlenecks.

Carson's model of founder-flush (Carson 1975) suggests that an increase in genetic diversity is associated with founder events (Carson and Wisotzkey 1989), and this increase in diversity allows organisms to adapt to new environments, resulting in speciation. If this is true then one could imagine that transposition which increases genetic diversity could function as a mechanism for this effect. The conventional view is that bottlenecks function to reduce genetic diversity and heterozygosity (Mayre 1963, Prakash 1972). However, mathematical models indicate that rapid population growth rate following bottlenecks may function to maintain heterozygosity and renew genetic diversity (Nei et al 1975). I suggest that activation of the Tc1-like Uhu elements in peripheral populations could function to increase genetic diversity of their hosts, thereby contributing to speciation. Recently, McDonald (1989, 1990) suggested that retroviral-like transposable elements (RLEs) may play a

significant role in macroevolution (speciation) by producing mutations effecting regulation and development. McDonald cautiously suggests that "rates of retroviral-element mediated mutations may be particularly labile in peripheral populations.". He bases this on the fact that environmental stress such as heat-shock and gamma radiation significantly increase transcription by copia (McDonald et al 1988, McDonald et al 1987), and, that peripheral populations may experience increased environmental stress. Although Tc1-like elements and RLEs are structurally unrelated, they are both transposable and may exibit similar evolutionary effects. Selection for variability has been argued against as a mechanism for maintaining transposable elements due to the likelihood of producing deleterious mutations. However, these arguments refer to the long-term maintenance of transposable elements in host genomes. Perhaps, in accordance with Carson's model, individuals with actively transposing elements would compete successfully in new environments due to increased genetic diversity.

The question has been raised as to how these elements have been distributed among such distantly related organisms. It was previously suggested that these elements may have been distributed by horizontal transmission similarly to the widely distributed retroposons and retrotransposons, which are thought to be horizontally transmitted into genomes as retroviruses (Bingham and Zachar 1989, Nocera and Sakaki

1990). However, the Tc1-like elements show no significant structural or sequence similarity to any known viruses. Furthermore, I have not found any evidence for independent horizontal transmission of the Tc1-like elements. If horizontal transmission had occured then I would expect to find discordance between the phylogeny of the transposable elements and their hosts. However, the phylogeny of Uhu is in agreement with the phylogeny of the Drosophila host species, and the phylogenetic relationship among Tc1, HB1 and Uhu is consistent with the phylogeny of their respective hosts, indicating that Uhu has been distributed among these species by vertical transmission.

The phylogenetic analysis of Uhu does indicate that it has been transmitted between species as a result of interspecific hybridization of the host flies. In most cases, these data do not allow the determination of the relationship of the sequences within or between closely related species because these sequences have not diverged sufficiently (figure XI). However, within the D. silvestris/D. heteroneura group there is one significant branch which groups SILV2 and HET1 together, indicating that they diverged from a common ancestor within this group (figure XI). However, D. silvestris and D. heteroneura, which are very closely related, are allopatric on the island of Hawaii, and interspecific hybridization is observed between them in the wild (Carson et al. 1989). The Uhu sequences from D. differens and D. planitibia are also

grouped together as indicated in figure XI. This is consistent with mitochondrial sequence data which suggests that, D. differens and D. planitibia have hybridized at some point since they diverged (DeSalle and Giddings 1986). Although these 2 species normally live on adjacent islands, DeSalle and Giddings suggest that they may have been sympatric during pleiocene when these islands were joined by a land bridge (Stearns 1966).

Within the Uhu genome there are regions with significantly different rates of divergence suggesting that there are functional constraints on certain portions of the element. The most variable region is the ITRs, but within each element, the ITRs are well conserved with respect to each other. Apparently, complementarity of the ITRs within an element is more important than the specific DNA sequence of the ITRs. This implies that their function is to recognize each other. Perhaps the ITRs within an element hybridize to form hairpin structures which act as substrates for cleavage by a transposase. There is evidence suggesting that ITRs are the sites of transposase activity in other elements (Federoff 1989, and Engels 1989); This is possible for the TC1-like elements as well. Although the ITRs are the most variable regions of Uhu, the 7 bp sequence TACAGTG and its inverse complement CACTGTA found at the 5' and 3' ends of Uhu, respectively are conserved among all of the Tc1-like elements. This sequence does not appear to be involved in

transcriptional regulation or replication (Soede et al. 1979, Soeda et al 1980, Mitchell and Tjian 1989, Murphy et al. 1989). Perhaps this short sequence is transposase binding site or regulatory sequence.

The most highly conserved region among Uhu-1, Uhu-3, and Uhu-4, (226-1380) includes the region which is conserved among all of the Tc1-like elements (558-1380), and aligns with the Tc1 large open reading frame (figure VI). However, within this conserved region, all three Uhu elements are frame-shifted relative to Tc1 due to a duplication at position 1096; This duplication is followed by an in-frame stop codon at position 1117, which is prior to the end of the conserved region. These elements have a conserved start codon at position 364, and a common stop codon in frame with the Tc1 alignment at position 1380. These data suggest that Uhu-1, Uhu-3, and Uhu4 are nonautonomous elements which share a common ancestor and may have been transposed by an autonomous element in the genome. Such an element would lack the 10 bp duplication and premature stop codons, and therefore, might express a 339 amino acid protein begining with the conserved start codon at position 364 and terminating at position 1388. The region from 261-364 which is also well conserved among the Uhu elements contains a possible TATA box and transcriptional initiation site suggesting that this region is required for correct transcriptional initiation. The Tc1-like elements are relatively small and it is unlikely that they encode proteins

that are not related to their own function. The only apparent function of these elements is transposition and therefore, this region probably encodes a transposase or transposase related protein in autonomous elements. The nature of this putative protein is unclear because it shows no significant identity at the DNA, or amino acid level, with any known sequences other than the homologous region in other Tc1-like elements.

ACKNOWLEDGEMENTS

I would like to thank Bailey Kessing for assistance with the phylogenetic analysis of these data and for helpful discussions of these results. This work was supported by RCMI grant #RR03061.

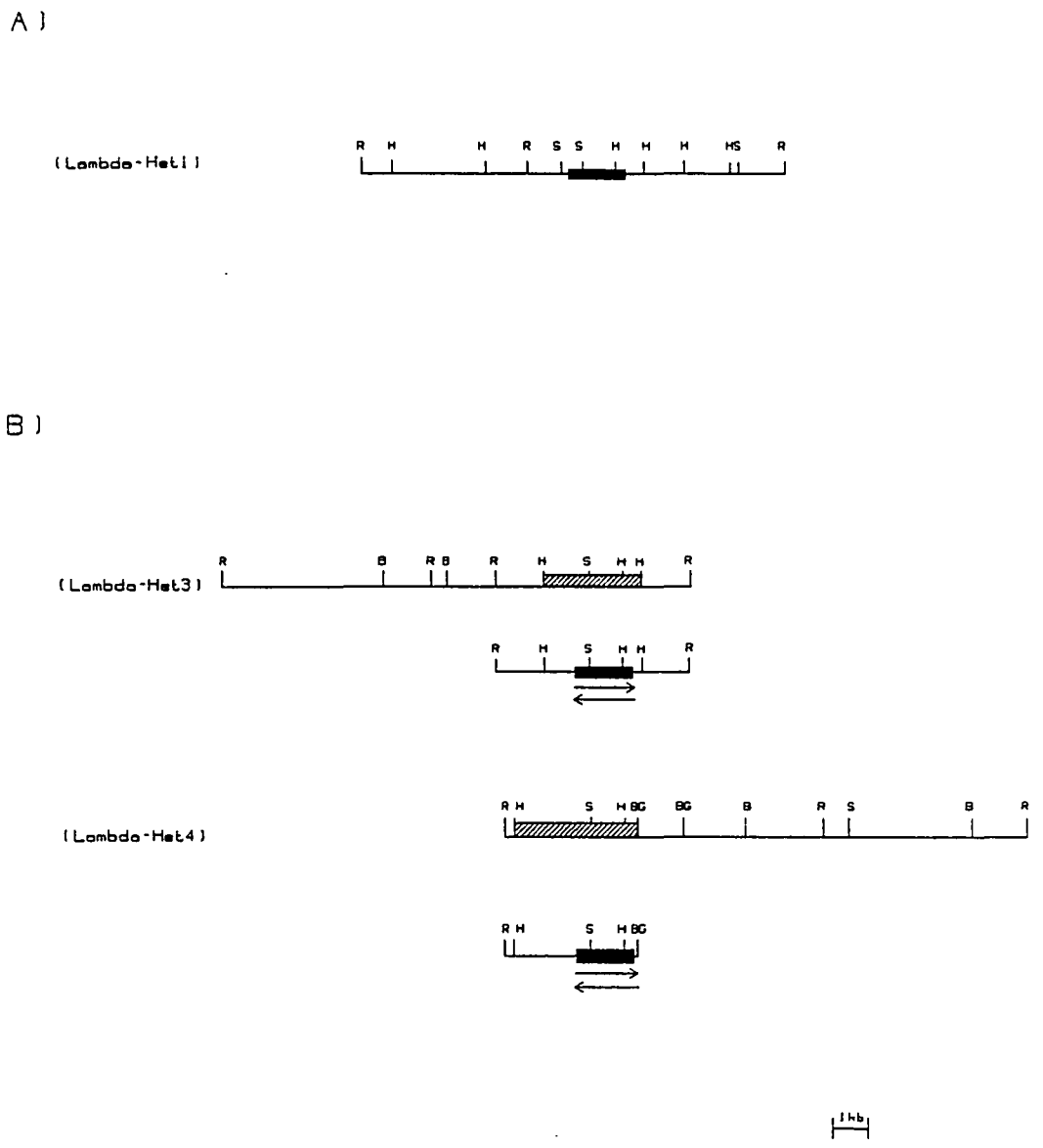


Figure IV Restriction maps of D. heteroneura DNA containing three independent Uhu isolates. Restriction enzymes are: BamHI (B), BglII(BG), EcoRI(R), HindIII(H), and SalI(S). Blackened boxed regions represent Uhu as determined by DNA sequence analysis. A) A restriction map of Lambda-HetI containing Uhu-1 B) Restriction maps of Lambda-Het3 and Lambda-Het4 containing Uhu-3 and Uhu-4 respectively. Hatched regions represent those fragments containing portions of Uhu-1 (as determined by southern blot analysis) which were subcloned in pzf vectors for sequencing. Arrows indicate sequencing strategy.

FIGURE V Comparison of the complete nucleotide sequence of three members of the Uhu family of transposable elements (Uhu-1, Uhu-2 and UHU-3). Positions identical to the Uhu1 sequence are indicated by (.). The inverse terminal repeats, possible start codons, stop codons, possible transcriptional initiation and polyadenylation sites, and duplications, are bolded and underlined, deletions are indicated by (-). The region that was PCR amplified is indicated by a perforated line (----) above the sequence. Precise locations of PCR primers are indicated.

```
                                1                                      45
UHU-1   TGTTTTATATATAAATATATACAGTGTCTTACAGCTCAACTGGACCAGTGCCTAGCAAAAAT
UHU-3   GC..CAGA..C..T...............C..............A................A
UHU-4   AT.....GTA...T...............CG....GT.TT..........T..........A

                                      80                               115
TTTAATTGCCTGCCATAAACTAATTATCCATTATTTTTCAAAAATTCCAAAGACCGATGGCAGGTACATA
.....AG..............A.......T..............................
.....

                                     150                               185
TATTAACCACCAAAATGAATATATGATCCCAATAAACTGGGGTTTCCCACCTGCTAGGTCGGGTTATGTA
.......T....T.................................T.....G......C..........
.....................................................................G....

                                     220                               255
AAAAAGTACCTTAATTTATGGTTACATATTATTTGGACCAGCGGCGTTATGGACACCTGGGTGCCATAAA
.C.....................T.............A..ATG.........................
.......................C..C.C..T.....................................

                                 TATA BOX                              325
ACCCGGATTTTTTACGTCAGGTTGATTATTTTCGGTATAAATAGACCAATCCTTCGTAGTCAGTTTAGTT
.....................................TATAAAT.........................
.....................................TATAAAT.........................

INITIATION                      360  START                             395
ATATCCTGCATCTCGGGTGCAACCAGCCAACAAGGCATATGGGCAAGCGGACTACCATTGAACAACGGAA
..A.....................................ATG.........................
..A.....................................ATG.........................

                                     430                               465
ACTGATCCTGGAACATTTCAAGATTGGATATTCATATCGCCAAATAGCTAAAATGGTAAATCTAAGTACC
T.....................................................................
T..C..........C........C.........C....................................

                                     500                               535
ACAACTGTATTCAACATCATTCGGCGCTTCGTCGACGAAAATCGGATAGAGGACAAGGGCAGAAAGGCAC
.....................................................................
.....................................................................

                                     570                               605
CAAACAAGATTTTCACCGAACAGGAGGAGCGGAGGATCATCAGGAAAATAAGGGAAAATCCCAAGCTATC
.........................A...........................................
.........................T.......G...................................

                                     640                               675
                        <   PCR PRIMER #1   >----------------------------
GGCTCCAAAACTGACTCAACAGGTGCAGGATGAAATGGGGAAAAAGTGCAGTGTGCAAACTGTGCGCCGG
.....................................................................
.....................................................................
```

```
                                  710                                   745
-------------------------------------------------------------------------
GTTCTGCACAACCATGACTTTAATGCCCGAGTACCACGGAAGAAGCCATTTATAAGCACAAAAAATAAAG
.............................A...........................................
.............................................................................

                                  780                                   815
-------------------------------------------------------------------------
GGACTAGGATGACGTTCGCCAAAACCCACTTGGACAAGGATTTGGAGTTCTGGAACACAATCATATTTGA
.....................................................G.....................
.............................................................................

                                  850                                   885
-------------------------------------------------------------------------
AGATGAGTCCAAATTCATAATTTTTGGCTCGGACGGACGGAATTATGTGCGGCGACAGTCCAATACTGAG
..................AC.......A.......................T.....................
..................C.............................T.....................

                                  920                                   956
-------------------------------------------------------------------------
CTGAATCCCAAAAACCTAAAGGCAACAGTGAAGCACGGCGGAGGAAGTGTCATGGTATGGGCATGTATCTC
........G...C................................................................
........G...C................................................................

                                  990                                  1026
-------------------------------------------------------------------------
GGCAGCCAGCGTCGGAAATTTGGTGTGTATTGAAACAACAACGGACAGGAATGTGGACCTCAGTATATTA
.......G.....................................................A......
.......G....A..G..........TC......C......T.....A......AT.T....A....C..

                                 1060                                  1096
--------------------<    PCR PRIMER #2     >                    COPY 1
AAGGAAAATTTACTCCAAAGTGCCGAGAAGCTAGGAATCCGACGTACTTTCCGGTTCTACCAGGACAACG
.............................................................................
.............................................................................

  COPY 2             STOP       1130                                  1166
ACCAGGACAACAACCAAGCATAAGTCCGGATTAGTACCGTCCTGGCTTATCTGGAACTGCCCCCACATGA
.....................................A.....................................
.....................................A.....................................

                                 1200                                  1236
TAATTTAACCGGCCCAGTCTCCAGATGTAAATGTTATTTAAAATTTGTGGGATCTGCTGGAAAATAACAT
.....CC......................................................................
.....CC......................................................................

                                 1270                                  1306
CCGGAATCACAGATCCAATCTCAAAAAATGCTTTGCTGGATGAGTGGAGCAAAATCAGTCCAGAAACTAC
...........................-..T.....................................
.................G..........-.....................................
```

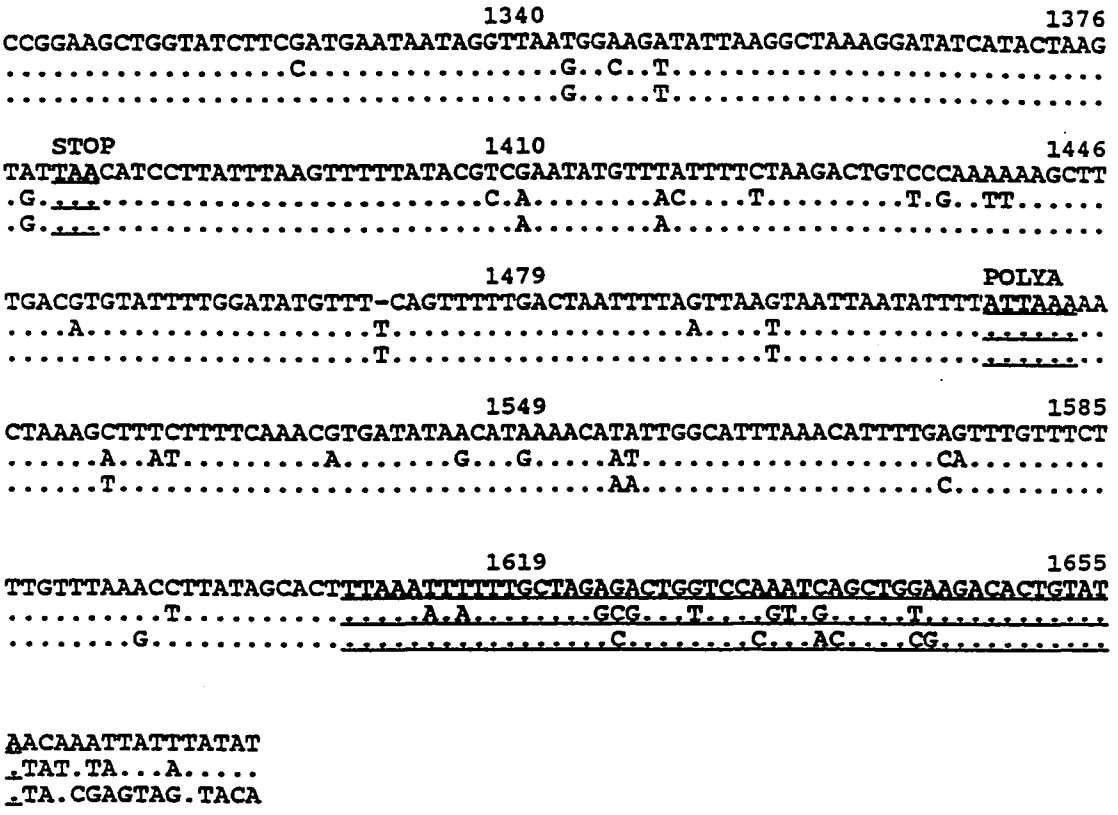

1340
1376
CCGGAAGCTGGTATCTTCGATGAATAATAGGTTAATGGAAGATATTAAGGCTAAAGGATATCATACTAAG
STOP
1410
1446
TATTAACATCCTTATTTAAGTTTTTATACGTCGAATATGTTTATTTTCTAAGACTGTCCCAAAAAAGCTT
1479
POLYA
TGACGTGTATTTTGGATATGTTT-CAGTTTTTGACTAATTTTAGTTAAGTAATTAATATTTTATTAAAAA
1549
1585
CTAAAGCTTTCTTTTCAAACGTGATATAACATAAAACATATTGGCATTTAAACATTTTGAGTTTGTTTCT
1619
1655
TTGTTTAAACCTTATAGCACTTTAAATTTTTTTGCTAGAGACTGGTCCAAATCAGCTGGAAGACACTGTAT
AACAAATTATTTATAT

Figure VI The complete DNA sequences of Uhu-1, aligned with Uhu-3, Uhu-4, Barney, Tc1, and Hb1. Translations of Uhu elements are given for the large open reading frame. Translations of all 6 elements are shown for the region of identity. Frame-shifted translations of Uhu are bolded. Uhu ITRs, and possible start codons, are bold faced and underlined, stop codons are indicated with (*), sequence identity with Uhu-1 is indicated as (.), insertions are indicated as (-).

```
UHU1 TATACAGTGTCTTACAGCTCAACTGGACCAGTGCCTAGCAAAAATTTTAATTGCCT
UHU3.............C..............A................A.....
UHU4.............CG....GT.TT..........T..........A.....

GCCATAAACTAATTATCCATTATTTTTCAAAAATTCCAAAGACCGATGGCAGGT    110
..AG..............A.......T.....................
.....................................................

ACATATATTAACCACCAAAATGAATATATGATCCCAATAAACTGGGGTTTCCCA    164
............T....T.................................T...
.....................................................

CCTGCTAGGTCGGGTTATGTAAAAAAGTACCTTAATTTATGGTTACATATTATT    218
..G......C............C.....................T.........
..............G.........................C..C.C..T..

TGGACCAGCGGCGTTATGGACACCTGGGTGCCATAAAACCCGGATTTTTACGTC    272
....A..ATG.....................................
.....................................................

AGGTTGATTATTTTCGGTATAAATAGACCAATCCTTCGTAGTCAGTTTAGTTAT    326
.....................................................
.....................................................

                                    START
ATCCTGCATCTCGGGTGCAACCAGCCAACAAGGCATATGGGCAAGCGGACTACC    380
.....................................ATG..............
.....................................ATG..............

                                      M  G  K  R  T  T
                                      .  .  .  .  .  .
                                      .  .  .  .  .  .

ATTGAACAACGGAAACTGATCCTGGAACATTTCAAGATTGGATATTCATATCGC    434
............T.....................................
............T..C..........C........C.........C.....

 I  E  Q  R  K  L  I  L  E  H  F  K  I  G  Y  S  Y  R
 .  .  .  .  N  .  .  .  .  .  .  .  .  .  .  .  .  .
 .  .  .  .  N  .  .  .  .  .  .  .  .  .  .  H  .
```

```
CAAATAGCTAAAATGGTAAATCTAAGTACCACAACTGTATTCAACATCATTCGG   488
.....................................................
.....................................................

 Q  I  A  K  M  V  N  L  S  T  T  T  V  F  N  I  I  R
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

CGCTTCGTCGACGAAAATCGGATAGAGGACAAGGGCAGAAAGGCACCAAACAAG   542
.....................................................
.....................................................

 R  F  V  D  E  N  R  I  E  D  K  G  R  K  A  P  N  K
    .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

ATTTTCACCGAACAGGAGGAGCGGAGGATCATCAGGAAAATAAGGGAAAATCCC   596
..................A..................................
.....................T.......G.........................
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

 I  F  T  E  Q  E  E  R  R  I  I  R  K  I  R  E  N  P
 .  .  .  .  .  .  K  .  .  .  .  .  .  .  .  .  .  .
 .  .  .  .  .  .  D  .  .  S  .  .  .  .  .  .  .  .
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
.....................................................A.
.....................................................
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
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  Q
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
 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..
AGT.........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
.............................................G.......
.....................................................
.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
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  V  .  .
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
 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  V  .  N  .  P  V  S  K  .  R  N  .  L  W
```

```
GAAGATGAGTCCAAATTCATAATTTTTGGCTCGGACGGACGGAATTATGTGCGG   866
.....................AC.......A.....................T..
.....................C.............................T..
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
 A  .  .  .  .  .  N  .  .  D  .  .  .  .  .  .  .  W
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  W
 S  .  .  .  .  .  N  M  .  .  T  .  .  I  Q  W  I  .
 S  .  .  .  .  .  N  L  .  .  .  .  .  N  S  W  .  .
 T  .  G  .  .  I  M  .  .  .  G  T  .  S  L  Q  Y  *

CGACAGTCCAATACTGAGCTGAATCCCAAAAACCTAAAGGCAACAGTGAAGCAC   920
.........................G...C.......................
.........................G...C.......................
..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
 .  .  .  .  .  .  .  .  .  .  H  .  .  .  .  .  .  .
 .  .  .  .  .  .  .  .  .  .  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  .

GGCGGAGGAAGTGTCATGGTATGGGCATGTATCTCCGCAGCCAGCGTC-GGAAA   972
.........................................G.....-.....
.........................................G....A-..G..
..A..T...TC...G.....T....G...CT......ACA.TTCTA.G-...CC
..A..T..G..C........G....GG..CT..A..AGCA.TTC.A.G-GGCCC
..T..ACCT.AAA.............T...T.T.TTTATAATG.TA.GA.TCAT

 G  G  G  S  V  M  V  W  A  C  I  S  A  A  S  V  G  N
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  G  .  .  .
 .  .  .  .  .  .  .  .  .  .  .  .  .  .  G  .  .  .
 .  .  .  .  .  .  .  .  G  .  F  .  D  T  .  M  .  P
 .  .  .  .  .  .  .  .  G  .  F  T  S  T  .  M  .  P
 .  .  P  K  I  .  .  .  .  .  F  F  Y  N  G  M  V  M
```

```
TTTGGTGTGTATTGAAACAACAACGGACAGGAATGTGGACCTCAGTATATTAAA  1026
.............................................A........
........TC......C......T.....A......AT.T....A....C....
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
  .  .  .  .  .  .  .  .  .  .  .  .  .  .  R  .  .  .
  .  .  F  .  .  P  .  M  .  K  .  .  Y  .  N  .  .  .
  .  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

GGAAAATTTACTCCAAAGTGCCGAGAAGCTAGGAATCCGACGTACTTTCCGGTTCTA1083
..........................................................
..........................................................
.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  T  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  O  N  V  G  .  G     .  V  .  Q
  D  V  I  L  .  Y  S  E  *  N  .  P  L  K     W  T  .  Q

CCAGGACAACGACCAGGACAACAACCAAGCATAAGTCCGGATTAGTACCGTCCTGGCTTA
..........................................................A.....
..........................................................A.....
AC.......T....C.----------.......CT..G..TCAT..CG.CAAT...T.C.
GA....T.....T.CT----------.......CT..TCTTCAT..G.GT..A...T..C
A.....T..T..T...----------..A.GC.GA.GTAA..CG.CTAA.AATA..T.C.

  Q  D  N  D  Q  D  N  N  Q  A  *
  .  .  .  .  .  .  .  .  .  .  .
  .  .  .  .  .  .  .  .  .  .  .
  .  .  .  .  P
  .  .  .  .  P
  .  .  .  .  .

                                 K  H  K  S  G  L  V  P  S  W  L  I
                                 .  .  .  .  .  .  .  Q  .  .  .  .
                                 .  .  .  .  .  .  .  Q  .  .  .  .
                                 .  .  T  .  .  H  .  A  N  .  F  R
                                 .  .  T  .  L  H  .  R  .  .  F  Q
                                 .  R  R  C  K  S  A  K  N  R  F  T
```

```
TCTGGAACTGCCCCCACATGATAATTTAACCGGCCCAGTCTCCAGATGTAAATGTTAT1201
.........................CC.............................
.........................CC.............................
GAC.TCG.C.TGTGA..C.CC..GAA.GG..AAGT..A.........CT.G...CCC..
AAC.TCGTCATGTG..TT..C.CGA..GG..AAGT.........G..CT.G...CCA..
C.CAA..TA.GATAG.TGCA..GCCGTGG.AA..A.CAC..T.CC..T....CCCG..

  W  N  C  P  H  M  I  I  *  P  A  Q  S  P  D  V  N  V  I
  .  .  .  .  .  .  .  .  P  .  .  .  .  .  .  .  .  .  .
  .  .  .  .  .  .  .  .  P  .  .  .  .  .  .  .  .  .  .
  R  R  R  V  N  L  L  E  W  .  S  .  .  .  .  L  .  P  .
  R  R  H  V  .  L  L  D  W  P  S  .  .  .  .  L  .  P  .
  Q  .  R  I  D  A  M  P  W  Q  .  P  P  S  H  L  .  P  .

TTAAAATTTGTGGGATCTGCTGGAAAATAACATCCGGAATCACAGA---TCCAA  1252
.....................................................---.....
.....................................................---...G.
CG.GC..A.......GGA.......CGCCG.C..AAAGGAGT....GCA.....
AG.GC.........AAGA.T......GACGTC.TG.AGG.ATTCGGGCT..A..
.G....CC...ATGGGGACATTA..C.GTTTG.GTC...GA.GTCCCCGA.GTC

  *  N  L  W  D  L  L  E  N  N  I  R  N  H  R     S  N
  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .     .  .
  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .     .  D
  E  H  M  .  E  E  .  .  R  R  L  K  G  V  .  A  .  .
  E  H  .  .  E  E  .  .  R  R  L  G  G  I  .  A  .  .
  E  .  .  Y  G  D  I  K  Q  F  V  S  K  K  S  P  T  S

TCTCAAAA---------AATGCTTTGCTGGATGAGTGGAGCAAAATCAGTCCAGA  1298
........---------..T.................................
........---------.....................................
.GC...TCAAAAGTTTGC.CAAC.CGAA.C..CT....AG.GT...CCGATGAC
.GCAG.TGCCAAATTC..CCAG...GAAA.C.CT....AAGCT...CCCATGTC
.AAG.CTCAGATTTGGC.A.T.G...A....ACA...GCA.....TCC..CCA.

  L  K
                    N  A  L  L  D  E  W  S  K  I  S  P  E
  .  .              .  V  .  .  .  .  .  .  .  .  .  .  .
  .  .              .  .  .  .  .  .  .  .  .  .  .  .  .
  A  N  Q  K  F  A  Q  .  E  A  A  .  K  S  .  P  M  T
  A  D  A  K  F  .  Q  .  E  N  A  .  K  A  .  P  M  S
  K  T  Q  I  W  Q  V  V  Q  .  T  .  A  .  .  P  .  K
```

```
AACTACCCGGAAGCTGGTATCTTCGATGAATAATAGGTTAATGGAAGATATTAA  1352
.......................C.................G..C..T......
.....................................................G.....T......
GGTGGTT.A..C...CC.GGAG......CCACG...A.GC.A..CT.T...CG.
.GT..TT.AC......A.CGACT.....CCACG.C.T.GTCAA.CT.T....G.
.C..TG.TA.GACT....GGACTTC...CCGCG.G...GTAA..CT.TGC.GGC

  T  T  R  K  L  V  S  S  M  N  N  R  L  M  E  D  I  K
  .  .  .  .  .  .  .  .  .  .  .  .  .  R  A  V  .  .
  .  .  .  .  .  .  .  .  .  .  .  .  .  R  .  V  .  .
  L  V  Q  T  .  L  E  .  .  P  R  .  C  K  A  V  .  D
  V  I  H  .  .  I  D  .  .  P  .  .  C  Q  A  V  .  D
  P  C  *  D  .  .  D  F  .  P  R  G  C  K  A  V  L  A

GGCTAAAGGATATCATACTAAGTATTAACATCCTTATTTAAGTTTTTATACGTC  1406
.......................G.............................C.
.......................G.............................
T..G........C.CA..G..A......
T..A..C.....CGCG..A.........
TAAC.....C....CAG.C........G

  A  K  G  Y  H  T  K  Y  *
  .  .  .  .  .  .  .  C  *
  .  .  .  .  .  .  .  C  *
  .  .  .  .  P  .  .  .  *
  .  N  .  .  A  .  .  .  *
  N  .  .  .  P  A  .  .  *

GAATATGTTTATTTTCTAAGACTGTCCCAAAAAAGCTTTGACGTGTATTTTGGA  1460
A........AC....T.........T.G..TT..........A...........
A........A.....................................

TATGTTT-CAGTTTTTGACTAATTTTAGTTAAGTAATTAATATTTTATTAAAAAC 1514
.......T.....................A....T.....................
.......T.....................T.....................

TAAAGCTTTCTTTTCAAACGTGATATAACATAAAACATATTGGCATTTAAACAT  1568
.....A..AT.........A.......G...G.....AT..............
.....T.................................A..............

TTTGAGTTTGTTTCTTTGTTTAAACCTTATAGCACTTTAAATTTTTTGCTAGAG  1622
....CA.....................T..........-----A.A--------GC
....C.....................G...........----------------C

ACTGGTCCAAATCAGCTGGAAGACACTGTATA                        1654
G...T....GT.G.....T.............
........C...AC....CG............
```

Figure VII Alignment of homologous sequences from 12 Uhu elements, HB1 from D. melanogaster and Tc1 from C. elegans. Identity to HET1 is indicated by (.), missing sequence data is indicated by (?), deletions are indicated by (-).

```
HET1    AAA AAG TGC AGT GTG CAA ACT GTG
HET3    ... ... ... ... ... ... ... ...
HET4    ... ... ... ... ... ... ... ...
SIL2    ??? ??. ... ... ..A G.. ... ...
SIL3    ... ... ... ... ..A G.. ... ...
PLAN2   ... ... ..G ... ... G.. ... ...
PLAN3   ... ... ... ... ... G.. ... ...
PLAN4   ... ... ... ... ... G.. ... ...
DIFF1   ... ... ... ... ... G.. ... ...
DIFF3   ... ... ... ... ... G.. ... ...
PICT1   ??? ??? ??? ??? ... G.. ... ...
PICT4   ... ... ... ... ... G.. ... ...
TC1     TTG GGA AT. ..C .AC GTT ... A.T
HB1     CCT GTA CCA ... AAA .G. ... ..T
```

```
HET1    CGC CGG GTT CTG CAC AAC CAT GAC TTT AAT GCC CGA GTA CCA CGG
HET3    ... ... ... ... ... ... ... ... ... ... ... .A. ... ... ...
HET4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SIL2    ... ... ... ... ... ... ... .G. ... ... ... ... ... ... ...
SIL3    ... .A. ... ... .G. ... ... ... ..C ... ... ... ... ... ...
PLAN2   T.. ... ... ... .G. ... T.. ... ... ... ... ... ... ... ...
PLAN3   T.. ... ... ... .G. ... T.. ... ... ... ... ... ... ... ...
PLAN4   T.. ... ... ... .G. ... T.. ... ... ... ... ... ... ... ...
DIFF1   T.. ... ... ... .G. ... ... ... ... ... ... ... ... G.. ...
DIFF3   T.. ... ... ... .G. ... T.. ... ... ... ... ... ... ... ...
PICT1   ... ... ... ... .G. ... ... ..___ ___ ___ ___ ___ ___ ___ ___
PICT4   ..T ... ... ... .G. G.. ... ... ___ ___ ___ ___ ___ ___ ___ ___
TC1     ..T A.A CGA ..A .TG ..T ..A A.T ..C .G. ..G A.G AGT ... ..A
HB1     ..T ..A CG. T.A ..G C.A GCA .GA C.A C.C .GA ... AAG ... GTC
```

```
HET1    AAG AAG CCA TTT ATA AGC ACA AAA AAT AAA GGG ACT AGG ATG ACG
HET3    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HET4    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SIL2    ... ... ... ... ... ... G.. ... ... ... ... ... ... ... .G.
SIL3    ... ... ..C ... ... ... G.. ... ... ... ... ... ... ... ...
PLAN2   ... ... ..C ... ... ... G.. .G. ... G.. ... .T. ... ... ...
PLAN3   ... ... ..___ ___ ___ ___ ___. .G. ... G.. ... .T. ... ... ...
PLAN4   ... ... ..C ... ... ... G.. .G. ... G.. ... .T. ... ... ...
DIFF1   ... ... ..C ... ... ... G.. .G. ... G.. ... .T. ... ... ...
DIFF3   ... ... ..C ... ... ... G.. .G. ... G.. ... .T. ... ... ...
PICT1   ... ... ..C ... ..T ... G.. ... ... ... .AA ... ... ..___ ___
PICT4   ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___
HB1     ... GTT ..C C.A CCT ... C.. .GG C.. .TT AA. G.A ... T.A .GC
TC1     ... ..A ..G ..C ..C ..T .AG ... ... CGC AT. G.. C.A G.T G..
```

```
HET1   TTC GCC AAA ACC CAC TTG GAC AAG GAT TTG GAG TTC TGG AAC ACA
HET3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
HET4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SIL2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SIL3   ... ... ___ ___ ... ... ... ... ... ... ... ... ... ... ...
PLAN2  ..G ... ... ... ... ... ... ... ... G.. ... ... ... ... ...
PLAN3  ..G ... ... ... ... ... ... ... ... G.. ... ... ... ... ...
PLAN4  ..G ... ... ... ... ... ... ... ... G.. ... ... ... ... ...
DIFF1  ..G ... ... ... ... ... ... ... ... G.. ... ... ... ... ...
DIFF3  ..G ... ... ... ... ... ... ... ... G.. ... ... ... ... ...
PICT1  ___ ___ ___ ___ ___ __. ... ... ... ... ..A ..G ... ... ...
PICT4
HB1    ..A ..T ... ... T.. C.A A.. TG. CCA G.C TCC AAA ... CGT .AT
TC1    .GG ..A ... G.G ..T C.T CGT TG. .GA CGT C.. GAA ... GCT .A.

HET1   ATC ATA TTT GAA GAT GAG TCC AAA TTC ATA ATT TTT GGC TCG GAC
HET3   G.. ... ... .C. ... ... ... ... ... .AC ... ... .A. ... ...
HET4   ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ...
SIL2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
SIL3   G.. ... ... ... ... ... ... ... ..T .AC ... ... ..A ... ...
PLAN2  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
PLAN3  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
PLAN4  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DIFF1  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
DIFF3  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
PICT1  G.. ... ... ... A.. ... ..A ... ..A .AC ... ... ... C.. .G.
PICT4
HB1    ... C.T .GG ACT ... .G. ..A ... A.. ..G C.A ... ..T GGA ACT
TC1    CA. ..C .GG TCT ..C ..A AG. ..G ... .AT T.G ..C ..G AGT ..T

HET1   GGA CGG AAT TAT GTG ___ CGG CGA CAG TCC AAT ACT GAG CTG AAT
HET3   ... ... ... ... ... ___ T.. ... ... ... ... ... ... ... ...
HET4   ... ... ... ... ... ___ T.. ... ... ... ... ... ... ... ...
SIL2   ... ... ... ... ... ___ ... ... ... ... ... ... ... ... ...
SIL3   ... .T. ... ... .G. ___ T.. ... ..A ... ... ... ... .C. ...
PLAN2  ... ... ... ... ... ___ ... ... ... ... ... ... ... ... ...
PLAN3  ... ... ... ... ... ___ ... ... ... ... ... ... ... ... ...
PLAN4  ... ... ... ... ... ___ ... ... ... ... ... ... .G. ... ...
DIFF1  ... ... ... ... ... ___ ... ... ... ... ... ... ... ... ...
DIFF3  ... ... ... ... ... ___ ... ... ... ... ... ... .G. ... ...
PICT1  ... ..A ... .G. ... ___ T.. ... ... ... ... CA. ... ... ...
PICT4                      ___                   ... ... ... G.. ... ..C ...
HB1    ..T TCA CTA C.G TAT ATC T.A ... .CT C.A ..C ..G ... TAT C.C
TC1    ... AAT TCC .GG ..A ___ ..T ..T .CT GTT GGC T.. AG. TAC TC.
```

```
HET1   CCC AAA AAC CTA AAG GCA ACA GTG AAG CAC GGC GGA GGA AGT GTC
HET3   ..G ... C.. ... ... ... ... ... ... ... ... ... ... ... ...
HET4   ..G ... C.. ... ... ... ... ... ... ... ... ... ... ... ...
SIL2   ... ... ... ... ... ... C.. ... ... ... ..G .AG ... ... ...
SIL3   ... ... GG. T.. ... C.. C.G ... ... ... ..A ... ... ... ...
PLAN2  ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ...
PLAN3  ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ...
Plan4  ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ...
DIFF1  ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ...
DIFF3  ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ...
PICT1  ... ... ... ... ... ... ... ..C ... ... ... ... ... ... T..
PICT4  ... ... ... ... ... .T. ..C .AC ... ... A.. A.C ... G.. ...
HB1    ..A ... C.. .C. GT. AAG ..T T.C ..T ... ..T ... CCT .AA A..
TC1    ..A ..G T.T .A. TGC C.. ..C ..T ... ..T ..A ..T ..G ..C ...

HET1   ATG GTA TGG GC_ ___ ATG TAT CTC CGC AGC CAG CGT C_G GAA ATT
HET3   ... ... ... ..._ ___ ... ... ... G.. ... .G. ... ._. ... ...
HET4   ... ... ... ..._ ___ ... ... ... G.. ... .G. ... A_. .G. ...
SIL2   ..T ... ... ..._ ___ ... ... ... G.. ... ... ... ._. ... ...
SIL3   ... ... ... ..G GGC ... ... ... G.. ... .G. ... ._. ... ...
PLAN2  ... ... ... ..._ ___ ..C ... ... G.. ... .G. ... ._. ... ...
PLAN3  ... ... ... ..._ ___ ..C ... ... G.. ... .G. ... ._. ... ...
PLAN4  ... ... ... ..._ ___ ..C ... ... G.. ... .G. ... ._. ... ...
DIFF1  ... ... ... ..._ ___ ..C ... ... G.. ..T .G. ... ._. ... ...
DIFF3  ... ... ... ..._ ___ ..C ... ... G.. ... .G. ... ._. ... ...
PICT1  ... ... ... ..A TCT ..A .C. A.A T?. T.T ATC TA. A_T CC. TC.
PICT4  ... ... G.. ..._ ___ ..C ... ... GCA ..T .GT ... ._. ... ...
HB1    ... ... ... ..._ ___ T.. .T. T.T TTA TAA TG. TA. GA. TC. TGC
TC1    ... ..G ... .G_ ___ G.. CT. .A. .AG CA. TTC .A. G_. .CC CAC

HET1   TGG TGT GTA TT_ __G AAA CAA CAA CGG ACA GGA ATG TGG ACC TCA
HET3   ... ... ... ..._ __. ... ... ... ... ... ... ... ... ... ...
HET4   ... ... TC. ..._ __. ..C ... ... T.. ... A.. ... .AT .T. ...
SIL2   ... ... ... ..._ __. ... ... ... ... ... ... ... ... ... ..?
SIL3   ... ... T.. ..._ __. ... ... ... TC. ... A.. ... ..T ... ...
PLAN2  ... ... TC. ..._ __. ... ... ... T.. ..T T.. ... ..T ... ...
PLAN3  ... ... TC. ..._ __. ... ... ... T.. ... T.. ..? ??? ??? ???
PLAN4  ... ... TC. ..._ __. ... ... ... T.. ... T.. ... ..T ... ...
DIFF1  ... ... TC. ..._ __. ... ... ... T.. ... T.. ... ..T ... ...
DIFF3  ... ... TC. ..._ __. ... ... ... T.. ... T.. ... ..T ... ...
PICT1  AAA .C. A.. ..._ __A C.. ... ... T.. ..C A.. ..A ..T .T. ...
PICT4  ... ... TC. C._ ___. ... ... ... T.. ... A.. ... ..T ... ...
HB1    .AT G.A T.. .GA TTT .TG GT. TT. TA. ..C AA. .C. CAT .TG .A.
TC1    .AA G.A .A. .C_ __C ... GC. TT. T.. .TC .TT T.C AAT ..G AA.
```

```
HET1    GTA TAT TAA AGG AAA ATT TAC TCC AAA GT
HET3    GA. ... ... ... ... ... ... ... ... ...
HET4    A.. ..C ... ... ... ... ... ... ... ...
SIL2    ??? ??? ??? ??? ??? ??? ??? ??? ??? ???
SIL3    A.. ... ... ... ... ..G ... ... ... ...
PLAN2   A.. ... ... ... ... .?? ??? ??? ??? ???
PLAN3   ??? ??? ??? ??? ??? ??? ??? ??? ??? ???
PLAN4   A.. ... ... ... ... ??? ??? ??? ??? ???
DIFF1   A.. ... ... ... ... ... ..? ??? ??? ???
DIFF3   A.. ... ... ... ... ... .?? ??? ??? ???
PICT1   A.. ... ... ... ... ... .?? ??? ??? ???
PICT4   A.. ... ... ... ... ..G ... ... ... ...
HB1     A.. ..C .T. GT. .TG TC. ..T .GT C.T A.T
TC1     AC. .C. .TG .AA CT. CAA .G. GA. CCT .G.
```

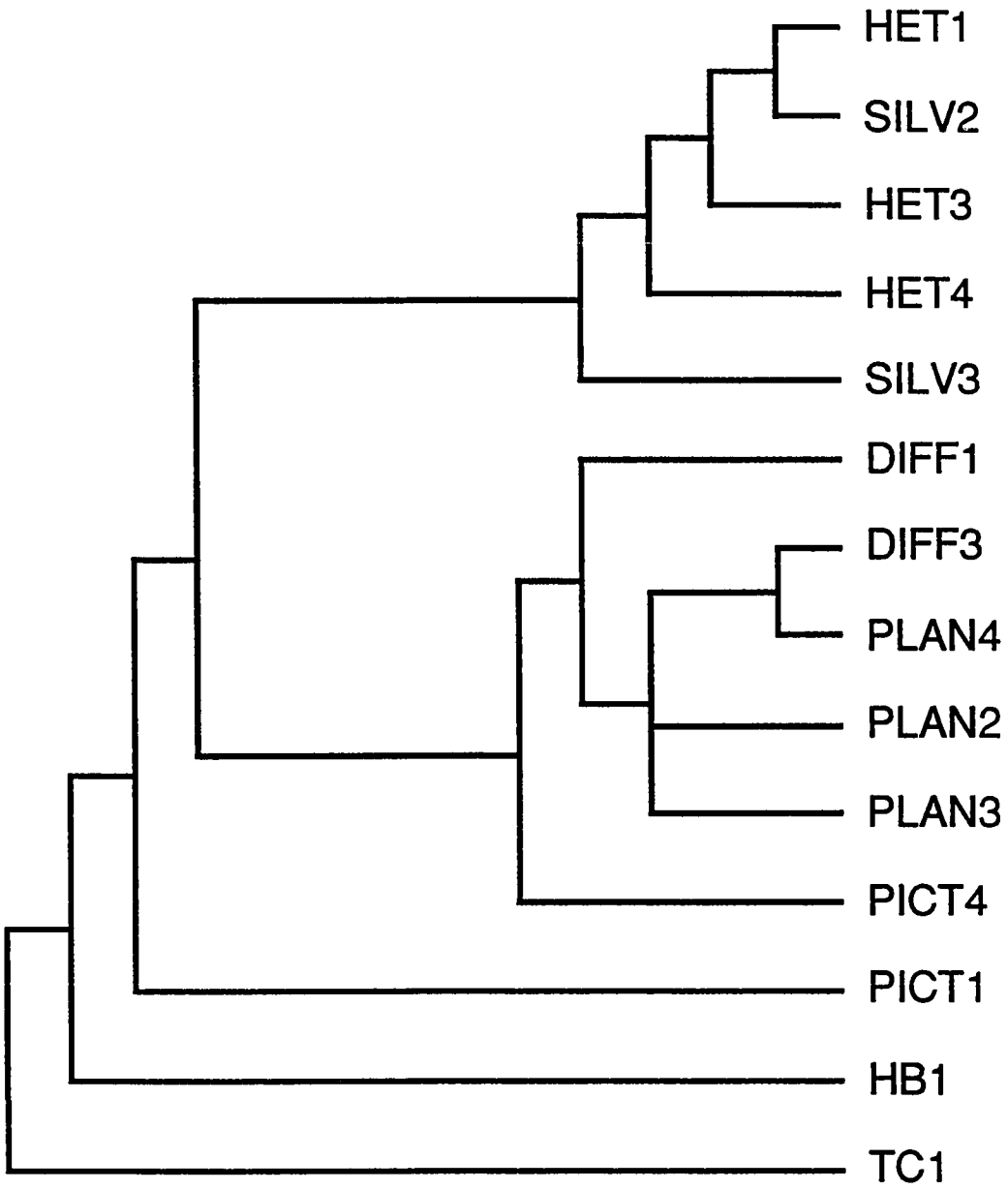


Figure VIII. Cladogram of the Tc1-like elements based on the DNA sequence data in figure VII. Tree was drawn using PAUP (maximum parsimony analysis) and represents the strict concensus of the shortest trees which include 1 tree with 564 steps and 13 trees with 565 steps. Tc1 was indicated as the outgroup. Branch lengths are not drawn to scale

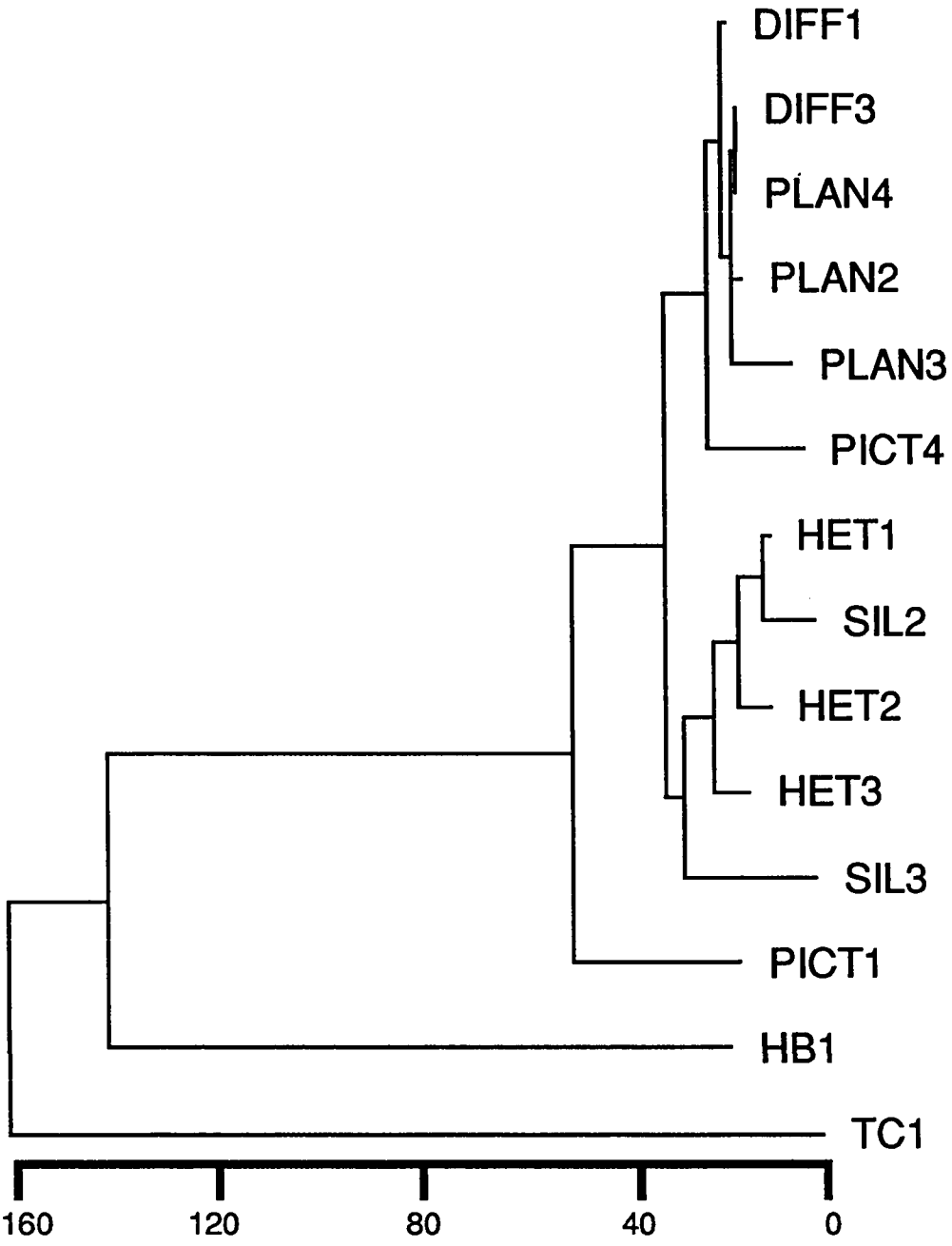


Figure IX Phylogram of Tc1-like elements based on the DNA sequence data in figure VI. Tree was drawn using PAUP (maximum parsimony analysis) and represents the strict consus of the shortest trees which includes one tree with 564 steps and 13 trees with 565 steps. Horizontal branch lengths are relative. Scale in steps is indicated below the tree.

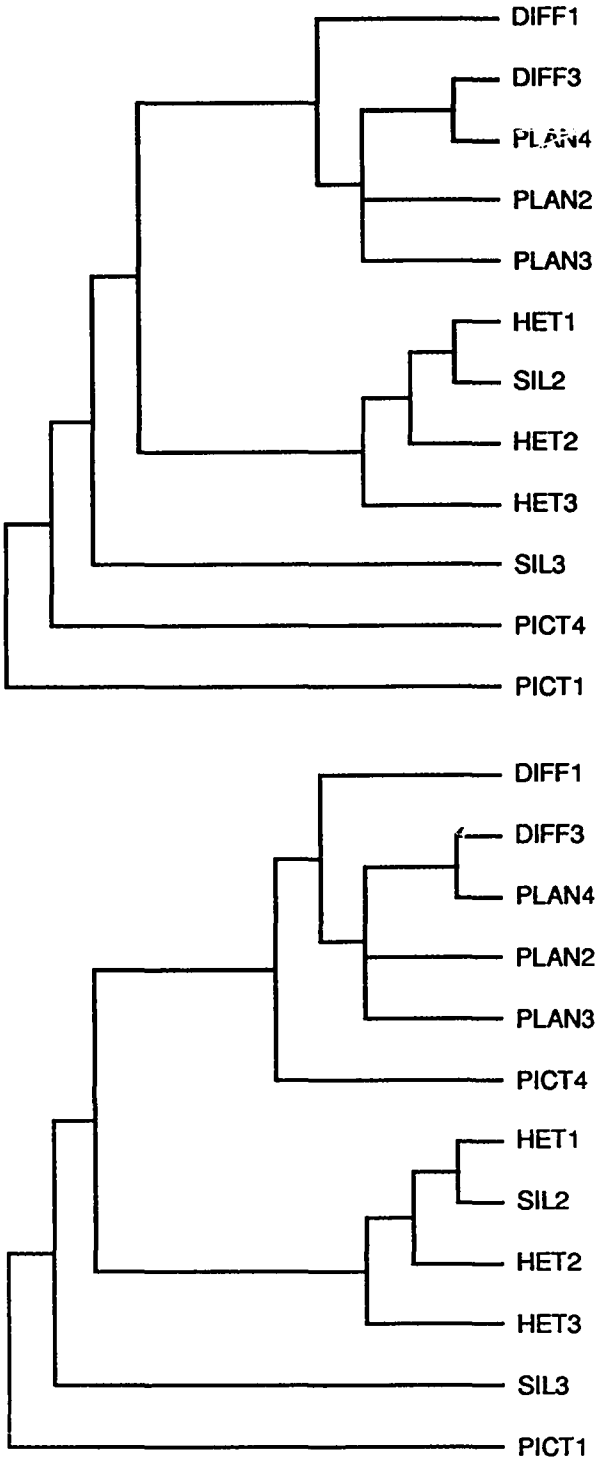


Figure X Cladograms of Uhu elements based on the sequence data in figure VI. Trees were constructed using PAUP (maximum parsimony analysis) and represent the 2 shortest trees which consist of 168 steps. PICT1 was indicated as the out-group. Trees are not drawn to scale.

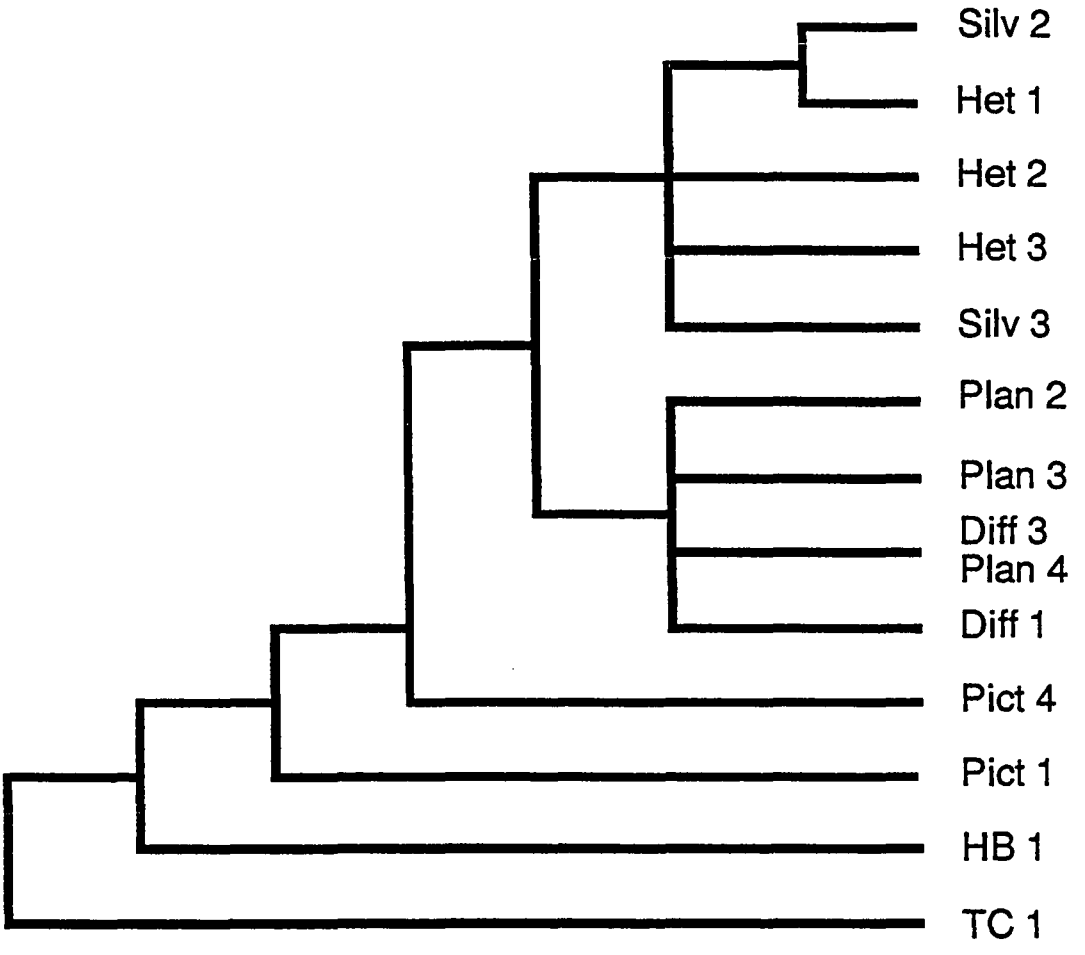


Figure XI. Cladogram showing relative relationships of Tc1-like elements based on the data in figure VII. Tree was constructed using FITCH with TC1 indicated as an outgroup. Only significant branches are indicated. Branches are not drawn to scale.

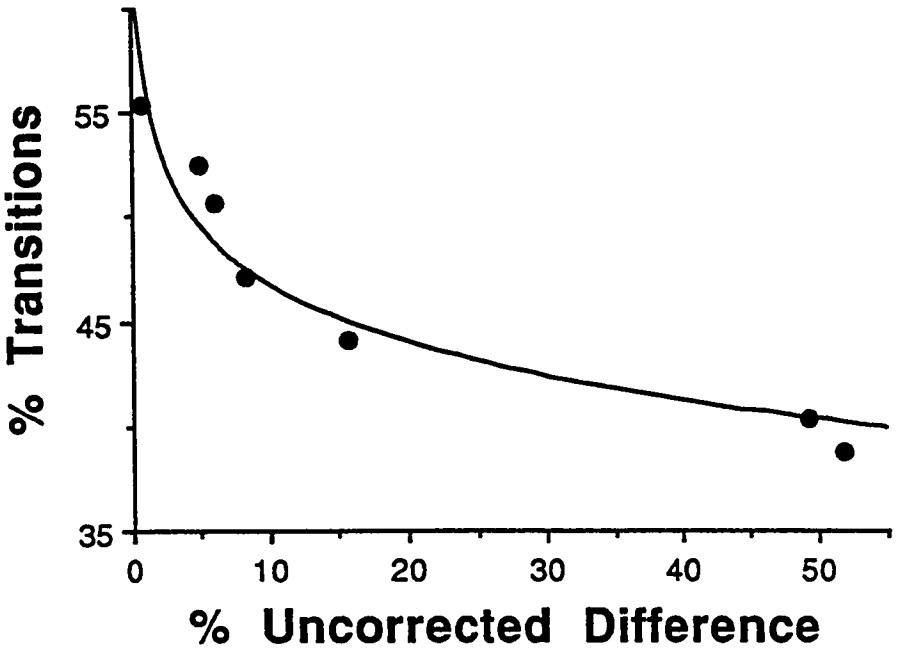


Figure XII Relationship between sequence divergence and percent transitions. Uncorrected percent differences are found in table II. All data points are averages of all possible comparisons as follows:Tc1/Hawaiian *Drosophila*, HB1/Hawaiian *Drosophila*, PICT1/all other Hawaiian *Drosophila*, PICT4/all other Hawaiian *Drosophila*, all possible combinations among HET,SILV, PLAN and DIFF, DIFF/PLAN, and HET/SILV.

	HET1	HET3	HET4	DIFF1	DIFF3	PICT1	PICT4
HET1		2.8±0.9(1.5)	4.2±1(1.8)	5.4±1.1(2.1)	5.4±1.1(2.1)	15.9±1.4(3.5)	9.3±1.3(3)
HET3			4.2±1(1.9)	7.4±1.1(2.4)	7.4±1.1(2.5)	16.2±1.4(3.5)	9.7±1.3(3.1)
HET4				6.1±1.1(2.2)	6.2±1.1(2.3)	15.2±1.4(3.4)	9.4±1.3(3)
DIFF1					1±0.7(1)	16.6±1.4(3.6)	6.7±1.2(2.7)
DIFF3						17.7±1.5(3.7)	7.3±1.2(2.9)
PICT1							17.6±1.5(4.2)
PICT4							
PLAN2							
PLAN3							
PLAN4							
SIL2							
SIL3							
HB1							
TC1							

	PLAN2	PLAN3	PLAN4	SIL2	SIL3	HB1	TC1
HET1	5.7±1.1(2.2)	4.4±1(2)	5.4±1.1(2.2)	2.8±0.9(1.6)	7.9±1.2(2.5)	93.0±2.6(2.1)	98±2.5(1.1)
HET3	7.8±1.2(2.5)	6.3±1.1(2.4)	7.5±1.2(2.5)	5.2±1.1(2.2)	7.6±1.1(2.4)	93.8±2.6(2)	100.1±2.6(NA)
HET4	6.5±1.1(2.3)	5±1.1(2.1)	6.2±1.1(2.3)	6.6±1.1(2.4)	7.7±1.1(2.4)	85±2.4(2.9)	101±2.6(NA)
DIFF1	1.3±0 .7(1.1)	0.8±0.7(0.9)	1.0±0.7(1.0)	7.3±1.2(2.6)	8.3±1.2(2.6)	81.7±2.4(3.3)	89.7±2.5(2.5)
DIFF3	0.8±0.7(0.9)	0.3±0.5(0.5)	0.0±0.0(0.0)	7.3±1.2(2.6)	8.3±1.2(2.6)	83.0±2.4(3.2)	86.4±2.4(2.8)
PICT1	17.5±1.5(3.7)	18±1.5(3.9)	17.9±1.5(3.9)	19.7±1.5(4)	18.5±1.5(3.7)	68.1±2.1(4.1)	78.2±2.4(3.6)
PICT4	7.8±1.2(2.9)	7.8±1.3(3)	7.5±1.2(2.9)	11.7±1.4(3.6)	9.7±1.3(3.1)	63.3±2.3(4.6)	69.5±2.3(4.5)
PLAN2		0.5±0.6(0.7)	0.8±0.7(0.9)	7.6±1.2(2.6)	8.4±1.2(2.6)	82.9±2.4(3.2)	87.2±2.4(2.8)
PLAN3			0.3±0.5(0.5)	6.7±1.2(2.5)	8.4±1.2(2.7)	77.1±2.3(3.7)	80.9±2.4(3.4)
PLAN4				7.3±1.2(2.6)	8.1±1.2(2.6)	82.6±2.4(3.2)	85.4±2.6(3.4)
SIL2					8.4±1.2(2.7)	79.9±2.3(3.5)	78.7±2.3(3.5)
SIL3						89.4±2.4(2.5)	100.7±2.6(NA)
HB1							104.5±2.9(NA)
TC1							

Table I Matrix of corrected percent difference between the indicated DNA sequences which are found in figure VI. Values have been corrected for convergence and reversals as described in the text. Standard deviation is indicated to the right of each corrected percent difference, 95% confidence intervals are indicated in parenthesis

	HET1	HET3	HET4	DIFF1	DIFF3	PICT1	PICT4
HET1		2.7	4.2	5.2	5.2	14.2	9.4
HET3	0.7		4.1	7.1	7.1	16.2	9.1
HET4	1.2	1.6		5.9	5.9	13.7	8.8
DIFF1	1.0	0.9	0.9		4.0	14.9	6.4
DIFF3	1.2	1.0	1.1	3.0		15.7	7.4
PICT1	0.9	0.8	0.7	0.7	0.8		15.7
PICT4	1.0	0.8	0.9	0.8	1.0	0.6	
PLAN2	0.9	0.8	0.9	0.7	0.5	0.8	0.8
PLAN3	1.1	1.0	0.7	2.0	—	0.7	1.0
PLAN4	1.2	1.0	1.1	3.0	0.0	0.8	1.1
SIL2	0.8	1.0	1.1	1.0	1.2	0.9	1.1
SIL3	1.2	1.0	1.1	1.4	1.5	0.8	1.2
HB1	0.7	0.7	0.6	0.7	0.9	0.5	0.7
TC1	0.6	0.6	0.6	0.6	0.6	0.8	0.7

	PLAN2	PLAN3	PLAN4	SIL2	SIL3	HB1	TC1
HET1	5.7	4.3	5.2	2.8	7.9	52.9	54.6
HET3	7.8	6.1	7.1	5.2	7.3	53.2	55.1
HET4	6.2	4.8	6.0	6.3	7.3	50.7	55.3
DIFF1	1.3	0.8	1.0	7.0	7.8	49.5	52.1
DIFF3	0.8	0.3	0.0	7.3	7.9	49.9	51.2
PICT1	15.6	16.0	15.9	17.3	16.4	44.7	48.1
PICT4	7.4	7.4	7.1	10.8	9.1	42.4	45.2
PLAN2		0.5	0.8	7.2	7.9	49.9	51.3
PLAN3	0.0		0.3	6.4	7.9	47.9	49.4
PLAN4	0.5	0.0		7.0	7.7	49.8	48.6
SIL2	0.9	1.3	1.2		7.9	49.0	55.3
SIL3	1.4	1.6	1.7	1.6		52.0	55.6
HB1	0.7	0.7	0.7	0.6	0.6		52.0
TC1	0.7	0.6	0.6	0.6	0.6	0.8	

Table II. Matrix of data derived from DNA sequence data in figure VI. Uncorrected percent difference is found above the diagonal and ratio of transitions:transversions is found below the diagonal.

APPENDIX

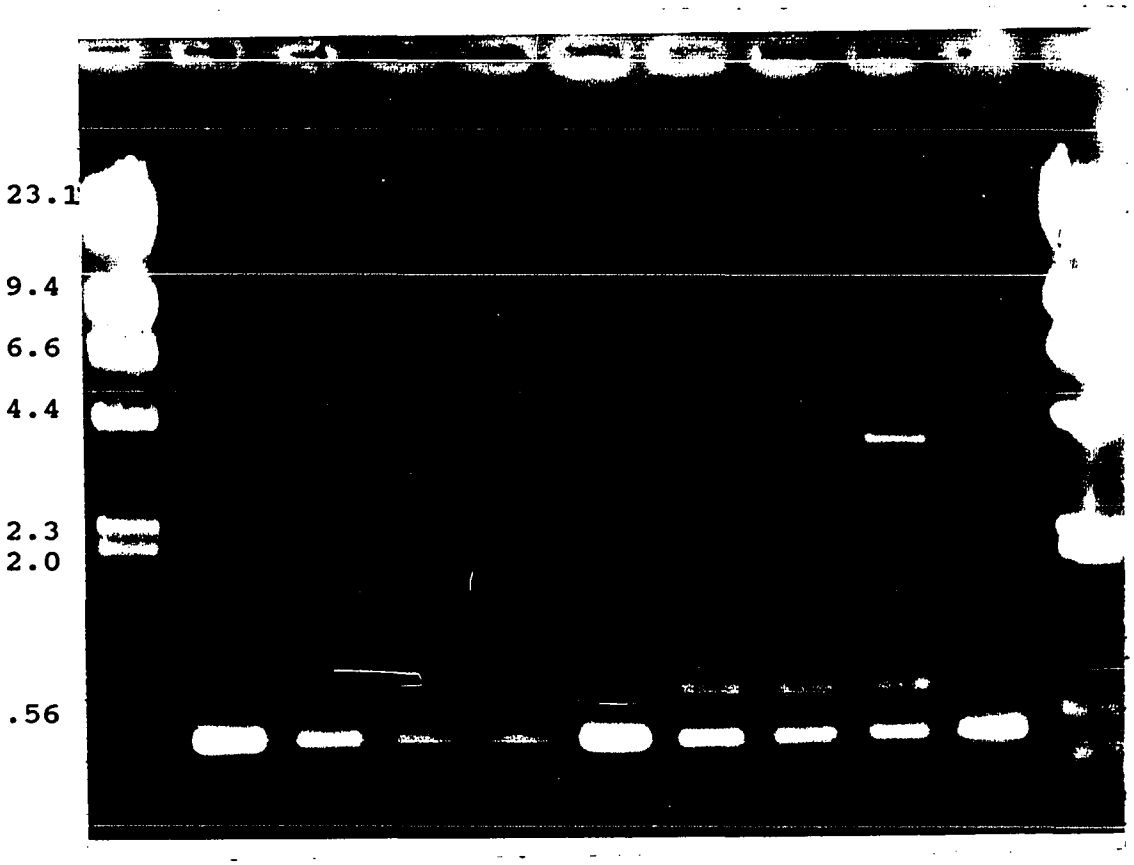
Asymetric PCR Amplification

The Polymerase chain reaction (PCR), results in the amplification of a region of a template DNA located between 2 oligonucleotide primers. This proceedure involves repeated cycles of template denaturation, primer annealing, and primer extension by Taq polymerase (a heat stabile DNA polymerase) (Saiki et al. 1988, Maniatis et al. 1989). Under standard conditions, PCR amplification produces double-stranded DNA. Asymetric PCR amplification is designed to produce predominantly single-stranded (ssDNA), which can be used as template for DNA sequencing (Maniatis et al. 1989). This is accomplished by using limiting amounts of one of the two primers in standard PCR reactions. The result is, that the limiting primer is rapidly used up, and the other primer continues to prime synthesis of the alternate strand, for the duration of the reaction. By using limiting amounts of each of the two primers, in separate reactions, both strands of the template can be amplified. The specific conditions used for PCR amplification have been described in the methods and materials section. The purpose of this appendix is to present an example of the physical results of asymetric PCR amplification.

The optimal ratio of one primer to the other was determined empirically. I conducted 4 reactions for each of the 2 primers in which I used, 100 pmoles of the nonlimiting primer, and: 10 pmoles, 2 pmoles, 1 pmole, and 0.5 pmoles of the limiting primer. PCR products were electrophoresed on agarose gels and visualized by ethidium bromide staining, using Lambda/HindIII, and double-stranded PCR amplification product as markers. Figure XIII shows the results of this experiment for the lambda clone, SILV3. Two prominant bands appear in each of the asymetric amplifications: The faster migrating band comigrates with the double-stranded product and the slower migrating band presumably represents the single-stranded product. However, a third, more slowly migrating band is apparent, which increases in intensity with decreasing ratios of primers. This band may represents single-stranded DNA, replicated from the original Lambda template with the nonlimiting primer. This would indicate that there was excess template in the reaction. Such a replication would have no specified termination position thus explaning why this band is smeared, and, is much larger than the presumed ssDNA product primed from the double-stranded PCR amplification product. It is possible that this band represents incorrect priming of the template. However, even if this were the case, it did not interfere with subsequent sequencing reactions; All sequences were nonambiguous.

To determine the optimal ratio of primers, I selected the reactions which exibited a decreased amount of double-stranded product and an increased amount of the presumed single stranded product, and sequenced them. In this case, I sequenced the reactions with the following ratios of primers: 1:50, 1:100, and 1:200 (lanes 3,4,5,7,8,and 9). The ratios of 1:100 and 1:200 yielded the best sequencing results. All 12 of the lambda clones amplified gave similar results.

Figure XIII Asymetric PCR amplification of the Lambda clone SILV3 using primers 1 and 2 (figure V). Samples are as follows (numbered from left to right): lanes 1 and 11, Lambda/HindIII, lane 10 double-stranded amplification of SILV3, lanes 2-9 are asymetric amplifications of SILV3 as follows: lane 2, 100 pmoles of primer #1 and 10 pmoles of primer #2 , lane 3, 100 pmoles of primer #1 and 2 pmoles of primer #2, lane 4, 100 pmoles of primer #1 and 1pmole of primer #2, lane 5, 100 pmoles of primer #1 and 0.5 pmoles of primer #2, lane 6, 100 pmoles of primer #2 and 10 pmoles of primer #1, lane 7, 100 pmoles of primer #2 and 2 pmoles of primer #1, lane 8, 100 pmoles of primer #2 and 1pmole of primer #1, lane 9, 100 pmoles of primer #2 and 0.5 pmoles of primer #1.


23.1
9.4
6.6
4.4
2.3
2.0
.56

BIBLIOGRAPHY

Berg, D. E., C. M. Berg, and C. Sasakawa, 1984 Bacterial transposon Tn: Evolutionary inferences. Mol. Biol. Evol. 1:411-422

Beverly, S. M. and A. C. Wilson, 1984 Molecular evolution in Drosophila and the higher diptera II. A time scale for foy evolution. J. Mol. Evol. 21:1-13

Bingham, P. M., M. G. Kidwell, and G. M. Rubin, 1982 The molecular basis of P-M dysgenesis: The role of the P element, a P-strain-specific transposon family. Cell 29:995-1004

Bingham, P. M., and Z. Zachar, 1989 Retrotrsnsposons and the FB transposon from Drosophila melanogaster. In: Mobile DNA Edited by D. Berg and M. M. Howe. American Society for Microbiology. Washington,D. C.

Birnboim, H. C. and J. Doly, 1979 A rapid alkaline extraction proceedure for screening recombinant plasmid DNA. Nuc. Acid. Res. 7:1513-1523

Bishop, J. G. B. III, 1982 Evolution of the alcohol dehydrogenase gene region in the Hawaiian Drosophila. doctoral thesis, University of Hawaii, Department of Genetics

Bishop, J. G. B. III and J. A. Hunt, 1988 DNA divergence in and around the alcohol dehydrogenase locus in five dolsely related species of Hawaiian Drosophila Mol. Biol. Evol. 5:415-531

Boeke, J., D. J.Garfinkel, C. A. Styles, and G. R. Fink, 1985 Ty elements transpose through an RNA intermediate. Cell 40:491-500

Breathrach, R., and P. Chambon, 1981 Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50:349-383

Brenner, S., 1974 The genetics of C. elegans. Genetics 77:71-94

Brezinsky, L., G. V. L. Wang, T. Humphreys, and J. Hunt, 1990 The transposable element Uhu from the Hawaiian Drosophila - A member of the widely dispersed class of Tc1-like transposons. Nuc. Acids. Res. 8:2053-2059

Brierly, H. L.,and S. S. Potter, 1985 Distinct characteristics of loop sequences of two Drosophila foldback transposable elements. Nucl. Acids Res. 13:485-500

Brookfield, J. F. Y., 1986 A model for DNA sequence evolution within transposable element families. Genetics 112:393-407

Brown W. M., E. M. Prager, A. Wang, A. C. Wilson, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Ecol. 18:225-239

Brown G. G., and M. V. Simpson, 1982 Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. Proc. Natl. Acad. Sci. USA 79:3246-3250

Bucheton, A., R. Paro, H. M. Sang, A. Pellison, and D. J. Finnegan, 1984 The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster:Identification cloning, and properties of the I factor. Cell 38:153-163

Cameron, J. R., E. Y. Loh, and R. W. Davis, 1979 Evidence for transposition of dispersed repetitive DNA families in Yeast. Cell 16:739-751

Capello, J. K. Handelsman, H Lodish, 1985 Sequence of Dictyostelium DIRS-1: An appparent retroposon with inverted terminal repeats and an internal circle junction sequence. Cell 43. 105-115

Carson, H. L. , 1975 The genetics of speciation at the diploid level. Am. Natur. 109:82-93

Carson, H.L. and K.Y. Kaneshiro, 1976 Drosophila of Hawaii: Systematics and ecological genetics. Annu. Rev. Ecol. Syst. 7:311-346

Carson, H. L., K. Y. Knaeshiro, and F. C. Val, 1989 Natural hybridization between the sympatric Hawaiian species Drosophila silvestris and Drosophila heteroneura. Evolution 43:190-203

Carson, H. L., F. C. Val, C. M. Simon, and J. W. Archie, 1982 Morphometric evidence for incipient speciation in Drosophila silvestris from the island of Hawaii. Evolution 36a:132-140

Carson, H. L., and J. S. Yoon, 1982 Genetics and evolution of Hawaiian Drosophila. In: The Genetics and Biology of Drosophila, Vol 3b, pp 297-344. Edited by M. Ashburner, H. L. Carson, and J.N. Thompson Jr. Academic Press. New York.

Carson, H. L., and R. G. Wisotzkey, 1989 Increase in genetic variance following a population bottleneck. The American Naturalist 134:668-673

Collins, J., B. Saari, and P. Anderson, 1987 Activation of a transposable element in the germ line but not the soma of Caenorhabditis elegans. Nature 328:726-728

Craddock, E. M. and W.E. Johnson, 1979 Genetic variation in Hawaiian Drosophila. Chromosomal and allozyme diversity in D. silvestris and its homosequential species. Evolution 33:137-155

Dale, R. M. K., B. A. McClure, and J. P. Houchins, 1985 A new proceedure for producing a sequential series of overlapping clones for use in DNA seqeuncing. Plasmid 13:31-40

Dalrymple, G. B., E. A. Silver, and E. D. Jackson, 1973 Origin of the Hawaiian Islands. Amer. Sci. 61:294-308

DeSalle, R., and V. Giddings, 1986 Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian Drosophila. Proc. Natl. Acad. Sci. USA 83:6902-6906

DeSalle, R., T. Freedman, E. M. Prager, and A. C. Wilson, 1987 Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian Drosophila. J. Mol. Evol. 26:157-164

Doolittle, W. F., 1980 Selfish genes, the phenotype, paradigm and genome evolution. Nature 284:601-603

Eide, D. and P. Anderson, 1985 Transposition of Tc1 in the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA 82:1756-1760

Eide, D., and P. Anderson, 1985 The gene structure of spontaneous mutations affecting a Caenorhabditis elegans myosin heavy chain gene Genetics 109:67-79

Eide, D., and P. Anderson, 1988 Insertion and Excission of Caenorhabditis transposable element Tc1. Mol. Cell Biol. 8:737-746

Emmons, S. W. and L. Yesner, 1984 High-frequency excission of transposable element Tc1 in the nematode Caenorhabditis elegans is limited to somatic cells. Cell 36:599-605

Engels, W. R., 1983 The P family of transposable elements in Drosophila. Ann. Rev. Genet. 17:315-344

Engels, W. R., 1989 P elements in Drosophila. In: Mobile Genetic Elements , Edited by D.Berg and M. Howe. American Society of Microbiology, Washington, D.C.

Felsenstein, J., 1988 Phylogenies from molecular sequences: Inferences and reliability. Ann. Rev. Genet. 22:521-565

Federoff, N. V., 1988 Maize transposable elements. In: Mobile Genetic Elements, Edited by D. Berg and M. Howe. American Society for Microbiology, Washington,D.C.

Flavell, A. J., and D. Ish-Horowicz 1981 Extrachromosomal circular copies of the eukaryotic transposable element in cultured Drosophila cells. Nature 292:591-595

Flavell, A. J., 1984 Role of reverse transcription in generation of Extrachromosomal Copia mobile genetic elements. Nature 310:514-516

Green, M. M., 1980 Transposable elements in Drosophila and other Diptera. Ann. Rev. Genetics 14,109-120

Hagenbuchle, O, B. Ramonde, and R. A. Younge, 1980 Tissue-specific expression of Mouse alpha-amylase genes:Nucleotide sequence of isozyme mRNAs from pancreas and salivary gland. Cell 21:179-187

Harris, L. J., and A. M. Rose, 1986 Somatic excission of transposable element Tc1 from the Bristol genome of Caenorhabditis elegans. Mol. Cell. Biol. 6:1782-1786

Harris, L. J., D. L. Baillie, and A. M. Rose, 1988 Sequence identity between an inverted repeat family of transposable elements in Drosophila and Caenorhabditis. Nucl. Acids Res. 16:5991-5999

Harris, L. J., S. Prasad, and A. M. Rose, 1989 Isolation and sequence analysis of C. briggsae repetitive elements related to the C. elegans transposon Tc1. J.Mol.Evol. In Press

Herman, H.K., and J.E. Shaw, 1987 The transposable genetic element Tc1 in the nematode Caenorhabditis elegans. Trends in Genetics 3:222-225

Henikoff, S., and R. H. A. Plasterk, 1988 Related transposons in C. elegans and D. melanogaster. Nucl. Acids Res. 16:6234

Hunt, J. A., and H.L. Carson, 1983 Evolutionary relationship of four species of Hawaiian Drosophila as measured by DNA reassociation. Genetics 104:353-364

Hunt, J. A., J. G. Bishop III, and H. L. Carson, 1984 Chromosomal mapping of a middle-repetitive DNA sequence in a cluster of five species of Hawaiian Drosophila. Proc. Natl. Acad. Sci. USA 81:7146-7150

Kaneshiro, K. Y., 1976 Ethological isolation and phylogeny of the planitibia subgroup of Hawaiian Drosophila. Evolution 30:740-745

Kaneshiro, K. Y., and F. C. Val, 1977 Natural hybridization between a sympatric pair of Hawaiian Drosophila. Amer. Nat. 111:897-902

Kimble, J., and D.Hirsh, 1979 The post-embryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. Dev. Biol. 70:396-417

Kimura, M, 1980 A simple method for estimating the evolutionary rates of base substitutions through comparative studies of nucleotide sequences.

Liao, L. W., B. Rosenzweig,and D. Hirsh, 1983 Analysis of a transposable element in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA 80:3585-3589

Kleckner, N., 1981 Transposable elements in prokaryotes. Ann. Rev. Genet. 15:341-404

Liao, W. L., Rosenzweig, B. , and Hirsh, D. 1983 Analysis of a transposable element in Caenohabditis elegans.

Mager, D. L. and P. S. Henthorn, 1984 Identification of a retrovorus-liike repetitive element in human. DNA. Proc. Natl. Acad. Sci. USA 7510-7514

Maniatis, T., E.F. Fritch, and J. Sambrook, 1982 Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbour. N. Y.

Maniatis, T., E.F. Fritch, and J. Sambrook, 1989 Molecular Cloning: A Laboratory Manual-second edition. Cold Spring Harbor Laboratory, Cold Spring Harbour. N. Y.

Martin, G.,D. Wiernasz, adn P Shcedl, 1983 Evolution of Drosophila repetitive-dispersed DNA. J. Mol. Evol. 19:203-213

Mayr, E., 1963 Animal species and evolution Harvard University Press

McClintock, B., 1951 Chromosome organization and genic expression. Cold Spring Harbor Symp. Quant. Biol. 16:13-47

McDonald, J. F. 1989 The potential evolutionary significance of retroviral-like transposable elements in peripheral populations. pp190-205 in: Evolutionary biology of transient unstable populations. edited by A. Fontdevilla. Springer-Verlag, New York

McDonald, J. F., D. J. Strand, M. R. Brown, S. M. Paskewitz, A. R. Csink, and S. H. Voss, 1988 Evidence of Host-mediated regulation of retroviral element expression at the post-transcriptional level. In: Eukaryotic transposable elements as mutagenic agents. edited by Lambert, McDonald,and Weinstein. Cold Spring Harbor Press, Cold Spring Harbor, NY

McDonald, J. F., D. J. Strand, M. E. Lambert, and I. B. Weinstein, 1987 The responsive genome: Evidence and evolutionary implications. Allan Liss Press, NY

McDonald, J. F., 1990 Macroevolution and retroviral elements. Bioscience 40:183-191

McDougall, I., 1969 Potassium-argon ages from lavas of the Hawaiian Islands. Bull. Geol. Soc. Am. 80:2597-2600

Mead, D. A. and B. Kemp, 1986 Chimeric single stranded DNA phage-plasmid cloning vectors. In: A survey of molecular cloning vestors and their uses. Butterworth publishers, Massachusetts

Messing, J., 1983 New M13 vectors for cloning. Meth. Enz. 101:20-79

Mitchell, P. J., and R. Tjian, 1989 Transcriptional regulation in mammalian cills by sequence -specific DNA binding proteins. Science 245:371-378

Moerman, D. G., and R. H. Waterston, 1984 Spontaneous unc-22 mutations in C. elegans var. Bergerac. Genetics 108:859-877

Moerman, R. H. Waterston, 1988 Analysis of a mutator activity necessary for germline transposition and excission of Tc1 transposable elements in Caenorhabditis elegans. Genetics 120:397-407

Moerman, D. G., and R. H. Waterston, 1988 Mobile elements in Caenorhabditis elegans and other nematodes. In : Mobile Genetic Elements, edited by D. Berge and M. Howe. American Society of Microbiology, Washington, D. C.

Mori, I., G. M. Benian, D. G. Moerman, and R. H. Waterston, 1988 Transposable element Tc1 of Caenorhabditis elegans recognizes specific target sites for integration. Proc. Natl Acad. Sci. USA 85:861-864

Mount, S., and G. M. Rubin, 1985 Complete nucleotide sequence of the Drosophila transposable element Copia: Homology between Copia and retroviral proteins. Mol. Cell Biol. July:1630-1638

Murphy, S., B. Moorefield, and T. Pieler, 1989 Common mechanisms of promotor recognician by RNA polymerases II and III. Trends in Genetics Vol 5, Number 4

Nakajima, N., M. Horikoshi,, and R. Roeder, 1988 Factors involved in specific transcription by mamalian RNA polymerase II: Purification, genetic specificity, and TATA box-promoter interactions of TFIID. Mol. Cell. Biol. 8:4028-4040

Nei, M., 1987 Phylogenetic trees. In:The molecular evolutionary genetics. pp 287-326 Columbia University Press

Nei, M., T. Maruyama, and R.Chakraborty, 1987 The bottleneck effect and genetic variability in populations. Evolution 29:1-10

Nocera, P. P. D., and Y. Sakiki, 1990 LINEs: a super family of retrotransposable ubiquitous DNA elements. Trends in genetics 6:29-67

OHare, K., and G. M. Rubin, 1983 Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell 34:25-35

Orgel, L. E, and F. H. Crick, 1980 Selfish DNA: The ultimate parasite. Nature 284:604-607

Panganiban, A. T., and H. M. Temin, 1984 The retrovirus pol gene encodes a product required for DNA integration:Identification of a retrovirus int locus. Proc. Natl Acad. Sci. USA 81:7885-7889

Prakash, S., 1972 Origin and reproductive isolation in the absence of apparent genic diferentiation in a geographic isolate of Drosophila pseudoobscura. Genetics 72:143-155

Proudfoot, N. J., and G. G. Brownlee 1976 3' Non-coding region seaquences in eukaryotic messanger RNA. Nature 263:211-214

Provine, W. B. 1897 Founder effects and genetic revolutions in microevoltion and speciation: An historical perspective. In: Genetics, Speciation and the Founder Principal, edited by L. V. Giddings,, K. Y. Kaneshiro, and W.W. Anderson. Oxford University Press, Oxford, N.Y.

Plasterk, R. H. A., 1987 Differences between Tc1 elements from the C. elegans strain Bergerac. Nucl. Acids Res. 15:10050
Reeder, R. H., 1984 Enhancers and ribosomal gene spacers. Cell 38:349-351Sheldon, F., 1987 Rates of single-copy DNA evolution in Herons. Mol. Biol. Evol. 4:56-69

Rose , A. M. and W. F. Doolittle, 1983 Molecular biological mechanism of speciation. Science 220:157-162

Rose, A. M. and T. P. Snutch, 1984 Isolation of the closed circular form of the transposable element Tc1 in Caenorhabditis elegans. Nature 311:485-486

Rosenzweig, B., L. W. Liao, and D. Hirsh, 1983 Sequence of the C. elegans transposable element Tc1. Nucl. Acids Res. 11:4201-4209

Rowan, R. G., and J. A. Hunt 1990 Rates of DNA change and phyologeny from the DNA sequence fo the Adh gene from five closely related species of Hawaiian Drosophila. (personal communication)

Ruan, K., and S. W. Emmons, 1984 Extrachromosomal copies of transposon Tc1 in the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA 81:4018-4022

Ruan, K. and S. W. Emmons, 1987 Precise and imprecise somatic excission of the transposon Tc1 in the nematode C. elegans Nucl. Acids Res. 15:6875-6881

Rubin, G. M., M. G. Kidwell, and P. M. Bingham, 1982 The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. Cell 29:987-994

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Mullis, and H. A. Erlich, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487

Saigo, K., W. Kugiyama, Y. Matsuo, S. Inouye, K. Yashioka, and S. Yuki, 1984 Identification of the coding sequence for a reverse transcriptase-like enzyme in a transo'posable genetic element in Drosophila melanogaster. Nature 312:659-663

Sanger, F., S. Nicklen, and A. R. Soulson, 1977 DNA

sequencing with chain terminating inhibitors. Proc Natl Acad. Sci. USA 74a;5463-5467

Schmidt, M., and W. R. Jelnick, 1982 The ALU family of dispersed repetitive sequences. Science 216:1065-1070

Sene, F. M. and H. L. Carson, 1977 Generic variation in Hawaiian Drosophila. IV Allozymic similarity between D. silvestris and D. heteroneura from the island of Hawaii Genetics 86:187-189

Sharp, P. M., and W-H Li, 1989 On the rate of DNA sequence evolution in Drosophila. J. Mol. Evol. 28:398-402

Shiba, T., and K. Saigo, 1983 Retrovirus-like particles containing RNA homologous to the transposable element copia in Drosophila melanogaster. Nature 302:119-124

Soeda, E., J. R. Arrand, N. Smolar, and B. E. Griffin 1979 Sequence from early region of polyoma virus DNA containing viral replication origon and enciding small, middle and (part of) large T antigens. Cell 17:357-370

Soeda, E., J. R. Arrand, N. Smolar J. E. Walsh and B. E. Griffin, 1980 Coding potential and regulatory signals of the polyoma virus genome. Nature 283:445-453

Southern, E. M., 1975 Detection of specific sequences among DNA fragments separated fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517

Southern, E. M., 1980 Gel electrophoresis of restriction fragments. Meth. Enz. 69:152

Spieth, H. T., 1981 Drosophila heteroneura and Drosophila silvestris: head shapes, behaviors and evolution. Evolution 35:9210930

Spradling, A. C. , and G. M. Rubin, 1981 Drosophila organization: conserved and dynamic aspects. Ann Rev. Genet. 15:219-264Sheldon, F. H., 1987 Rates of single-copy DNA evolution in Herons. Mol. Biol. Evol. 4:56-69

Stearns, H. T., 1966 Geology of the state of Hawaii. Pacific Books. Palo Alto, Ca.

Sulston, J. E. and H. R. Horvitz, 1977 Post-embryonic cell lineages of the nematode Caenorhabditis elegans.Dev. Biol. 56:110-156

Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson, 1983 The embryonic cell lineages of the nematode

Caenorhabditis elegans. Dev. Biol. 100:64-119

Temin, H. M., 1982 Function of the retrovirus long terminal repeat. Cell 28:3-5

Throckmorton, L. H., 1975 The phylogeny, ecology, and geography of Drosophila. Handbook of Genetics. Vol 3,17

Truett, M. A., R. S. Jones, and S. S. Potter, 1981 Unusual structure of the FB family of transposable elements in Drosophilla. Cell 24:753-763

Wu C-I, and W-H Li, 1985 Evidence for higher rates of substitution in rodents than in man. Proc. Natl. Acad. Sci. 82:1741-1745

Yoon, S. J., 1989 Chromosomal Evolution and Speciation in Hawaiian Drosophila. In: Genetics, Speciation and the Founder Principal, edited by L. V. Giddings,, K. Y. Kaneshiro, and W.W. Anderson. Oxford University Press, Oxford, N.Y.

Zagursky, R. J. and M. L. Berman, 1984 Cloning vectors that yield high levels of single-stranded DNA for rapid DNA sequencing. Gene. 27:183-191

Zeletsova, E. S., R. O. Vashakidze, A. S. Krayev, and M. B. Evgen'ev, 1986 Dispersed repeats in Drosophila Virilis: Elements mobilized by interspecific hybridization. Chromosoma 93:469-476

Zinder, N. D. and J. D. Boeke, 1982 The filamentous phage (Ff) as vectors for recombinant DNA-a review. Gene. 19:1-10

Zuker, C., and H. Lodish, 1981 Repetitive DNA sequences cotranscribed with developementally regulated Dictyostelium disciodium mRNAs. Proc. Natl. Acad. Sci. USA 78:5386-5390